

Perceptual Learning of Object Shape

Doruk Gölcü and Charles D. Gilbert

The Rockefeller University, New York, New York 10065

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 near-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 also transferred to the components of the object when these components were embedded in an array of distractors of similar complexity. 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, regardless of how well they distinguish the object from distractors. This suggests that a joint involvement of multiple components was necessary for full performance.

Introduction

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 keep getting combined as they travel upstream, until the full information about the object is assembled together. This information is compared with a previously stored template of the object. According to some of these models, this process of complexification culminates with cells in the monkey inferotemporal cortex (IT) or human lateral occipital cortex (LOC) that are specifically responsive to images of entire objects. One of the most prominent criticisms of such holistic 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

complexification stops at an earlier stage, and instead of having a single template that stores object information, objects are coded as a combination of smaller, simpler parts that are largely viewpoint invariant (Marr and Nishihara, 1978; Marr, 1980; Hoffman and Richards, 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 (for review, see 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 used 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.

Materials and Methods

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

Received June 4, 2009; revised Sept. 8, 2009; accepted Sept. 27, 2009.

This study was sponsored by National Institutes of Health Grant EY007968. We thank Dr. Gerald Westheimer from University of California, Berkeley for his commentary throughout this study.

Correspondence should be addressed to Dr. Charles D. Gilbert, The Rockefeller University, 1230 York Avenue, New York, NY 10065. E-mail: gilbert@rockefeller.edu.

DOI:10.1523/JNEUROSCI.2612-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2913621-09\$15.00/0

Board and gave written informed consent. All subjects except one (author D.G.) were naive on the specific task used when they started the study and had good or corrected vision.

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 (cathode ray tube) 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 chin rest 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° 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 (Fig. 1). The first setup used was a rectangular 5×5 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 center so that all objects were equidistant from the fovea. The objects were placed with equal separation along the circumference of the circle, at 3° eccentricity, with the meridional positions on the circle circumference 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 (Fig. 2). 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 1-s-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.

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 s to remind the subjects of the target shape. After this display, there was a 6 s period during which only the fixation point was present to enable subjects to maintain fixation. Consecutive trials were separated by a 1500 ms interval (Fig. 2). The subjects were allowed to rest between rounds and to start each round at a time of their own choice. Sessions took ~ 1 h, and the subjects did three to five sessions per week. Whenever possible, the sessions were scheduled for the same time of the day 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 with 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

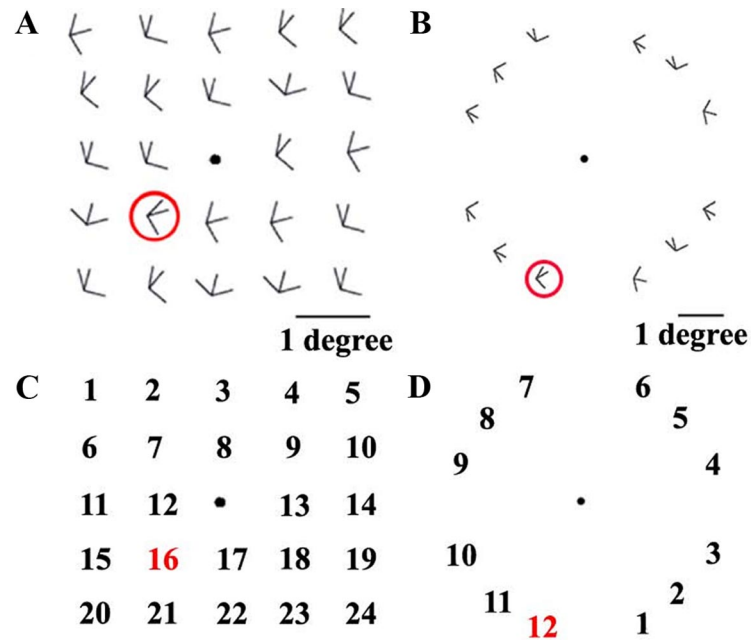


Figure 1. Stimulus array. The stimuli consisted of arbitrary three-line shapes distributed in either a square 5×5 array (**A**) or a circular array of 12 objects (**B**; target shapes are encircled in red). Subjects were asked to report if they had seen the target shape or not. In trials where they responded positively, they were also asked to report the location of the target object by entering a number corresponding to 1 of 24 positions within the square array (**C**) and 12 positions within the circular array (**D**).

over time are shown for individual subjects, with the error bars corresponding to the variation of performance between blocks.

Results

Detectability of target

We conducted a set of experiments with changing stimulus parameters 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, even if at a low level of performance. 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 found that, for small differences in orientation, naive 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 (Fig. 3).

