Perceiving visually presented objects: recognition, awareness, and modularity

Anne M Treisman* and Nancy G Kanwisher†

Object perception may involve seeing, recognition, preparation of actions, and emotional responses – functions that human brain imaging and neuropsychology suggest are localized separately. Perhaps because of this specialization, object perception is remarkably rapid and efficient. Representations of componential structure and interpolation from view-dependent images both play a part in object recognition. Unattended objects may be implicitly registered, but recent experiments suggest that attention is required to bind features, to represent three-dimensional structure, and to mediate awareness.

Addresses
*Department of Psychology, Princeton University, Princeton, New Jersey 08544-1010, USA;
e-mail: treisman@phoenix.princeton.edu
†Department of Brain and Cognitive Sciences, E10-243, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, USA; e-mail: ngk@psyche.mit.edu

Current Opinion in Neurobiology 1998, **8**:218–226

http://biomednet.com/elecref/0959438800800218

© Current Biology Ltd ISSN 0959-4388

Abbreviations

ERP	event-related potential
fMRI	functional magnetic resonance imaging
IT	inferotemporal cortex

Introduction

It is usually assumed that perception is mediated by specific patterns of neural activity that encode a selective description of what is seen, distinguishing it from other similar sights. When we perceive an object, we may form multiple representations, each specialized for a different purpose and therefore selecting different properties to encode at different levels of detail. There is empirical evidence supporting the existence of six different types of object representation. First, representation as an 'object token' — a conscious viewpoint-dependent representation of the object as currently seen. Second, as a 'structural description' — a non-visually-conscious object-centered representation from which the object's appearance from other angles and distances can be predicted. Third, as an 'object type' — a recognition of the object's identity (e.g. a banana) or membership in one or more stored categories. Fourth, a representation based on further knowledge associated with the category (such as the fact that the banana can be peeled and what it will taste like). Fifth, a representation that includes a specification of its emotional and motivational significance to the observer. Sixth, an 'action-centered description', specifying its "affordances" [1], that is, the properties we need in order to program appropriate motor responses to it, such as its location, size and shape relative to our hands. These different representations are probably formed in an interactive fashion, with prior knowledge facilitating the extraction of likely features and structure, and *vice versa.*

Evidence suggests that the first four types of encoding depend primarily on the ventral (occipitotemporal) pathway, the fifth on connections to the amygdala, and the sixth on the dorsal (occipitoparietal) pathway; however, object tokens have also been equated with action-centered descriptions [2•]. Dorsal representations appear to be distinct from those that mediate conscious perception; for example, grasping is unaffected by the Titchener size illusion [3]. Emotional responses can also be evoked without conscious recognition (e.g. see [4••]). Object recognition models differ over whether the type or identity of objects is accessed from the view-dependent token or from a structural description; in some cases, it may also be accessed directly from simpler features.

The goal of perception is to account for systematic patterning of the retinal image, attributing features to their real world sources in objects and in the current viewing conditions. In order to achieve these representations, multiple sources of information are used, such as color, luminance, texture, relative size, dynamic cues from motion and transformations, and stereo depth; however, the most important is typically shape. Many challenges arise in solving the inverse problem of retrieving the likely source of the retinal image: information about object boundaries is often incomplete and noisy; and three-dimensional objects are seen from multiple views, producing different two-dimensional projections on the retina, and objects in normal scenes are often partially occluded. The visual system has developed many heuristics for solving these problems. Continuity is assumed rather than random variation. Regularities in the image are attributed to regularities in the real world rather than to accidental coincidences. Different types of objects and different levels of specificity require diverse discriminations, making it likely that specialized modules have evolved, or developed through learning, to cope with the particular demands of tasks such as face recognition, reading, finding our way through places, manipulating tools, and identifying animals, plants, minerals and artifacts.

Research on object perception over the past year has made progress on a number of issues. Here, we will discuss recent advances in our understanding of the speed of object recognition, object types and tokens, and attention and awareness in object recognition. In addition, we will review evidence for cortical specializations for particular components of visual recognition.

The speed of object recognition

Evolutionary pressures have given high priority to speed of visual recognition, and there is both psychological and neuroscientific evidence that objects are discriminated within one or two hundred milliseconds. Behavioral studies have demonstrated that we can recognize up to eight or more objects per second, provided they are presented sequentially at fixation, making eye movements unnecessary [5]. Although rate measurements cannot tell us the absolute amount of time necessary for an individual object to be recognized, physiological recordings reveal the latency at which the two stimulus classes begin to be distinguished. Thorpe *et al.* [6••] have demonstrated significant differences in event-related brain potential (ERP) waveforms for viewing scenes containing animals versus scenes not containing animals at 150 ms after stimulus onset. Several other groups [7,8•,9–11] have found face-specific ERPs and magnetoencephalography (MEG) waveforms with latencies of 155–190 ms. DiGirolamo and Kanwisher (G DiGirolamo, NG Kanwisher, abstract in *Psychonom Soc* 1995, 305) found ERP differences for line drawings of familiar versus unfamiliar three-dimensional objects at 170 ms (see also [5]).

Parallel results were found in the stimulus selectivity of early responses of cells in inferotemporal (IT) cortex in macaques, initiated at latencies of 80–100 ms. On the basis that IT cells are selective for particular faces even in the first 50 ms of their response, Wallis and Rolls [12] conclude that "visual recognition can occur with largely feed-forward processing". The duration of responses by these face-selective cells was reduced from 250 ms to 25 ms by a backward mask appearing 20 ms after the onset of the face, a stimulus onset asynchrony at which human observers can still just recognize the face. The data suggest that "a cortical area can perform the computation necessary for the recognition of a visual stimulus in 20–30 ms". Thus, a consensus is developing that the critical processes involved in object recognition are remarkably fast, occurring within 100–200 ms of stimulus presentation. However, it may take another 100 ms for subsequent processes to bring this information into awareness.

Object tokens

How then does the visual system solve the problems of object perception with such impressive speed and accuracy? A first stage must be a preliminary segregation of the sensory data that form separate candidate objects. Even at this early level, familiarity can override bottom-up cues such as common region and connectedness, supporting an interactive cascade process in which "partial results of the segmentation process are sent to higher level object representations", which, in turn, guide the segmentation process [13•].

