2009

Perceptual Learning Of Object Shape

Doruk Golcu

Recommended Citation

Golcu, Doruk, "Perceptual Learning Of Object Shape" (2009). *Student Theses and Dissertations.* Paper 114.

PERCEPTUAL LEARNING OF OBJECT SHAPE

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Doruk Golcu

June 2009

PERCEPTUAL LEARNING OF OBJECT SHAPE

Doruk Golcu, Ph.D.

The Rockefeller University 2009

Recognition of objects is accomplished through the use of cues that depend on internal representations of familiar shapes. We used a paradigm of perceptual learning during visual search to explore what features human observers use to identify objects. Human subjects were trained to search for a target object embedded in an array of distractors, until their performance improved from chance levels to over 80% of trials in an object specific manner. We determined the role of specific object components in the recognition of the object as a whole by measuring the transfer of learning from the trained object to other objects sharing components with it. Depending on the geometric relationship of the trained object with untrained objects, transfer to untrained objects was observed. Novel objects that shared a component with the trained object were identified at much higher levels than those that did not, and this could be used as an indicator of which features of the object were important for recognition. Training on an object transferred to the less complex components of the object when these components were embedded in an array of distractors of similar complexity. There was transfer to the different components of object, regardless of how

well they distinguish the object from distractors. These results suggest that objects are not represented in a holistic manner during learning, but that their individual components are encoded. Transfer between objects was not complete, and occurred for more than one component, suggesting that a joint involvement of multiple components was necessary for full performance. The sequence of this learning indicated a possible underlying mechanism of the learning. Subjects improved first in a single quadrant of the visual field, and the improvement then spread out sequentially to the other quadrants. This location specificity of the improvement suggests that, with training, encoding information about object shape occurs in early, retinotopically mapped cortical areas. fMRI work suggests that the learning of novel objects in this manner involves a reciprocal switch between two cortical networks, one that involves the normally object-sensitive regions of LOC, and one that involves the temporal and parietal cortices.

This work is dedicated to my parents Abdulvahap and Semra, my sister Ezgi, and my wife Erin.

Acknowledgements

Many people contributed to this work. Here, I would like to thank some of the most important: My advisor, Dr. Charles Gilbert for his guidance, Valentin Piëch for his work at adapting the fMRI methods to our experiments, Mr. Muc Du and Mr. Josefino Borja from the Weill Cornell Biomedical Imaging Core Facility for their assistance with the fMRI experiments, and my personal friend Ben Kelly for his instructions and corrections in C++ programming. Most importantly, I would like to thank all the people who volunteered their time and effort as subjects in these experiments.

Table of Contents: **Page:**

Dedication iii

Acknowledgements iv

Table of Contents v

List of Figures viii

CHAPTER 1: INTRODUCTION 1

- 1.1 Models of Object recognition 2
- 1.2 Neural Substrates of object recognition 8
- 1.3 Visual Search 16

CHAPTER 2: HUMAN PSYCHOPHYSICS 23

- 2.1 Methods 25
  - 2.1.1 Subjects 25
  - 2.1.2 Task 26
- 2.2 Results 31
  - 2.2.1 Detectability of target 31
  - 2.2.2 Pre-and post-training performance 35
  - 2.2.3 Effects of Position 38

2.2.4 Transfer between objects that share components 40

2.2.5 Transfer from objects to components 44

2.2.6 Effect of distractor similarity on performance 48

2.3 Discussion 59

CHAPTER 3: FUNCTIONAL MRI 65

3.1 Methods 68

3.1.1 Subjects 68

3.1.2 Stimulus generation, scan and analysis 68

3.1.3 Retinotopic Mapping of early visual cortices 70

3.1.4 Mapping of LOC 72

3.1.5 Psychophysical Task 74

3.1.6 Analysis 76

3.2 Results 82

3.2.1 Psychophysical performance 82

3.2.2 Effect of object familiarity on cortical activity 83

3.2.3 Correlation of activity between brain cortices 86

3.3 Discussion 90

References 95

List of figures:

Figure 1	Flow of visual information in macaque monkey brain	10
Figure 2	Easy (parallel) and hard (serial) search task	18
Figure 3	Switch from serial to parallel search task	19
Figure 4	Stimulus array	28
Figure 5	Stimulus timeline	30
Figure 6	Detectability of target	28
Figure 7	Increase in detection performance	37
Figure 8	Sequence of learning based on location	39
Figure 9	Transfer of learning to objects	42
Figure 10	Transfer of learning to 4-line objects	43
Figure 11	Transfer of learning from objects to components	45
Figure 12	Transfer of learning from components to objects	47
Figure 13	Effect of distractor similarity on performance	50
Figure 14	Transfer after training with distractors that shared components with target	54

Figure 15 Transfer after training when components of target appear at different frequencies in the distractors 58
Figure 16 Stimulus images for mapping 71
Figure 17 Functional map of visual areas 73
Figure 18 Fluctuations in fMRI signal over time 80
Figure 19 Differences in fMRI activation by trained and untrained targets 85
Figure 20 Changes in connectivity of visual cortices as a result of training 88

CHAPTER 1: INTRODUCTION

Visual identification and categorization of objects one encounters is a widespread cognitive ability amongst a range of species. Humans, being predominantly visual creatures, are especially adept at this process. They develop the ability to distinguish object categories and identify specific objects as early as 30 weeks in infancy (Cohen 1979, Strauss 1979), and adults are extremely rapid at accomplishing these tasks (Thorpe et al. 1996, Delorme et al. 2004).

One of the most important questions posed about the object recognition system is how the brain makes sense of the very complex and variable information that arrives from the eyes. So far, there has not been great success in duplicating this ability using artificial systems, which illustrates the complexity of the problem. Computer simulations often fail at recognition of objects under natural viewing conditions, where the objects need to be segmented from a background that can be very similar to the object in basic visual properties like brightness and color. Even once objects are segmented, their appearance is heavily influenced by viewing conditions such as angle, distance, and illumination. Again, artificial systems have great difficulty dealing with all these variations, whereas the brain usually overcomes them very efficiently. Below, we will discuss some of the more prominent models devised to explain the ability of the brain at overcoming these difficulties so easily, and the neural mechanism that are potentially responsible. We will also discuss visual search, which can be a powerful method at investigating the properties of object recognition behaviorally.

1.1 Models of object recognition:

A variety of models have been put forward to explain how the recognition is achieved. The main query of the majority of these models has been what cues and mental representations the brain uses to match an object that is being observed to the memory of previously encountered objects. Even though in certain instances cues such as characteristic motion (the motion of a fly), color (quickly finding a red shirt in a pile of clothes), or texture (identifying a piece of tree bark) can help with the recognition of an object, most of the time such properties of an object are secondary to its shape in terms of facilitating recognition. Modeling efforts using shape cues are divided into two major schools of thought: one that proposes that objects are recognized holistically, and one that claims they are analyzed in a parts-based fashion.

The holistic models postulate that information about an object is represented in the brain in the form of complete two-dimensional images of the object. These models propose that transformation invariant recognition of an object can be accomplished by approximating all appearances of the object from a limited number of stored viewpoints (Poggio and Edelman 1990, Bülthoff and Edelman 1992, Ullman 1989, 1992, 1996). A number of studies show that performance at recognizing novel views of objects degrades with increasing distance from a familiar view (Booth and Rolls 1998, Tarr and Gauthier 1998, Vogels 1999), which supports this type of model. A severe criticism of holistic models of object recognition comes from the variation of the appearance of objects in real-life conditions. There are many factors that contribute to variability in object appearance, including viewing angle, distance, visual field position, and illumination. The combinatorial explosion of all possible object attributes makes storing internal representations of each appearance unlikely, in addition to the need to recognize the object when seen in a novel viewing condition. The internal representation of object shape must therefore be invariant to all of these attributes. Parts-based models of recognition are generally seen as an alternative to holistic models that bypass this problem more efficiently. The parts-based models propose that information about objects is represented in the brain as parts or fragments that can be extracted from objects. According to these models, specific combinations of a relatively small number of components can represent a very high number of objects. This sort of representation can thus dramatically reduce the computational load the recognition of a high number of objects and their variations. In these models, effects of viewpoint dependence are often explained through the occlusion of informative parts. Most of the discussion on parts-based object recognition models has been what parts have the necessary properties to be useful for this process.

One of the first models to propose a set of shapes to define the pool of real world objects that need to be recognized is that of Marr (Marr and Nishihara 1978, Marr 1980). He proposes three criteria that parts must fulfill to be useful for object recognition. These are a) accessibility: parts need to be such that they can be easily extracted from whole objects; b) scope and uniqueness: parts need to be able to define all objects, and a definition of an object through parts needs to be unique to that object; and c) stability and sensitivity: parts need to define the similarities between objects, but they also need to define an object specifically enough that it can be distinguished from other objects that are very similar. The parts he proposes following these criteria are generalized cones defined through specific axes of orientation with respect to an object-centered coordinate system. This results in a representation of the objects in a manner reminiscent of stick figures. A study by Kovács and Julesz (1994) shows that contrast sensitivity is enhanced within enclosed spaces, in a manner determined by the global properties of the enclosing shape. These data support a model of object recognition where ‘skeletons’ are extracted from the shape of object and used for storage of object information, similar to the model of Marr.

Marr’s generalized cones are criticized as being good for defining objects of the animal kingdom, especially humans and most other mammals, but they are often insufficient for defining objects outside this group. Hoffman, while not offering an alternate class of objects to replace Marr’s, suggests a method by which the objects can be segmented into informative parts based specifically on their geometric properties. He claims that natural segmentations of objects occur at points of concavity, or at lines of curvature minima for smooth surfaces (Hoffman and Richards 1984). Using this sort of segmentation, Biederman offers his own set of object parts that he called ‘geons’ (Biederman 1987). According to this model, such volumetric parts can be inferred from two-dimensional images using the existence of five non-accidental properties. These are: i) colinearity, ii) curvilinearity, iii) symmetry, iv) parallelness, and v) vertices. When observed in two-dimensional images, these properties indicate the existence of similar relationships in the three-dimensional structure of the object as well. The relative placement of components is significant in recognition with this type of model. This is supported by behavioral studies. For example, the recognition of three bars embedded in complex background improves if the bars are placed in a manner that is reminiscent of a face (Gorea and Julesz 1990). With the ‘geon’ model, a small number of visible components are sufficient to recognize an object, but for more complex objects performance increases with increased number of available components. Furthermore, structural information through the parts takes precedence over information of color or texture; line drawings of objects are recognized as efficiently as color photographs (Biederman 1987). The importance of the recognizability of components is supported by studies on face recognition. Recognition of faces is reduced by coarse quantization, which turns the structure of the face into apparent blocks. In this form, components of an object are transformed into simple blocks due to the coarse sampling, thus can no longer be recognized as what they really are, and performance drops significantly (Harmon and Julesz 1973). The recognition of the face improves if noise is added to the images since the increased noise prevents the components from appearing as blocks (Morrone et al. 1983). Additionally, different objects of similar appearance can be more easily differentiated through the presence of different geons than they can be through their metric properties (Biederman and Bar 1999).

Two other important models make use of object parts. The nonlinear maximum operation uses a cortical response scheme where the postsynaptic response of a neuron is determined by the strongest of the incoming information, and therefore shows the best match of parts of the stimulus to the preferred features of the cells earlier in the cortical hierarchy (Riesenhuber and Poggio 1999). The model of fragment-based hierarchy uses a different method of extracting parts from an object than most other models (Ullman 2002, 2007). Instead of being an internal property of objects, the components in this model are generated by the brain through the interaction of images of objects with the images of the environment of the object. Fragments of an object that are most informative about the identity and category of the object are used for recognition. Fragments are most informative when they occur with reliability within the object, and very rarely or never in the environment of the object. Computer simulations are able to classify images of objects at above human levels using this type of computation.

1.2 Neural Substrates of object recognition:

With evidence existing for both sides of the issue, there has been significant discussion in the field about whether a parts-based or a holistic mechanism is predominantly used by the human brain for object recognition (Biederman et al. 1995, Tarr et al. 1995). A look into the selectivity of neurons in the brain and how they are organized can be helpful in determining what type of mechanism the brain is better set up to use for recognition.

The flow of visual information in the brain is generally accepted to be divided into two distinct streams; of these, the dorsal stream is mainly involved with the processing of visually guided movement, while the ventral stream is involved with analyzing object shape, color, and texture (Van Essen 1979, Felleman and Van Essen 1991, Van Essen and Gallant 1994, Figure 1). The analysis of shape in the ventral stream is believed to be hierarchical, i.e. cells in the early cortices of the stream such as V1 are sensitive to very simple properties of an object such as contour orientation, position, or curvature, and they have small receptive fields. As one progresses further along the stream to higher centers of processing, the cells are found to have increasingly larger receptor fields and be sensitive to increasingly more complex stimulus properties (Hubel and Wiesel 1962, 1965, 1968, Gallant et al. 1993, Pasuphaty and Connor 1999, Ito and Komatsu 2004). These observations have led to the conclusion that individual cells in a specific visual cortex receive input from multiple cells in an earlier cortex and are sensitive to a combination of stimulus properties that trigger these earlier cells. Early studies postulate that this complexification process will continue to the point where at the top of the ventral stream there will be cells that are individually sensitive to the entirety of the visual stimulus, i.e. to a whole object. Such hypothetical cells have come to be known as ‘grandmother cells’, due to the notion that among them there will be cells that are sensitive to the appearance of one’s grandmother (Gross et al. 1972, Perrett et al. 1982).

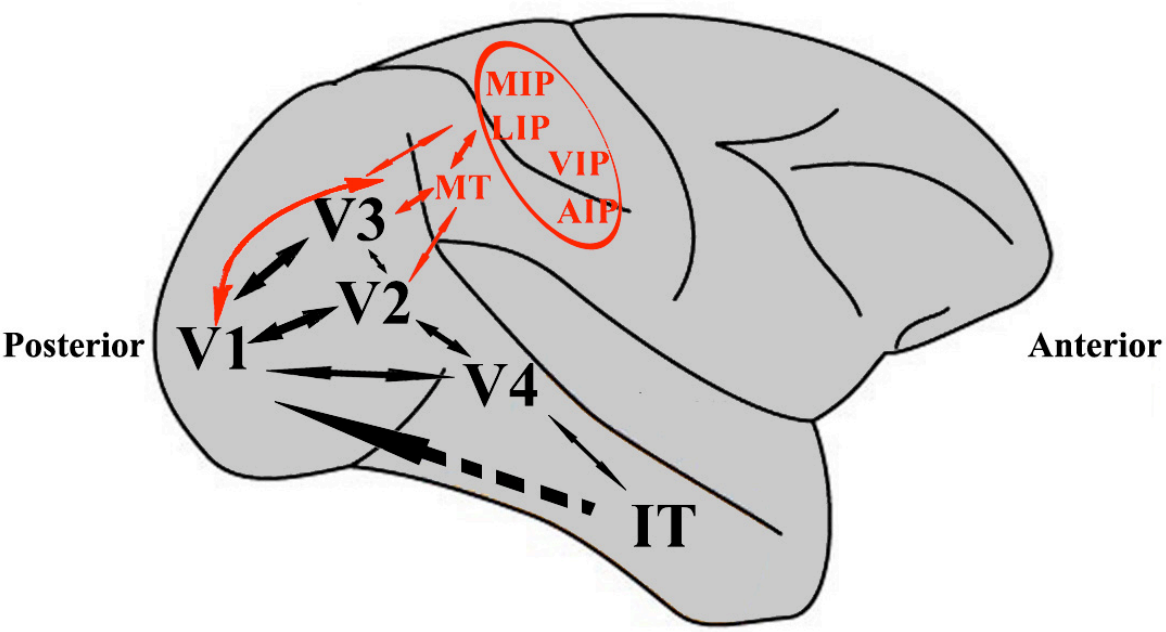


Figure 1. Flow of visual information in macaque monkey brain. The dorsal pathway (red) regulates visual attention and visually directed movement while the ventral pathway (black) is responsible for the analysis of the shape of the stimulus.

Cases of brain lesion in animals and humans with cognitive defects point to the possible seat of the representation of objects in the brain. In monkeys, ablations of inferotemporal cortex (IT) lead to significant defects in object recognition, implicating this as the region responsible for computations necessary for object recognition. For the same reason, in humans, parts of the fusiform and occipito-temporal junction appear to be involved in this process (Farah et al. 1989, Damasio 1990, Damasio et al. 1990, Goodale et al. 1991, Farah 1992). Electrophysiological recordings of neurons in monkey IT confirm that neurons in this region have the properties necessary for performing object recognition. Initial results have shown that IT neurons had large receptive fields and had very specific and complex triggers. Most of these triggers were then identified as complex three-dimensional objects, most notably small populations of cells extremely selective for faces and hands (Desimone et al. 1984). Optical imaging studies reveal that pictures of a single head from different viewpoints activate distinctly separate patches of the brain tissue (Wang et al. 1998). Another evidence for such sparse coding of objects comes from studies done using single-cell recordings in humans during surgical procedures to treat epilepsy. These studies found that cells in medial temporal lobe responded very specifically to different images of a single individual (Quiroga et al. 2005, Conner 2005).

The presence of such object-specific cells leads to the assumption that objects are represented in the brain as entire units, which correlates well with holistic models of recognition. This view, however, is contested with more detailed studies of the specificity of IT neurons. The study of Tanaka et al. (1991) is one of the most detailed, aimed to determine the exact triggers of the IT cells. In this study, they use a number of three-dimensional objects and simple geometric shapes as stimuli, and examine how well these images stimulate the activation of IT neurons. When they found cells that responded well to objects, they use progressively more simplified two-dimensional models of the object to see if the cells responded better to whole objects or to features extracted from the objects. They find that most of the IT neurons that were sensitive to objects actually responded to features of these objects. On the average, the neurons responded best to stimuli that are more complex than simple bars, but not complex enough to be real objects. In humans, the cortical representation of objects is extensively studied using fMRI, first by Malach et al. (1995) and many others following him (Grill-Spector et al. 1998, 2001, Kourtzi and Kanwisher 2000). These studies looked at the brain regions that are activated more strongly by objects compared to scrambled images of objects. Such regions are activated specifically by the presence of the object and not by the physical properties of the stimuli since scrambled images retained the same properties as intact objects. The lateral occipital cortex came out as the region most strongly associated with object recognition from these studies. Furthermore, occipital, fusiform, and superior temporal sulcus, anterior collateral sulcus, and some regions anterior to the fusiform gyrus are found to be specifically responsive to faces (Sergent et al. 1992, Haxby et al. 1994, Puce et al. 1996, Kanwisher 1997, Hadjikhani and de Gelder 2003,Tsao et al. 2003, 2006, 2008a,b). Other specialized centers exist for other object categories like body parts (Downing et al. 2001, Peelen and Downing 2005, Schwarzlose et al. 2005), and places (Epstein and Kanwisher 1998). Objects are sparsely coded in the monkey IT, with object-selective cells organized in a columnar fashion, i.e. neurons that are activated by a specific object are clustered into a number of small discrete patches. Cells within one patch all share the same specificity and respond to a simplified model of the object (Fujita et al. 1992, Tsunoda et al. 2001, Brincat and Connor 2004, 2006). fMRI studies both in humans and non-human primates show that there is significant overlap between regions activated by different types of objects. This overlap is an indicator of the presence of cells sensitive to components shared between objects, and therefore activated by objects belonging to different categories (Ishai et al. 1999, 2000, Grill-Spector et al. 2001, Op de Beeck et al. 2008, Bell et al. 2009). In general, these properties of the neurons appear to favor a parts-based mechanism of object recognition.

It is argued that the conflicting data about how objects are represented in the brain can be a result of the fact that different objects are represented through different mechanisms (Farah et al. 1998). One potential source of this difference is expertise (Logothetis 2000). Experts are usually better at detecting fine distinctions between objects relating to their expertise and categorizing them faster and more accurately than non-experts. Sparser coding can be beneficial for such rapid distinction by storing exact representations of each possible category. Therefore, it is possible that the representation of an object changes as a result of expertise, becoming more sparse and holistic. Faces are one type of object that appears to be more holistically represented in the studies mentioned above. For both humans and monkeys, the ability to recognize and distinguish faces is of extreme social importance. Most members of these species thus already are experts at recognizing faces. Evidence for the special treatment of faces by the brain comes from patients of prosopagnosia, who, due to damage to certain regions of the brain can no longer distinguish individual faces, but remain largely normal at recognizing other objects (reviewed in Damasio et al. 1990, Farah 1996). In addition to this functional evidence for special treatment of faces, there also exists significant behavioral evidence, such as the effect of inversion, which severely impairs recognition of faces. This impairment is noticeably higher for faces compared to other objects, suggesting that there is a significant role of expertise in the recognition of the upright-oriented faces, which is the more common orientation, much more so than for non-face objects (Yin 1969, Valentin and Bruce 1986, Valentin 1998). It is argued that these effects are due to a holistic representation of the faces where the specific configuration of the features is important (Rhodes et al 1993, Farah et al. 1995, Farah 1996, Tanaka and Sengco 1997). All these studies are seen as evidence that faces and possibly other objects of expertise are represented differently and more holistically than common objects (however, see also Wright and Barton 2008, McKone et al. 2006).