Next, we tested the effect of changing the number of types of distractors on target detection by naive subjects. Experimental conditions with 4, 8, 16, and 33 different distractors were compared. In these experiments, during every stimulus presentation, the distractors that were to be displayed within the array were chosen randomly from sets of 4, 8, 16, or 33 objects, depending on the condition to be tested. The stimulus array remained otherwise exactly the same as the ones previously used. There was a

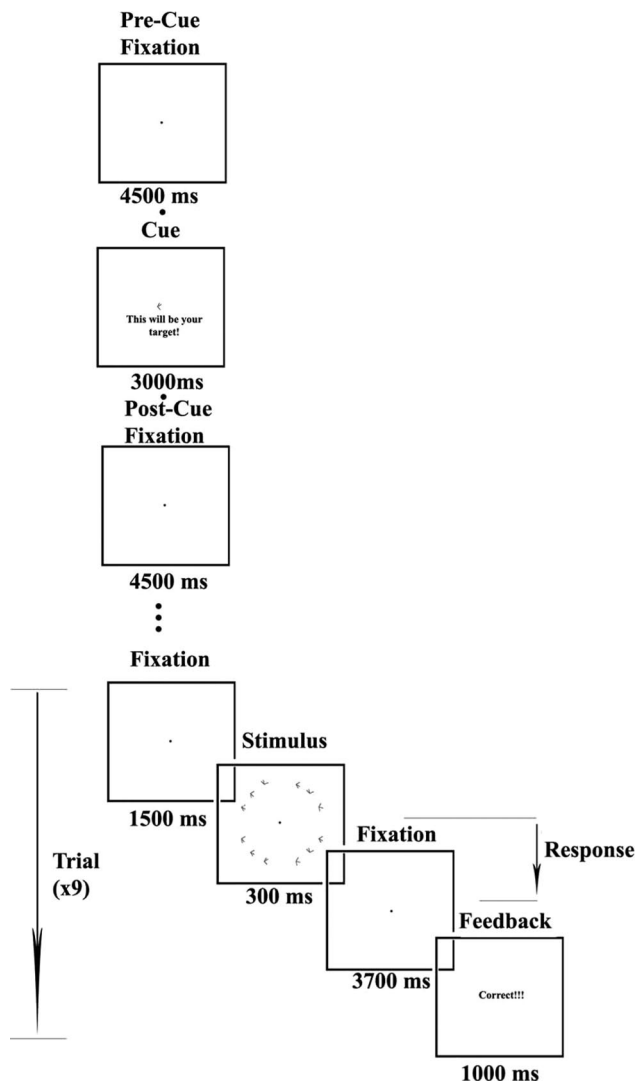


Figure 2. The stimulus timeline. The cue shape is displayed once every nine trials.

visible overall trend of lower level of performance at conditions with more distractors. The subjects performed significantly better when there were fewer types of distractors (performance with 4 distractors, $37.4 \pm 12.0\%$; performance with 33 distractors, $19.4 \pm 12.8\%$; $p < 0.0008$, two-tailed paired t test, average of three subjects). This difference was maintained after training (performance of $85.0 \pm 2.8\%$ vs $33.0 \pm 2.8\%$, respectively; $p < 0.05$, two-tailed paired t test, one subject).

Pretraining and posttraining 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: $p' = (p - fp)/(1 - fp)$,

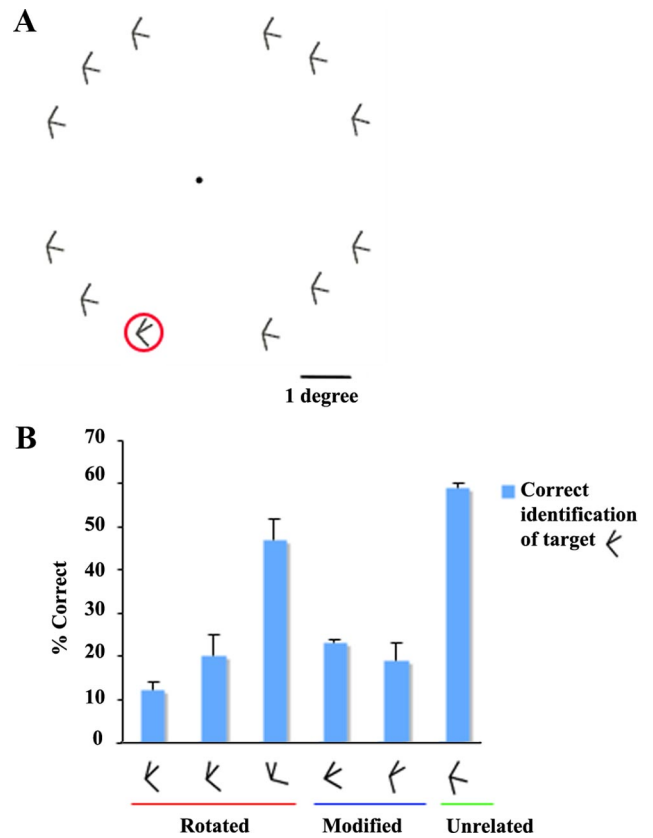


Figure 3. *A*, Search setup with a single type of distractor. The target object is encircled in red. *B*, Performance on target identification when embedded in an array of a single-type of distractor. The distractors were rotated (underlined red) or modified (underlined blue) versions of the target, or an unrelated object (underlined green), and are illustrated underneath the relevant bar in the graph.

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 d. Subjects improved from a near chance level of performance before training (correct detection, $16.1 \pm 5.4\%$) to a performance level of 70–80% correct responses after training, at which point we stopped training (correct detection, $71.3 \pm 5.5\%$; significance of the change $p < 10^{-51}$, two-tailed paired t test, average of 38 subjects). This process took 10–15 d (Fig. 4). Longer periods of training resulted in further improvement above this level (data not shown).