Kahneman, Treisman, and Gibbs [14] have proposed that conscious seeing is mediated by episodic 'object files' within which the object tokens defined earlier are constructed. Information about particular instances currently being viewed is selected from the sensory array, accumulates over time, and is 'bound' together in structured relations. Evidence for this claim came partly from the observation of 'object-specific' priming — that is, priming that occurs only, or more strongly, when the prime and probe are seen as a single object. This occurs even when they appear in different locations, if the object is seen in real or apparent motion between the two. Object-specific priming occurs between pictures and names when these are perceptually linked through the frames in which they appear (RD Gordon, DE Irwin, personal communication), suggesting that object files accumulate information not only about sensory features but also about more abstract identities. However, priming between synonyms or semantic associates is not object specific [15], that is, it occurs equally whether they are presented in the same perceptual object or in different objects. It appears that object files integrate object representations with their names, but maintain a distinct identity from other semantically associated objects. Priming at this level would be between object types rather than tokens. Irwin [16] has reviewed evidence on transsaccadic integration, suggesting that it is limited to about four object files.

A similar distinction between tokens and types has emerged from the study of repetition blindness, a failure to see a second token of the same type, which was attributed to refractoriness in attaching a new token to a recently instantiated type [17]. Recent research has further explored this idea. One role of object tokens is to maintain spatiotemporal continuity of objects across motion and change. Chun and Cavanagh [18••] confirmed that repetition blindness is greater when repeated items are seen to occur within the same apparent motion sequence and hence are integrated as the same perceived object. They suggest that perception is biased to minimize the number of different tokens formed to account for the sensory data. Objects that appear successively are linked whenever the spatial and temporal separations make this physically plausible. This generally gives veridical perception because in the real world, objects seldom appear from nowhere or suddenly vanish. Arnell and Jolicoeur [19] have demonstrated repetition blindness for novel objects for which no pre-existing representations existed. According to Kanwisher's account [17], this implies that a single presentation is sufficient to establish an object type to which new tokens will be matched.

The 'attentional blink' [20] describes a failure to detect the second of two different targets when it is presented soon after the first. Chun [21•] sees both repetition blindness and the attentional blink as failures of tokenization, although for different reasons, because they can be dissociated experimentally. Attentional blinks (reduced by target–distractor discriminability) reflect a general limit on the speed with which successive stimuli can be individuated as separate tokens, whereas repetition blindness (reduced by increased discriminability between the two repeated items) reflects the token-minimizing constraint described above. One important finding is that a blinked item can prime a subsequent related item [22], implying that the blink must, at least on some proportion of trials, follow rather than precede identification of the missed target. Subjects are also more likely to detect their own names in normally blinked locations [23], and the meanings of blinked stimuli have differential effects on the N400 ERP [24•]. The results imply that objects may activate matching types even before object files have been completed and stabilized for their conscious access. There may, however, be limits to the accuracy of identification achieved through this route.

When attention is divided across a display, the onset of four dots around a recently presented object often erases it from experience. This new phenomenon of attentional masking [25] may reflect the disruption of object tokens. Attentional masking can also occur with simultaneous onset but delayed offset of the mask (V Di Lollo, JT Enns, personal communication). The account proposed is that awareness depends on a match between re-entrant information and the current sensory input at early visual levels. A mismatch erases the initial tentative representation. "It is as though the visual system treats the trailing configuration as a transformation or replacement of the earlier one." Conversely, repetition blindness for locations (R Epstein, NG Kanwisher, abstract in *Psychonom Soc* 1996, 593) may result when the representation of an earlier-presented letter prevents the stable encoding of a subsequently presented letter appearing at the same location.

Attention and awareness in object perception

Attention seems, then, to be necessary for object tokens to mediate awareness. However, there is evidence (see [24•]) that objects can be identified without attention and awareness. If this is so, do the representations differ from those formed with attention? Activation (shown by brain-imaging) in specialized regions of cortex for processing faces [26] and visual motion [27] is reduced when subjects direct attention away from the faces or moving objects (respectively), even when eye movements are controlled to guarantee identical retinal stimulation (see also [28]), consistent with the effects of attention on single units in macaque visual cortex. Unattended objects are seldom reportable. However, priming studies suggest that their shapes can be implicitly registered [29,30••], although there are clear limits to the number of unattended objects that will prime [31]. Representations formed without attention may differ from those that receive attention: they appear to be viewpoint-dependent [32•], two-dimensional, with no interpretation of occlusion or amodal completion [30••]. On the other hand, in clinical neglect, the 'invisible' representations formed in a patient's neglected field include illusory contours and filled-in surfaces [33•], suggesting that neglect arises at stages of processing beyond those that are suppressed in normal selective attention. With more extreme inattention, little explicit information is available beyond simple features such as location, color, size, and gross numerosity; even these simple features may not be available, producing 'inattentional blindness' [34•]. Again, however, some implicit information is registered: unseen words may prime word fragment completion, and there is clear selectivity for emotionally important objects such as the person's own name and happy (but not sad) faces.

Binding of features to objects is often inaccurate unless attention is focused on the relevant locations [35]. Although the parietal lobes are usually thought to be associated with the processing of space and of action, they may also be intimately involved, through spatial attention, in binding and individuating object tokens in displays with more than one object present, and therefore in allowing conscious access to normal scenes [36]. Bilateral damage to the parietal lobes results in Balint's syndrome, with its accompanying simultanagnosia (i.e. an inability to see more than one object at a time) and dramatic failures in binding features correctly. Binding is also disrupted by transcranial magnetic stimulation of the parietal lobes [37]. Extinction following unilateral parietal lesions may result from a similar attentional problem [2•,38]; there is often evidence of implicit knowledge of extinguished items, perhaps through direct access from features to types. Individuating objects in 'crowded' displays is more difficult in the lower than upper visual field [39••], consistent with the greater parietal projection from the lower visual field.

