What fMRI has taught us about human vision
Susan M Courtney* and Leslie G Ungerleider†

The recent application of functional magnetic resonance imaging (fMRI) to visual studies has begun to elucidate how the human visual system is anatomically and functionally organized. Bottom-up hierarchical processing among visual cortical areas has been revealed in experiments that have correlated brain activations with human perceptual experience. Top-down modulation of activity within visual cortical areas has been demonstrated through studies of higher cognitive processes such as attention and memory.

Addresses
Laboratory of Brain and Cognition, Building 10, Room 4C104, 10 Center Drive, Bethesda, Maryland 20892-1366, USA
*E-mail: susan.courtney@nih.gov
†E-mail: lgu@ltunihl.nih.gov

Current Opinion in Neurobiology 1997; 7:554–561
http://biomednet.com/elecref/0959438800700554
© Current Biology Ltd ISSN 0959-4388

Abbreviations
fMRI functional magnetic resonance imaging
LO lateral occipital (area)
MT middle temporal (area)
PET positron emission tomography
V1 primary visual cortex

Introduction
In an earlier review [1], we summarized progress in understanding the organization and functional properties of visual areas in the human cortex from studies using positron emission tomography (PET). Here, we describe new insights gained from studies using functional magnetic resonance imaging (fMRI). As with any new technology, most initial fMRI experiments were devised to validate the method, and the findings largely confirmed those obtained with PET. However, in the past few years, fMRI studies have provided new and valuable information as a result of fMRI’s enhanced spatial and temporal resolution in comparison to PET and the ability both to study activations in single subjects and to study subjects longitudinally in repeated scan sessions. In this review, we describe progress in our understanding of three areas: first, homologies between visual cortical areas in monkeys and those in humans; second, functional specialization in the human cortex; and finally, bottom-up processing and top-down influences within the visual cortex.

Functional magnetic resonance imaging
Functional MRI is a noninvasive method for measuring hemodynamic responses to changes in neural activity in the brain. When neural activity increases, regional cerebral blood flow increases, causing the local concentration of deoxygenated hemoglobin to decrease [2]. This decrease results in a localized increase in the fMRI signal in the brain [3–5]. Therefore, fMRI signal intensity is correlated with localized changes in neural activity (typically averaged over 2–6 s and over 1–27 mm³ of cortex). At least within the primary visual cortex (V1), the fMRI signal increases monotonically with stimulus contrast [6*].

Organization of visual cortical areas in monkeys
Vision is the most richly represented sensory modality in primates. Visual information is processed in over 30 functional cortical areas. In Old World monkeys [7]—our seemingly closest evolutionary ancestors, aside from apes—these cortical areas cover about one-half of the total cortex. Visual cortical areas are organized into two processing pathways, or ‘streams’, both of which originate in area V1 [8]. The ventral stream, projecting from area V1 through areas V2 and V4 to the inferior temporal cortex, processes the physical attributes of stimuli that are important for object identification, such as color, shape, and pattern. The dorsal stream, projecting from V1 through areas V2 and V3 to the middle temporal area (MT) and thence to additional areas in superior temporal and parietal cortex, processes attributes of stimuli important for localizing objects in space and for the visual guidance of movement towards them, such as the direction and velocity of stimulus motion [9].

Early processing areas within both pathways are retinotopically organized, such that each contains a representation of all or part of the visual field; adjacent locations within the visual field are represented in adjacent locations in cortex. At progressively later stages of processing, however, the retinotopic organization of visual areas becomes increasingly coarse. This loss in retinotopic specificity is accompanied by an increasing selectivity for more complex stimuli. For example, whereas neurons in V1 respond to oriented bars and edges placed in their small receptive fields, those in the inferior temporal cortex respond to complex objects, such as faces, placed anywhere in the visual field [10,11]. Both processing streams have further projections to regions within the frontal lobe, which, in turn, send reciprocal projections back to the visual processing areas (see e.g. [12–14]). These reciprocal, or feedback, projections are thought to play a ‘top-down’ role in vision, such as in the allocation of attention to selected visual stimuli [15].