Effects of position

The performance levels in Figures 3 and 4 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 (Fig. 5). 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.

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 (Fig. 6). There was significant improvement in the recognition of objects that shared components with the trained target (before training, $27.7 \pm 10.0\%$; after training, $54.0 \pm 6.4\%$; $p < 10^{-3}$, 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 (Fig. 7). We examined the changes in reaction times occurring with training (Fig. 8). Although reaction times went down for all objects, the most substantial decreases were seen for trained objects and for objects sharing a component with the trained object (trained objects before training, 1200 ± 300 ms; after training, 820 ± 200 ms; $p < 10^{-32}$; objects sharing components with trained objects, before training, 1140 ± 290 ms; after training, 870 ± 230 ms; $p < 10^{-15}$). The changes were more modest for the unrelated object (before training, 1230 ± 390 ms; after training, 1080 ± 320 ms; $p > 9 \times 10^{-4}$). This drop held for the correct trials, so it was not a result of correct trials being responded to faster. For the correct trials, the decrease in reaction times was much more significant for the trained object. The fact that reaction times decreased together with improved performance indicates that our results were not an artifact of a trade off between accuracy and reaction time.

Transfer from objects to components

If the components were 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

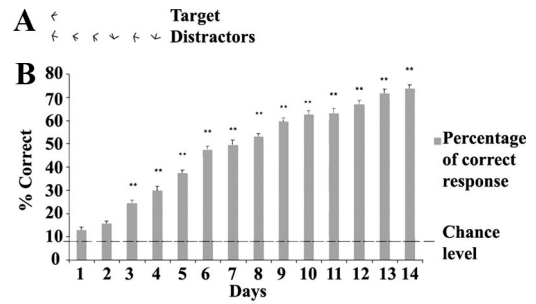


Figure 4. A, Target and distractors used for training. B, Increase in performance at detection of a target object embedded in an array of distractors of similar shape, through several days of training. Performances are given as the percentage of correct detection of the target against the total number of appearances of the target. Dashed line represents chance level. Single subject, Error bars represent SEs across individual blocks. * $p < 0.01$; ** $p < 0.001$, compared with the performance of the first day.

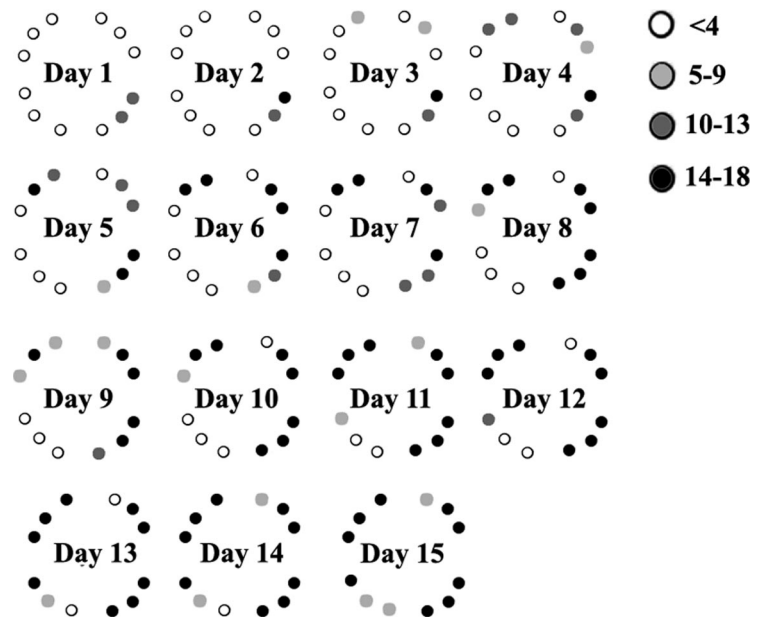


Figure 5. Point-by-point learning within the array. The target position was changed from trial to trial, in a random block design, for a total of 18 presentations per position. The shading of the squares indicates the level of performance at each day of training. Although the sequence of target presentation 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.

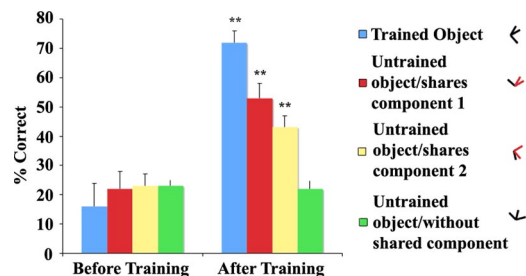


Figure 6. Performance on recognition of trained (blue) versus untrained shapes that either shared (red and yellow) or did not share (green) a component with the trained shape. For the purposes of this illustration, the shared components are highlighted in red. There was significant improvement in recognizing untrained shapes that shared a component with the trained shape but not for shapes with no shared components. One subject, Error bars represent SEs across subjects. * $p < 0.01$; ** $p < 0.001$, compared with the pretraining levels of performance.