Other studies have investigated what is perceived with attention distributed globally rather than specifically excluding the critical object. Global attention allows amodal completion for homogeneous displays [40]. Studies of visual search suggest that displays are automatically parsed into preattentive object files, acting as holders for collections of attributes but not for their structural relations (with the exception of the part-whole relation; [41•]). Wolfe [42] has collected surprising evidence that previously attended object tokens revert to a similar unstructured state once attention is withdrawn, concluding that "Vision exists in the present tense. It remembers nothing". Experiments on change detection in natural scenes show that focused, rather than global, attention is necessary for the identification of even quite dramatic changes between saccades ([43]; RD Gordon, DE Irwin, personal communication) or between alternating versions of a scene with one object changed, added, or deleted [44,45••,46]. Thus, attention seems critical at least for the explicit voluntary storage and retrieval of objects.

Striking dissociations between conscious access and implicit measures of object processing are found in patients with localized brain injuries. These dissociations suggest multiple systems, each forming representations of objects for specific purposes, only some of them conscious. For example, damage to the fusiform gyrus results in loss of conscious face recognition, or prosopagnosia, whereas emotional assessment depends on the amygdala, and may be selectively impaired in Capgras syndrome, where patients show normal face recognition but no emotional skin conductance responses [47]. Conversely, functional magnetic resonance imaging (fMRI) activation of the amygdala for emotionally expressive faces compared to neutral ones occurs even when the emotional expressions are masked and unseen [12]. Separate pathways may be responsible for conscious perception of objects and for the object representations that control actions, including the metric information necessary for grasping and manipulating [3]. For example, patient D.F. has severe agnosia as a result of damage in ventral visual areas, but can still manipulate objects appropriately, presumably through an intact dorsal route. Survival of action-related object coding has also been shown by a hemianopic patient in his blind field [48]. Another patient, with damage in the ventral route, shows a striking dissociation in expressing his perceptual knowledge, interpreting a picture of a clarinet verbally as "Perhaps a pencil" while at the same time his fingers clearly mimic playing a clarinet (D Margolin *et al.*, abstract in *J Clin Exp Neuropsychol* 1985, 6). Recent findings with patient D.F. suggest, however, that shape processing in the dorsal route may be restricted to measures of orientation, size and motion [49]. Positron emission tomography (PET) studies have also failed to find the sharp dissociation between areas involved in grasping and in perceptual matching that would be predicted [50] for a complete segregation of perceptual and action-based processes.

Object types

Formal theories of object perception have dealt primarily with object recognition — that is, the identification of object types, rather than the formation of object tokens. They fall into two classes: those that base recognition on a structural description specifying parts and their relationships (e.g. see [51]), and those that use more holistic viewpoint-dependent representations [52–55]. Structural descriptions specify the relations between volumetric parts or 'geons' (e.g. 'above', 'smaller than', or 'perpendicular to'), which, in turn, are defined by features signaling their cross section, axis shape, rough aspect ratio and whether they are truncated. View-dependent models differ in how they solve the recognition problem for novel views, whether by interpolation between stored views [56], by 'blurred' template-matching [55,57], by linear combinations of stored views [58], or by mental rotation [59].

The debate between those supporting the 'structural descriptions' model versus those supporting the view-dependent models continued over the past year; recent evidence suggests that both accounts play a role and clarifies the conditions in which each may be used. View-based representations predict the observed specificity of learning, with gradients of generalization around the particular views experienced [60•], even when the objects were novel and clearly composed of geons. Learned views were shown also to influence the appearance of an object in motion, determining whether or not it was seen as rigid [61•]. Apparent motion between rotated views of novel objects demonstrated the psychological reality of an induced interpolation process [62••]: both intermediate views and views just beyond the second view were primed, but not views that preceded the first. Priming was abolished when the interval between the two views was too long to induce apparent motion.

Outside the laboratory, we normally experience dynamically changing views of objects, through either our own motion or the motion of the object. This could be an important perceptual learning mechanism in object recognition. Physiological evidence consistent with the view-based account comes from single-unit recordings in IT of macaque monkeys [63], showing neurons that respond selectively to different views of novel objects, firing most to one view, with a gradually decreasing response as the object rotates away from the preferred view. The results closely parallel the generalization gradients shown in human priming experiments. Only a few cells were found to respond selectively to one object regardless of the view from which it was seen. The existence of IT columns systematically coding similar object components [64] may contribute to perceived invariance across different views and locations of the same object.

The geon-based account has also received considerable empirical support (reviewed in [51]). Its proponents have shown that simple filters cannot account for the types of errors that humans make [65]. In recent applied research on distinguishing military vehicles in infra-red photos [66], a geon-based conditional tree predicted perceptual confusions much better than a deformable template account [67], although the latter did better with faces. Identification can be dissociated from the conscious perception of orientation: two studies have reported that three patients with right or bilateral parietal lesions correctly identified objects or letters without being able to name or copy their orientations [39••,68].

Studies comparing priming and recognition also suggest that both structural descriptions and more specific viewpoint-dependent representations are retained in visual memory. Whereas implicit priming suggests invariance across changes in location, color, orientation and size, explicit tests of recognition show much more specificity [69,70]. Srinivas [71] confirmed that for attended objects, priming was invariant with left–right orientation, although it was reduced by changes in size if the task made size relevant. Short-term matching of temporally contiguous stimuli suggested equivalence across views and seems, like priming, to tap an invariant representation [72]. Similarly, repetition blindness for pictures across very short lags shows complete invariance to size, orientation, and viewpoint [73].

The general conclusion is emerging that both mechanisms are used at different stages of processing, and/or on different classes of objects [74]. A recent model of object perception [75•] combines an initial view-dependent representation of geons followed by a 'dynamic binding' process that creates a structural description of their relations while retaining their independence as separable parts. Distinctive features or parts contribute when they are present, ruling out a pure template-matching mechanism [76]. Structural descriptions based on geons may be good for accessing basic level categories for the many objects that are naturally decomposable into distinct parts, but cannot succeed for discriminations within classes of objects that share parts and differ only in metric properties. Faces are a clear case where more holistic template models can capture subtle differences between instances, all of which share the same basic geon structure. The task may also play a part in determining the kind of analysis that is carried out; in speeded naming, subtle differences within categories are irrelevant, whereas in same-different matching tasks, metric comparison processes may be invoked. Finally, there may also be a shift with experience. Experts with extensive encounters with different instances may base their recognition on matching to multiple stored views, giving the impression of invariant representation. Gauthier and Tarr [77] gave subjects prolonged training in recognizing novel objects with shared parts ('greebles') varying along a few specified dimensions, and found that with experience, they became sensitive to configural qualities as well as to specific features.