Retinotopic mapping of early visual areas in human cortex
Several groups [16–18,19**,20] have mapped retinotopic visual areas in human cortex, which, presumably, are homologous to monkey areas V1, V2, V3, VP (or ventral V3), V3a and V4. Retinotopic mapping in humans would not have been possible without the fine spatial resolution
and the ability to study individual subjects afforded by fMRI. These mappings have demonstrated that humans and monkeys have a very similar topological organization of early visual areas: firstly, areas V1, V2, V3/VP, and V4 are bounded by alternating representations of the vertical and horizontal meridians; secondly, in areas V1, V2, and V3/VP, the upper visual field is represented ventrally in the cortex and the lower visual field more dorsally; and finally, areas V2 and V3/VP form concentric bands partially surrounding V1, with the isoeccentricity lines continuous from one area to the next.

Even in these early visual areas, however, a few differences between monkeys and humans have been observed. It has been reported, for example, that there is a greater emphasis on central vision in human V1, as demonstrated by a larger cortical magnification factor near the center-of-gaze (i.e., a greater extent of cortex representing the central portion of the visual field) [18]. This finding, though intriguing, has been questioned [20]. Indeed, an expansion in the central visual field representation in human V1 would be surprising given that humans and macaque monkeys are thought to have comparable spatial acuity. A second difference between human and monkey brains is that human areas V3 and VP appear to be wider relative to the width of V1 [18,21*]; however, the functional significance of this finding is as yet unclear. Finally, it has been reported that both the lower and upper visual field representations of human V4 are located ventrally in the hemisphere on the fusiform gyrus (DJ McKeefry, S Zeki, abstract in *Neuroimage* 1997, 5:4:S1), whereas in the monkey, the lower field is represented dorsally in the hemisphere and the upper field is represented ventrally. It is important to note, however, that the topological relationship between upper and lower field representations in V4 seems to be preserved from monkeys to humans. The location of the upper visual field representation in presumed human V4 agrees with that identified in earlier fMRI studies (see [18,19**]), although those studies did not find a corresponding lower visual field representation.

Thus, what we have learned from fMRI is the retinotopic organization of early visual areas in humans, their spatial extents, and their borders. The establishment of areal borders is critical for interpreting the results of studies examining functional specialization.

**Functional specialization**

Along with retinotopic mapping, functional specialization may be used to identify distinct visual cortical areas, on the basis of their selective activation by motion, color, or other stimulus properties. To identify homologous regions, the fMRI activation of human cortical areas may be compared to the average responses of single cells in the candidate regions in monkey cortex.

**Motion**

One of the first such regions to be identified was the presumed human homologue of area MT (also known as V5). A high proportion of the neurons in monkey MT are selective for the direction of visual motion and velocity [22,23], and the responses of these neurons can be modulated by attention [24]. Human MT was originally identified in PET studies on the basis of its selective activation by stimulus motion and attention to motion [25–27], and these results have been confirmed using fMRI [28,29*,30*,31]. In addition, fMRI studies have shown that human MT, unlike V1, is as effectively driven by low-contrast stimuli as by high-contrast stimuli [28]. This is in accord with the finding that the inputs to MT in monkeys derive from cells in the magnocellular layers of the lateral geniculate nucleus [32], which have greater contrast sensitivity than those in the parvocellular layers.

There is evidence from both psychophysical [33] and anatomical [34] studies for an anomaly in the magnocellular system in dyslexics. In support of these observations, an fMRI study found that dyslexics, compared to normal controls, have an almost complete lack of activation in MT during motion perception [35**].

By implication, disorders in reading and phonological awareness may reflect a general deficit in the processing of the temporal properties of stimuli, an example of which is motion. An fMRI study in normal subjects [29*] found that moving stimuli activated, in addition to MT, an area in the superior temporal gyrus that overlapped with regions activated when the same subjects were listening to speech, providing further evidence for a link between motion processing and language. The link between motion and language may be related to action words. It has been shown that generating action words in response to visually presented static objects activates a region, including a part of the superior temporal gyrus, just anterior to the area that processes visual motion (i.e., area MT) [36].