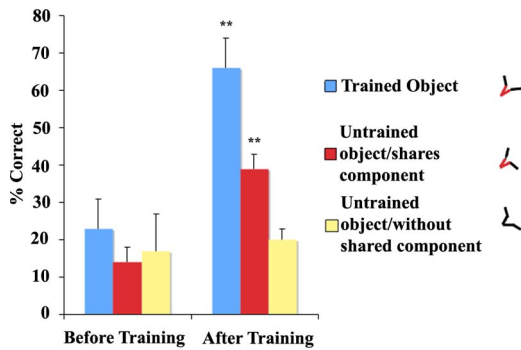


Figure 7. Performance on recognition of trained (blue) versus untrained shapes that either shared (red) or did not share (yellow) a component with the trained shape, for four-line shapes. For the purposes of this illustration, the shared components are highlighted in red. There was significant improvement in recognizing untrained shapes that shared a component with the trained shape but not for shapes without shared components. Here, training and transfer for four-line shapes followed the same pattern as for three-line shapes. Single subject, Error bars represent SEs across individual blocks. * $p < 0.01$; ** $p < 0.001$, compared with the pretraining levels of performance.

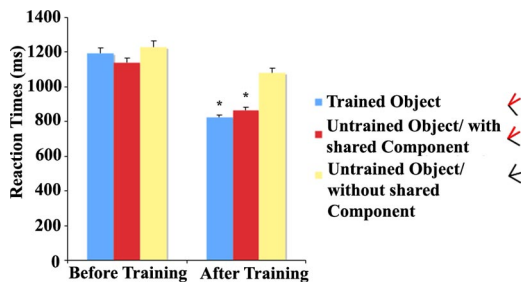


Figure 8. Average reaction times for the correct trials while looking for trained (blue) object versus untrained shapes that either shared (red) or did not share (yellow) a component with the trained shape. For the purposes of this illustration, the shared components are highlighted in red. There was significantly higher improvement in reaction times both when looking for trained shapes and for untrained shapes that shared a component with the trained shape than for shapes with no shared components. * $p < 0.0001$, compared with the pretraining levels of reaction times.

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 (Fig. 9). We observed that components of the trained objects were recognized at a higher performance level (before training, $7.4 \pm 2.3\%$; after training, $45.0 \pm 6.3\%$; two-tailed paired t test, two subjects) by the subjects, than two-line objects that were not components of the trained object.

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 (Fig. 10). 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

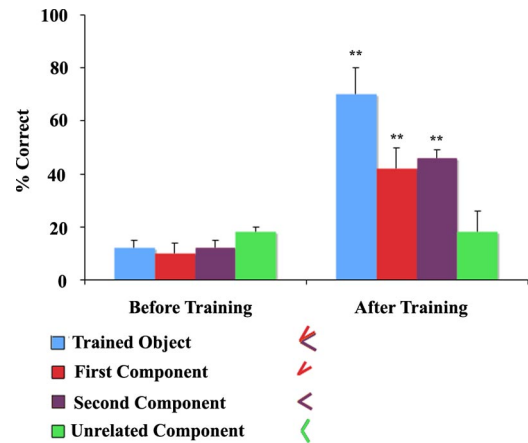


Figure 9. Performance on recognition of trained (blue) shape versus components that were either part (red and purple) or were not part (green) of the trained shape. For the purposes of illustration, the components that were part of the trained shape are highlighted in red and yellow. There was significant transfer to both components of the shape but not to the unrelated component. Single subject, Error bars represent SEs across individual blocks. * $p < 0.01$; ** $p < 0.001$, compared with the pretraining levels of performance.

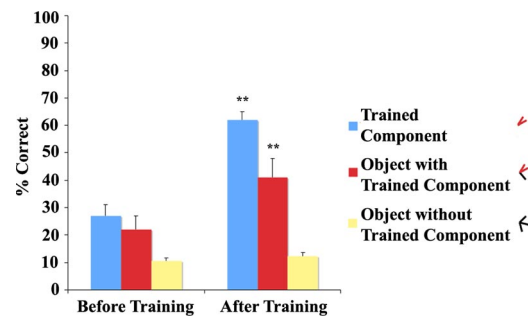


Figure 10. Performance on recognition of trained (blue) component versus untrained object that contains the trained component (red). For the purposes of illustration, the component is highlighted in red. There was significant transfer of training from the component to the object. Single subject, Error bars represent SEs across individual blocks. * $p < 0.01$; ** $p < 0.001$, compared with the pretraining levels of performance.

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).

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. 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

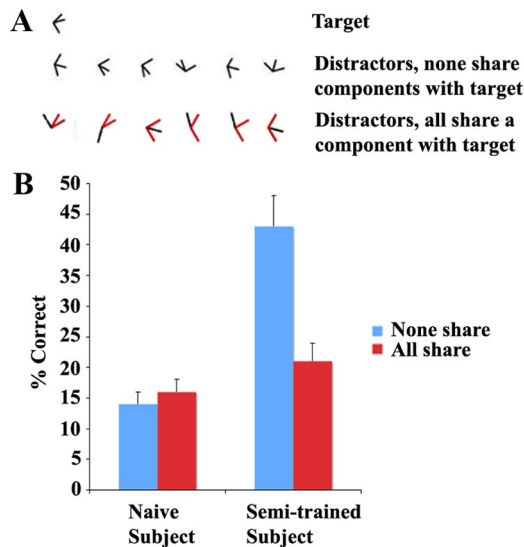


Figure 11. *A*, Target and distractors. In the first condition, none of the distractors shared components with the target. In the second condition, each distractor shared one component with the target. For the purposes of this illustration, the shared components are highlighted in red. Each component 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 SEs across individual blocks.