Striking examples of perceptual plasticity in form perception have recently been reported. Implicit traces can mediate priming for novel nonsense shapes across several weeks delay after a single presentation [29,30••]. Analogously, rapid learning has been demonstrated in single-unit recordings in monkeys [78••]: when exposed to binarized faces, face-sensitive cells gave little response, but after the animal was given a few seconds of viewing gray-scale versions of the same faces, the cells responded equally to the binarized images. A similar result has been shown in humans using fMRI [79]. Logothetis and Pauls [80] found IT cells that, with experience, became selective for novel objects that previously did not excite them; these cells also showed some viewpoint dependency. Other examples of very rapid perceptual learning have been reported [81,82], and a reverse hierarchical system, to account for perceptual learning effects, has been proposed [81].

Cortical specializations for visual recognition

Evidence from neuropsychology, cognitive psychology, and brain imaging suggests that the remarkable speed and accuracy of visual recognition are achieved through the operation of a set of special-purpose mechanisms instantiated in at least partially segregated brain regions.

The shape of an object is usually the most important cue to its identity. Humphrey *et al.* [83] have reported that although patient D.F. could discriminate the apparent three-dimensional structure of shapes defined by shading gradients, she was unable to discriminate similar shapes in which the edges were depicted as luminance discontinuities or lines, suggesting that extracting shape from shading is a distinct process from extracting shape from edges. Humphrey *et al.* [84] used fMRI on normal subjects to show that shape-from-shading processes produce activation in primary visual cortex. Evidence from a variety of sources indicates that a large region of lateral occipital cortex just anterior to retinotopic cortex (but posterior to the visual motion area MT) responds more strongly to stimuli depicting shapes than to stimuli with similar low-level features that do not depict shapes [85,86]. Common areas within this lateral occipital region are activated by structure from motion, structure from texture, and luminance silhouettes (K Grill-Spector *et al.*, *Soc Neurosci Abstr* 1997, 23:868.12). Whereas simple forms defined by differences in luminance, color, or direction of motion largely activate regions in retinotopic cortex, stereoscopic and illusory-contour displays primarily activate the lateral occipital region (J Mendola *et al.*, *Soc Neurosci Abstr* 1997, 23:550.11). Thus, although some of the necessary computations take place in retinotopic cortex, lateral occipital cortex may contain regions specialized for some aspect of visual shape analysis. However, three important questions remain to be answered. First, what specific aspect of shape analysis is computed in this region (e.g. edge extraction or figure–ground segmentation or implied depth)? Second, would the areas activated by different shape cues in different studies overlap exactly if run on an individual subject, or would different but adjacent regions within lateral occipital cortex be activated by different shape cues? Third, might the activations, in part, reflect attentional artifacts, as all of the stimuli depicting shapes are likely to be more attention-capturing than the control stimuli depicting random texture fields?

Shape analysis can be carried out on virtually any visually presented object. Other processing mechanisms appear to be recruited by exemplars of just one stimulus class. Evidence has been presented for special-purpose cortical machinery for the recognition of words, tools, biological motion [87,88], and other object categories. In the past year, the already strong evidence for the case of face perception [89] has received further support. First, a recent study of patient C.K. [90••] presents perhaps the most compelling evidence that face and object recognition are separated at a relatively early stage. C.K.'s general visual abilities are drastically disrupted, and he has great difficulty recognizing objects and words, yet he is absolutely normal at face recognition. Second, intracranial recordings from epileptic patients have demonstrated single cells in the human hippocampus, amygdala, and entorhinal cortex that respond selectively to faces, particular facial expressions, or gender [91], or to familiar versus unfamiliar faces [91,92]. Third, human brain imaging studies have shown that regions within the fusiform gyrus are not only responsive to faces [93–95], but also respond in a highly specific fashion to faces compared to a wide range of other kinds of objects [96•,97].

The accumulating evidence for cortical specialization for specific components of visual recognition raises a number of important questions. Does this fine-grained specialization of function arise from experience-dependent self-organizing properties of cortex [98], or are cortical specializations innately specified? For the case of faces, this question is hard to answer because both experiential and evolutionary arguments are plausible. However, evidence for cortical specializations for visually presented words (T Polk *et al.*, *Soc Neurosci Abstr* 1996, 22:291.2) and letters (M Farah *et al.*, *Soc Neurosci Abstr* 1996, 22:291.1) suggests that experience may be sufficient, at least in some cases. Further evidence for experience-induced cortical specialization comes from Logothetis and Pauls [80], who found that after training monkeys with a specific class of stimuli, small regions in anterior IT (AIT) contained cells selectively responsive to these stimuli.

What are the implications of cortical specialization for theories of visual recognition? Does the selectivity of certain cortical areas for the recognition of different stimulus classes imply that qualitatively distinct processing mechanisms are involved in each? Connectionist researchers have noted the computational efficiency gained by the decomposition of a complex function into natural parts [99]. Cortical specializations for components of visual recognition are plausible candidates for such task decomposition. On the other hand, a shallower account might argue that cells selective for particular specialized features happen to land together in a cortical surface organized by feature columns [100]. Support for this interpretation comes from a recent report that localized regions in human extrastriate cortex are selectively responsive to apparently arbitrary categories, such as chairs and houses (A Ishai *et al.*, abstract in *Neuroimage* 1997, 5.4:S149). It remains for future research to determine whether the functional organization of visual recognition is better characterized as 'shallow specialization' or a deeper form of modularity in which a small number of functionally specific regions each carries out a qualitatively distinct computation in the service of an evolutionarily or experientially fundamental visual process.