1.3 Visual Search:

Another possible way of making inferences about how the brain processes incoming information is based on perceptual behavior. The behavioral performance at recognition of objects under different conditions can give valuable clues about what cues the brain is using to perform this role. One of the behavioral paradigms frequently used for this type of study is the visual search. A simple visual search task involves finding a specific target object within a field of related but different distractors. Original examination of visual search tasks have shown that performance in such tasks followed a clear dichotomy: under certain conditions, the targets pop out from the distractors and are easily found, while in others, there is no such clear distinction between target and distractors. These two conditions were named parallel and serial search, respectively (Treisman and Gelade 1980, Sagi and Julesz 1985, Figure 2). Performance in a serial search task diminishes with increasing number of distractors while it remains constant in a parallel search task. The prevailing theory of the time stated that the difference between the two conditions was in how attention was used. Parallel search tasks are pre-attentive, and happen simultaneously across the visual field. Serial search tasks on the other hand require the attention to be shifted to each individual location of the stimulus array sequentially to find the target. The most elaborate explanation of why this difference exists comes from Treisman and Gelade (1980) in the form of the feature integration theory of attention. According to this theory, the visual field is mapped in the early visual cortex with respect to the elementary features, such as color, orientation, direction of motion, etc. If an object stands out from the rest of the objects in the visual field with respect to an elementary feature, then it is perceived in a parallel, pre-attentive fashion, and pops out. If instead it differs from the rest of the objects in terms of the combination of multiple elementary features, then this combination needs to be analyzed at a higher center where information from early visual cortices is combined. Therefore the object no longer pops-out as easily and is instead processed in a serial fashion.

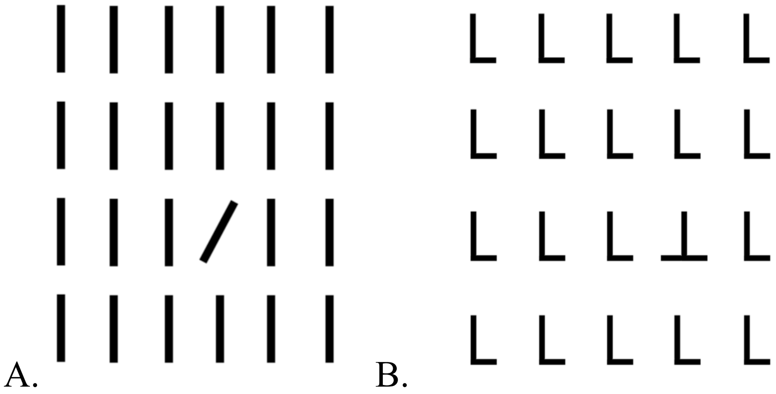


Figure 2. A) Easy (parallel) and B) hard (serial) search task. In the parallel search task the target pops out due to being different from the distractors by a single elementary feature, whereas in the serial search task a conjunction of features need to be used, resulting in a less efficient search (based on Treisman and Gelade 1980).

It was soon realized, however, that performance in visual search is not solely dependent on the physical properties of an object. Wang et al. (1994, Figure 3) have shown that number 2's among 5's (written with straight lines similar to seven-segment-display) pop out, yet they lose this pop-out quality when the stimulus is rotated by ninety degrees, rendering the images unfamiliar. Essentially, the loss of familiarity without changing any other visual property of the object causes the type of search required to complete the task from parallel to serial. Later studies have made an even stronger case by using characters from different alphabets (Malinowski and

Hubner 2001, Shen and Reingold 2001). Those individuals familiar with the alphabet that the stimuli are based one performed the task using parallel search, yet those whom were unfamiliar with the symbols used serial search, despite the stimulus being exactly the same without even the change associated with a rotation. These results have clearly shown that familiar objects were found more efficiently than unfamiliar objects.

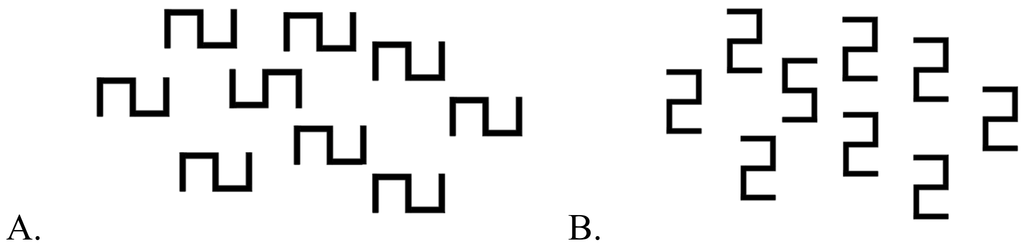


Figure 3. A) A serial search task that can be rendered parallel by B) rotating each object within the array by ninety degrees, forming the more familiar 2 and 5 shapes (based on Wang et al. 1994).

The dichotomy of parallel and serial search was also challenged later by a series of studies by Ken Nakayama and colleagues (Bravo and Nakayama 1992, Maljkovic and Nakayama 1994, Joseph et al. 1997). Using a number of different, attentionally demanding tasks concurrently with the visual search task, they have shown conclusively that attention was still required for parallel search. Performances suffer significant drops when attention for the search task was thus depleted. Furthermore, priming heavily influences the performance. These results prove that parallel and serial search tasks were extreme ends of a continuous spectrum instead of two discreet mechanisms.

Studies involving perceptual learning in visual search task have shown that the recognition of an object can be taken from one end of this continuum to the other (Sigman and Gilbert 2000, Leonards et al. 2002). This change simulates the effects of familiarity on the visual search. In these experiments a search task involving an unfamiliar object as a target is used. Under normal conditions, such an object does not pop out among similar distractors and performance levels were low. However, the performance steadily improves across several days of repetition to reach much more reliable levels. These experiments suggest that there can be a top-down reorganization of the object recognition pathway so that objects can become elementary features. Several lines of evidence exist to support this hypothesis. It is now known that neurons in V1 can respond to much more complex features than originally thought (Kapadia et al. 1995, Sillito et al., Das and Gilbert1999, Posner and Gilbert 1999, Gilbert et al. 2000, 2001, Li and Gilbert 2002, Li et al. 2004, 2006, 2008). There is a considerable measure of plasticity in adult V1, both at the level of receptive field properties (Kaas et al. 1990, Kaas 1991, Garraghty and Kaas 1992, Kapadia et al. 1994, Gilbert and Wiesel 1992, Crist et al. 2001) and in the capacity of local circuits in V1 to undergo sprouting and synaptogenesis (Darian-Smith and Gilbert 1994, Gilbert et al. 1996, Obata et al. 1999, Stettler et al. 2006, Yamahachi et al. 2008), which creates a plausible mechanism by which such reorganization can take place.

fMRI studies were conducted to provide functional support of these behavioral observations. Sigman et al. (2005) have used a task of searching for T’s of a specific orientation between T’s of other orientations. Using a block-based design, they have investigated the changes in brain activity when searching for a familiar orientation compared to the brain activity when searching for an unfamiliar orientation. Their findings show that for untrained orientations, an extended network mainly consisting of parietal and frontal cortices and lateral occipital cortex, which was consistent with the commonly accepted localization of object recognition in humans. Trained shapes on the other hand more highly activated a smaller region in the middle occipital cortex that corresponds to early visual regions, possibly V1. These studies show that a functional reorganization of brain activity was indeed occurring during learning. However, the study localizes the activity changes anatomically, and therefore the exact cortices involved remain speculative.

In this study we use a visual search task to disambiguate if information about objects are represented in the brain as whole units or in the form of combination of parts. We look at the transfer of learning between objects that share components to see if the components of objects have a role at the recognition of objects. We investigate the transfer of learning between objects and their components in order to determine if the learning of objects occurs through the learning of components. We also use different sets of distractors and examine how geometric relationships between target and distractors influence the use of components for the recognition of targets. We make certain inferences about the changes of cortical representation of objects as a result of training based on the sequence of learning, and test these inferences using functional MRI.

CHAPTER 2: HUMAN PSYCHOPHYSICS

To understand the brain's mechanisms of object recognition, a key question is what object features are used for recognition, how these features interact with each other, and how the characteristics of the background influence which features contribute perceptually to object identification. There are two major theories about how object recognition takes place. Of these, the first is a holistic model, where the whole object is learned and recognized as a single independent entity. These models are based on the hierarchical nature of the visual stream of information processing, and assume that pieces of visual information about an object keeps getting combined as they travel upstream, until the full information about the object is assembled together. This information is compared to a previously stored template of the object. One of the most prominent criticisms of such template-based models is the potential explosion of the number of transformational variants that appears to be needed to account for all the visual variations of all possible objects that are known by an individual (Gray 1999, von der Malsburg 1999). This is often thought to constitute an implausibly large load on the available neurological resources. The second type of model that is offered as an alternative to holistic models is the parts-based model of object recognition. These models postulate that instead of having a single template that stores object information, objects are instead coded as a combination of smaller, simpler parts that are largely viewpoint invariant (Marr 1978, Marr et al. 1980, Hoffman et al. 1984, Biederman 1987). This allows different combinations of a finite number of parts to code for large numbers of objects and their variations, reducing the required amount of storage significantly. Computer simulations support the possibility of a parts-based object recognition mechanism that makes use of parts of medium complexity as very good indicators of both identity and category of an object (reviewed in Ullman 2007). There has been significant discussion in the field about which one of these two kinds of mechanisms is used for object recognition in the human brain (Biederman and Gerhardstein 1995, Tarr and Bülthoff 1995).

To obtain a psychophysical measure of what is encoded by the brain in object recognition, we have employed perceptual learning in a visual search paradigm. Recognition of an object embedded in an array of distractors can, with practice, improve from chance levels to much more reliable performance (Sigman and Gilbert 2000). We can measure what is learned by looking at the transfer of perceptual learning between objects related through shared parts, and thereby determine which of the two models are predominantly used in the recognition of objects. We used a variety of different search conditions to simulate and investigate the effects of the visual characteristics of the environment on the recognition of an object.

2.1 Methods:

2.1.1 Subjects:

Fifty-one subjects (34 female, 17 male, 31 of these subjects were asked to report their handedness, of these, 29 were right-handed and 2 were left-handed) that were adults ranging in age from 18 to 70 participated (median age=29). They were recruited according to the regulations set forward by the Rockefeller University Institutional Review Board, and gave written informed consent. All subjects except one (author D.G.) were naïve on the specific task used when they started the study, and had good or corrected vision.

2.1.2 Task:

Psychophysical experiments were designed to study the transfer of training between objects via shared components. Stimuli were presented on a SONY Trinitron flatscreen CRT monitor with a refresh rate of 60 Hz. Objects were created using Inkscape open source vector editor, and displayed using E-Prime 1.1 (Schneider et al. 2002a, b). Subjects were seated at 180 cm distance from the monitor. A chinrest was used to stabilize head position relative to the monitor.

The search task involved a set of arbitrary shapes consisting of three connected line segments. The size of each object was 0.3 degrees of visual angle along each of the three component lines. For each study, one object, at a specific orientation, was chosen as a target. In each trial, the object was embedded in an array of distractors, which bore similarities to the target, in that they consisted of three connected lines, but differed from the target in their orientation or the angles between the constituent line segments. Two variations of the stimulus setup were used (Figure 4). The first setup used was a rectangular 5x5 grid, with the central position taken by a fixation point in the form of a white dot. A single object was presented in each of the other positions of the grid, for a total of 24 objects in the stimulus. The second stimulus configuration was a circular grid with the fixation spot placed at the centre, so that all objects were equidistant from the fovea. The objects were placed with equal separation along the circumference of the circle, at three degrees eccentricity, but with the horizontal and vertical meridians left empty. For all objects, the point where the three lines intersected was placed on the circumference, and the separation distances between objects was measured from these points. Circular grids with 8 or 12 objects were used in different experiments, with the lower number of objects intended to reduce task difficulty. The stimulus array was displayed as white objects (187 cd/m^2) on a black background (34 cd/m^2) at high contrast. It was presented for 300 ms, followed by a 3700 ms blank period, during which the subjects were asked to report the presence or absence of the target object within the array (Figure 5). If the subjects reported seeing the object, they were also asked to report its location within the array by entering a number corresponding to the array position where they think they have seen the target shape. The responses were collected using an Ergodex DX-1 Input System. A one second long visual feedback was given at the end of each trial. The degree and rate of learning did not noticeably change between the rectangular and circular grids.

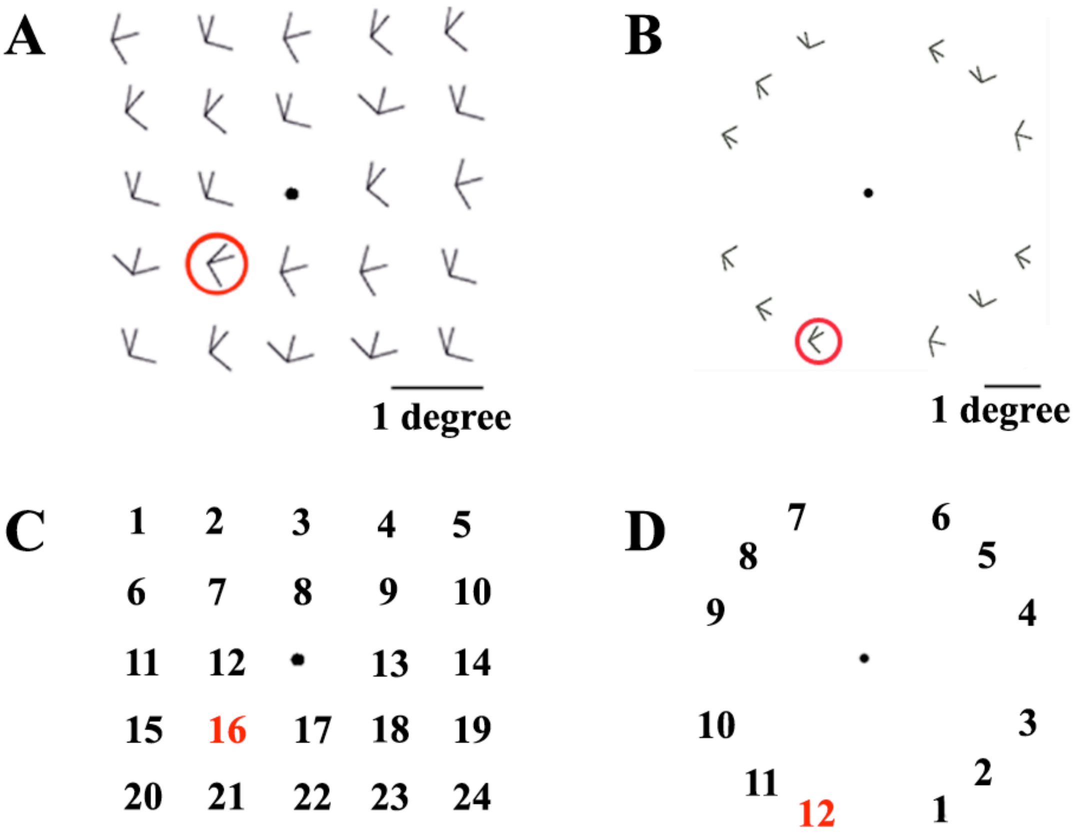
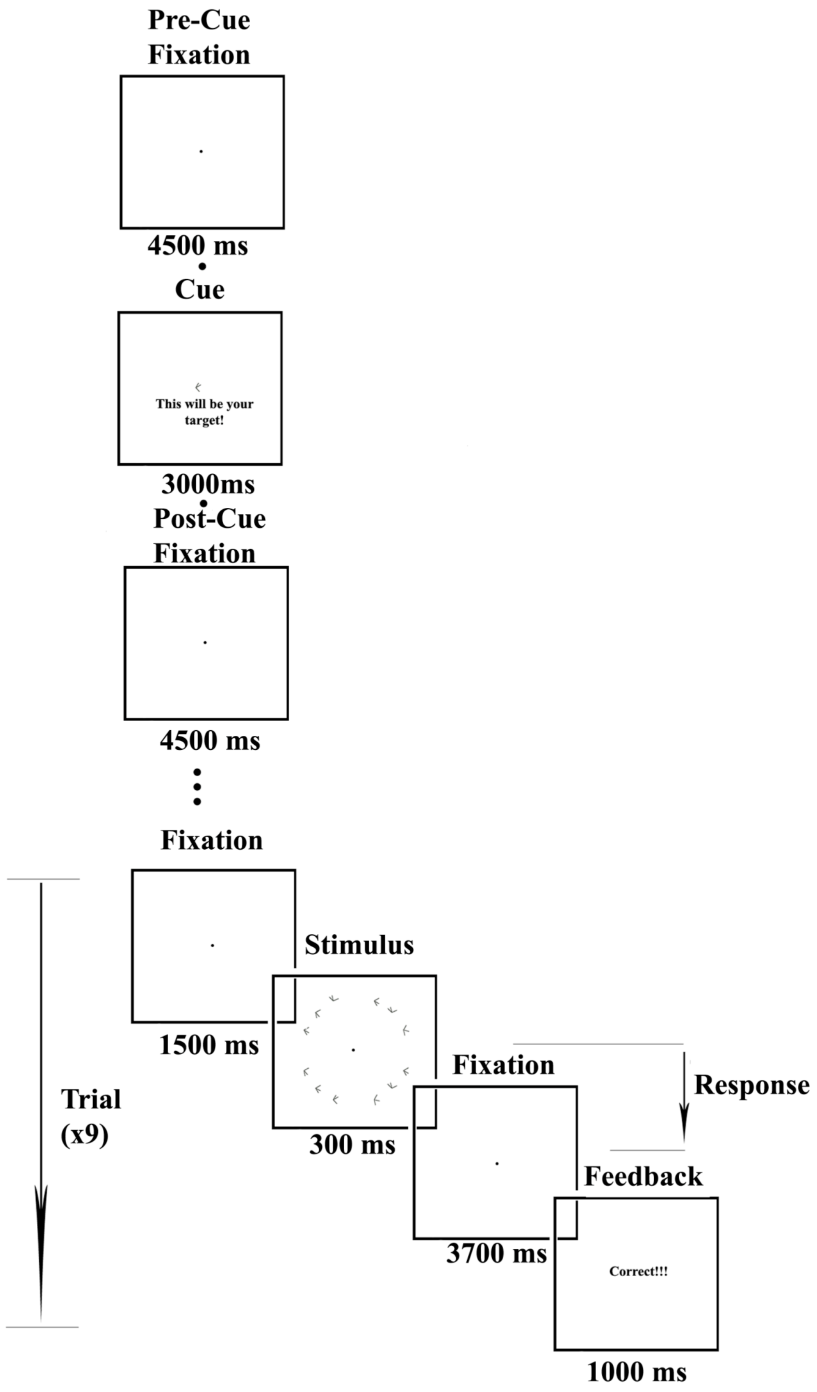


Figure 4. Stimulus array. The stimuli consisted of arbitrary 3-line shapes distributed in either A) a square 5x5 array or B) a circular array of 12 objects (target shapes are encircled in red). Subjects were asked to report if they have seen the target shape or not. In trials where they responded positively, they were also asked to report the location of the target object by entering a number corresponding to one of C) 24 positions within the square array and D) 12 positions within the circular array.

The total number of trials per session ranged from 500 to 1500. Sessions were divided into rounds of 60 trials, and each round divided into blocks of 10 trials, at the beginning of which the target object was displayed in isolation for 3 seconds to remind the subjects of the target shape. After this display, there was a 6 second period during which only the fixation point was present to enable subjects to maintain fixation. Consecutive trials were separated by a 1500ms interval (Figure 5). The subjects were allowed to rest between rounds and to start each round at a time of their own choice. Sessions took about one hour, and the subjects did three to five sessions per week. Whenever possible, the sessions were scheduled for the same time of the day in order to reduce the impact of external factors on performance. We analyzed the data using a two-tailed, paired student's t-test when comparing performance levels before and after training. Performances are given as the percentage of correct responses, including the correct location, compared to the total number of trials where the target was present. Since different subjects showed different rates of learning and different starting performance levels, the plots of changes in performance over time are shown for individual subjects, with the error bars corresponding to the variation of performance between blocks.



E.

Figure 5. The stimulus timeline. The cue shape is displayed once every nine trials.