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, naive subjects performed at chance level with little observable difference. However, subjects that had some training in searching for the target shape performed much more poorly at detecting this shape when it was embedded in an array of distractors with shared components with the target than in an array with no shared components (performance with shared components in distractors, $20.9 \pm 14.8\%$; performance without shared components in distractors, $43.8 \pm 14.9\%$; $p < 10^{-4}$, two-tailed paired *t* test) (Fig. 11, single subject).

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 (Treisman and Gelade, 1980; Bergen and Julesz, 1983; Steinman, 1987), 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 ($42.2 \pm 6.7\%$, 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 ($72.3 \pm 6.7\%$, 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. Although the components of the trained object were shared with the distractors used during training, we nevertheless observed a significant transfer to

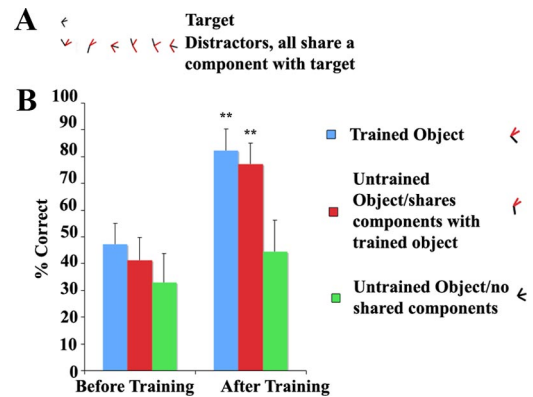


Figure 12. *A*, Target and distractors. During training, each distractor shared one component with the target. For the purposes of this illustration, the shared components are highlighted in red. Each component of the target object appeared in two of the six distractors. *B*, Performance on recognition 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 SEs across individual blocks. $*p < 0.01$; $**p < 0.001$, compared with the pretraining levels of performance.

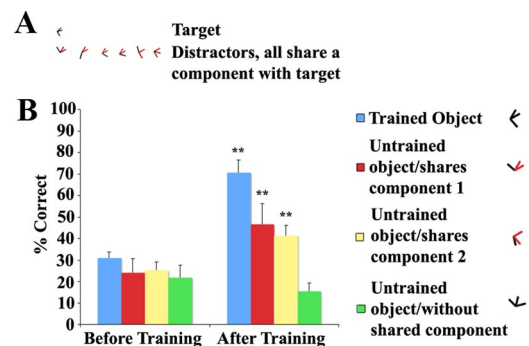


Figure 13. *A*, Target and distractors. During training, each distractor shared one component with the target. For the purposes of this illustration, the shared components are highlighted in red. One component appeared in four of the six distractors, while the other component appeared in the remaining 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 SEs across individual blocks. $*p < 0.01$; $**p < 0.001$, compared with the pretraining levels of performance.

the objects that shared components with the target (Fig. 12). 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, $30.2 \pm 5.9\%$; after training, $41.1 \pm 8.8\%$; $p > 0.01$, two-tailed paired *t* test).

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 (Fig. 13). 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 as the other component within the stimuli. We trained our subjects under this condition until their performance reached an arbitrary chosen $\sim 70\%$ level (performance before training, $30.8 \pm 5.0\%$; performance after training, $70.7 \pm 10.2\%$; $p < 0.0058$, 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, $24.1 \pm 11.5\%$; performance after training with different distractors, $32.3 \pm 12.6\%$; with the same distractors, $46.5 \pm 16.6\%$; average of three subjects, $p > 0.16$ and $p < 0.023$, respectively, two-tailed paired *t* test). There was also significant posttraining 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, $25.2 \pm 6.6\%$; performance after training, $41.4 \pm 8.0\%$; $p < 0.02$, 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. The level of transfer to either component did not appear to depend on the frequency with which the component appeared in the distractors.

Discussion

We studied what is learned in object recognition by training subjects on a visual search task and looking for transfer to untrained objects. In this study, we used search as a probe to measure an object's recognizability. The search task simulates the constraints of normal viewing conditions where objects have to be identified in complex environments. Although rapid recognition of objects may require specialized strategies, complex objects can be recognized after brief exposures (Thorpe et al., 1996). Our stimuli do not have some existing information for real objects such as color, texture, and shading. We confined our shapes to cues based on linear contour elements, since objects composed only of line contours (such as black and white drawings) can be recognized. While our objects do not contain every kind of contour shape, many objects contain edges and corners of the sort used here. Identifiable components that could be shared between objects allowed us to investigate transfer and create a population of objects with shared features. The experimental paradigm can be adapted for psychophysical and physiological studies of the role of other features and of shape representation in the visual pathway.