Conclusions

Behavioral and physiological work has provided a rich characterization of the multiple representations that are extracted in the first quarter of a second of viewing a complex visual stimulus. Both structural descriptions and viewpoint-dependent representations sufficient for discriminating between objects are extracted within about 200 ms. The phenomena of repetition blindness, attentional blink, attentional masking, and inattentional blindness reveal some of the heuristics by which the visual system decides which of these representations to incorporate into the developing stable representation of visual experience. Functional imaging and patient studies complement this picture by revealing some of the fundamental components of the machinery of visual recognition. Persuasive evidence exists for a special-purpose 'module' mediating face perception, and ongoing research suggests the existence of several other dissociable components of object perception.

Acknowledgements

Preparation of this review was supported by a National Science Foundation grant #SBR-9511633 to AM Treisman, and a Human Frontiers Grant and National Institute of Mental Health grant 56037 to NG Kanwisher.

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Gibson JJ: *The Ecological Approach to Visual Perception.* Boston: Houghton Mifflin; 1979.

2. • Driver J: **What can visual neglect and extinction reveal about the extent of 'preattentive' processing?** In *Converging Operations in the Study of Visual Selective Attention.* Edited by Kramer AF, Coles M, Logan GD. Washington, DC: American Psychological Association; 1996:193-223.

Reviews evidence of implicit knowledge of stimuli in neglect and extinction resulting from brain damage, and suggests that the function of the ventral pathway might be primarily to recognize object types, whereas recognition of object tokens may depend on the dorsal pathway. If tokens are necessary for conscious experience, priming of object types in the absence of object tokens could explain the observed phenomena of neglect.

3. Milner AD, Goodale MA: *The Visual Brain in Action.* Oxford: Oxford University Press; 1995.

4. •• Whalen PJ, Rauch SL, Etcoff NL, McInery SC, Lee MB, Jenike MA: **Masked presentations of emotional expressions modulate amygdala activity without explicit knowledge.** *J Neurosci* 1998, **18**:411-418.

The authors report that the fMRI response from the amygdala to unseen emotionally expressive faces provides strong evidence for high-level perception without awareness.

5. Potter MC: **Short term conceptual memory for pictures.** *J Exp Psychol [Hum Learn Mem]* 1976, **2**:509-522.

6. •• Thorpe S, Fize D, Marlot C: **Speed of processing in the human visual system.** *Nature* 1996, **381**:520-522.

Subjects decided whether each of 4000 previously unseen photographs contained an animal or not; ERPs specific to negative responses occurred at 150 ms after stimulus onset, suggesting that much of human object recognition is based on feed-forward mechanisms.

7. Jeffreys DA: **Evoked potential studies of face and object processing.** *Vis Cogn* 1996, **3**:1-38.

8. • Bentin S, Allison T, Puce A, Perez E, McCarthy G: **Electrophysiological studies of face perceptions in humans.** *J Cogn Neurosci* 1996, **8**:551-565.

Face-specific ERPs at 172 ms (N200) were delayed but of the same amplitude for inverted versus upright faces. The ERPs were larger for eyes alone than for whole faces. Neither animal faces nor human hands elicited N200s.

9. Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD: **Face recognition in human extrastriate cortex.** *J Neurophysiol* 1994, **71**:821-825.

10. Sams M, Hietanen JK, Hari R, Ilmoniemi RJ, Lounasmaa OV: **Face-specific responses from the human inferior occipito-temporal cortex.** *Neuroscience* 1997, **1**:49-55.

11. Schendan HE, Ganis G, Kutas M: **Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms.** *Psychophysiology* 1998, in press.

12. Wallis G, Rolls ET: **Invariant face and object recognition in the visual system.** *Prog Neurobiol* 1997, **51**:167-194.

13. • Vecera SP, Farah MJ: **Is visual image segmentation a bottom-up or an interactive process?** *Percept Psychophys* 1997, **59**:1280-1296.

Explored object segmentation process in which the subjects were given the task of deciding whether two Xs were on the same one of two overlapping shapes. The subjects showed better performance with familiar shapes (letters), even when sensory cues such as common region and connectedness favored unfamiliar shapes. The results support an interactive cascade model of segmentation, in which partial bottom-up information is sent to higher level object representations that, in turn, feed back to guide the segmentation process.

14. Kahneman D, Treisman A, Gibbs B: **The reviewing of object files: object-specific integration of information.** *Cogn Psychol* 1992, **24**:175-219.

15. Gordon RD, Irwin DE: **What's in an object file? Evidence from priming studies.** *Percept Psychophys* 1996, **58**:1260-1277.

16. Irwin DE: **Integration and accumulation of information across saccadic eye movements.** In *Attention and Performance,* vol XVI: *Information Integration in Perception and Communication.* Edited by McClelland J, Inui T. Cambridge, Massachusetts: MIT Press; 1996:125-156.

17. Kanwisher N: **Repetition blindness: type recognition without token individuation.** *Cognition* 1987, **27**:117-143.

18. •• Chun MM, Cavanagh P: **Seeing two as one: linking apparent motion and repetition blindness.** *Psychol Sci* 1997, **8**:74-78.

In this cleverly designed study, two letters were made to appear as part of the same versus different motion streams by varying only the trajectories of nontarget items; the data are inconsistent with most alternative accounts and argue strongly for the token individuation explanation of repetition blindness.

19. Arnell KM, Jolicoeur P: **Repetition blindness for pseudoobject pictures.** *J Exp Psychol [Hum Percept Perform]* 1997, **23**:999-1013.

20. Raymond JE, Shapiro KL, Arnell KM: **Temporary suppression of visual processing in an RSVP task: an attentional blink?** *J Exp Psychol [Hum Percept Perform]* 1992, **18**:849-860.

21. • Chun MM: **Types and tokens in visual processing: a double dissociation between the attentional blink and repetition blindness.** *J Exp Psychol [Hum Percept Perform]* 1997, **23**:738-755.

Shows that different factors affect the attentional blink (item discriminability) and repetition blindness (episodic distinctiveness of repeated targets), suggesting that the two reflect different limitations on the formation of object tokens.

22. Shapiro K, Driver J, Ward R, Sorensen R: **Priming from the attentional blink: a failure to extract visual tokens but not visual types.** *Psychol Sci* 1997, **8**:95-100.