2.2 Results:

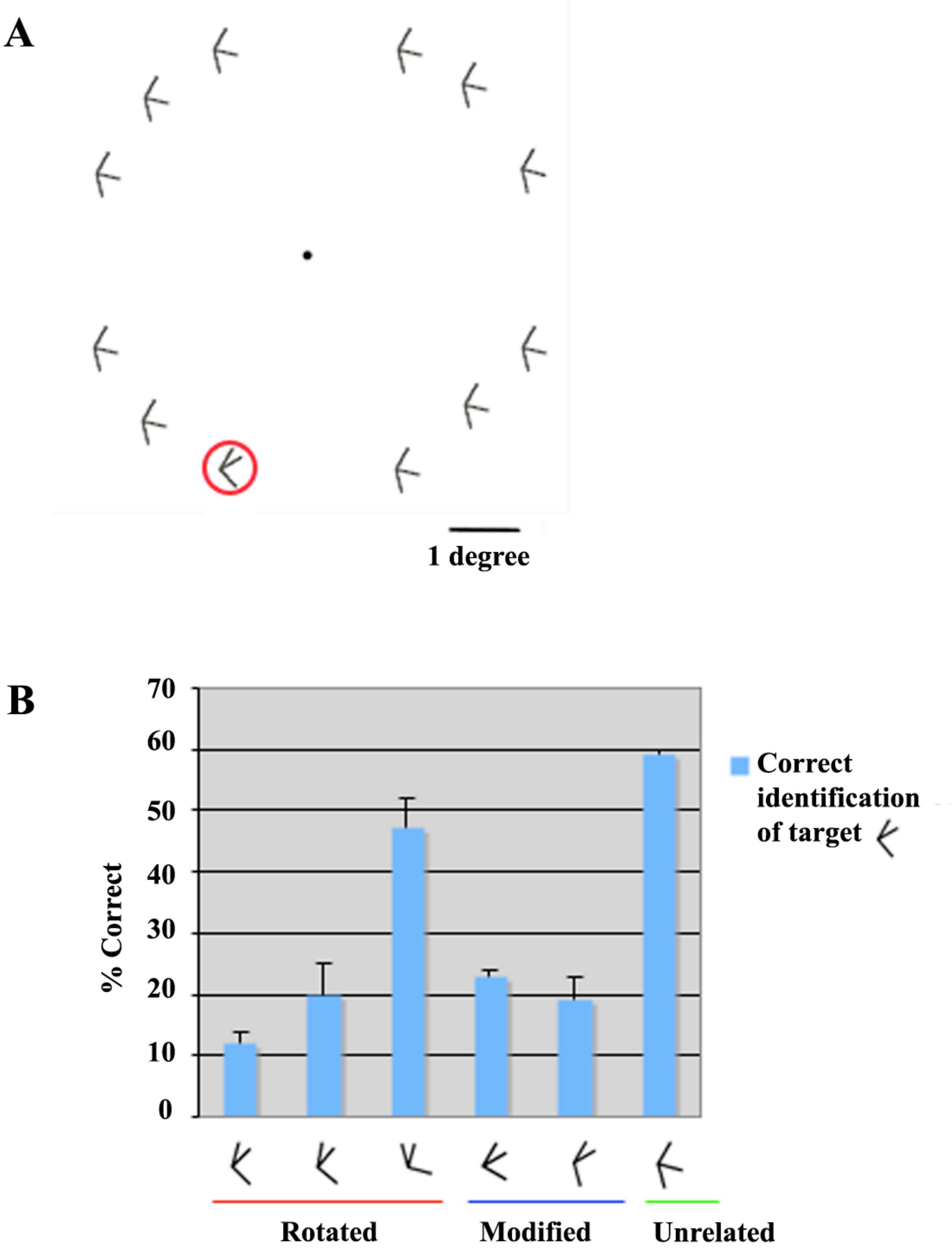
2.2.1 Detectability of target:

We conducted a set of experiments with changing stimulus parameters in order to determine the parameters best suited for our study. First, we looked at the properties of distractor shape. For any perceptual learning to take place, the object needs to be detectable among the distractors, if even at a low level of performances. Therefore, we examined how different from the target the distractors need to be for the target to be detected. For this purpose we used multiple copies of the same object as distractors. In the trials where the target object was present, it was displayed together with 11 copies of one object in the other positions of the stimulus array. We used this setup to display the target object with distractors bearing similarities with the trained object. The distractors used were similar to the target in one of the following two ways: either they were a rotated version of the target object; or they were composed of a modified form of the target, with changes in the angle between the three line segments of the target. We have found that, for small differences in orientation, naïve subjects were able to discriminate the target object from the distractors with difficulty, and therefore performed at very low levels. For large orientation differences the performance was higher. The performance was also highest when the distractor showed the greatest geometric differences from the target (Figure 6).

Next, we tested the effect of changing the number of types of distractors on target detection by naïve subjects. Experimental conditions with 4, 8, 16, and 33 different distractors were compared. There was a visible overall trend of higher level of performance at conditions with more distractors. The subjects performed significantly better when there were fewer types distractors (performance with 4-distractors 37.4±12.0%, performance with 33-distractors 19.4±12.8%, $p<0.0008$, two-tailed paired t-test, average of three subjects). This difference was maintained after training (performance of 85.0±2.8% vs. 33.0±2.8%, respectively, $p<0.05$, two-tailed paired t-test, one subject).

Figure 6. A) Search setup with a single type of distractor. The target object is encircled in red. B) Performance on target identification when embedded in an array of a single type of distractor. The distractors were rotated (underlined red) or modified (underlined blue) versions of the target, or an unrelated object (underlined green), and are illustrated underneath the relevant bar in the graph.

Figure 6


A
1 degree
B
70
60
50
40
30
20
10
0
% Correct
Correct identification of target
Rotated
Modified
Unrelated

2.2.2 Pre-and post-training performance:

For the purpose of the perceptual learning experiments, we chose a set of targets and distractors that were similar enough in appearance and sufficiently unfamiliar so that the subjects performed at chance level at the beginning of training. The level of performance was measured as the fraction of trials when they detected the target correctly relative to the total number of trials where the target was present. Trials were marked as "correct" when the subjects properly indicated the object location. Thus, trials where the object was present and was reported as being seen, but whose location was not correctly indicated, were marked as error trials. For experiments where indication of object location was not required, the proportion of correct responses was corrected for false positives by using the following formula: $p' = (p-fp)/(1-fp)$ where p is the percentage of positive responses, fp is the rate of false positives (rate of trials where the subject reported seeing the object when the object was not present) and p' is the 'real' percentage of correct responses. By repeating the task daily, subjects' performance steadily increased over a period of 10-15 days. Subjects improved from a near chance level of performance before training (correct detection=16.1±5.4%) to a performance level of 70-80% correct responses after training, at which point we stopped training (correct detection=71.3±5.5%;

significance of the change $p<10^{-51}$, two-tailed paired t-test, average of 38 subjects). This process took 10-15 days (Figure 7). Longer periods of training resulted in further improvement above this level (not shown).

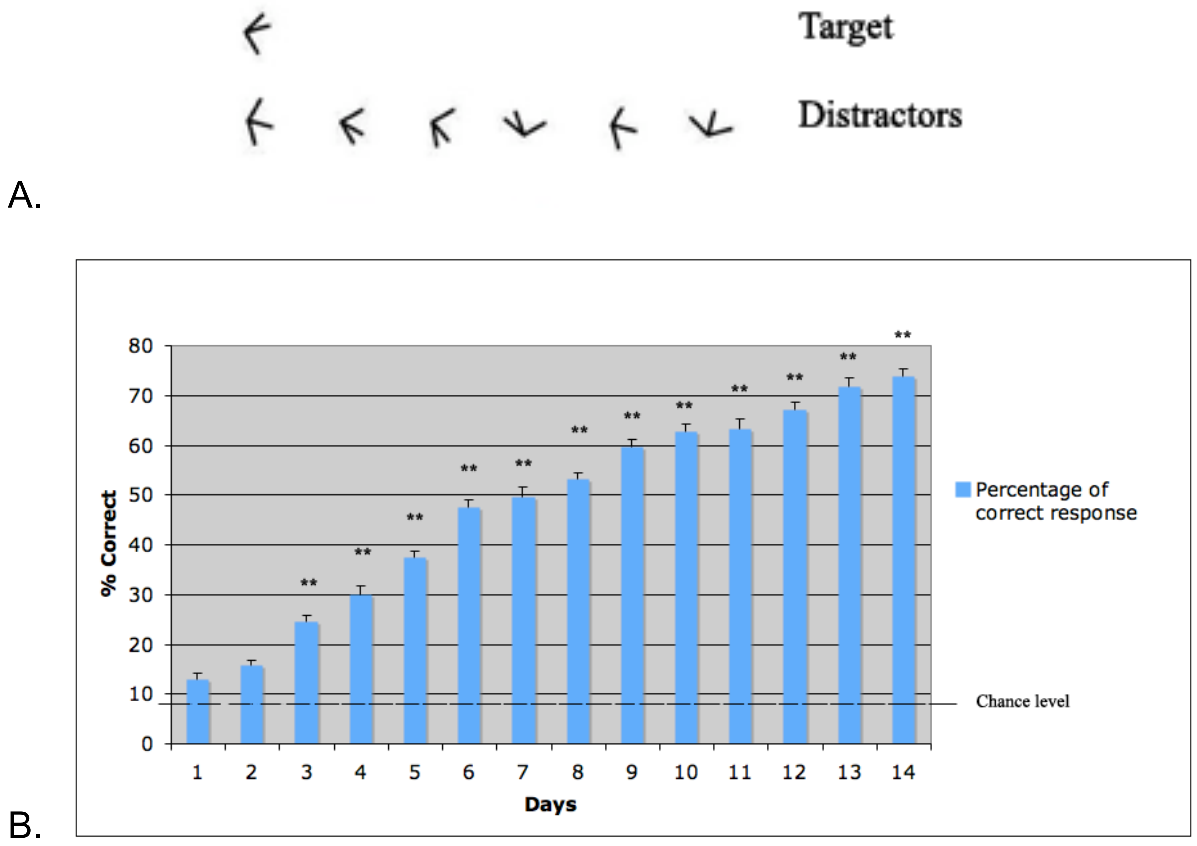


Figure 7. A) Target and distractors used for training. B) Increase in performance at detection of a target object embedded in an array of distracters of similar shape, through several days of training. Performances are given as the percentage of correct detection of the target against the total number of appearances of the target. Dashed line represents chance level. Single subject, error bars represent standard errors across individual blocks. *= $p<0.01$, **=$p<0.001$, in comparison to the performance of the first day.

2.2.3 Effects of Position:

The performance levels in figures 6 and 7 reflect averages across all positions in the array. We wanted to determine the visuotopic specificity of the learning, in particular whether it occurred globally across the entirety of the visual field or if it happened over a sequence of locations. We analyzed the improvement in performance on object recognition at each location of the array as the training progressed (Figure 8). The target object appeared randomly and an equal number of times at each location of the array per session to avoid biasing learning to any specific location. Despite this, the increase in performance occurred over a sequence of locations, with the subject initially detecting the target correctly in a small number of nearby positions, and then gradually spread to the whole array.

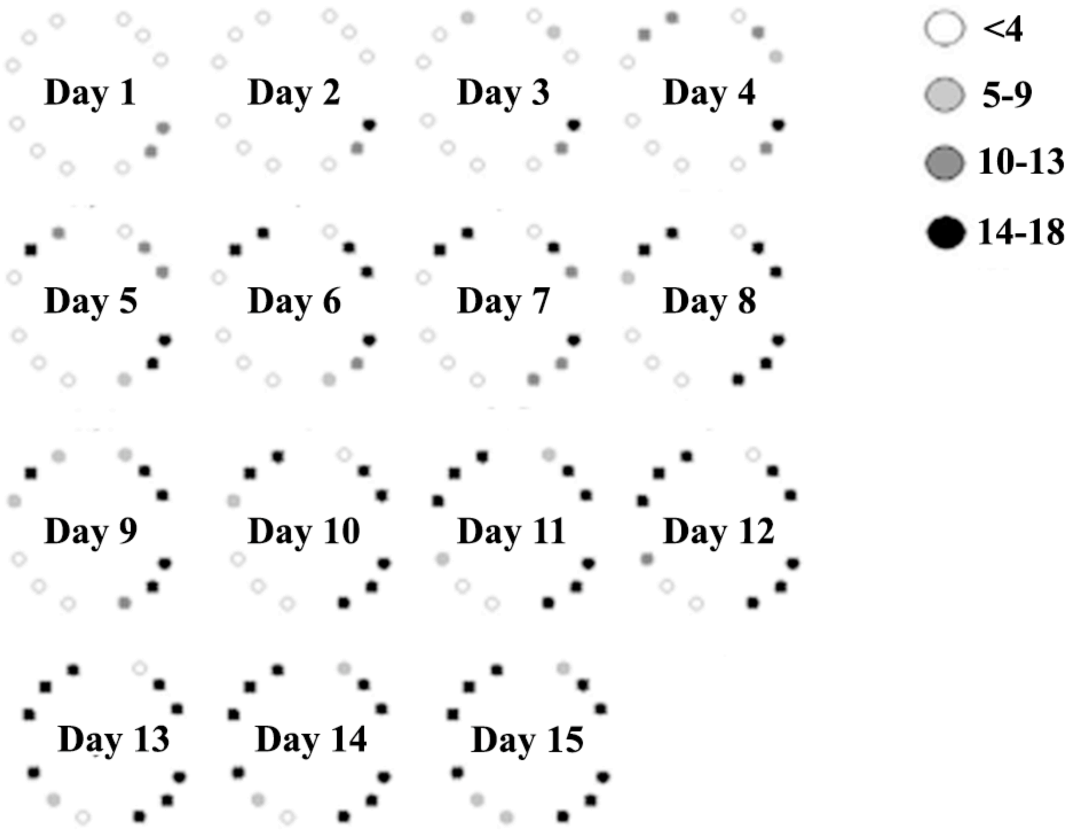


Figure 8. Point by point learning within the array. The target position was changed from trial to trial, in a random block design, for a total of 18 presentations per position. The shading of the squares indicates the level of performance at each day of training. Although the sequence of target presentation was random, the learning did not emerge evenly at all positions, but tended to develop in a sequence of positions over the training period. Single subject.

2.2.4 Transfer between objects that share components

One of the central questions concerning the mechanisms of object recognition is whether the brain stores information in the form of whole objects or as parts of objects. Visual psychophysics can help us determine the answer to this question by showing what is being learned during the perceptual learning of a novel object. To accomplish this, we have looked at the transfer of learning between objects. Based on two prevalent models of object recognition, there are two alternative possibilities for how objects are represented in the brain. If a holistic system of object recognition were at work, one would expect that the training would be specific to the trained shape. A parts-based mechanism on the other hand would result in a transfer of training from trained to untrained objects that share those components that contribute to the recognition of the trained objects. We therefore measured performance before training on the object to be trained as well as on several other objects that either shared or did not share components with the trained object. We then measured the performance of the subjects on recognizing both the trained and untrained objects after the period of training on the target (Figure 9). There was significant improvement in the recognition of objects that shared components with the trained target (before training=27.7±10.0%; after training=54.0±6.4%;

$p<10^{-3}$, two-tailed paired t-test, average of eight subjects) while objects that did not share any components with the trained target did not show significant improvement (before training=28.1±8.3, after training=28.7±13.1, $p>0.8$, two-tailed paired t-test). After training, subjects recognized objects sharing components with the trained target significantly better than those that did not (significance $p=10^{-16}$, two-tailed paired t-test). This effect was seen for a variety of object types, for repeating the same experiment with more complex objects yielded similar results (Figure 10).

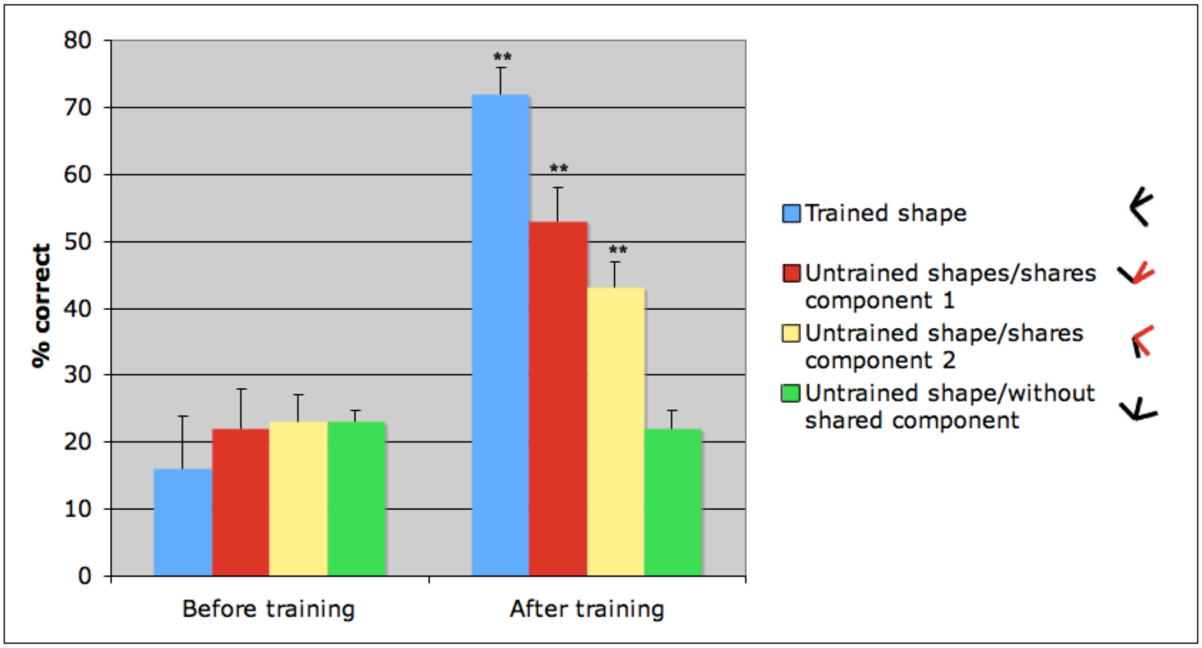


Figure 9. Performance on recognition of trained (blue) versus untrained shapes that either shared (red and yellow) or did not share (green) a component with the trained shape. For the purposes of this illustration the shared components are highlighted in red. There was significant improvement in recognizing untrained shapes that shared a component with the trained shape, but not for shapes with no unshared components. One subject, error bars represent standard errors across subjects. *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

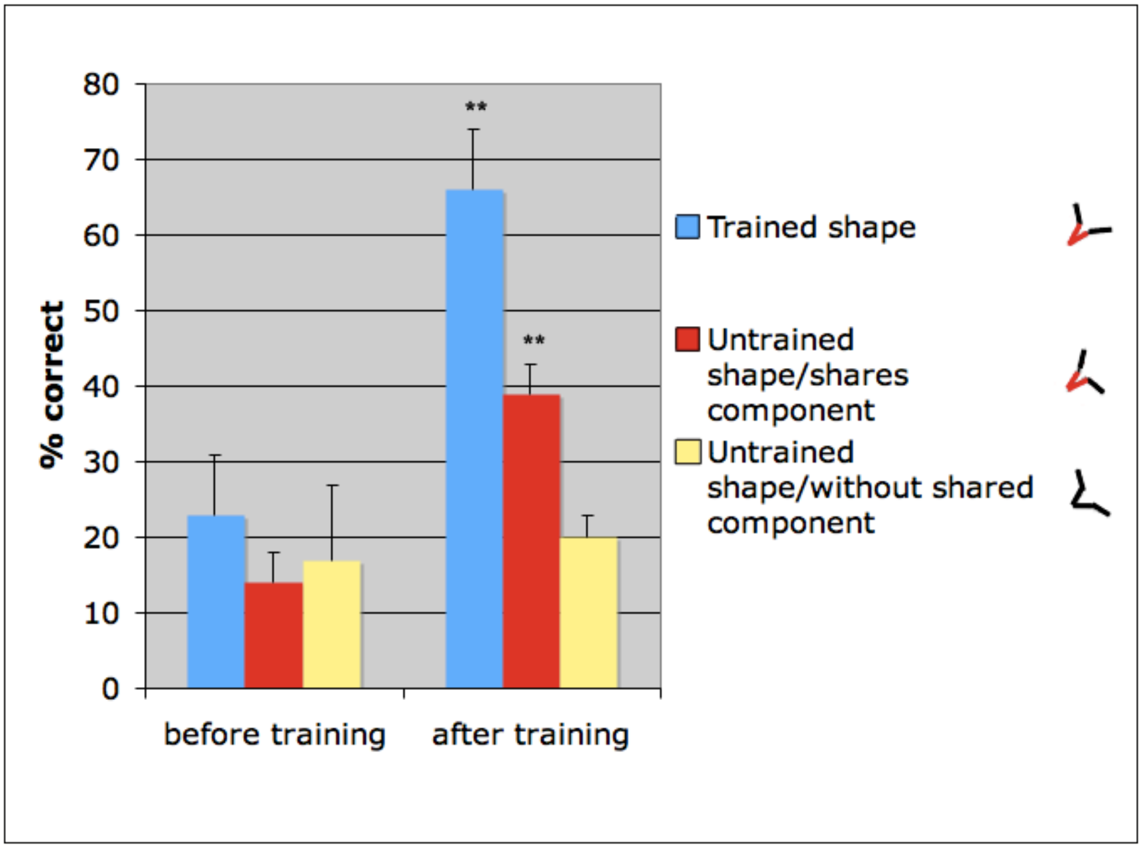


Figure 10. Performance on recognition of trained (blue) versus untrained shapes that either shared (red) or did not share (yellow) a component with the trained shape, for 4-line shapes. For the purposes of this illustration the shared components are highlighted in red. There was significant improvement in recognizing untrained shapes that shared a component with the trained shape, but not for shapes with no unshared components. Here training and transfer for 4-line shapes followed the same pattern as for 3-line shapes. Single subject, error bars represent standard errors across individual blocks. *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

2.2.5 Transfer from objects to components

If the components are indeed important for the transfer we observed, then it is likely that training in an object would increase the subjects' performance in recognizing objects composed of only of a single component of the trained object. To test this, we trained subjects in the recognition of a target object made up of three lines. Once they reached to ~70% performance, we tested their ability to recognize two-line components of this object within arrays of two-line distractors (Figure 11). We observed that components of the trained objects were recognized at a higher performance level (before training=7.4±2.3%; after training=45.0±6.3%, two-tailed paired t-test, two subjects) by the subjects, than two-line objects that were not components of the trained object.