Previous research indicates that object recognition is subject to perceptual learning, the improvement of performance in discriminating a perceptual attribute through repetition. In vision, individuals can improve their performance for several stimulus attributes, such as orientation (Vogels and Orban, 1985; Shiu and Pashler, 1992; Schoups et al., 1995), depth (Fendick and Westheimer, 1983), motion direction (Ball and Sekuler, 1982, 1987), segmentation through texture cues (Karni and Sagi, 1991, 1993), and visual hyperacuity (Poggio et al., 1992; Fahle and Edelman, 1993; Kapadia et al., 1995). Perceptual learning is specific to simple stimulus attributes like location, orientation, eye, or task-relevant

attributes. Earlier studies emphasize perceptual learning of simple stimulus attributes; the current study examines the learning of more complex forms, whereby object recognition becomes more rapid and is done in parallel with other objects. Sigman and Gilbert (2000) have shown that perceptual learning of objects is specific to the learned object. Here, we expand upon these results.

While in our study object learning was comparable with earlier studies, with steady, location-specific improvement over several thousand trials, we observed significant exceptions to specificity. There was significant transfer to objects sharing geometric components with the trained object. Since there was no transfer to unrelated but similar control shapes, the transfer was unlikely to reflect a nonspecific improvement in task. This suggests that objects are learned in a parts-based manner. Earlier psychophysical experiment studies show that components are necessary for object recognition, objects can be identified by partial exposure to a subset of their components, and similar objects can be differentiated through differing parts (Biederman, 1987; Biederman and Gerhardstein, 1993; Biederman and Bar, 1999; Gosselin and Schyns, 2001). However, the finding that not all components are required for identification does not in itself imply a parts-based representation, only that features may be redundant. Although evidence exists for holistic representation of certain objects, such as faces (Young et al., 1987; Tanaka and Farah, 1993; Wang et al., 1998; Schiltz and Rossion, 2006; Freiwald et al., 2009), this may result from expertise (Gauthier et al., 2003) or the specialized requirements for identifying members of a large set with subtle interindividual differences. We provide additional evidence for parts-based object representation. In our study, individual components were also learned when the whole object is learned. Conversely, learning a simple object improved recognition of more complex objects containing that object as a component. Furthermore, we showed that individual parts can be used to recognize multiple shapes. These results show that object parts are processed as individual units. Performance is also affected by very small variations in the objects' shape, which were sufficient to abolish learning. Therefore, transfer in our setup appears to rely on the actual components rather than the geometric "distance" between objects, consistent with a parts-based model of recognition rather than a holistic mechanism of template matching through mental transformations or interpolations (Riesenhuber and Poggio, 2000). The transfer of learning through one shared component was incomplete, and all components of an object were learned, indicating the necessity of combinations of components for recognition.

Our results resonate with the feature selectivity of neurons in monkey IT. There, a large fraction of neurons are sensitive to simplified parts of objects, and objects activate cortical columns 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 (Grill-Spector et al., 2001). The classical view of a cortical hierarchy, with ascending levels representing increasing complexity, would include a representation of parts at earlier stages. But the key issue is the level of complexity at the highest level, where the parts-based representation would involve distributed activity across multiple columns. Transfer between objects sharing components suggests that these components become represented in a way that involves access to the parts at whatever level of the hierarchy they are represented.

It is plausible that learning might change the representation of object parts. Past studies suggest a sharpening of tuning of IT neurons for diagnostic features upon learning (Sigala, 2004).

However, learning occurs in a location-specific manner in our study, which is also consistent with past reports (Crist et al., 1997; Dill and Fahle, 1997). Location specificity is a property of earlier areas in the ventral visual stream rather than LOC or IT. Therefore, a top-down reorganization of the object processing to early visual cortices is more likely (Sigman and Gilbert, 2000; Sigman et al., 2005). V1 cells 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) and have a considerable measure of plasticity in the receptive field properties (Gilbert and Wiesel, 1992; Obata et al., 1999; Crist et al., 2001) and the capacity to undergo sprouting and synaptogenesis (Darian-Smith and Gilbert, 1994; Gilbert et al., 1996; Stettler et al., 2006; Yamahachi et al., 2008). This plasticity can be used to reorganize elementary feature maps to represent object parts. One potential advantage of such a shift is an increased ability to perform rapid, parallel recognition. The representation of object features along the visual pathway and how experience changes it will continue to be elaborated.

Learning in our task requires top-down influences, because only features of attended objects are learned, as distinct from learning of unattended attributes (Watanabe et al., 2001; Seitz and Watanabe, 2003). Feedback was not required, since subjects that were not given feedback still learned the targets, but feedback made learning more rapid and robust, consistent with past literature (Fahle and Edelman, 1993; Herzog and Fahle, 1997, 1999). Physiological studies show the requirement of cholinergic input from nucleus basalis, which presumably represents the physiological basis for reward signals (Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998). Phasic dopamine signals aid learning as well by signaling mismatches between expected and reward outcome and by attracting attention to novel, unexpected visual events (Schultz and Dickinson, 2000; Nakahara et al., 2004; Dommett et al., 2005; Redgrave et al., 1999, 2008). Even without explicit feedback, there may be an implicit reward signal when subjects see the target, enabling learning.