23. Shapiro KL, Caldwell J, Sorensen RE: **Personal names and the attentional blink: a visual 'cocktail party' effect.** *J Exp Psychol [Hum Percept Perform]* 1997, **23**:504-514.

24. • Luck SJ, Vogel EK, Shapiro KL: **Word meanings can be accessed but not reported during the attentional blink.** *Nature* 1996, **382**:616-618.

Target words either related or unrelated to a context word were presented at several intervals after another target. Although accuracy of relatedness judgment fell sharply for targets appearing 166 ms (but not 0 or 500 ms) after the first target (the 'attentional blink'), the N400 related–unrelated difference wave was not affected by lag. Thus, even though the word meaning is not available, it was apparently extracted, suggesting a postperceptual locus of the attentional blink.

25. Enns JT, Di Lollo V: **Attentional masking: visual interference by object substitution.** *Psychol Sci* 1997, **8**:135-139.

26. Wojciulik E, Kanwisher N, Driver J: **Covert visual attention modulates face-specific activity in the human fusiform gyrus: an fMRI study.** *J Neurophysiol* 1998, **79**:1574-1579.

27. O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL: **Voluntary attention modulates fMRI activity in human MT-MST.** *Neuron* 1997, **18**:591-598.

28. Clark VP, Parasuraman R, Keil K, Kulansky R, Fannon S, Maisog JM, Ungerleider LG, Haxby JV: **Selective attention to face identity and color studied with fMRI.** *Hum Brain Mapp* 1997, **5**:293-291.

29. DeSchepper B, Treisman A: **Visual memory for novel shapes: implicit coding without attention.** *J Exp Psychol [Learn Mem Cogn]* 1996, **22**:27-47.

30. •• Treisman A, DeSchepper B: **Object tokens attention and visual memory.** In *Attention and Performance,* vol XVI: *Information Integration in Perception and Communication.* Edited by Inui T, McClelland J. Cambridge, Massachusetts: MIT Press; 1996:15-46.

Explored the implicit memory representations that are formed for unattended novel objects and events. Using a negative priming paradigm, showed that long-lasting memory traces could be formed in a single trial, independently of attention. The traces are stored at a level that precedes the allocation of a shared contour to the figure rather than the ground, and the interpretation of occlusion. The results suggest a surprising combination of plasticity and permanence in the visual system.

31. Neumann E, DeSchepper BG: **An inhibition-based fan effect: evidence for an active suppression mechanism in selective attention.** *Can J Psychol* 1992, **46**:1-40.

32. • Stankiewicz BJ, Hummel JE: **The role of attention in priming for left-right reflections of object image: evidence for a dual representation of object shape.** *J Exp Psychol [Hum Percept Perform]* 1998, in press.

Measured priming from attended and from unattended pictures. Found evidence for two separate processes, one viewpoint-dependent but independent of attention and one requiring attention, invariant with reflection, and longer lasting. The authors interpret the results in terms of the two representations generated in their model.

33. • Mattingley JB, Davis G, Driver J: **Preattentive filling-in of visual surfaces in parietal extinction.** *Science* 1997, **275**:671-674.

The authors found that extinction patient V.R., who has right parietal damage, is more likely to detect removal of segments in disks in the contralesional field when they are combined with those on the ipsilesional side to create an illusory surface. The results suggest that object surfaces are created preattentively and that visual extinction affects only later conscious levels of processing.

34. • Mack A, Rock I: *Inattentional Blindness: Perception Without Attention.* Cambridge, Massachusetts: MIT Press; 1998.

Reports a large number of studies using a paradigm to explore how much information is extracted from ignored stimuli when attention is focused elsewhere and the ignored stimuli are completely unexpected. Although only simple features appear to be explicitly reportable, there is evidence of implicit processing of words and of pictures with emotional significance. The conclusion drawn is that attention selects only after considerable perceptual analysis, "to highlight relevant stimulus information" for conscious awareness.

35. Treisman A, Gelade G: **A feature integration theory of attention.** *Cogn Psychol* 1980, **12**:97-136.

36. Robertson L, Treisman A, Friedman-Hill S, Grabowecky M: **The interaction of spatial and object pathways: evidence from Balint's syndrome.** *J Cogn Neurosci* 1997, **9**:254-276.

37. Ashbridge E, Walsh V, Cowey A: **Temporal aspects of visual search studied by transcranial magnetic stimulation.** *Neuropsychologia* 1997, **35**:1121-1131.

38. Baylis GC, Driver J, Rafal RD: **Visual extinction and stimulus repetition.** *J Cogn Neurosci* 1993, **5**:453-466.

39. •• He S, Cavanagh P, Intrilligator J: **Attentional resolution and the locus of visual awareness.** *Nature* 1996, **383**:334-337.

The authors demonstrate orientation-specific adaptation effects under conditions that do not permit awareness of the orientation (flanking by other similar gratings); this 'crowding' effect occurs when different objects cannot be attentionally resolved. Attentional resolution is greater in the lower than upper visual field, and acts as a filter restricting the availability of visual information to awareness.

40. Rensink RA, Enns JT: **An object completion process in early vision.** *Vision Res* 1998, in press.

41. • Wolfe JM, Bennett SC: **Preattentive object files: shapeless bundles of basic features.** *Vision Res* 1997, **37**:25-44.

This extensive set of experiments on visual search suggests that preattentive processing sets up an array of object tokens to which the relevant features have been assigned, but without any specification of their structured relations except for the part-whole assignment. Attention is required to determine the arrangement and the global shape of the elements in the search array.

42. Wolfe JM: **Inattentional amnesia.** In *Fleeting Memories.* Edited by Coltheart V. Cambridge, Massachusetts: MIT Press; 1998:in press.

43. McConkie GW, Currie C: **Visual stability across saccades while viewing complex pictures.** *J Exp Psychol [Hum Percept Perform]* 1996, **22**:563-581.

44. Rensink RA, O'Regan JK, Clark JJ: **To see or not to see: the need for attention to perceive changes in scenes.** *Psychol Sci* 1997, **8**:368-373.

45. Simons DJ: **In sight out of mind.** *Psychol Sci* 1996,
•• **7**:301-305.

In the most striking of many similar demonstrations, subjects approached by a stranger asking directions do not notice when the stranger is replaced by a completely different person (while two confederates carry a door between the two conversants). Apparently, the contents of current awareness are less detailed than introspection suggests.