**Color**

Functional MRI studies have reported that an area in ventral occipitotemporal cortex, most often observed in the collateral sulcus and lingual gyrus, is activated selectively by the perception of (or attention to) colored stimuli [37,38,39*], confirming earlier results from PET studies [25,26,40]. Lesions that cause achromatopsia, or color-blindness, in humans appear to include this area [41]. Because V4 in monkeys has a high proportion of cells that show color selectivity [42,43], it has been proposed that this color-selective area is the human homologue of area V4 [26]. However, neurons in monkey V4 also show form selectivity [44,45], but there is, as yet, no evidence regarding the response of the human color-selective area to form. More compelling evidence to support the idea that the color-selective area in humans is the homologue of monkey V4 would come from a demonstration that the
Figure 1 legend Experimental design and results from an fMRI study of working memory for faces. (a) The design of the task. For each series of fMRI scans, the subjects performed three baseline-activation task cycles, each consisting of 30 s of a sensorimotor control task followed by 30 s of a working memory task. Each task period consisted of two items for that task. During the memory task, subjects saw a picture of a face, a delay, and then another picture of a face. Subjects were asked to hold an image of the first face in mind during the delay and to respond with a left or right button press to indicate whether the second face matched the first. During the control task, subjects simply looked at the scrambled pictures and then pressed both buttons when the second scrambled picture appeared. Three time series are shown that represent the different cognitive components of the task: a transient, nonselective response to visual stimuli; a transient, selective response to faces; and sustained activity during memory delays. These time series (smoothed and delayed by convolution with a model of the hemodynamic response) were used as regressors in a multiple regression analysis of the time course of activation in each area. (b) Results from a single subject performing the working memory task for faces overlaid onto that subject’s anatomical images. The three images shown are axial slices taken approximately (i) 18 mm below, (ii) 12 mm below, and (iii) 6 mm above the plane connecting the anterior and posterior commissures. Activations are color-coded according to the relative sizes of the three regression coefficients described above. Areas that responded transiently and nonselectively to any visual stimulus, such as the posterior occipital cortex (A in white arrow), are shown in green. Areas that responded transiently and showed a selective response to faces over scrambled faces, such as the fusiform gyrus (B in white arrow), are shown in blue. Areas that showed sustained activation during the memory delay after the stimulus was removed from view, such as the inferior frontal cortex (C in white arrow), are shown in red. Areas that showed a combination of these types of responses are shown in a blend of colors. These data show both functional specialization in inferior occipitotemporal cortex and the presence of top-down influences (e.g., the sustained memory activity in temporal cortex). Adapted from [51*].

color-selective area coincides with the one identified as human V4 through retinotopic mapping. However, such a demonstration would still leave unanswered questions, such as why retinotopically defined human V4 (at least its upper field representation) does not seem to be form selective [21*] and why lesions of V4 in monkeys do not produce the same color impairments as those in humans [46,47].

Objects, faces, and letters: functional modules? The discovery of 'face' cells in the inferior temporal cortex and within the banks of the superior temporal sulcus of the monkey suggested the existence of subpopulations of cells selective for particular classes of behaviorally significant objects [10]. As face-selective cells seem to be concentrated in particular cortical regions, attempts have been made, using fMRI, to identify a face-specific module within the human brain. Several studies have activated an area in ventral occipitotemporal cortex using faces as stimuli [48,49**,50,51**] (see Figure 1), but it has only recently been shown that this area is selective for faces over other objects ([52*,53*]; JV Haxby et al., abstract in Neuroimage 1997, 5:4:S4). This same area is also more activated by faces than by inverted faces, suggesting that the area is involved in recognizing facial identity, a function impaired by the inversion of faces (JV Haxby et al., abstract in Neuroimage 1997, 5:4:S4). Lesions that produce prosopagnosia, the inability to recognize familiar faces, seem to include this same cortical region [54,55].