Past studies suggest that objects in a visual search task pop-out if they differ from the distractors by an elementary feature; the search is inefficient if they differ by combinations of elementary features (Treisman and Gelade, 1980). Our study shows that object parts can become elementary features through learning. Possibly, performance in visual search is not a dichotomy of parallel and serial search but a continuum (Wolfe et al., 1989; Wolfe, 1997; Joseph et al., 1997; Wolfe, 2003). Our observation of a gradual improvement rather than an abrupt switch from poor to good performance supports this possibility.

The presence of distractors also affected performance. Smaller stimulus arrays improved untrained performance, training becoming possible in conditions where it was not. This result agrees with the properties of difficult (serial) search tasks where increasing number of distractors reduces performance (Treisman and Gelade, 1980; Bergen and Julesz, 1983; Sagi and Julesz, 1985; Steinman, 1987). The shape of the distractors affected target recognition as well. According to the fragment-based hierarchy model of recognition (Ullman, 2007), components are most useful for recognition when they are highly informative, i.e., when they appear within the object and rarely in the environment. Using distractors that contain components of the target, we reduced the informativeness of those components. This impaired performance and learning; however, parts were still used for recognition, since transfer still occurred. No single component identified the target; therefore, even components occurring more frequently among the distractors contributed to recognition. It is

likely that the visual system finds the target by performing an “and” operation, using multiple components. Our findings show how familiar shapes, reflecting the regularities of the visual environment, become assimilated during perceptual learning and suggests how these shapes may be represented.

References

- Bakin JS, Weinberger NM (1996) Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc Natl Acad Sci U S A* 93:11219–11224.
- Ball K, Sekuler R (1982) A specific and enduring improvement in visual motion discrimination. *Science* 218:697–698.
- Ball K, Sekuler R (1987) Direction-specific improvement in motion discrimination. *Vision Res* 27:953–965.
- Bergen JR, Julesz B (1983) Parallel versus serial processing in rapid pattern discrimination. *Nature* 303:696–698.
- Biederman I (1987) Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94:115–147.
- Biederman I, Bar M (1999) One-shot viewpoint invariance in matching novel objects. *Vision Res* 39:2885–2899.
- Biederman I, Gerhardstein PC (1993) Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *J Exp Psychol Hum Percept Perform* 19:1162–1182.
- Biederman I, Gerhardstein PC (1995) Viewpoint-dependent mechanisms in visual object recognition: reply to Tarr and Bühlhoff (1995) *J Exp Psychol Hum Percept Perform* 21:1506–1514.
- Crist RE, Kapadia MK, Westheimer G, Gilbert CD (1997) Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol* 78:2889–2894.
- Crist RE, Li W, Gilbert CD (2001) Learning to see: experience and attention in primary visual cortex. *Nat Neurosci* 4:519–525.
- Darian-Smith C, Gilbert CD (1994) Axonal sprouting accompanies functional reorganization in adult cat striate cortex. *Nature* 368:737–740.
- Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature* 399:655–661.
- Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062.
- Dill M, Fahle M (1997) The role of visual field position in pattern-discrimination learning. *Proc Biol Sci* 264:1031–1036.
- Dommett E, Coizet V, Blaha CD, Martindale J, Lefebvre V, Walton N, Mayhew JE, Overton PG, Redgrave P (2005) How visual stimuli activate dopaminergic neurons at short latency. *Science* 307:1476–1479.
- Fahle M, Edelman S (1993) Long-term learning in vernier acuity: effects of stimulus orientation, range and of feedback. *Vision Res* 33:397–412.
- Fendick M, Westheimer G (1983) Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Res* 23:145–150.
- Freiwald WA, Tsao DY, Livingstone MS (2009) A face feature space in the macaque temporal lobe. *Nat Neurosci* 12:1187–1196.
- Gauthier I, Curran T, Curby KM, Collins D (2003) Perceptual interference supports a non-modular account of face processing. *Nat Neurosci* 6:428–432.
- Gilbert C, Ito M, Kapadia M, Westheimer G (2000) Interactions between attention, context and learning in primary visual cortex. *Vision Res* 40:1217–1226.
- Gilbert CD, Wiesel TN (1992) Receptive field dynamics in adult primary visual cortex. *Nature* 356:150–152.
- Gilbert CD, Das A, Ito M, Kapadia M, Westheimer G (1996) Spatial integration and cortical dynamics. *Proc Natl Acad Sci U S A* 93:615–622.
- Gilbert CD, Sigman M, Crist RE (2001) The neural basis of perceptual learning. *Neuron* 31:681–697.
- Gosselin F, Schyns PG (2001) Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res* 41:2261–2271.
- Gray CM (1999) The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron* 24:31–47, 111–125.
- Grill-Spector K, Kourtzi Z, Kanwisher N (2001) The lateral occipital complex and its role in object recognition. *Vision Res* 41:1409–1422.
- Herzog MH, Fahle M (1997) The role of feedback in learning a vernier discrimination task. *Vision Res* 37:2133–2141.
- Herzog MH, Fahle M (1999) Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Res* 39:4232–4243.