46. Simons DJ, Levin DT: **Change blindness.** *Trends Cogn Sci* 1997, **1**:261-267.

47. Ellis HD, Young AW, Quayle AH, De Pauw KW: **Reduced autonomic responses to faces in Capgras delusion.** *Proc R Soc Lond [Biol]* 1997, **264**:1085-1092.

48. Perenin M-T, Rossetti Y: **Grasping without form discrimination in a hemianopic field.** *Neuroreport* 1996, **7**:793-797.

49. Carey DP, Harvey M, Milner AD: **Visuomotor sensitivity for shape and orientation in a patient with visual form agnosia.** *Neuropsychologia* 1996, **34**:329-337.

50. Faillenot I, Toni I, Decety J, Gregoire MC, Jeannerod M: **Visual pathways for object-oriented action and object recognition: functional anatomy with PET.** *Cereb Cortex* 1997, **7**:77-85.

51. Biederman I: **Recognition by components: a theory of human image understanding.** *Psychol Rev* 1987, **94**:115-147.

52. Tarr MJ, Bulthoff HH: **Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein 1993.** *J Exp Psychol [Hum Percept Perform]* 1995, **21**:1494-1505.

53. Bulthoff HH, Edelman SY, Tarr MJ: **How are three-dimensional objects represented in the brain?** *Cereb Cortex* 1995, **3**:247-260.

54. Vetter T, Hurlbert A, Poggio T: **View-based models of 3D object recognition: invariance to imaging transforms.** *Cereb Cortex* 1995, **3**:261-269.

55. Poggio T, Edelman S: **A network that learns to recognize three-dimensional objects.** *Nature* 1990, **343**:263-266.

56. Bulthoff HH, Edelman S: **Psychophysical support for a two-dimensional view interpolation theory of object recognition.** *Proc Natl Acad Sci USA* 1992, **89**:60-64.

57. Poggio TA, Hurlbert A: **Observations on cortical mechanisms for object recognition and learning.** In *Large-Scale Neuronal Theories of the Brain.* Edited by Koch C, Davis JL. Cambridge, Massachusetts: MIT Press; 1994:153-182.

58. Ullman S, Basri R: **Recognition by linear combinations of models.** *IEEE Trans Patt Anal Mach Intel* 1991, **13**:992-1006.

59. Tarr MJ, Pinker S: **Orientation-dependent mechanisms in shape recognition: further issues.** *Psychol Sci* 1991, **2**:207-209.

60. Hayward WG, Tarr MJ: **Testing conditions for viewpoint**
• **invariance in object recognition.** *J Exp Psychol [Hum Percept Perform]* 1997, **23**:1511-1521.

The authors used a sequential same–different matching or a naming paradigm to explore the degree of viewpoint invariance in coding of one- and two-part novel objects. They found no difference in latency up to 10 degrees, then progressive increase up to 30 degrees, questioning the achievement of viewpoint-independent structural descriptions, even for geon-based objects.

61. Sinha P, Poggio T: **I think I know that face.** *Nature* 1996,
• **384**:404.

Describes a test for learning in the perception of three-dimensional structure, based on the perception of rigidity or nonrigidity. A rigid wire object is shown rocking through 20 degrees, followed by a test object with the same mean-angle projection, also rocking. The learned interpretation of the first is imposed on the second, resulting in a nonrigid interpretation, consistent with the suggestions of view-based representation for recognizing three-dimensional structures.

62. Kourtzi Z, Shiffrar M: **One-shot view invariance in a moving**
•• **world.** *Psychol Sci* 1997, **8**:461-466.

The authors used a priming paradigm to probe the representation that is formed when an object is seen rotating in apparent motion. They found view-dependence of priming, but generalization within and just beyond the path of the apparent motion, as though the visual system links successive orientations when they are created by apparent motion and extrapolates some distance beyond the final view.

63. Logothetis NK, Sheinberg DL: **Recognition and representation of visual objects in primates: psychophysics and physiology.** In *The Mind-Brain Continuum.* Edited by Llinas RR, Churchland PS. Cambridge, Massachusetts: MIT Press; 1996:147-172.

64. Fujita I, Tanaka K, Ito M, Chang K: **Columns for visual features of objects in monkey inferotemporal cortex.** *Nature* 1992, **360**:343-346.

65. Fiser J, Biederman I, Cooper EE: **To what extent can matching algorithms based on direct outputs of spatial filters account for human object recognition?** *Spatial Vision* 1996, **10**:237-271.

66. O'Kane BL, Biederman I, Cooper EE, Nystrom B: **An account of object identification confusions.** *J Exp Psychol [Applied]* 1997, **3**:21-41.

67. Lades M, Vortbruggen JC, Buhmann J, Lange J, Von der Malsburg C: **Distortion invariant object recognition in the dynamic link architecture.** *IEEE Trans Comput* 1993, **42**:300-311.

68. Turnbull OH, Beschin N, Della Sala S: **Agnosia for object orientation: implications for theories of object recognition.** *Neuropsychologia* 1997, **35**:153-163.

69. Biederman I, Cooper EE: **Size invariance in visual object priming.** *J Exp Psychol [Hum Percept Perform]* 1992, **18**:121-133.

70. Cooper LA: **Probing the nature of the mental representation of visual objects: evidence from cognitive dissociations.** In *Cognitive Approaches to Human Perception.* Edited by Ballesteros S. Hillsdale, New Jersey: Erlbaum; 1994:199-221.

71. Srinivas K: **Size and reflection effects in priming: a test of transfer-appropriate processing.** *Mem Cogn* 1996, **244**:441-452.

72. Srinivas K: **Representation of rotated objects in explicit and implicit memory.** *J Exp Psychol [Learn Mem Cogn]* 1995, **21**:1019-1036.

73. Kanwisher N, Yin C, Wojciulik E: **Repetition blindness for pictures: evidence for the rapid computation of abstract visual descriptions.** In *Fleeting Memories.* Edited by Coltheart V. Cambridge, Massachusetts: MIT Press; 1998:in press.