The evidence is accumulating, therefore, that there exists a face-selective region in human cortex and that this region is within the ventral object vision stream. This does not necessarily mean, however, that this region is a 'module' in the strict sense [56], for it is activated by other objects, though less strongly, and its selectivity for faces may not be innate. Indeed, there appear to be other regions within the ventral stream that show functional specializations that could not be innate. For example, a region lateral to the face-selective area seems to be differentially activated by letterstrings [49**], and another, located more medially, is activated by viewing words and pronounceable pseudo-words [57]. These areas could not have evolved to process these kinds of information but must surely have developed over time with reading experience. The notion that experience, even in adulthood, can lead to functional specialization, or remodeling, within cortex has been convincingly demonstrated through behavioral [58] and lesion (for a review, see [59]) studies. Patients with ventral temporal lesions can lose specific, acquired visual recognition skills: for example, a car salesman who could no longer distinguish different car models, and a farmer who could no longer identify specific cows in his herd. Because fMRI is noninvasive, involves no radioactivity exposure, and is sensitive enough to enable single-subject analysis, it could be used to measure remodelling of the cortex over the course of weeks, or even years, as expert knowledge is acquired.

Not all objects, of course, can be processed selectively by dedicated neural machinery. A region in the lateral occipitotemporal cortex, termed area LO, appears to respond to all objects, regardless of meaning. This region generates a larger fMRI signal in response to pictures of objects compared to either textured patterns or degraded pictures of objects [60]; the same region appears to have also been identified in a PET study as being activated by pictures of both real and nonsense objects compared to visual noise patterns or scrambled pictures of objects [61,62]. Thus, area LO is involved in form processing independent of object recognition. There are additional regions that respond preferentially to some categories of meaningful objects over others ([52*,53*]; A Ishai, LG Ungerleider, A Martin, JM Maisog, JV Haxby, abstract in Neuroimage 1997, 5:4:S149; JV Haxby et al., abstract in Neuroimage 1997, 5:4:S4), but it is still unclear whether these activations depend on the physical or the semantic properties of the stimuli. It may be that there is an
arrangement of feature space within the ventral stream, similar to that proposed in the monkey [11], whose topological organization remains to be characterized.

**Observing distributed neural systems**

**Bottom-up hierarchical processing within visual cortical areas**

As one progresses from lower-level to higher-level visual cortical areas, there is a progressive loss in retinotopy and a progressive increase in selectivity, such that more useful representations of the visual input can be seen at each successive stage of processing. A strong correlation between neuronal activity in a given area and perception, as measured behaviorally, has been taken as evidence for that area's contribution to the perceptual function under investigation. This approach has been used to argue for the involvement of V1 and V2 in the perception of illusory contours [63–65], of V3 in perceptual 'filling-in' [66], of MT in motion perception [67,68], and of V4 in color perception [69].

A similar approach, of correlating activity with perception, has been applied in several fMRI studies. For example, illusory contours produced by Kanizsa figures have been shown to most consistently activate regions beyond V1 ([70]; JD Mendola, AM Dale, AK Liu, NR Karp, RBH Tootell, abstract in *Neuroimage* 1997, 5:456; see also [71] for corresponding PET results), suggesting that the linking of local features within these figures is accomplished after V1.

Another perceptual phenomenon, called an 'aftereffect', occurs after prolonged viewing of certain adapting stimuli. For example, prolonged viewing of a stimulus moving in one direction will cause a subsequently viewed stationary stimulus to appear to move in the opposite direction. A second example occurs after prolonged viewing of a colored stimulus: a uniform achromatic field will appear to be colored in the opponent colors of the adapting stimulus. Areas that show increased activity during aftereffects may be the sites responsible for the perceptual experience. Consistent with this idea, Tootell et al. [72] showed a larger increase in activity in MT relative to V2 and V3a during a motion aftereffect, even after correcting for the greater motion selectivity of MT. Similarly, Sakai et al. [37] showed increased activity in the posterior fusiform gyrus, the supposed human homologue of V4, during a color aftereffect. Because aftereffects have been shown to be spatially localized and occur only for a few seconds immediately following the presentation of a stimulus, such observations would not have been possible without the good spatial and temporal resolution achieved by fMRI.