- Hoffman DD, Richards WA (1984) Parts of recognition. *Cognition* 18:65–96.
- Joseph JS, Chun MM, Nakayama K (1997) Attentional requirements in a 'preattentive' feature search task. *Nature* 387:805–807.
- Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–856.
- Karni A, Sagi D (1991) Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc Natl Acad Sci U S A* 88:4966–4970.
- Karni A, Sagi D (1993) The time course of learning a visual skill. *Nature* 365:250–252.
- Kilgard MP, Merzenich MM (1998) Cortical map reorganization enabled by nucleus basalis activity. *Science* 279:1714–1718.
- Li W, Gilbert CD (2002) Global contour saliency and local colinear interactions (2002). *J Neurophysiol* 88:2846–2856.
- Li W, Piëch V, Gilbert CD (2004) Perceptual learning and top-down influences in primary visual cortex. *Nat Neurosci* 7:651–657.
- Li W, Piëch V, Gilbert CD (2006) Contour saliency in primary visual cortex. *Neuron* 50:951–962.
- Li W, Piëch V, Gilbert CD (2008) Learning to link visual contours. *Neuron* 57:442–451.
- Marr D (1980) Visual information processing: the structure and creation of visual representations. *Philos Trans R Soc Lond B Biol Sci* 290:199–218.
- Marr D, Nishihara HK (1978) Representation and recognition of the spatial organization of three-dimensional shapes. *Proc R Soc Lond B Biol Sci* 200:269–294.
- Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. *Neuron* 41:269–280.
- Obata S, Obata J, Das A, Gilbert CD (1999) Molecular correlates of topographic reorganization in primary visual cortex following retinal lesions. *Cereb Cortex* 9:238–248.
- Poggio T, Fahle M, Edelman S (1992) Fast perceptual learning in visual hyperacuity. *Science* 256:1018–1021.
- Posner MI, Gilbert CD (1999) Attention and primary visual cortex. *Proc Natl Acad Sci U S A* 96:2585–2587.
- Redgrave P, Prescott TJ, Gurney K (1999) Is the short-latency dopamine response too short to signal reward error? *Trends Neurosci* 22:146–151.
- Redgrave P, Gurney K, Reynolds J (2008) What is reinforced by phasic dopamine signals? *Brain Res Rev* 58:322–339.
- Riesenhuber M, Poggio T (2000) Models of object recognition. *Nat Neurosci* 3 Suppl:1199–204.
- Sagi D, Julesz B (1985) "Where" and "what" in vision. *Science* 228:1217–1219.
- Schiltz C, Rossion B (2006) Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage* 32:1385–1394.
- Schneider W, Eschman A, Zuccolotto A (2002a) E-prime reference guide. Pittsburgh: Psychology Software Tools.
- Schneider W, Eschman A, Zuccolotto A (2002b) E-prime user's guide. Pittsburgh: Psychology Software Tools.
- Schoups AA, Vogels R, Orban GA (1995) Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *J Physiol* 483:797–810.
- Schultz W, Dickinson A (2000) Neuronal coding of prediction errors. *Annu Rev Neurosci* 23:473–500.
- Seitz AR, Watanabe T (2003) Psychophysics: is subliminal learning really passive? *Nature* 422:36.
- Shiu LP, Pashler H (1992) Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept Psychophys* 52:582–588.
- Sigala N (2004) Visual categorization and the inferior temporal cortex. *Behav Brain Res* 149:1–7.
- Sigman M, Gilbert CD (2000) Learning to find a shape. *Nat Neurosci* 3:264–269.
- Sigman M, Pan H, Yang Y, Stern E, Silbersweig D, Gilbert CD (2005) Top-down reorganization of activity in the visual pathway after learning a shape identification task. *Neuron* 46:823–835.
- Steinman SB (1987) Serial and parallel search in pattern vision? *Perception* 16:389–398.
- Stettler DD, Yamahachi H, Li W, Denk W, Gilbert CD (2006) Axons and synaptic boutons are highly dynamic in adult visual cortex. *Neuron* 49:877–887.
- Tanaka JW, Farah MJ (1993) Parts and wholes in face recognition. *Q J Exp Psychol A* 46:225–245.
- Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *J Neurophysiol* 66:170–189.
- Tarr MJ, Bühlhoff HH (1995) 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* 21:1494–1505.
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature* 381:520–522.
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136.
- Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci* 4:832–838.
- Ullman S (2007) Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn Sci* 11:58–64.
- Vogels R, Orban GA (1985) The effect of practice on the oblique effect in line orientation judgments. *Vision Res* 25:1679–1687.
- von der Malsburg C (1999) The what and why of binding: the modeler's perspective. *Neuron* 24:95–104, 111–125.
- Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neurosci Res* 32:33–46.
- Watanabe T, Náñez JE, Sasaki Y (2001) Perceptual learning without perception. *Nature* 413:844–848.
- Wolfe JM (1997) Experimental psychology. In *A blink of the mind's eye*. *Nature* 387:756–757.
- Wolfe JM (2003) Moving towards solutions to some enduring controversies in visual search. *Trends Cogn Sci* 7:70–76.
- Wolfe JM, Cave KR, Franzel SL (1989) Guided search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform* 15:419–433.
- Yamahachi H, Marik S, McManus JN, Denk W, Gilbert CD (2008) Axonal sprouting and pruning accompany functional reorganization in primary visual cortex. *Soc Neurosci Abstr* 34:16.6.
- Young AW, Hellawell D, Hay DC (1987) Configurational information in face perception. *Perception* 16:747–759.