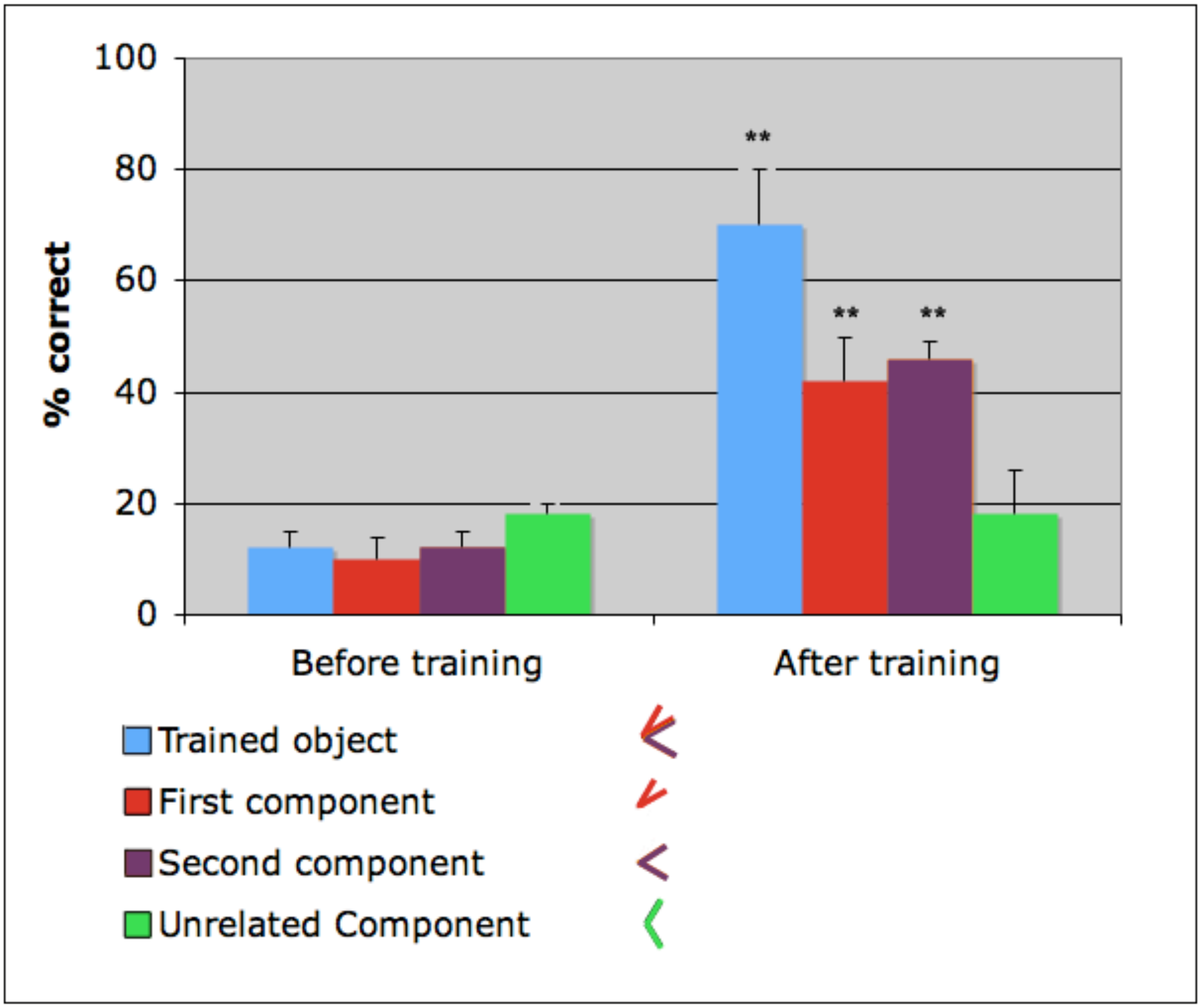


Figure 11. Performance on recognition of trained (blue) shape versus components that were either part (red and purple) or were not part (green) of the trained shape. For the purposes of illustration, the components that were part of the trained shape are highlighted in red and yellow. There was significant transfer to both components of the shape, but not to the unrelated component. Single subject, error bars represent standard errors across individual blocks *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

Similarly, if the components are instrumental in the transfer of learning between objects, then one would expect that improvement in the ability to recognize a trained simple shape that is a component of a more complex shape would improve a subject's ability to recognize the more complex shape. We have tested this by training subjects to recognize two-line objects among an array of objects of similar complexity (Figure 12). In these experiments the distractors were chosen to match the complexity of the target, e.g. two-line distractors for the trained two-line shape, three line distractors for the untrained three-line shape. This ensured that the targets did not automatically pop-out from the distractors by making the target/distractor difference too obvious. The degree of improvement in the components was comparable to that observed when training subjects on the more complex three-line objects (before training=21.3±4.5, after training=61.8±2.7, $p<10^{-17}$, two-tailed paired t-test, single subject). After training was completed, we looked for improvements in the recognition of three-line objects. The subjects showed increased performance at detecting objects that contained the trained components (before training=22.2±5.1, after training=41.8±6.1; $p<10^{-7}$, two-tailed paired t-test), but not at detecting objects without the trained component (before training=10.6±3.3, after training=12.3±3.8; $p>0.6$, two-tailed paired t-test).

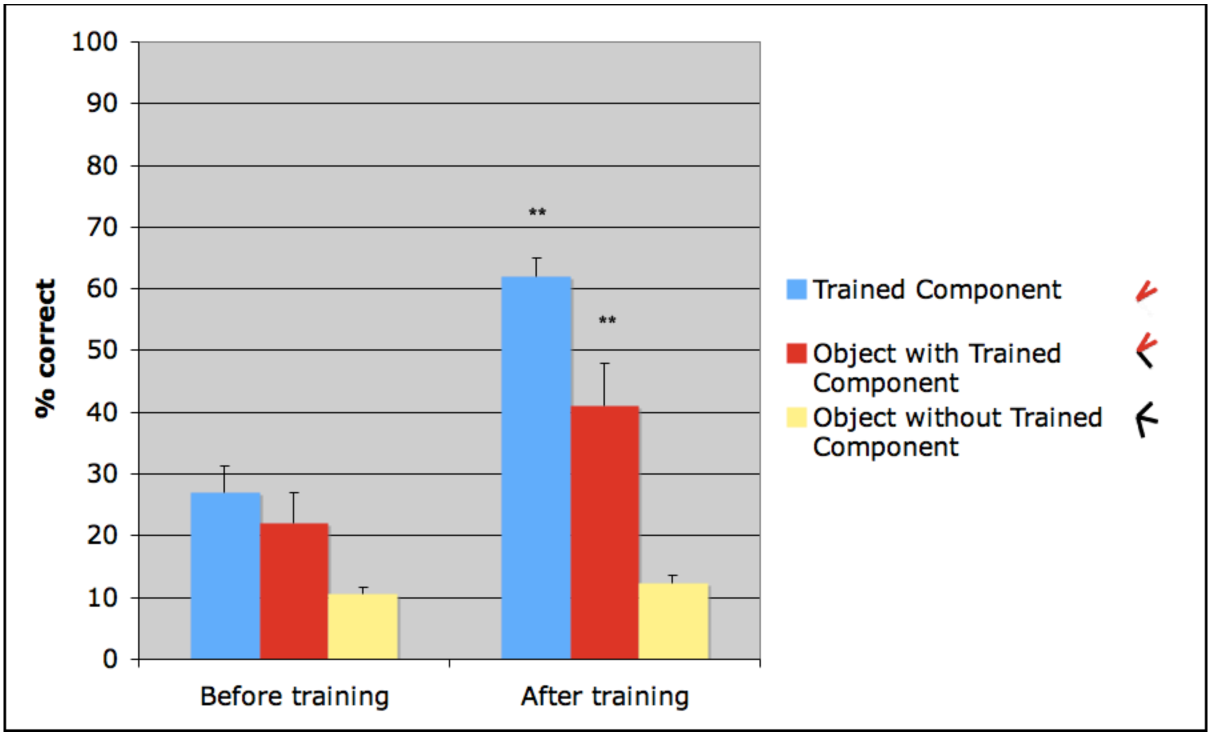


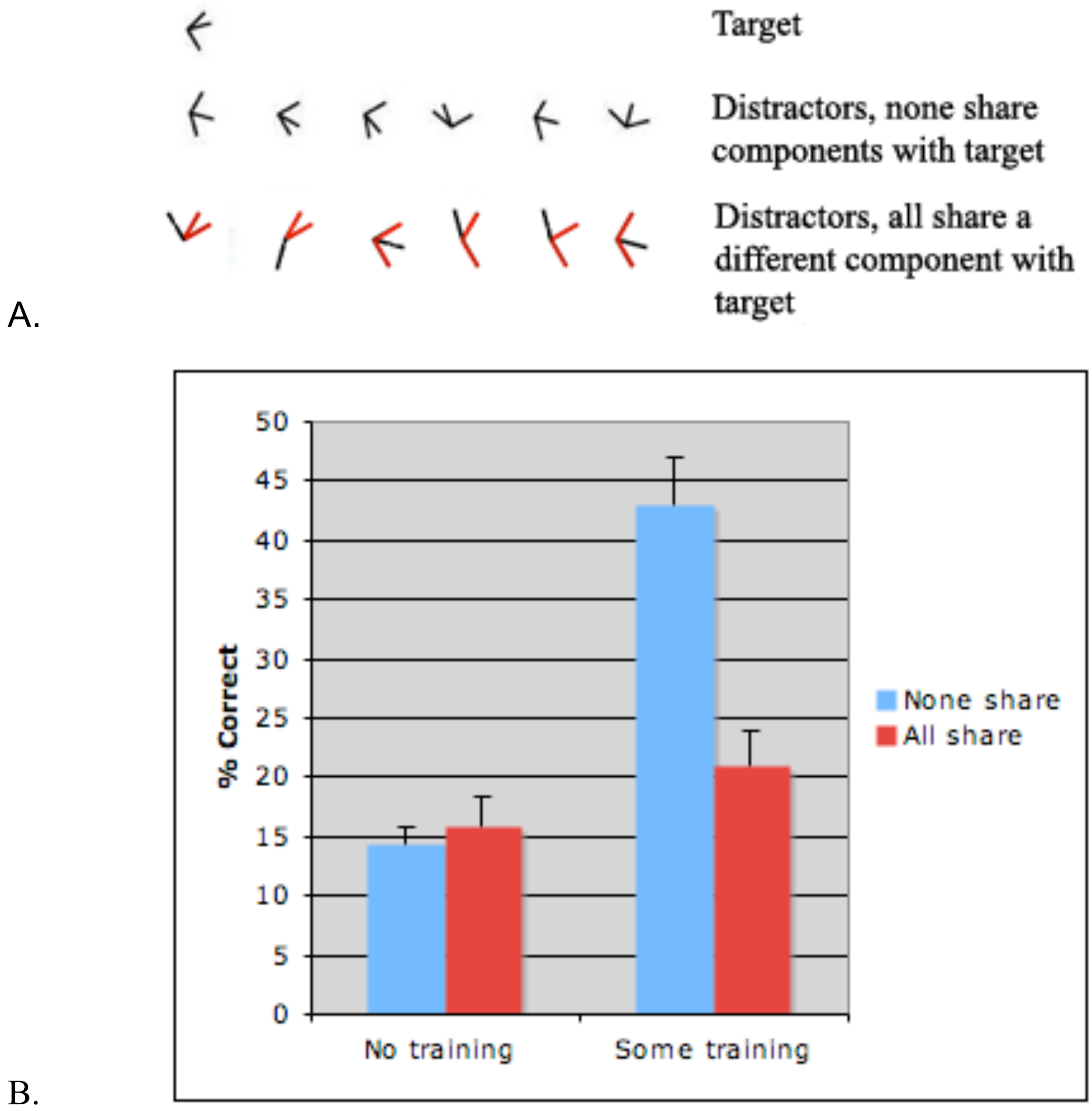
Figure 12. Performance on recognition of trained (blue) component versus untrained object that contains the trained component (red). For the purposes of illustration, the component is highlighted in red. There was significant transfer of training from the component to the object. Single subject, error bars represent standard errors across individual blocks. *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

2.2.6 Effect of distractor similarity on performance:

Objects do not appear in isolation in natural environment, but together with numerous other objects that bear a variety of relationships to the target object. In order to investigate the effects of such an environment on object recognition, we studied how relationships of the shape of distractors to that of the target influenced recognition. For this experiment, we compared the performance of subjects to recognize target objects under two different conditions. The first condition was one where none of the distractors shared components with the target, to simulate a situation where the object was present in a background that shared no features with the target. The next condition was one where all distractors shared a component with the target. In each condition six different distractors were used. In the second condition each of the three components of the target were shared with two of the distractors. Under both conditions, naïve subjects performed at chance level with little observable difference. However, there was a significant reduction of performance in the condition of shared components for subjects that had some experience with the target shape (performance with shared components in distractors: 20.9± 14.8%, performance without shared components in distractors: 43.8± 14.9%, $p<10^{-4}$, two-tailed paired t-test, Figure 13, single subject).

Figure 13. A) Target and Distractors. In the first condition, none of the distractors shared components with the target. In the second condition, each distractor shared one component with the target. For the purposes of this illustration the shared components are highlighted in red. Each component of the target object appeared in two of the six distractors. B) Performance on recognition of a target shape when no distractor shared components with it (blue) versus when all of them did (red). Performances are shown when the subject was untrained (left) and partially trained (right). The components that the distractors shared with the target are highlighted in red. Performance was at chance level for both conditions without training, but was reduced for the condition where the distractors shared components with the target with training. Single subject, error bars represent standard errors across individual blocks.

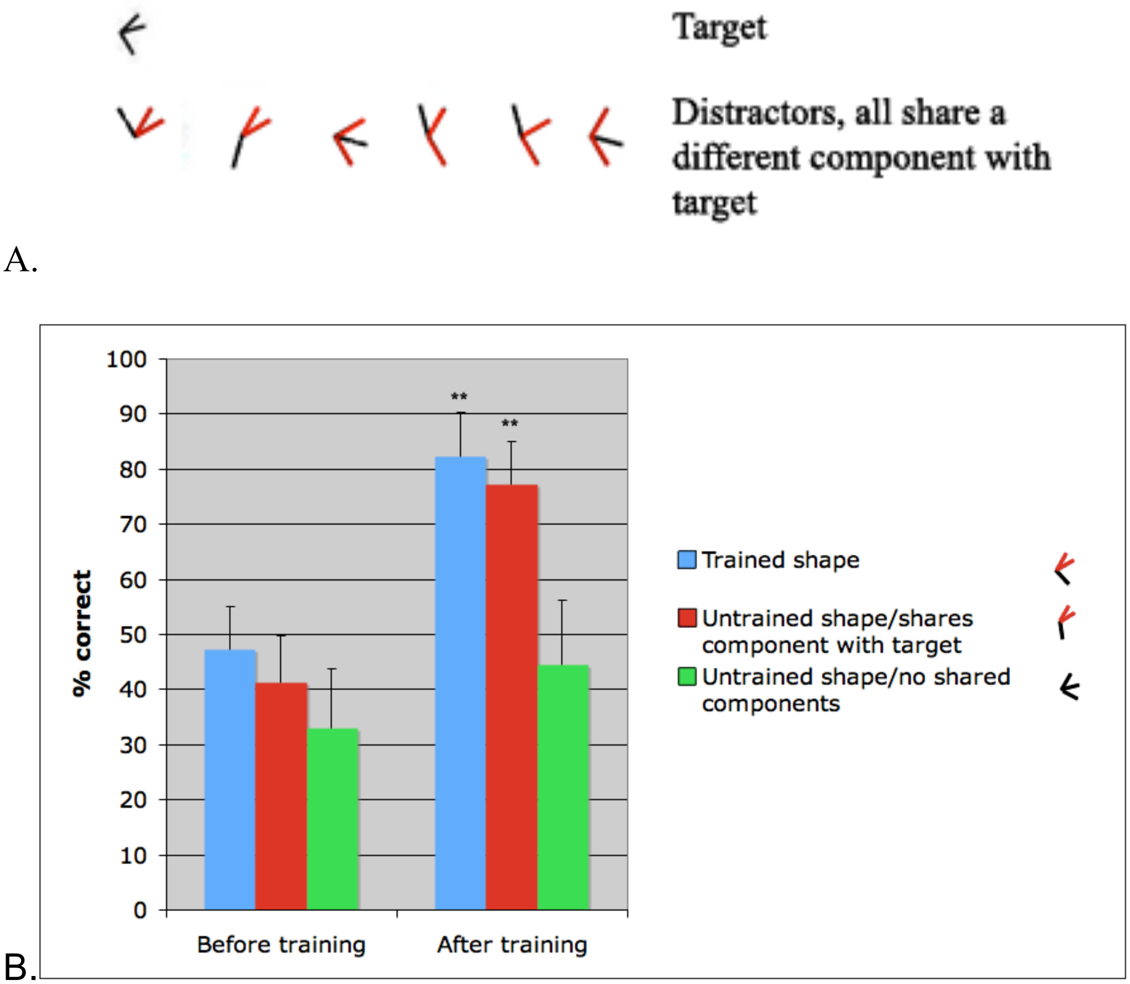
Figure 13


Target
Distractors, none share components with target
Distractors, all share a different component with target
A.
50
45
40
35
30
25
20
15
10
5
0
% Correct
No training
Some training
🟦 None share
🟥 All share
B.

Since there was such a significant effect on performance, one might expect that perceptual learning of the object would be affected as well. To test how learning is affected by distractors sharing components with the target, we trained subjects under the condition where all distractors shared components with the target object. Even after extended training, none of our subjects showed appreciable improvement in their levels of recognition. Since performance in difficult search tasks is proportional to the number of distractors (Bergen and Julesz 1983, Steinman 1987, Treisman and Gelade 1980); we reduced task difficulty by reducing the number of shapes present in the array from twelve to eight. This had the effect of increasing performance in recognizing the target before training (42.2% ±6.7%, average of five), and also made it possible for the subject to increase performance as a result of training. After successful training to a performance of 65% or higher correct detection, (72.3%±6.7%, significance of change after training $p<10^{-5}$, two-tailed paired t-test, average of five) we looked at the transfer of this training to objects sharing components with the trained target. Even though the components of the trained object were shared with the distractors used during training, we nevertheless observed a significant transfer to the objects that shared components with the target (Figure 14). Furthermore, transfer was seen both for objects that served as distractors, as well as to those that did not. As before, no transfer was observed to a control shape that shared no components with the trained target (correct detection before training 30.2%,±5.9, after training 41.1%±8.8%, $p>0.01$, two-tailed paired t-test).

Figure 14. A) Target and Distractors. During training, each distractor shared one component with the target. For the purposes of this illustration the shared components are highlighted in red. Each component of the target object appeared in two of the six distractors. B) Performance on recognition of trained (blue) versus untrained shapes that either shared (red) or did not share (green) a component with the trained shape, after training in a condition where all distractors shared components with the trained target. For the purposes of this illustration the shared components are highlighted in red. There was significant improvement in recognizing untrained shapes that shared a component with the trained shape, but not for shapes with no unshared components. Single subject, error bars represent standard errors across individual blocks. *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

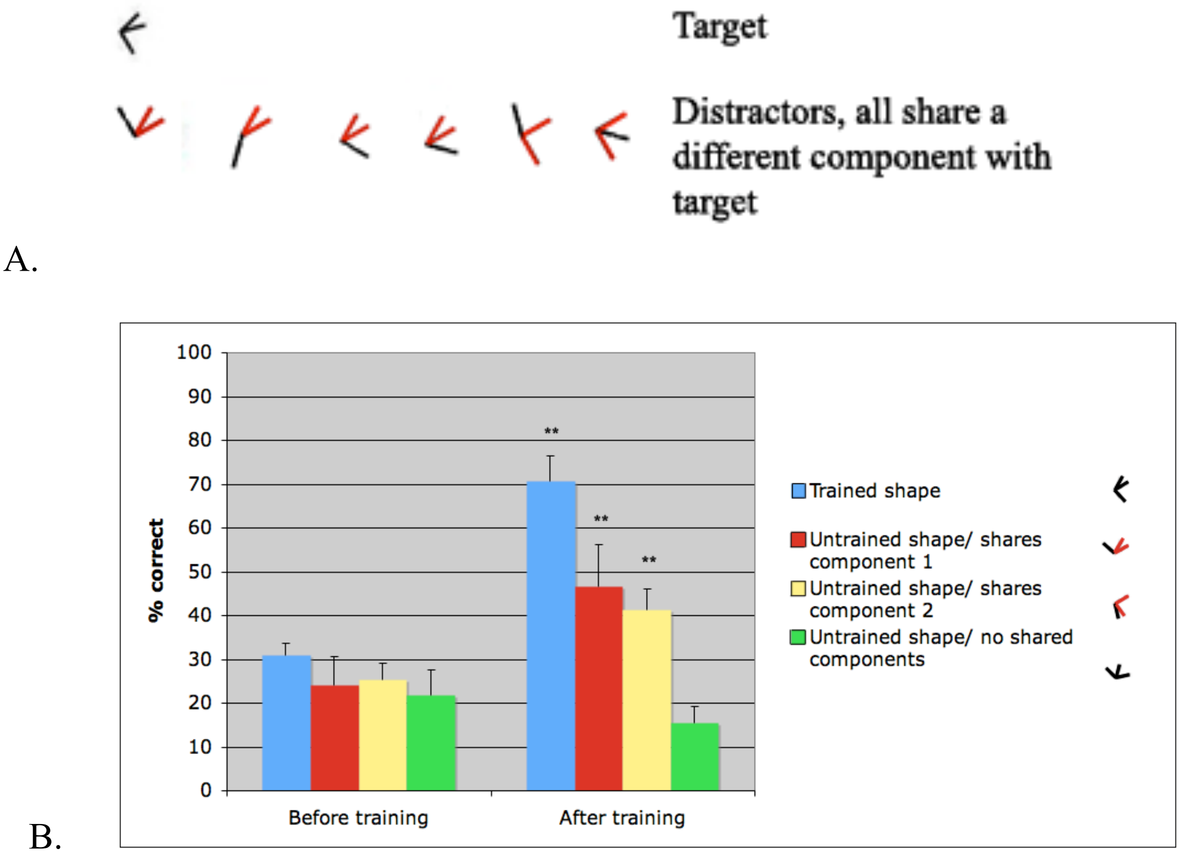
Figure 14


Target
Distractors, all share a different component with target
A.
100
90
80
70
60
50
40
30
20
10
0
% correct
**
**
Before training
After training
Trained shape
Untrained shape/shares component with target
Untrained shape/no shared components
B.