**Top-down influences on visual cortical areas**

Some effects observed at the level of single-cell activity in visual cortical areas reflect not only bottom-up processing, but also the influences of attention and memory [24,73,74]. It is thought that these influences are mediated by feedback inputs from cortical areas outside the traditional visual processing streams, such as those in prefrontal cortex. The contributions of prefrontal cortex to the processing of visual information have recently been examined in an fMRI study of working memory [51].

Working memory is the process of maintaining an active representation of information so that it is available for further use. One example of a simple working memory task is delayed match-to-sample in which the subject is shown a sample stimulus followed by a brief delay, and then is required to make a choice response on the basis of information 'held in mind' during the delay period. Courtney et al. [51] found that occipitotemporal areas in the ventral object vision pathway had mostly transient responses to stimuli, indicating their predominant role in perceptual processing, whereas three prefrontal areas demonstrated sustained activity over memory delays, indicating their predominant role in working memory (see also [75]). This distinction, however, was not absolute. The more anterior visual areas demonstrated some sustained activity during the memory delays as well, suggesting a distributed neural system for working memory. Single neurons in both the inferior temporal [76,77] and prefrontal [78,79] cortex also demonstrate sustained activity during memory delays. However, in the presence of intervening stimuli, this sustained activity is maintained only in the prefrontal cortex [80]. Therefore, it is thought that the sustained activity in the temporal cortex is a reflection of the top-down influences from prefrontal cortex, as this activity is dissociable from successful performance of the working memory task.

Functional MRI studies have shown that other higher cognitive functions also are associated with activation of visual cortical areas, independent of, or even in the absence of, visual stimulus input. Mental imagery of objects activates occipitoparietal cortex [81] and, apparently, V1 and V2 [82]. Mental rotation of objects activates occipitoparietal cortex and, apparently, MT [83]. Finally, working memory (SM Courtney; JM Maisog, LG Ungereider, JV Haxby, Soc Neurosci Abstr 1996, 22:968) and long-term memory retrieval [84] have been shown to produce dissociable activations in occipitotemporal and parietal cortex for object appearance and spatial location tasks, respectively, using identical stimuli. (See also PET studies on imagery [85,86], mental rotation [87,88], and working memory [89].)

**Conclusions**

Thus far, fMRI has revealed the locations, spatial extents, and borders of early human visual areas. In addition, evidence has been obtained for hierarchical processing in the human visual cortex and for the existence of functionally specialized areas. Areas specialized for motion perception, color perception, and for such complex object categories as faces and letterstrings have been demonstrated. Functional MRI has also provided evidence
for a strong coupling between activity in specific cortical locations and perceptual experience. Finally, top-down modulation of activity in visual processing areas has been observed in complex cognitive tasks, such as working memory, selective attention, and imagery. It is currently unknown precisely what the ultimate spatial or temporal resolution of fMRI will turn out to be, as the technology and methods for data analysis are still being developed. In addition, the temporal relationship between neural activity and the hemodynamic changes measured by fMRI is still poorly characterized. If this relationship were understood, then one could make inferences about the time course of brief cognitive events ([90*]; VP Clark, JM Maisog, Jv Haxby, abstract in Neuroimage 1997, 5:4:S50).

Acknowledgements

We thank Robert Desimone, James V Haxby, Alex Martin, and Cherri Wiggs for helpful comments on the manuscript.

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


The authors describe a linear systems analysis of fMRI responses in V1 and demonstrate that the response is a separable function of stimulus duration and stimulus contrast. Their analysis demonstrates that the fMRI signal, at least in V1, increases monotonically with stimulus contrast and that the response to a long, periodic stimulus is slightly overestimated by a linear model based on the response to a brief pulse. These findings help us begin to understand the quantitative relationship between neural activity and the fMRI signal.