74. Logothetis NK, Sheinberg DL: **Visual object recognition.** *Annu Rev Neurosci* 1996, **19**:577-621.

75. Hummel JE, Stankiewicz BJ: **An architecture for rapid**
• **hierarchical structural description.** In *Attention and Performance,* vol 16. Edited by Inui T, McClelland J. Cambridge, Massachusetts: MIT Press; 1996:93-121.

Describes a model for object recognition that represents shapes in a hybrid fashion early on by forming a fast viewpoint-dependent estimate of object identity and more slowly by using synchronized firing to establish a structural description.

76. Tarr MJ, Bulthoff HH, Zabinski M, Blanz V: **To what extent do unique parts influence recognition across changes in viewpoint?** *Psychol Sci* 1997, **8**:282-289.

77. Gauthier I, Tarr MJ: **Becoming a 'Greeble' expert: exploring mechanisms for face recognition.** *Vision Res* 1997, **37**:1673-1682.

78. Tovee MJ, Rolls ET, Ramachandran VS: **Rapid visual learning in**
•• **neurones of the primate temporal visual cortex.** *Neuroreport* 1996, **7**:2757-2760.

Recorded from 21 face-selective neurons in the superior temporal sulcus and area IT in monkeys. Seven of the 21 cells showed a large increase in response to binarized (hard to recognize) faces after just ten presentations of the full grey-scale versions. The increase was specific to the particular face shown, suggesting rapid learning in single neurons.

79. Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RSJ, Friston KJ: **How the brain learns to see objects and faces in an impoverished context.** *Nature* 1997, **389**:596-599.

80. Logothetis NK, Pauls J: **Psychophysical and physiological evidence for viewer-centered object representations in primates.** *Cereb Cortex* 1995, **3**:270-288.

81. Ahissar M, Hochstein S: **Task difficulty and the specificity of perceptual learning.** *Nature* 1997, **387**:401-406.

82. Rubin N, Nakayama K, Shapley R: **Abrupt learning and retinal size specificity in illusory-contour perception.** *Curr Biol* 1997, **7**:461-467.

83. Humphrey KG, Symons LA, Herbert AM, Goodale MA: **A neurological dissociation between shape from shading and shape from edges.** *Behav Brain Res* 1996, **76**:117-125.

84. Humphrey KG, Goodale MA, Bowen CV, Gati JS, Vilis T, Rutt BK, Menon RS: **Differences in perceived shape from shading correlate with activity in early visual areas.** *Curr Biol* 1997, **7**:144-147.

85. Malach R, Reppas JB, Benson RB, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RBH: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex.** *Proc Natl Acad Sci USA* 1995, **92**:8135-8138.

86. Kanwisher N, Woods R, Ioacoboni M, Mazziotta J: **A locus in human extrastriate cortex for visual shape analysis.** *J Cogn Neurosci* 1996, **91**:133-142.

87. Bonda E, Petrides M, Ostry D, Evans A: **Specific involvement of human parietal systems and the amygdala in the perception of biological motion.** *J Neurosci* 1996, **16**:3737-3744.

88. McLeod W, Dittrich J, Driver J, Perrett D, Zihl J: **Preserved and impaired detection of structure from motion by a 'motion-blind' patient.** *Vis Cog* 1996, **3**:363-392.

89. Puce A, Allison T, Spencer SS, Spencer DD, McCarthy G: **Comparison of cortical activation evoked by faces measured by intracranial field potentials and functional MRI: two case studies.** *Hum Brain Mapp* 1997, **5**:298-305.

90. •• Moscovitch M, Winocur G, Behrmann M: **What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition.** *J Cogn Neurosci* 1997, **9**:555-604.

Investigated a relatively isolated face processing mechanism in patient C.K.; found a reduction in accuracy for identification of upside down or configurally disrupted ('fractured') faces that was much larger than the cost seen in normal subjects. Inferred that normal face recognition depends on both orientation-sensitive face-specific mechanisms and a part-based object recognition system that is damaged in patient C.K..

91. Fried I, MacDonald K, Wilson C: **Single neuron activity in human hippocampus and amygdala during recognition of faces and objects.** *Neuron* 1997, **18**:753-765.

92. Seeck M, Michel CM, Mainwaring N, Cosgrove R, Blume H, Ives J, Landis T, Schomer DL: **Evidence for rapid face recognition from human scalp and intracranial electrodes.** *Cog Neurosci Neuropsychol* 1997, **8**:2749-2754.

93. Puce A, Allison T, Spencer SS, Spencer DD, McCarthy G: **Comparison of cortical activation evoked by faces measured by intracranial field potentials and functional MRI: two case studies.** *Hum Brain Mapp* 1997, **5**:298-305.

94. Courtney SM, Ungerleider LG: **What fMRI has taught us about human vision.** *Curr Opin Neurobiol* 1997, **7**:554-561.

95. Clark VP, Keil K, Maisog JM, Courtney S, Ungerleider S, Haxby JV: **Functional magnetic resonance imaging of human visual cortex during face matching: a comparison with positron emission tomography.** *Neuroimage* 1996, **4**:1-15.

96. • Kanwisher N, McDermott J, Chun M: **The fusiform face area: a module in human extrastriate cortex specialized for face perception.** *J Neurosci* 1997, **17**:4302-4311.

The authors used multiple fMRI tests of the same cortical region (the fusiform face area) within individual subjects to demonstrate a high degree of selectivity of this region for faces and to rule out alternative accounts of the face activation (e.g. luminance confounds, subordinate-level categorization of any stimulus class, attentional biases toward faces, etc.).

97. McCarthy G, Puce A, Gore J, Allison T: **Face-specific processing in the human fusiform gyrus.** *J Cogn Neurosci* 1997, **9**:605-610.

98. Jacobs RA: **Nature, nurture, and the developmental of functional specializations: a computational approach.** *Psych Bull Rev* 1997, **4**:299-309.

99. Jacobs RA, Jordan MI, Barto AG: **Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks.** *Cogn Sci* 1991, **15**:219-250.

100. Tanaka K: **Mechanisms of visual object recognition: monkey and human studies.** *Curr Opin Neurobiol* 1997, **7**:523-529.