If components of an object were learned solely on the basis of how informative they were, then one might expect there to be significantly more transfer of learning to the more informative components. We manipulated the "informativeness" of individual components in distinguishing the trained object from its distractors by changing how frequently the components appeared among the distractors, and then measuring the influence of the frequency with which components were shared with the distractors on transfer of training (Figure 15). Of the total of six objects used as distractors, four shared one component with the target, and the remaining two shared the other component. As a result, one component was on the average displayed twice as often the other component within the stimuli. We trained our subjects under this condition until their performance reached an arbitrary chosen ~70% level (performance before training= 30.8±5.0%, performance after training= 70.7±10.2%, $p<0.0058$, average of three subjects). After subjects reached saturation in their performance, we measured transfer of detection to other objects sharing either component with the trained object. The average performances of the subjects were significantly higher for objects sharing the more commonly occurring component with the trained target after training, but only if the objects were tested with the same distractors used during training (performance before training= 24.1±11.5%, performance after training with different distractors= 32.3±12.6%, with the same distractors= 46.5±16.6% average of three subjects, $p>0.16$ and $p<0.023$, respectively, two-tailed paired t-test). There was also significant post-training transfer of learning to an object that shared the less commonly occurring component with the trained target when presented with the distractors used during training (performance before training=25.2±6.6, performance after training=41.4±8.0%, $p<0.02$, two-tailed paired t-test, average of three subjects). This experiment was also repeated with the frequencies of the target components among the distractors flipped, i.e. the component that appeared in two distractors now appeared in four and vice versa. After training there was again a high degree of transfer to objects sharing either component with the target, but this did not appear to depend on the frequency with which the component appeared in the distractors.

Figure 15. A) Target and Distractors. During training, each distractor shared one component with the target. For the purposes of this illustration the shared components are highlighted in red. One component appeared in four of the six distractors while the other component appeared in the remaining two. B) Performance on recognition of trained (blue) versus untrained shapes that either shared (red and yellow) or did not share (green) a component with the trained shape. There was significant improvement in recognizing untrained shapes that shared either component with the trained shape, but not for shapes with no shared components. Single subject, error bars represent standard errors across individual blocks. *=p<0.01, **=p<0.001, in comparison to the pre-training levels of performance.

Figure 15


Target
Distractors, all share a different component with target
A.
100
90
80
70
60
50
40
30
20
10
0
% correct
**
**
**
Before training
After training
Trained shape
Untrained shape/ shares component 1
Untrained shape/ shares component 2
Untrained shape/ no shared components
B.

2.3 Discussion:

We studied what is being learned in object recognition by training subjects on a visual search task and looking for transfer from trained to untrained objects. Previous research has shown that performance in recognizing objects in this type of task is subject to perceptual learning. The learning in our task was comparable to that seen in early studies, with steady improvement over several days that progressed in a location-specific manner (Sigman and Gilbert 2000). In the current study we saw significant transfer of learning between objects that shared components. This suggests that novel objects are learned in a parts-based fashion. As further support of this idea, learning of an object was accompanied by learning of its individual components. Conversely, improvement in recognition of a simple object improved the ability to recognize more complex objects that contained the trained object as a component. Both of these observations further support the notion that parts are important for object recognition.

Psychophysical experiments show that components are necessary for recognition of objects, objects can be identified by partial exposure to a subset of their components, and similar objects can easily be differentiated through differing parts (Biederman 1987, Biederman and Gerhardstein 1993, Biederman and Bar 1999). Here we showed that object components were used for the learning of novel objects, and same components could be used to recognize multiple objects. Furthermore we observed that transfer of learning between objects that shared one component was not complete. This observation, coupled with the evidence that all components of an object were learned during training, indicates that combinations of multiple components were necessary for recognition.

It should be noted that in certain instances there was no transfer of training to an object that shared components with the trained target. In these cases the object either had an extremely different orientation from the trained target, or the third line of the object was placed in the middle of the shared component. Two conclusions can be drawn from these observations: the orientation of an object is used in conjunction with its components for recognition, and a component needs to be clearly visible and undivided to be used.

The visual shape of the distractors affected the recognition of the target as well. The presence of components of the target within distractors severely reduced both performance and learning. According to the fragment-based hierarchy model of recognition by Ullman (2007), components of an object are most useful for recognition when they are highly informative about the object, i.e. when they appear often within the object and rarely in the environment. Therefore by using distractors that contain components of the target shape, we reduced the usefulness of those components for the recognition of the target. Even in this condition, however, it appeared that parts were still being used for recognition. Although each component appeared among the distractors, it could still be used as a feature in recognition of the object, as evidenced by the transfer of learning to objects sharing that component. In this experiment no single component uniquely identified the object, and therefore even components that occurred less frequently among the distractors contributed to identifying the target. It is likely that the visual system picks out the target by performing an “and” operation, requiring the presence of multiple components to recognize the object.

It has been suggested that objects in a visual search task pop-out if they differ from the distractors by an elementary feature. If, on the other hand, they differ by specific combinations of elementary features, then the search is inefficient (Treisman and Gelade 1980). Our study shows that with practice specific combinations of object parts can become elementary features. It has been also suggested that performance in visual search is not a dichotomy of parallel and serial search, but a continuum (Wolfe et al. 1989, 1997, Joseph et al. 1997, Wolfe 2003). The gradual improvement of performance rather than an abrupt switch from poor to good performance in this study supports this possibility. In our experiments the distractors appear to have a significant effect on how easily the target can be recognized. We have seen that the number of distractors has a direct effect on task difficulty. Untrained performance increased noticeably when the size of the stimulus array was reduced from 12 objects to 8 objects, and training became possible in conditions where it wasn't with the larger array. This result is in good agreement with earlier studies that show that in an inefficient (serial) search task the difficulty increases with increasing number of distractors (Bergen and Julesz 1983, Steinman 1987, Treisman and Gelade 1980). It should be noted that on some instances, pre-training performance was noticeably higher than the predicted chance level performance, especially with lower number of distractors. Even during the early stages of the training, it is possible for the subjects to limit their attention at multiple locations of the array, such as the entirety of a specific quadrant, enabling them to detect the target at these locations somewhat reliably. This potentially allows for initial performances above chance level. However, there was still a statistically significant improvement in performance as a result of training.

The results here, supporting the representation of objects by their component parts, resonates with findings on the feature selectivity of neurons in monkey inferotemporal cortex. There, a large fraction of neurons are sensitive to simplified parts of objects, and objects activate cortical columns that are selective for their components (Desimone et al. 1984, Tanaka et al. 1991, Tsunoda et al. 2001). fMRI studies suggest a similar organization within the human LOC (reviewed in Grill-Spector et al. 2001). These properties fit very well with object recognition mechanisms that are parts-based. However, our results show that learning of new objects occurs in a location-specific fashion, which is generally thought not to be a property of LOC or of inferotemporal cortex, but of areas at earlier stages in the ventral visual stream. A top-down reorganization of the object processing to early visual cortices was suggested by earlier studies (Sigman and Gilbert 2000, Sigman et al. 2005), and can account for the pattern of learning we observe. Cells in V1 can respond to more complex features than originally believed (Das and Gilbert 1999, Posner and Gilbert 1999, Gilbert et al. 2000, 2001, Li and Gilbert 2002, Li et al. 2004, 2006, 2008). Furthermore, we have seen a considerable measure of plasticity in adult V1, both at the level of receptive field properties (Gilbert and Wiesel 1992, Obata et al. 1999, Crist et al. 2001) and in the capacity of local circuits in V1 to undergo sprouting and synaptogenesis (Darian-Smith and Gilbert 1994, Gilbert et al. 1996, Stettler et al. 2006, Yamahachi et al. 2008). In the context of the current study, this plasticity can be used to reorganize elementary feature maps to represent object parts and their specific combinations. How different areas along the visual form pathway contribute to the representation of object features will continue to be elaborated. But one potential advantage of shifting feature representation towards earlier cortical stages is an increased ability to recognize objects rapidly and in parallel with other, similar objects.

CHAPTER 3: FUNCTIONAL MRI

In the classical view of the brain's representation of visual form, early visual cortices analyze local attributes, such as orientation, and subsequent levels of visual cortical processing are selective for more complex shapes. Along with this "complexification" of receptive field properties along a hierarchy of visual cortical areas, there is presumed to be a decrease in retinotopic specific, with small, retinotopically organized receptive fields at early areas, and large receptive fields showing translational invariance at higher level areas (Hubel and Wiesel 1962, 1965, 1968, Tanaka et al. 1991). The brain regions involved with a task can be inferred by how much the performance in the task is affected by the location of the task within the visual field. The brain regions with matching retinotopic specificity are usually responsible for the computations necessary to accomplish the task (Berardi and Fiorentini 1987, Karni and Sagi 1991, Sigman and Gilbert 2000). Complex objects contain numerous features that need to be integrated for their recognition, and this recognition is usually independent of the visual field location. Under normal circumstances, both integration of features and locational invariance are properties of regions high in the visual processing hierarchy, which are therefore implicated as being responsible for object recognition. Studies using electrophysiology in monkeys and fMRI in humans have implicated the inferotemporal cortex (Desimone et al. 1984, Tanaka et al. 1991, Tsunoda et al. 2001) and lateral occipital cortex (Malach et al. 1995, Grill-Spector et al. 1998, 2001, Kourtzi and Kanwisher 2000) as the areas in which object shape is represented.

Recent studies have questioned the classical view of object recognition. Early visual centers like V1 are shown to be sensitive to more complex stimuli than originally thought (Das and Gilbert1999, Posner and Gilbert 1999, Gilbert et al. 2000, 2001; Li and Gilbert 2002, Li et al. 2004, 2006, 2008). Furthermore, familiar objects can be recognized within arrays of distractors efficiently, independent of the size of the array (Wang et al. 1994, Malinowski and Hubner 2001, Shen and Reingold 2001), which is a property of features mapped in early visual cortices (Treisman and Gelade 1980). Individuals can gain such efficiency at the recognition of novel objects through perceptual learning that is retinotopically specific, again suggesting a role of earlier visual cortices in the process (Sigman and Gilbert 2000). Based on the results of these studies it is suggested that training causes the visual processing system to reorganize itself so that the processing of the learned objects are remapped to early visual cortices. fMRI studies offer functional support for these behavioral observations. In a block-based design of detecting T’s of a specific orientation among T's of other orientations, an extended network mainly consisting of parietal and frontal cortices and lateral occipital cortex is activated for untrained shapes. After training, on the other hand, the trained shapes lead to higher activation of a region in the middle occipital cortex, corresponding to early visual regions (Sigman et al. 2005).

In the current study we combine retinotopic mapping and the use of functional localizers to identify the visual cortical areas in human subjects (Malach et al. 1995, Grill-Spector et al. 1998, 2001, Tootel et al. 1998Kourtzi and Kanwisher 2000) enabling us to determine which areas show differential activation and connectivity as a result of perceptual learning trained on an object recognition task. This allows us to investigate how the cortical representation of objects changes as a result of perceptual learning by using a visual search task.

3.1 Methods:

3.1.1 Subjects:

Subjects were three adults (two female, one male, all right-handed) of below 30 years of age. They were recruited according to the regulations set forward by the Rockefeller University Institutional Review Board, and gave written informed consent. All subjects had good or corrected vision. Each subject participated in two scanning sessions, one for functional mapping of the brain and one for functional scan during visual search task.

3.1.2 Stimulus generation, scan and processing:

All the stimuli were generated using E-Prime program for visual stimulus generation (Schneider et al. 2002), which was synchronized with the MRI scanner using the Integrated Functional Imaging System (IFIS). The stimuli were displayed using an arrangement of mirrors, with the final mirror placed in front of the subjects' faces. The screen size was ca. 12x9 degrees. Scans were performed with a GE-Sigma 3-Tesla MRI scanner (maximum gradient strength, 50 mT/m; maximum gradient slew rate, 150 T/m/s) with an 8-channel headcoil at the Bioimaging Core Facility of Weill Medical College of Cornell University in New York. 27 slices of 4mm thickness were acquired for the functional scan, with a sampling rate of TR=1.5 s/volume. The whole brain was sampled in an anterior-posterior direction. Immediately before the functional scanning an anatomical image of the subject's brain was acquired using a sagittal 3D-MPRAGE sequence with acquisition matrix size of 256 × 192 × 120, a field of view of 24 cm, and slice thickness of 1.5 mm.

For analysis of the fMRI signal acquired from the scanner, we used SPM5 (see http://www.fil.ion.ucl.ac.uk/spm/software/spm5/) software in the MATLAB (MathWorks, Inc., Sherborn, MA) environment**.** Standard image processing procedures of SPM5 were performed on the data: The functional images were realigned to correct for small head movement artifacts and coregistered with the anatomical scans of the corresponding subjects. The image acquisition time differences between slices were corrected with an interleaved slice sampling, using the first acquired slice as reference. The Marsbar toolbox (Brett et al. 2002) was used for the ROI analysis of the functional data ROI's generated from functional maps (see below).

3.1.3 Retinotopic mapping of early visual cortices:

To define the areas engaged in the object recognition task, we mapped early visual cortex and lateral occipital cortex (LOC). Early visual cortices are organized in a repeating retinotopic fashion, and can therefore be mapped by determining where each position in the visual field corresponds on the cortex. For this retinotopic mapping we used the following stimuli: a ring to map retinal eccentricity and a rotating double-wedge to map polar angle (Figure 16). Two blocks of each stimulus condition were used, with the ring expanding in one block and contracting in the other, and the wedges rotating clockwise in one block and counter-clockwise in the other. The width of the ring took up a quarter of the maximum radius of the ring. A full expansion or contraction of the ring took 24 seconds and happened 10 times per block. Each wedge described a 45-degree arc, for a total of 90-degrees visual angle for the two wedges. A half rotation of the two wedges took 24 seconds (a half rotation resulted in an image identical to the starting point since there were two wedges) and happened 10 times per block. Both stimuli were constructed of a flickering (4 Hz) checkerboard pattern to avoid visual adaptation. To maintain the attention of the subjects on the screen, they were asked to fixate to a short horizontal red line at the center of the screen and asked to respond when the line changed orientation from horizontal to vertical. This orientation change had a 50% chance of happening every 1.5 seconds, and lasted for 250 ms. The anatomic location in the visuotopic map that corresponds to each pixel on the screen was determined by convolving the activities generated by the two stimuli, as described by Dougherty et al. 2003 (Figure 17).

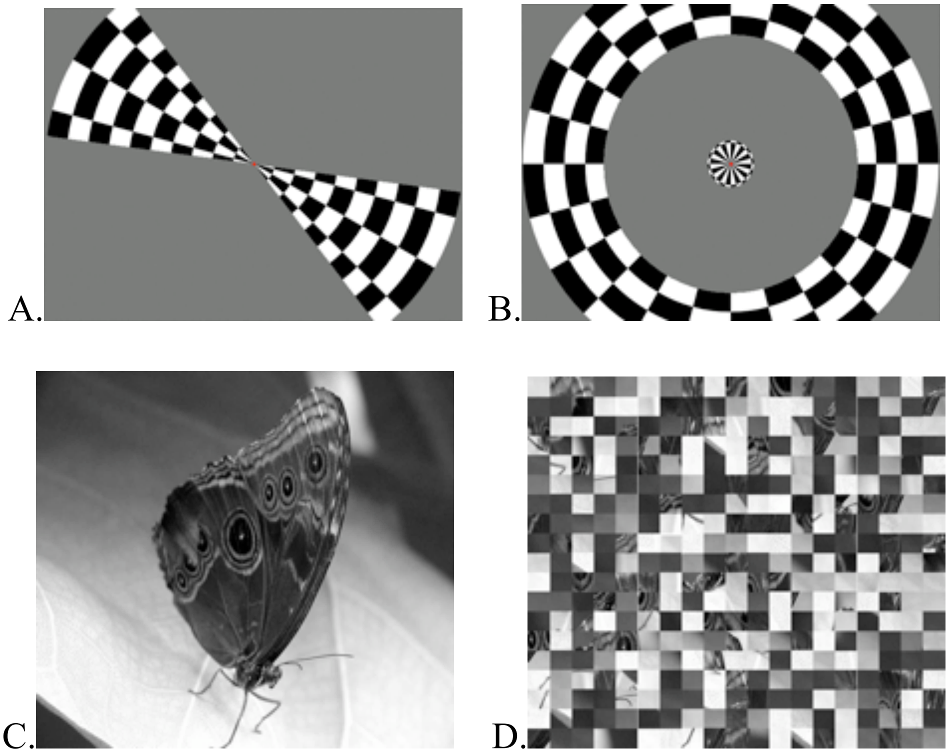


Figure 16. Stimulus images for mapping. A) Rotating wedges for eccentricity mapping B) expanding/contracting ring for polar angle mapping C) objects and D) scrambled objects for functional mapping of LOC

3.1.4 Mapping of LOC:

For the mapping of LOC, retinotopic localizers cannot be used since this region does not have a very specific retinotopic organization. Instead, a functional localizer is used. Since LOC is sensitive to objects, voxels that are activated more strongly by objects compared to scrambled images of objects will belong to LOC. Two types of images were presented at the center of the screen in separate blocks (Figure 16). The first group of stimuli consisted of grayscale photographs of various kinds of both natural and man-made objects, obtained royalty-free from the Internet. The second group was images that were scrambled versions of the same photographs. The images were represented for 250 ms, with 750 ms intervals between them, for a total of 24 images per block. Two runs were conducted, with five alternating blocks per condition per run. Between two blocks there were 12-second rest periods with no stimuli. The subjects were asked to fixate a central location marked by a red cross. The attention of the subjects was engaged by asking them to report if any image was repeated twice in succession, which happened four times per block. The LOC was defined as the regions that showed significantly higher signal in the fMRI data as a response to objects compared to scrambled objects (Kourtzi and Kanwisher 2000, Figure 17).

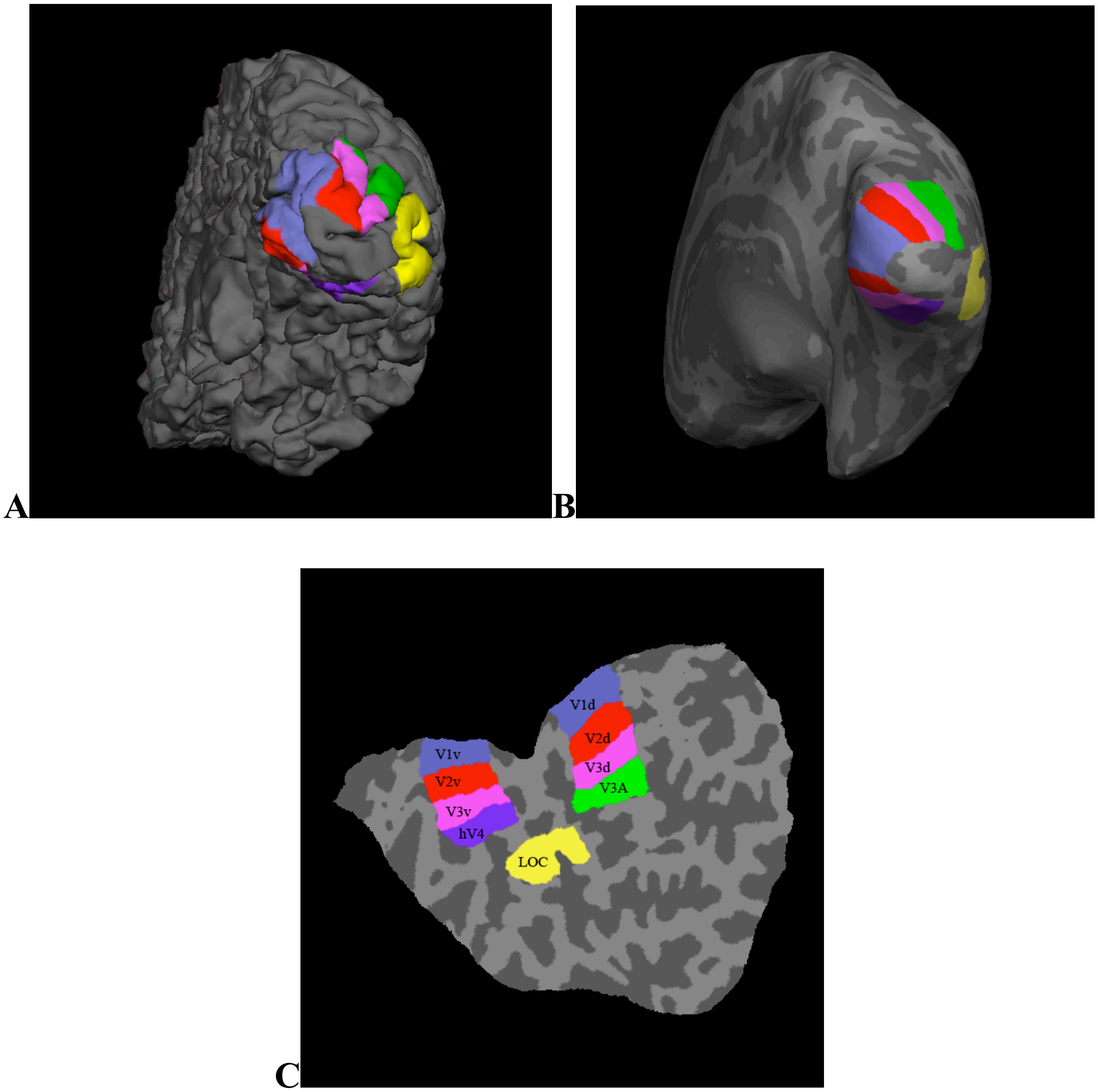


Figure 17. Retinotopically mapped cortices projected onto the A) normal B) inflated and C) flattened images of the right hemisphere of the brain. Both ventral (v) and dorsal (d) segments of V1 (blue), V2 (red), and V3 (pink) are depicted, along with the entire V3A (green), hV4 (purple), and LOC (yellow).