The authors used the same method as the one presented in [17] and a cortical unfolding algorithm, different from [18], to map the locations, spatial extents, and borders of several retinotopic visual areas in human cortex. Their findings agree well with those presented in [18].


This review summarizes much of the imaging, histological, and physiological data regarding the organization of early visual areas in human cortex. It also includes additional data that clarify or support previously published fMRI results on retinotopic mapping, contrast sensitivity in MT and object selectivity in LO.


This paper reports that moving stimuli activate, in addition to MT, a presumably multimodal area in the superior temporal gyrus. The area appears to overlap with regions activated when the same subjects were listening to speech. This study, along with those presented in [35*,36], provides evidence for a link between motion processing and language.

In this study, subjects viewed a peripheral, presented annulus containing colored, moving dots while maintaining central fixation. The results showed that activation in MT was greatest when spatial attention was directed toward the annulus and featural attention was directed toward the speed of the dots. This activation was reduced when spatial attention was directed toward the color of the dots, and further reduced when spatial attention was directed toward the central fixation point. These results demonstrate that top-down influences, such as spatial and featural attention, have a powerful effect on motion processing in MT.


This paper presents an fMRI study demonstrating an almost complete lack of activation in MT during motion perception in dyslexics, unlike in normal controls. Dyslexics, however, showed normal activation to textured patterns in early visual areas. These results support the idea that disorders in reading may reflect a general deficit in the visual processing of the temporal properties of stimuli, an example of which is motion.


This paper reports results from selective attention tasks to either color or face identity, using identical stimuli (color-washed grayscale images of faces). The results showed that selective attention to either faces or color activates a region along the fusiform gyrus, whereas only selective attention to color activates an additional area in the collateral sulcus. This latter area corresponds to the color-selective area found by previous fMRI and PET studies comparing passive viewing of colored and grayscale stimuli (25,26,37,38,40).


The authors measured the differential sensitivity of human visual cortex to faces, letters, and letterstrings. The locations of the texture and face activations agree with the results of other studies [50,51*]. The letter-string area seems to be just posterior to an area associated previously with recognition of visually presented words [52]. If the letter-string area proves to be specific to letters, it would provide further evidence for a remodeling of cortex with experience.


This study of working memory for faces used multiple regression analysis in order to distinguish activity related to transient perceptual processing during stimulus presentation and sustained activity during working memory delays. The results show a gradual decrement in the amount of perception-related activity and a corresponding increase in the amount of memory-related activity in six different cortical regions, from posterior extrastriate through to anterior prefrontal cortices.


This paper presents the results of several fMRI experiments demonstrating the face selectivity of the middle fusiform gyrus.


This fMRI study shows a separate region of activation for faces over other objects in a rapid presentation, passive viewing paradigm. The area showing face selectivity seems to be the same as that found with very different paradigms in other studies [49,54*].


What fMRI has taught us about human vision Courtney and Ungerleider 561


This experiment demonstrates the activation of extrastriate visual areas during imagery of the referents of auditorily presented concrete nouns. By showing that visual areas are activated in the complete absence of visual input, the study provides evidence both for top-down processing and for an interaction between the auditory and visual modalities through language.


In this study, subjects made same/different responses to pictures of three-dimensional block figures that were shown at different orientations. Activations during the task were compared to those during a control condition in which the pairs of pictures were either identical or mirror images of each other. Activations for the task requiring mental rotation included Brodmann areas 7 and 19, and presumed human MT, providing evidence for top-down processing specific to parts of the dorsal spatial vision pathway.


In this study, subjects made judgments regarding the appearance and position of familiar locations within a virtual reality environment. The fMRI data revealed a double dissociation of regional activation, with dorsal activation for the position of environmental landmarks and ventral activation for their appearance. The study lends further support for the dichotomy between dorsal and ventral visual streams.


This paper provides evidence that it is possible to obtain reliable fMRI signals from single-trial paradigms, laying the way for greater flexibility in task designs that could take advantage of the high temporal resolution afforded by fMRI.