3.1.5 Psychophysical Task:

For the functional imaging, we used a visual search task that involved responding to the presence or absence of a specific target object composed of three lines within a circular array of 12 objects of similar appearance, with the vertical and horizontal meridians of the circle left empty. The approximate radius of the array was 3 degrees and the length of individual line segments of the objects was 0.3 degrees. The stimulus array was presented for 300 ms and subjects were asked to respond within 4 seconds. No visual feedback was provided. Both target and distractors used in this experiment were three short lines of different orientations that connect at one single point. They were displayed in white on a uniform black background, with a circular fixation spot in the center of the array. The array was designed to maintain the same eccentricity for all objects within the array. The target object was present in two thirds of the trials. The location of the target each time it appeared was semi-randomized in a fashion that fulfills the following conditions: a) the target appeared in each individual location an equal number of times and b) the appearance of the target in each quadrant was balanced, i.e. an appearance in each quadrant was followed by an appearance in each other quadrant an equal number of times. Each different object used was presented in blocks of nine trials. The target object for an upcoming block was presented for 3 seconds at the center of the screen at the beginning of each block. The onset of the first stimulus of the block was 6 seconds after this display. The subjects responded to the stimuli by pressing a button under their right-hand thumb when they saw the target, and under their right-hand index finger when they did not. Due to the requirements of synchronizing the task presentation computer with the fMRI scanner, they were not asked to report the location of the objects during the scan session. The total block length was 45 seconds. Each run contained two blocks with each test objects and the target objects alternated in each block. A total of six runs were conducted in each session. At the end of each run there was a 16.5 second long blank screen with only the fixation spot present to allow for the BOLD response to dissipate. The first three volumes of each run are omitted from the analysis to reduce scanner artifacts.

Subjects were trained for the recognition of one target object before any scans were performed. The training consisted of 54 blocks of 9 trials per session, for a total of 486 trials. Sessions were run for 3 to 5 five times per week. In addition to the task setup detailed above, subjects were also asked to identify the location of the target to eliminate false positives, and were given visual feedback. They were trained until their performance at correctly detecting the target object reached an arbitrarily set cut-off point of 70%. During training, stimuli were presented on a Viewsonic Graphics Series G90fB CRT monitor with a refresh rate of 60 Hz. Subjects were seated at 180 cm distance from the monitor. A chinrest was used to ensure consistent distance and reduce head movement.

3.1.6 Analysis:

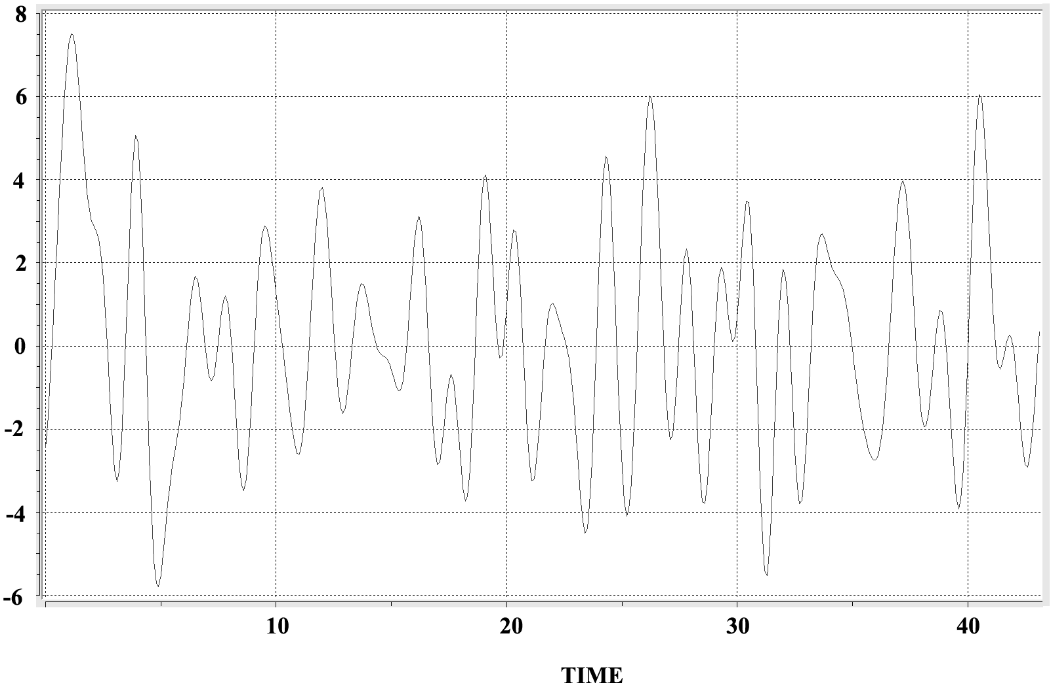
We determined the changes in fMRI activity as a result of training by comparing activity levels in the trained and untrained conditions using the standard statistical procedures of SPM. For this analysis, we used a whole-brain, voxel-by-voxel multiple linear regression model (general linear model). General linear models explain the blood oxygenation level dependent response (BOLD) measured by the fMRI as a linear summation of multiple regressors. The weight of each regressor is fitted in such a way as to minimize the error term of the model. Regressors that are of actual interest (i.e. the variables that are being studied) are called principal regressors. The principal regressors were set to a model of neural activity convolved with a prototypical hemodynamic response function (HRF). The 'canonical HRF' that is built in to the SPM5 software was chosen, which is a mixture of two γ functions that peeks at 6 seconds and later undershoots the baseline with a minimum at 16 seconds, with a time derivative that allows the peak response to vary by plus or minus one second. Trained and Untrained conditions were modeled as principal regressors, weighted negative and positive, respectively. The changes are displayed using TkSurfer tool (linear opaque threshold, min=2.5, max=5, see http://surfer.nmr.mgh.harvard.edu/fswiki/TkSurfer) of FreeSurfer toolset. To improve statistical power, a second analysis was performed using the same statistical approach, but limited to restricted regions of interest (ROI's) rather than the whole brain. The regions previously mapped as retinotopic cortex and LOC were used as ROI's for this analysis. The marsbar toolbox of SPM5 is used to restrict the analysis to these ROI's.

Within the time frame of one block the fMRI activity shows significant fluctuations. If two distinct regions of the brain are involved with the processing of the same task, these regions most likely communicate with each other. As a result, the fluctuations of activity are expected to correlate. To analyze this correlation of activity between brain regions we used the voxels mapped as belonging to V1 as reference. We calculated the average time-course of activity of these voxels, and compared the fluctuations of activity in each individual voxel of the brain to the fluctuations in this average (Fox et al.

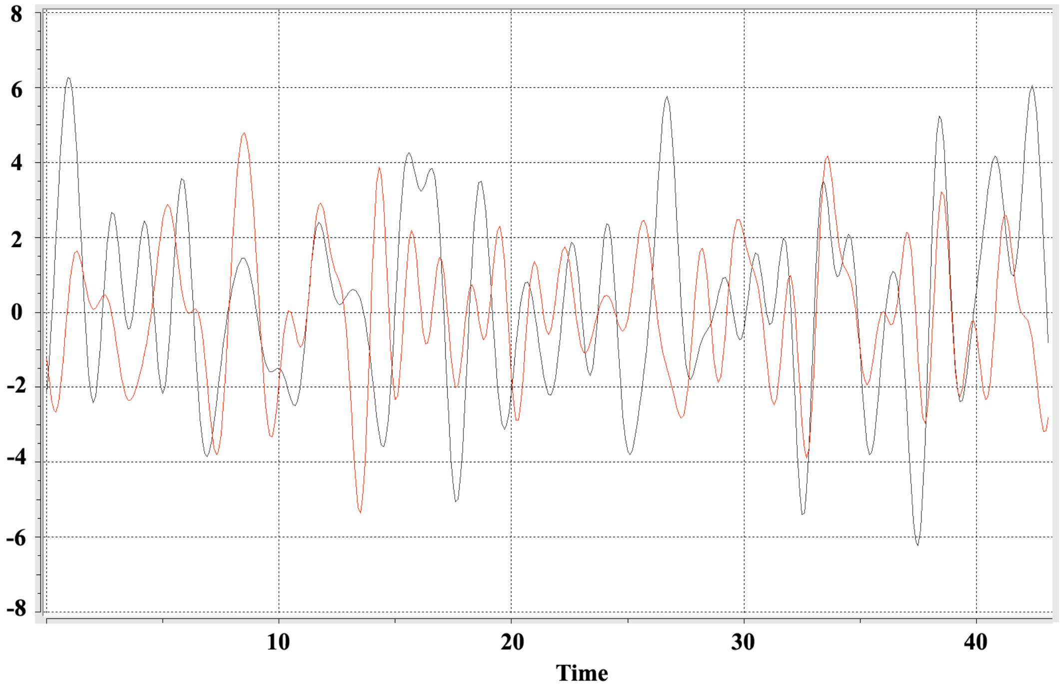
2005, 2006, for a sample comparison, see Figure 18). We determined a correlation coefficient ranging from -1 to 1 for each voxel depending on how well these fluctuations correlated. High levels of correlation corresponded to a high absolute value of correlation coefficient. Positive values indicated a positive correlation while negative values indicated a negative correlation. The results are displayed on brain slices using FSL View v3.0 (threshold min=0.2, max=0.8, see http://www.fmrib.ox.ac.uk/fsl/fslview/index.html), and on flattened brain images using TkSurfer tool (linear opaque threshold, min=0.2, max=0.8) of FreeSurfer toolset. For the comparison of different brain regions, the average correlation coefficient of each region was calculated. For this analysis, the regions that were not mapped functionally were determined through automated anatomical segmentation during initial image processing.

Figure 18. Fluctuations in fMRI activity of V1 over one block. x-axis represents time in seconds, y-axis represents a unitless value corresponding to the activity measured by the MRI scanner. Timecourse of activity in A) V1 B) V1 (black) compared to a voxel in LOC (red) in untrained condition C) V1 (black) compared to a voxel in LOC (red) in trained condition, are shown. The plots of the timecourse of V1 and LOC are better correlated in the untrained condition compared to the trained condition

Figure 18

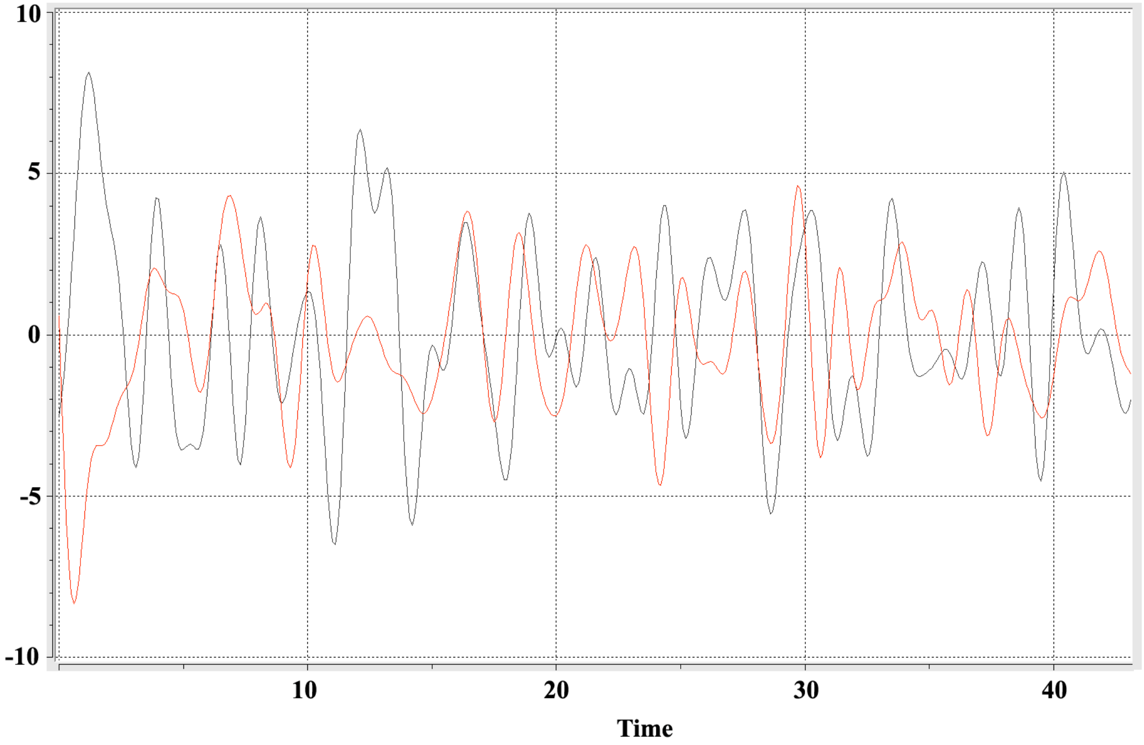


A.



B.

Figure 18 continued



C.

3.2 Results:

3.2.1 Psychophysical performance:

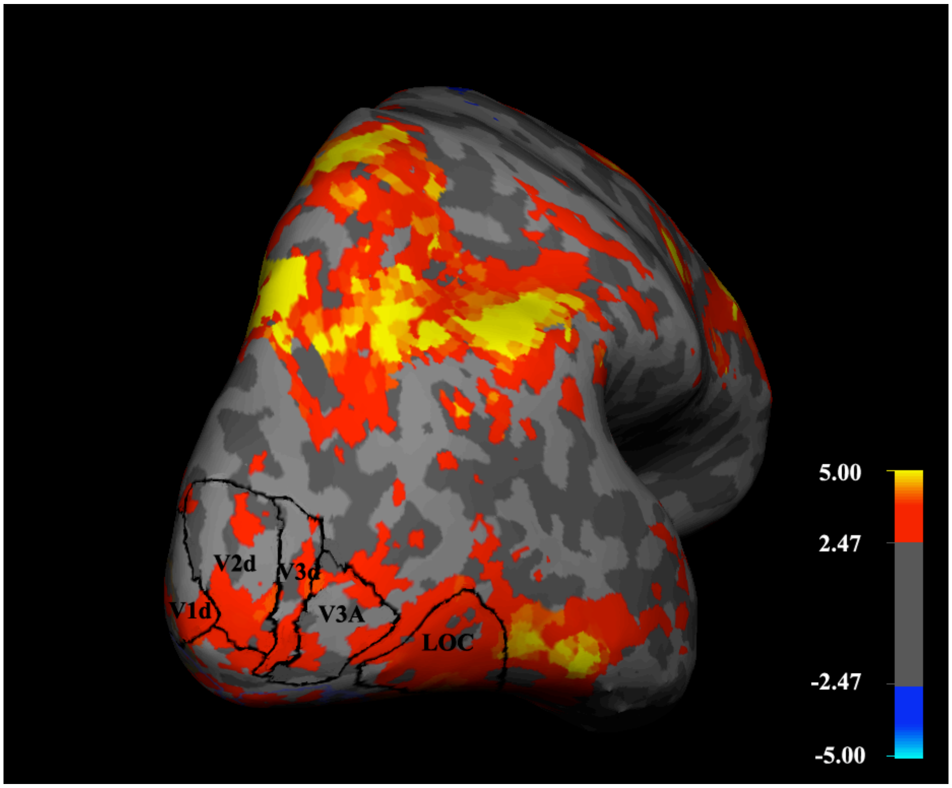
In order to be able to compare cortical activation when searching trained and untrained objects, we needed a perceptual task where the performance showed a strong effect of learning**.** The task we chose was a visual search task that involved the detection of a target object in an array of distractors of similar appearance. In this task, the level of performance was measured as the fraction of trials when the subjects detected the target correctly relative to the total number of trials where the target was present. The performance was corrected for false positives by using the following formula: $p' = (p-fp)/(1-fp)$ where p is the percentage of positive responses, fp is the rate of false positives (rate of trials where the subject reported seeing the object when the object was not present) and p' is the 'real' percentage of correct responses. Subjects were extensively trained in one object before the fMRI scan session as described for the psychophysics experiments until they reached at least 70% correct detection ratio. During the imaging session, subjects were asked to look for the trained shape and an untrained shape in alternating blocks of nine trails each, for a total of 18 blocks per object. Within the MRI scanner, the improvement of performance through learning was maintained for the trained shape, showing that the training was robust to the changing physical conditions (correct detection performance 77.3+-11.6%). The subjects performed more poorly at detecting the untrained target (correct detection performance 21.3+-12.7%).

3.2.2 Effect of object familiarity on cortical activity:

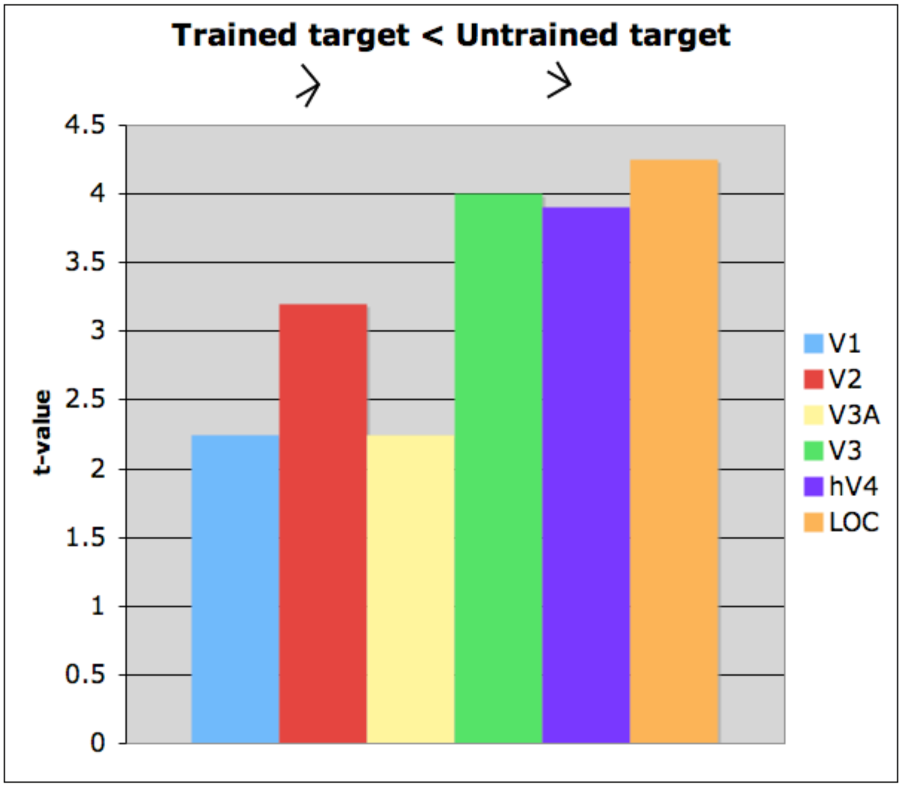
To determine whether cortical representation changed with training we compared the fMRI activity levels when subjects were looking for trained and untrained objects. To increase statistical power, we limited the analysis to specific regions of interests that were early visual cortices in the retinotopic cortex as determined by our mapping procedure, that were implicated as the location of object recognition post-training in earlier publications (Sigman 2005). In all the regions of retinotopic cortex tested, we found that activities are higher for the untrained target compared to the trained target (Figure 19). A similar reduction was observed in the LOC.

Figure 19. A) Cortical regions activated more strongly by untrained shapes compared to trained shapes (yellow), given in t-values of the statistical significance of the difference. In the expected vicinity of the visual cortices, a large area near the lateral posterior side of the brain (marked with blue crosshairs) appeared as more strongly activated in untrained condition. This area of high activity overlapped with the areas that were functionally mapped as LOC. B) fMRI Activity difference between trained and untrained target in V1 (blue), V2 (red), V3A (yellow), V3A (green), and hV4 (purple). For the purposes of this illustration, a positive t-value indicates higher activity in the untrained condition whereas a negative t-value indicates higher activity in the trained condition. For all brain regions analyzed, there was higher activity in the untrained condition.

Figure 19


V2d
V3d
V1d
V3A
LOC
5.00
2.47
-2.47
-5.00

A.


Trained target < Untrained target
>
>
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
t-value
V1
V2
V3A
V3
hV4
LOC

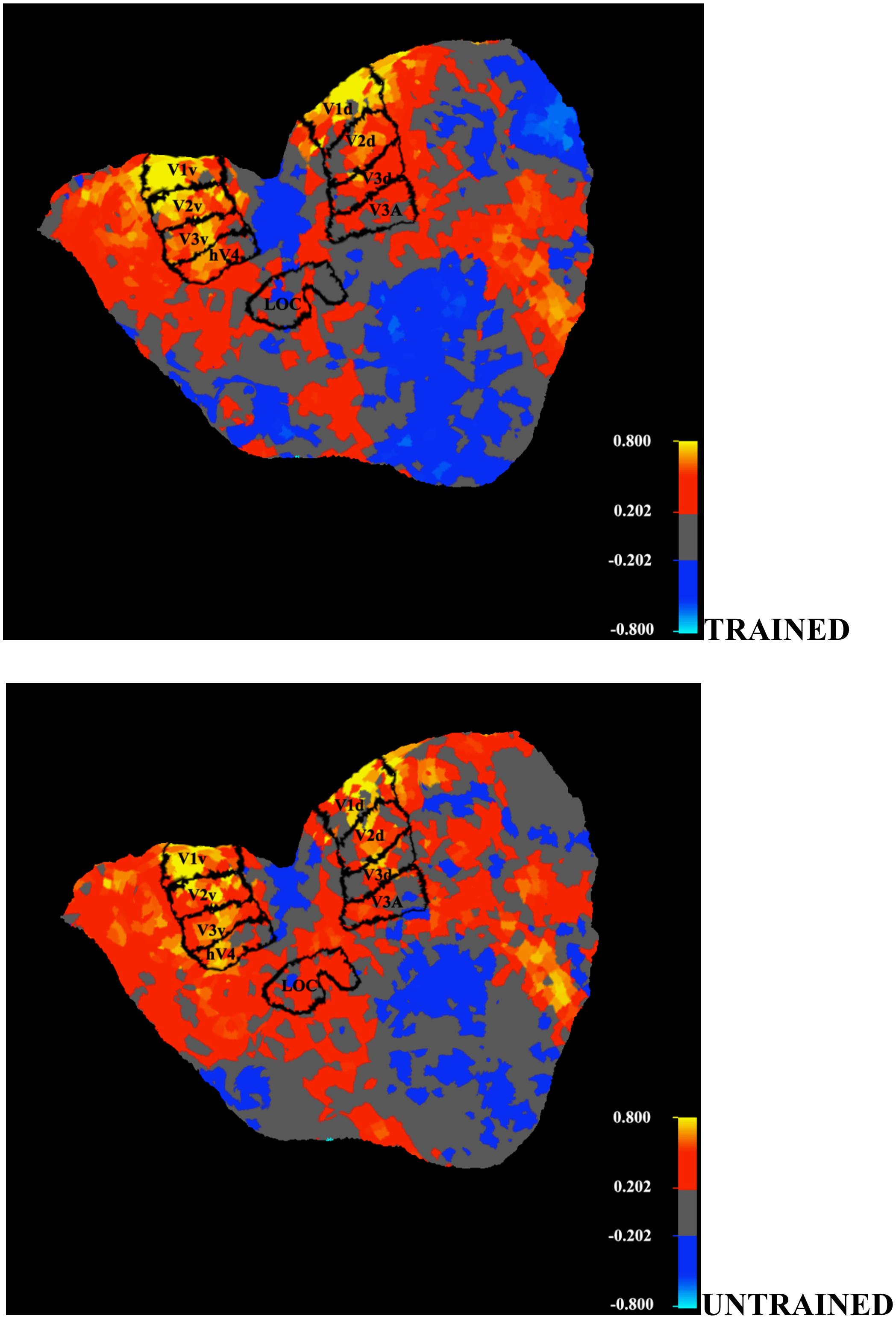
B.

3.2.3 Correlation of activity between brain cortices:

The correlation of activity between different brain cortices is seen as a measure of the connectivity between these regions (Biswal et al. 1995, Lowe et al. 1998, McIntosh et al. 1998, Bhattacharya et al. 2006, He et al. 2007, Marreiros et al. 2008). It is possible that changes in functional localization due to training will also change the communication, and therefore connectivity, between the brain cortices involved. Such a change will be visible as changes in the correlation of fMRI activity between these cortices. To analyze this possibility, we examined how well the fMRI activity at several functionally or anatomically determined brain regions correlated with the fMRI activity at V1. This analysis was done by correlating the average time course of activity in V1 to those of the other regions (Biswal et al. 1995 and He et al. 2007, Figure 17). We observed that there was a distinct difference in correlation of activity between trained and untrained conditions (Figure 20). With training there was a change in the correlation between V1 and LOC activity, and an opposite change in the correlation between V1 and superiorparietal, inferiortemporal and middle temporal cortices. Although all subjects showed this reciprocal relationship between these two networks, the direction of the change was not the same for all subjects.

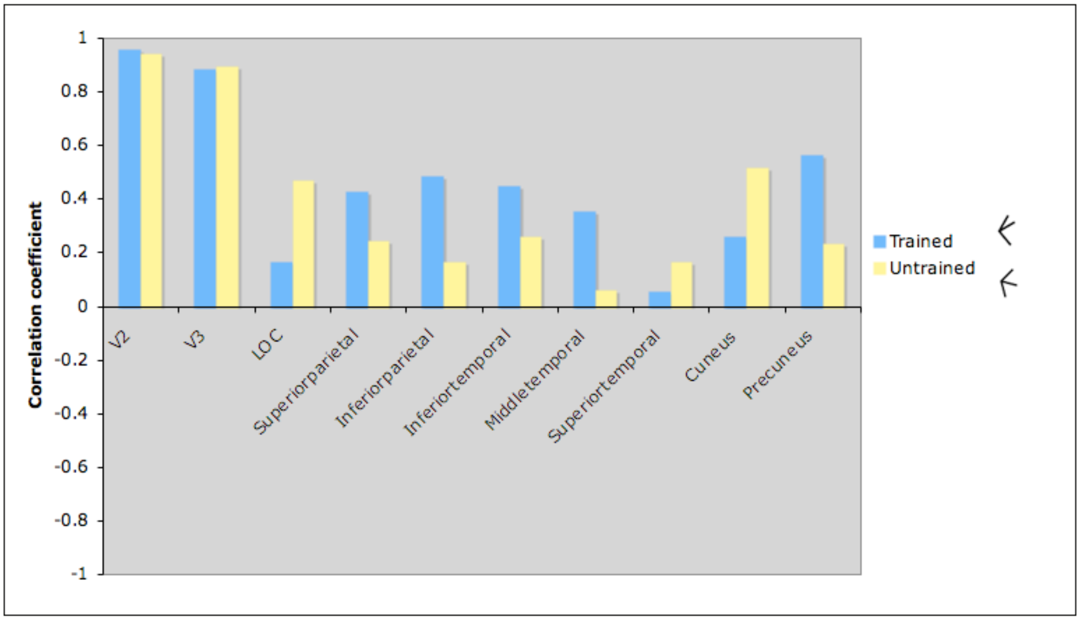
Figure 20. A) Correlation of fMRI activity with V1 when looking for trained (left) and untrained (right) object projected on an image of the flattened occipital pole, with the early visual cortex and LOC marked by enclosing black boundaries based on the functional maps. Shades of red indicate a positive correlation, shades of blue indicate a negative correlation with V1. A larger area of the brain was correlated with V1 in the untrained condition. Single subject. B) Correlation coefficients of average fMRI activity of different brain cortices to the average fMRI activity of V1 for two subjects, when the subjects searched for trained (blue), or untrained (yellow) objects. Two subjects. Of the regions tested, V2, V3, and LOC were determined functionally while the rest were determined anatomically. Values of correlation coefficients ranged from 1 (perfect correlation) to -1 (perfect negative correlation). In this instance all regions tested were positively correlated with V1. The correlation of V2 and V3 with V1 remained relatively constant between the three conditions, whereas that of LOC changed significantly. This change was accompanied by a reciprocal change for superiorparietal, inferiortemporal, inferiortemporal, middletemporal cortices, and cuneus. Both subjects showed this reciprocal relationship between the two networks; however, the direction of the change was opposite.

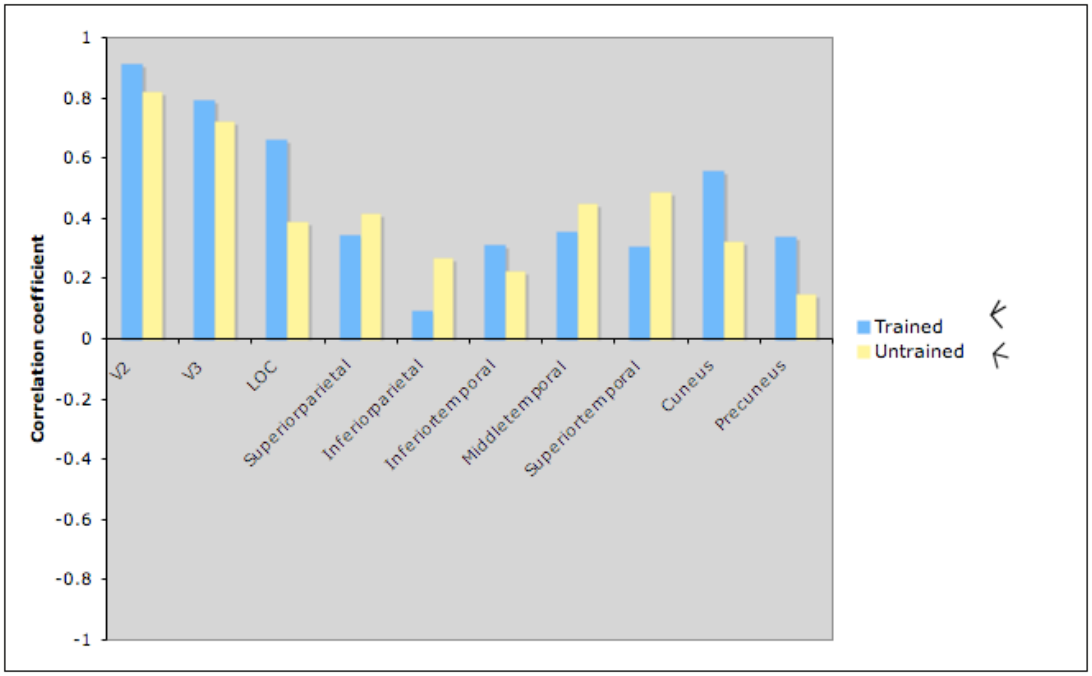
Figure 20


V1d
V2d
V3d
V3A
V1v
V2v
V3v
hV4
LOC
0.800
0.202
-0.202
-0.800
TRAINED
V1d
V2d
V3d
V3A
V1v
V2v
V3v
hV4
LOC
0.800
0.202
-0.202
-0.800
UNTRAINED

A.

Figure 20 continued





B.

3.3 Discussion:

Our results in general showed higher fMRI activity when subjects searched for untrained objects compared to trained objects. Furthermore, activity levels in early retinotopic cortices were also better correlated for untrained objects compared to trained objects. These patterns of change in fMRI activity stand in partial contrast with results previously reported by our lab that also show a reorganization of cortical activity following training, but with higher activity in the retinotopic cortex for trained objects, which was also better correlated with performance in this region compared to untrained objects (Sigman et al. 2005). There is, however, a diversity of findings concerning the direction of change of cortical activation with training. Some studies show that an increase in performance during training is correlated with an initial increase followed by a decrease in cortical activity (Mukai et al. and Yotsumoto et al.). One can imagine either outcome could lead to an improved performance, depending on the effect of training on the tuning properties of cortical neurons. An increase in activation could reflect a process of cortical recruitment, where more neurons become engaged in the task, and as a result of probability summation, this leads to an improvement in psychophysical performance (Recanzone et al. 1992, 1993, Nudo et al. 1996, Xerri et al. 1998). A decrease in activation could represent a sharpening of the tuning of neurons to the trained attributes (Schoups et al. 2001, Faber et al. 1999, Kobatake et al. 1998, Logothetis et al. 1995, Miyashita and Hayashi 2000), which would lead to an improvement in the threshold of the task and a decrease in the number of neurons involved. The differences in the fMRI studies might be attributed to differences in the duration of training. The Sigman et al. study involved shorter periods of training, while here the training extended for longer periods of time. One might therefore speculate that as the learning becomes consolidated over time, there are increased efficiencies in the representation of the learned information, and as a consequence fewer, more sharply tuned neurons become activated during task execution.

A complicating factor affecting the level of cortical activation is the change in task difficulty with perceptual learning. Arguably, searching for untrained objects requires more effort than searching for trained objects, which tend to pop-out more readily from the distractors. The untrained condition may therefore be accompanied by a larger attentional load (Joseph et al. 1997). Increases in neural activity due to attention have been shown both with electrophysiological (Luck et al. 1997) and fMRI studies (Brefczynski and DeYoe 1999, Gandhi et al.1999, Martinez et al.1999, Somers et al. 1999). Such activity is independent of the presence of the stimulus and is purely due to attention to stimulus location (Kastner et al. 1999, Ress et al. 2000, Silver et al. 2007) and is very strong in V1 for fMRI data. At this stage it is difficult to disambiguate changes in activation with training that are due to changes in the tuning characteristics of neurons from those that are due to changes in task difficulty.

We observed two major centers of change in activity levels as a result of training that did not fall into any of the visual centers that we mapped functionally. Anatomically, this observed activity fell into the parietal lobe and into the inferior temporal cortex. Of these, the parietal lobe is a part of the dorsal stream of processing and is mostly involved in spatial attention. The change in activity observed in this region likely reflects the different levels of attention required to complete the search task in the trained and untrained conditions. The probable reason for the activity changes in the IT is, however, less obvious. Unlike its anatomical counterpart in monkeys, this region is not necessarily involved in object recognition in humans (except perhaps faces, see Allison et al. 1994); this role is instead taken over by LOC, which lies more posterior and lateral. Discussion with the subjects reveals a possible reason why a change in this region is observed. Most subjects mentioned that they tried to liken the target objects to objects with which they were already familiar to aid them during the search task. One type of object that was often mentioned as being used was letters. There is some evidence that the human IT might be involved in the processing of letters and letter strings, even when they are nonsensical (Nobre et al. 1994, Allison et al. 1994). Therefore, such a strategy can potentially involve this region. It is likely that this strategy is used more heavily for unfamiliar shapes compared to familiar shapes, which can be recognized by themselves without such mental aids. This difference can explain why there is stronger activation of the IT in the untrained condition.

The changes in correlations between different brain regions we observed suggest that under different circumstances, communication between V1 to higher brain centers could be diverted to a different pathway that bypasses LOC and instead goes through temporal and parietal cortex and cuneus. However, it should be noted that both cuneus and parietal cortex (Kertzman et al. 1996, Kusunoki and Goldberg 2003) are involved in visual guidance of motion representation of spatial location. During the training, subjects were asked to report the location of the target object to correct for false positive responses. It is possible that the subjects were continuing to judge the position of the target even within the MRI scanner where they were not required to report it. The subjects were accustomed to look for target location in the trained condition and not in the untrained condition. Therefore, it is possible that the differences of connectivity between V1 and the location-sensitive cuneus and parietal cortices in trained and untrained conditions were due to this habit of the subjects. The change of connectivity between V1 and LOC surprisingly does not happen in the same condition in all subjects. Instead, two subjects have their LOC better correlated with V1 when they are looking for the trained target whereas the third subject has a higher correlation between these two regions when looking for the untrained target. This difference suggests that the brains of different people might adopt different strategies to solve the same perceptual problem. Because the correlation analysis involves smoothing of the fMRI signal at the voxel level, a statistical bias was introduced that prevented us from determining the statistical significance of the observed changes at the single subject level in a meaningful manner. Further studies are needed so that the statistical significance of the changes may be determined at the level of multiple subjects, and also to show if the differences between individuals we observed are a common occurrence.

References:

Ahissar M, Hochstein S. Attentional control of early perceptual learning. Proc Natl Acad Sci U S A. 1993 Jun 15;90(12):5718-22

Ahissar M, Hochstein S. Learning pop-out detection: specificities to stimulus characteristics. Vision Res. 1996 Nov;36(21):3487-500.

Allison T, McCarthy G, Nobre A, Puce A, Belger A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. Cereb Cortex. 1994 Sep-Oct;4(5):544-54. Review.

Bell AH, Hadj-Bouziane F, Frihauf JB, Tootell RB, Ungerleider LG. Object representations in the temporal cortex of monkeys and humans as revealed by functional magnetic resonance imaging. J Neurophysiol. 2009 Feb;101(2):688-700.

Berardi N, Fiorentini A. Interhemispheric transfer of visual information in humans: spatial characteristics. J Physiol. 1987 Mar;384:633-47.

Bergen JR, Julesz B. Parallel versus serial processing in rapid pattern discrimination. Nature. 1983 Jun 23-29;303(5919):696-8.

Bhattacharya S, Ringo Ho MH, Purkayastha S. A Bayesian approach to modeling dynamic effective connectivity with fMRI data. Neuroimage. 2006 Apr 15;30(3):794-812.

Biederman I. Recognition-by-components: a theory of human image understanding. Psychol Rev. 1987 Apr;94(2):115-47.

Biederman I, Bar M. One-shot viewpoint invariance in matching novel objects. Vision Res. 1999 Aug;39(17):2885-99.

Biederman I, Cooper EE. Evidence for complete translational and reflectional invariance in visual object priming. Perception. 1991;20(5):585-93.

Biederman I, Gerhardstein PC. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. J Exp Psychol Hum Percept Perform. 1993 Dec;19(6):1162-82.

Biederman I, Gerhardstein PC. Viewpoint-Dependent Mechanisms in Visual Object Recognition: Reply to Tarr and Bülthoff (1995) J Exp Psychol Hum Percept Perform. 1995 Dec; 21(6):1506-14

Biswal B, Yetkin FZ, Haughton VM, Hyde JS. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn Reson Med. 1995 Oct;34(4):537-41.

Booth MC, Rolls ET. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex. 1998 Sep;8(6):510-23.

Brady MJ, Kersten D. Bootstrapped learning of novel objects. J Vis. 2003;3(6):413-22.

Braun J. Vision and attention: the role of training. Nature. 1998 Jun 4;393(6684):424-5.

Bravo MJ, Nakayama K. The role of attention in different visual-search tasks. Percept Psychophys. 1992 May;51(5):465-72.

Brincat SL, Connor CE. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci. 2004 Aug;7(8):880-6.

Brefczynski JA, DeYoe EA. A physiological correlate of the 'spotlight' of visual attention. Nat Neurosci. 1999 Apr;2(4):370-4.

Brett M, Anton J, Valabregue R, Poline J. Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, Vol 16, No 2.

Brincat SL, Connor CE. Dynamic shape synthesis in posterior inferotemporal cortex. Neuron. 2006 Jan 5;49(1):17-24.

Bülthoff HH, Edelman S. Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc Natl Acad Sci U S A. 1992 Jan 1;89(1):60-4.

Crist RE, Li W, Gilbert CD. Learning to see: experience and attention in primary visual cortex. Nat Neurosci. 2001 May;4(5):519-25.

Chun MM, Jiang Y. Contextual cueing: implicit learning and memory of visual context guides spatial attention. Cogn Psychol. 1998 Jun;36(1):28-71.

Cohen LB, Strauss MS. Concept acquisition in the human infant. Child Dev. 1979 Jun;50(2):419-24.

Connor CE. Neuroscience: friends and grandmothers. Nature. 2005 Jun 23;435(7045):1036-7.

Damasio AR. Category-related recognition defects as a clue to the neural substrates of knowledge. Trends Neurosci. 1990 Mar;13(3):95-8.

Damasio AR, Tranel D, Damasio H. Face agnosia and the neural substrates of memory. Annu Rev Neurosci. 1990;13:89-109. Review.

Darian-Smith C, Gilbert CD. Axonal sprouting accompanies functional reorganization in adult cat striate cortex. Nature. 1994 Apr 21;368(6473):737-40.

Das A, Gilbert CD. Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature. 1999 Jun 17;399(6737):655-61.

Delorme A, Rousselet GA, Macé MJ, Fabre-Thorpe M. Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. Brain Res Cogn Brain Res. 2004 Apr;19(2):103-13.

Desimone R, Albright TD, Gross CG, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci. 1984 Aug;4(8):2051-62.

DiCarlo JJ, Maunsell JH. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. J Neurophysiol. 2003 Jun;89(6):3264-78.

Dill M, Fahle M. The role of visual field position in pattern-discrimination learning. Proc Biol Sci. 1997 Jul 22;264(1384):1031-6.

Dougherty RF, Koch VM, Brewer AA, Fischer B, Modersitzki J, Wandell BA. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J Vis. 2003;3(10):586-98.

Downing PE, Jiang Y, Shuman M, Kanwisher N.
A cortical area selective for visual processing of the human body.
Science. 2001 Sep 28;293(5539):2470-3.

Edelman, S., Grill-Spector, K., Kushnir, T. & Malach, R. Toward direct visualization of the internal shape representation space by fMRI. Psychobiology ***26****, 309–321 (1998).*

Edelman S, Intrator N. Visual processing of object structure. The Handbook of Brain Theory and Neural Networks (2nd ed.), M.A. Arbib,ed., MIT Press, 2002

Epstein R, Kanwisher N. A cortical representation of the local visual environment. Nature. 1998 Apr 9;392(6676):598-601.

Faber T, Joerges J, Menzel R. Associative learning modifies neural representations of odors in the insect brain. Nat Neurosci. 1999 Jan;2(1):74-8.

Farah MJ. Agnosia. Curr Opin Neurobiol. 1992 Apr;2(2):162-4. Review.

Farah MJ, Hammond KM, Mehta Z, Ratcliff G. Category-specificity and modality-specificity in semantic memory. Neuropsychologia. 1989;27(2):193-200.

Farah MJ. Is face recognition 'special'? Evidence from neuropsychology. Behav Brain Res. 1996 Apr;76(1-2):181-9. Review.

Farah MJ, Tanaka JW, Drain HM. What causes the face inversion effect? J Exp Psychol Hum Percept Perform. 1995 Jun;21(3):628-34.

Farah MJ, Wilson KD, Drain M, Tanaka JN. What is "special" about face perception? Psychol Rev. 1998 Jul;105(3):482-98. Review.

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex. 1991 Jan-Feb;1(1):1-47.

Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. Proc Natl Acad Sci U S A. 2006 Jun 27;103(26):10046-51.

Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc Natl Acad Sci U S A. 2005 Jul 5;102(27):9673-8.

Fujita I, Tanaka K, Ito M, Cheng K. Columns for visual features of objects in monkey inferotemporal cortex. Nature. 1992 Nov 26;360(6402):343-6.

Furmanski CS, Engel SA. Perceptual learning in object recognition: object specificity and size invariance. Vision Res. 2000;40(5):473-84.

Gallant JL, Braun J, Van Essen DC. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science. 1993 Jan 1;259(5091):100-3.

Gandhi SP, Heeger DJ, Boynton GM. Spatial attention affects brain activity in human primary visual cortex. Proc Natl Acad Sci U S A. 1999 Mar 16;96(6):3314-9.

Garraghty PE, Kaas JH. Dynamic features of sensory and motor maps. Curr Opin Neurobiol. 1992 Aug;2(4):522-7. Review.

Gilbert CD, Das A, Ito M, Kapadia M, Westheimer G. Spatial integration and cortical dynamics. Proc Natl Acad Sci U S A. 1996 Jan 23;93(2):615-22.

Gilbert CD, Ito M, Kapadia M, Westheimer G. Interactions between attention, context and learning in primary visual cortex. Vision Res. 2000;40(10-12):1217-26.

Gilbert CD, Sigman M, Crist RE. The neural basis of perceptual learning. Neuron. 2001 Sep 13;31(5):681-97.

Gilbert CD, Wiesel TN. Receptive field dynamics in adult primary visual cortex. Nature. 1992 Mar 12;356(6365):150-2.

Goodale MA, Milner AD, Jakobson LS, Carey DP. A neurological dissociation between perceiving objects and grasping them. Nature. 1991 Jan 10;349(6305):154-6.

Gorea A, Julesz B. Context superiority in a detection task with line-element stimuli: a low-level effect. Perception. 1990;19(1):5-16.

Gray CM. The temporal correlation hypothesis of visual feature integration: still alive and well. Neuron. 1999 Sep;24(1):31-47, 111-25. Review.

Grill-Spector K, Kourtzi Z, Kanwisher N. The lateral occipital complex and its role in object recognition. Vision Res. 2001;41(10-11):1409-22. Review.

Grill-Spector K, Kushnir T, Edelman S, Itzchak Y, Malach R. Cue-invariant activation in object-related areas of the human occipital lobe. Neuron. 1998 Jul;21(1):191-202.

Gross CG, Rocha-Miranda CE, Bender DB. Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol. 1972 Jan;35(1):96-111.

Hadjikhani N, de Gelder B. Seeing fearful body expressions activates the fusiform cortex and amygdala. Curr Biol. 2003 Dec 16;13(24):2201-5.

Harmon LD, Julesz B. Masking in visual recognition: effects of two-dimensional filtered noise. Science. 1973 Jun 15;180(91):1194-7.

Haxby JV, Hoffman EA, Gobbini MI. The distributed human neural system for face perception. Trends Cogn Sci. 2000 Jun;4(6):223-233.

He BJ, Snyder AZ, Vincent JL, Epstein A, Shulman GL, Corbetta M. Breakdown of functional connectivity in frontoparietal networks underlies behavioral deficits in spatial neglect. Neuron. 2007 Mar 15;53(6):905-18.

Hoffman DD, Richards WA. Parts of recognition. Cognition. 1984 Dec;18(1-3):65-96.

Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol. 1962 Jan;160:106-54.

Hubel DH, Wiesel TN. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol. 1965 Mar;28:229-89.

Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. J Physiol. 1968 Mar;195(1):215-43.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A. 1999 Aug 3;96(16):9379-84.

Ishai A, Ungerleider LG, Martin A, Haxby JV. The representation of objects in the human occipital and temporal cortex. J Cogn Neurosci. 2000;12 Suppl 2:35-51.

Ito M, Komatsu H. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. J Neurosci. 2004 Mar 31;24(13):3313-24.

Ito M, Tamura H, Fujita I, Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol. 1995 Jan;73(1):218-26.

Ito M, Westheimer G, Gilbert CD. Attention and perceptual learning modulate contextual influences on visual perception. Neuron. 1998 Jun;20(6):1191-7.

Joseph JS, Chun MM, Nakayama K. Attentional requirements in a 'preattentive' feature search task. Nature. 1997 Jun 19;387(6635):805-7.

Kaas JH. Plasticity of sensory and motor maps in adult mammals. Annu Rev Neurosci. 1991;14:137-67. Review.

Kaas JH, Krubitzer LA, Chino YM, Langston AL, Polley EH, Blair N. Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. Science. 1990 Apr 13;248(4952):229-31.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci. 1997 Jun 1;17(11):4302-11.

Kapadia MK, Ito M, Gilbert CD, Westheimer G. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron. 1995 Oct;15(4):843-56.

Karni A, Sagi D. Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proc Natl Acad Sci U S A. 1991 Jun 1;88(11):4966-70.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron. 1999 Apr;22(4):751-61.

Kertzman C, Schwarz U, Zeffiro TA, Hallett M. The role of posterior parietal cortex in visually guided reaching movements in humans. Exp Brain Res. 1997 Mar;114(1):170-83.

Kobatake E, Tanaka K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol. 1994 Mar;71(3):856-67.

Kobatake E, Wang G, Tanaka K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. J Neurophysiol. 1998 Jul;80(1):324-30.

Kourtzi Z, Erb M, Grodd W, Bülthoff HH. Representation of the perceived 3-D object shape in the human lateral occipital complex. Cereb Cortex. 2003 Sep;13(9):911-20.

Kourtzi Z, Kanwisher N. Cortical regions involved in perceiving object shape. J Neurosci. 2000 May 1;20(9):3310-8.

Kovács I, Julesz B. Perceptual sensitivity maps within globally defined visual shapes. Nature. 1994 Aug 25;370(6491):644-6.

Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nat Neurosci. 2000 Sep;3(9):946-53.

Kusunoki M, Goldberg ME. The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. J Neurophysiol. 2003 Mar;89(3):1519-27.

Leonards U, Rettenbach R, Nase G, Sireteanu R. Perceptual learning of highly demanding visual search tasks. Vision Res. 2002 Aug;42(18):2193-204.

Li W, Gilbert CD. Global contour saliency and local colinear interactions. J Neurophysiol. 2002 Nov;88(5):2846-56.

Li W, Piëch V, Gilbert CD. Perceptual learning and top-down influences in primary visual cortex. Nat Neurosci. 2004 Jun;7(6):651-7.

Li W, Piëch V, Gilbert CD. Contour saliency in primary visual cortex. Neuron. 2006 Jun 15;50(6):951-62.

Li W, Piëch V, Gilbert CD. Learning to link visual contours. Neuron. 2008 Feb 7;57(3):442-51.

Logothetis NK. Object recognition: holistic representations in the monkey brain. Spat Vis. 2000;13(2-3):165-78. Review.

Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. Curr Biol. 1995 May 1;5(5):552-63.

Logothetis NK, Pauls J, Bülthoff HH, Poggio T. View-dependent object recognition by monkeys. Curr Biol. 1994 May 1;4(5):401-14.

Lowe MJ, Mock BJ, Sorenson JA. Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations. Neuroimage. 1998 Feb;7(2):119-32.

Luck SJ, Chelazzi L, Hillyard SA, Desimone R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol. 1997 Jan;77(1):24-42.

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 U S A. 1995 Aug 29;92(18):8135-9.

Malhotra P, Coulthard EJ, Husain M. Role of right posterior parietal cortex in maintaining attention to spatial locations over time. Brain. 2009 Jan 21.

Malinowski P, Hubner R. The effect of familiarity on visual-search performance: evidence for learned basic features. Percept Psychophys. 2001 Apr;63(3):458-63.

Maljkovic V, Nakayama K. Priming of pop-out: I. Role of features. Mem Cognit. 1994 Nov;22(6):657-72.

Marr D. Visual information processing: the structure and creation of visual representations. Philos Trans R Soc Lond B Biol Sci. 1980 Jul 8;290(1038):199-218.

Marr D, Nishihara HK. Representation and recognition of the spatial organization of three-dimensional shapes. Proc R Soc Lond B Biol Sci. 1978 Feb 23;200(1140):269-94.

Marreiros AC, Kiebel SJ, Friston KJ. Dynamic causal modelling for fMRI: a two-state model. Neuroimage. 2008 Jan 1;39(1):269-78.

Martínez A, Anllo-Vento L, Sereno MI, Frank LR, Buxton RB, Dubowitz DJ, Wong EC, Hinrichs H, Heinze HJ, Hillyard SA.Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat Neurosci. 1999 Apr;2(4):364-9.

McIntosh AR, Cabeza RE, Lobaugh NJ. Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. J Neurophysiol. 1998 Nov;80(5):2790-6.

McKone E, Kanwisher N, Duchaine BC. Can generic expertise explain special processing for faces? Trends Cogn Sci. 2007 Jan;11(1):8-15. Review.

Mishkin M, Ungerleider LG. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. Behav Brain Res. 1982 Sep;6(1):57-77.

Miyashita Y, Hayashi T. Neural representation of visual objects: encoding and top-down activation. Curr Opin Neurobiol. 2000 Apr;10(2):187-94.

Morrone MC, Burr DC, Ross J. Added noise restores recognizability of coarse quantized images. Nature. 1983 Sep 15-21;305(5931):226-8.

Mukai I, Kim D, Fukunaga M, Japee S, Marrett S, Ungerleider LG. Activations in visual and attention-related areas predict and correlate with the degree of perceptual learning. J Neurosci. 2007 Oct 17;27(42):11401-11.

Nielsen KJ, Logothetis NK, Rainer G. Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. J Neurosci. 2006 Sep 20;26(38):9639-45.

Nobre AC, Allison T, McCarthy G. Word recognition in the human inferior temporal lobe. Nature. 1994 Nov 17;372(6503):260-3.

Nudo RJ, Milliken GW, Jenkins WM, Merzenich MM. Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. J Neurosci. 1996 Jan 15;16(2):785-807.

Obata S, Obata J, Das A, Gilbert CD. Molecular correlates of topographic reorganization in primary visual cortex following retinal lesions. Cereb Cortex. 1999 Apr-May;9(3):238-48.

Op de Beeck HP, Deutsch JA, Vanduffel W, Kanwisher NG, DiCarlo JJ. A stable topography of selectivity for unfamiliar shape classes in monkey inferior temporal cortex. Cereb Cortex. 2008 Jul;18(7):1676-94.

Pasupathy A, Connor CE. Responses to contour features in macaque area V4. J Neurophysiol. 1999 Nov;82(5):2490-502.

Peelen MV, Downing PE. Selectivity for the human body in the fusiform gyrus. J Neurophysiol. 2005 Jan;93(1):603-8.

Perrett DI, Rolls ET, Caan W. Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res. 1982;47(3):329-42.

Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. Nature. 1990 Jan 18;343(6255):263-6.

Posner MI, Gilbert CD. Attention and primary visual cortex. Proc Natl Acad Sci U S A. 1999 Mar 16;96(6):2585-7.

Posner MI, Petersen SE. The attention system of the human brain. Annu Rev Neurosci. 1990;13:25-42. Review.

Puce A, Allison T, Asgari M, Gore JC, McCarthy G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. J Neurosci. 1996 Aug 15;16(16):5205-15.

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature. 2005 Jun 23;435(7045):1102-7.

Recanzone GH, Merzenich MM, Jenkins WM, Grajski KA, Dinse HR. Topographic reorganization of the hand representation in cortical area 3b owl monkeys trained in a frequency-discrimination task. J Neurophysiol. 1992 May;67(5):1031-56.

Recanzone GH, Schreiner CE, Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci. 1993 Jan;13(1):87-103.

Ress D, Backus BT, Heeger DJ. Activity in primary visual cortex predicts performance in a visual detection task. Nat Neurosci. 2000 Sep;3(9):940-5.

Rhodes G, Brake S, Atkinson AP. What's lost in inverted faces? Cognition. 1993 Apr;47(1):25-57. Review.

Ribeiro S, Cecchi GA, Magnasco MO, Mello CV. Toward a song code: evidence for a syllabic representation in the canary brain. Neuron. 1998 Aug;21(2):359-71.

Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nat Neurosci. 1999 Nov;2(11):1019-25.

Rock I, DiVita J. A case of viewer-centered object perception. Cogn Psychol. 1987 Apr;19(2):280-93.

Rolls ET, Treves A, Tovee MJ. The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. Exp Brain Res. 1997 Mar;114(1):149-62.

Sagi D, Julesz B. "Where" and "what" in vision. Science. 1985 Jun 7;228(4704):1217-9.

Salinas E, Abbott LF. Invariant visual responses from attentional gain fields. J Neurophysiol. 1997 Jun;77(6):3267-72.

Schneider W, Eschman A, Zuccolotto A. E-Prime Reference Guide, Pittsburg: Psychology Software Tools Inc. 2002a

Schneider W, Eschman A, Zuccolotto A. E-Prime User's Guide, Pittsburg: Psychology Software Tools Inc. 2002b

Schoups A, Vogels R, Qian N, Orban G. Practising orientation identification improves orientation coding in V1 neurons. Nature. 2001 Aug 2;412(6846):549-53.

Schwarzlose RF, Baker CI, Kanwisher N. Separate face and body selectivity on the fusiform gyrus. J Neurosci. 2005 Nov 23;25(47):11055-9.

Sheinberg DL, Logothetis NK. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J Neurosci. 2001 Feb 15;21(4):1340-50.

Shen J, Reingold EM. Visual search asymmetry: the influence of stimulus familiarity and low-level features. Percept Psychophys. 2001 Apr;63(3):464-75.

Sigala N, Logothetis NK. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature. 2002 Jan 17;415(6869):318-20.

Sigala N, Gabbiani F, Logothetis NK. Visual categorization and object representation in monkeys and humans. J Cogn Neurosci. 2002 Feb 15;14(2):187-98.

Sigala N. Visual categorization and the inferior temporal cortex. Behav Brain Res. 2004 Feb 4;149(1):1-7.

Sigman M, Gilbert CD. Learning to find a shape. Nat Neurosci. 2000 Mar;3(3):264-9.

Sigman M, Pan H, Yang Y, Stern E, Silbersweig D, Gilbert CD. Top-down reorganization of activity in the visual pathway after learning a shape identification task. Neuron. 2005 Jun 2;46(5):823-35.

Sillito AM, Grieve KL, Jones HE, Cudeiro J, Davis J. Visual cortical mechanisms detecting focal orientation discontinuities. Nature. 1995 Nov 30;378(6556):492-6.

Silver MA, Ress D, Heeger DJ. Neural correlates of sustained spatial attention in human early visual cortex. J Neurophysiol. 2007 Jan;97(1):229-37.

Sireteanu R, Rettenbach R. Perceptual learning in visual search: fast, enduring, but non-specific. Vision Res. 1995 Jul;35(14):2037-43.

Sireteanu R, Rettenbach R. Perceptual learning in visual search generalizes over tasks, locations, and eyes. Vision Res. 2000;40(21):2925-49.

Somers DC, Dale AM, Seiffert AE, Tootell RB. Functional MRI reveals spatially specific attentional modulation in human primary visual cortex. Proc Natl Acad Sci U S A. 1999 Feb 16;96(4):1663-8.

Steinman SB. Serial and parallel search in pattern vision?
Perception. 1987;16(3):389-98.

Stettler DD, Yamahachi H, Li W, Denk W, Gilbert CD. Axons and synaptic boutons are highly dynamic in adult visual cortex. Neuron. 2006 Mar 16;49(6):877-87.

Strauss MS. Abstraction of prototypical information by adults and 10-month-old infants. J Exp Psychol [Hum Learn]. 1979 Nov;5(6):618-32.

Tanaka JW, Sengco JA. Features and their configuration in face recognition. Mem Cognit. 1997 Sep;25(5):583-92.

Tanaka K. Inferotemporal cortex and object vision. Annu Rev Neurosci. 1996;19:109-39.

Tanaka K, Saito H, Fukada Y, Moriya M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol. 1991 Jul;66(1):170-89.

Tarr MJ, Bülthoff 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 Dec;21(6):1494-505.

Tarr MJ, Gauthier I. Do viewpoint-dependent mechanisms generalize across members of a class? Cognition. 1998 Jul;67(1-2):73-110.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature. 1996 Jun 6;381(6582):520-2.

Tootell RB, Hadjikhani NK, Vanduffel W, Liu AK, Mendola JD, Sereno MI, Dale AM. Functional analysis of primary visual cortex (V1) in humans. Proc Natl Acad Sci U S A. 1998 Feb 3;95(3):811-7.

Treisman AM, Gelade G. A feature-integration theory of attention. Cognit Psychol. 1980 Jan;12(1):97-136.

Treisman A, Paterson R. Emergent features, attention, and object perception. J Exp Psychol Hum Percept Perform. 1984 Feb;10(1):12-31.

Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB. Faces and objects in macaque cerebral cortex. Nat Neurosci. 2003 Sep;6(9):989-95.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. A cortical region consisting entirely of face-selective cells. Science. 2006 Feb 3;311(5761):670-4.

Tsao DY, Moeller S, Freiwald WA. Comparing face patch systems in macaques and humans. Proc Natl Acad Sci U S A. 2008 Dec 9;105(49):19514-9a.

Tsao DY, Schweers N, Moeller S, Freiwald WA. Patches of face-selective cortex in the macaque frontal lobe. Nat Neurosci. 2008 Aug;11(8):877-9b.

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci. 2001 Aug;4(8):832-8.

Ullman S. Artificial intelligence and the brain: computational studies of the visual system. Annu Rev Neurosci. 1986;9:1-26.

Ullman S. Aligning pictorial descriptions: an approach to object recognition. Cognition. 1989 Aug;32(3):193-254.

Ullman S. Low-level aspects of segmentation and recognition. Philos Trans R Soc Lond B Biol Sci. 1992 Sep 29;337(1281):371-8

Ullman S. High-level Vision, Object Recognition and Visual Cognition, MIT Press, 1996

Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. Nat Neurosci. 2002 Jul;5(7):682-7.

Ullman S. Object recognition and segmentation by a fragment-based hierarchy. Trends Cogn Sci. 2007 Feb;11(2):58-64. Review.

Valentine T, Bruce V. The effect of race, inversion and encoding activity upon face recognition. Acta Psychol (Amst). 1986 Apr;61(3):259-73.

Valentine T. Upside-down faces: a review of the effect of inversion upon face recognition. Br J Psychol. 1988 Nov;79 (Pt 4):471-91. Review.

Van Essen DC. Visual areas of the mammalian cerebral cortex. Annu Rev Neurosci. 1979;2:227-63. Review.

Van Essen DC, Gallant JL. Neural mechanisms of form and motion processing in the primate visual system. Neuron. 1994 Jul;13(1):1-10. Review.

Vetter T, Hurlbert A, Poggio T. View-based models of 3D object recognition: invariance to imaging transformations. Cereb Cortex. 1995 May-Jun;5(3):261-9.

Vogels R. Categorization of complex visual images by rhesus monkeys. Part 1: behavioural study. Eur J Neurosci. 1999 Apr;11(4):1223-38.

Vogels R. Categorization of complex visual images by rhesus monkeys. Part 2: single-cell study. Eur J Neurosci. 1999 Apr;11(4):1239-55.

Von der Malsburg C. The what and why of binding: the modeler's perspective. Neuron. 1999 Sep;24(1):95-104, 111-25. Review.

Wang G, Tanifuji M, Tanaka K. Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. Neurosci Res. 1998 Sep;32(1):33-46.

Wang Q, Cavanagh P, Green M. Familiarity and pop-out in visual search. Percept Psychophys. 1994 Nov;56(5):495-500.

Wilson HR, Wilkinson F, Asaad W. Concentric orientation summation in human form vision. Vision Res. 1997 Sep;37(17):2325-30.

Wolfe JM, Cave KR, Franzel SL. Guided search: an alternative to the feature integration model for visual search. J Exp Psychol Hum Percept Perform. 1989 Aug;15(3):419-33.

Wolfe JM. Experimental psychology. In a blink of the mind's eye. Nature. 1997 Jun 19;387(6635):756-7.

Wolfe JM. Moving towards solutions to some enduring controversies in visual search. Trends Cogn Sci. 2003 Feb;7(2):70-76.

Wright A, Barton JJ. Viewpoint invariance in the discrimination of upright and inverted faces. Vision Res. 2008 Nov;48(25):2545-54.

Xerri C, Merzenich MM, Peterson BE, Jenkins W. Plasticity of primary somatosensory cortex paralleling sensorimotor skill recovery from stroke in adult monkeys. J Neurophysiol. 1998 Apr;79(4):2119-48.

Yamahachi H, Marik S, McManus JN, Denk W, Gilbert CD. Axonal sprouting and pruning accompany functional reorganization in primary visual cortex, Society for Neuroscience Abstract, 2008.

Yin R. Looking at upside-down faces. Journal of Experimental Psychology. 1969 Jul; 81(1):141–45.

Yotsumoto Y, Watanabe T, Sasaki Y. Different dynamics of performance and brain activation in the time course of perceptual learning. Neuron. 2008 Mar 27;57(6):827-33.

Young MP, Yamane S. Sparse population coding of faces in the inferotemporal cortex. Science. 1992 May 29;256(5061):1327-31.

Yund EW, Efron R. Guided search: the effects of learning. Brain Cogn. 1996 Aug;31(3):369-86.