PERGAMON

Vision Research 40 (2000) 473–484

Vision Research

www.elsevier.com/locate/visres

Perceptual learning in object recognition: object specificity and size invariance

Christopher S. Furmanski *, Stephen A. Engel

Department of Psychology, University of California, Los Angeles, Box 951563, Franz Hall, Los Angeles, CA 90095-1563, USA

Received 10 March 1999; received in revised form 7 June 1999

Abstract

A series of four experiments measured the transfer of perceptual learning in object recognition. Subjects viewed backward-masked, gray-scale images of common objects and practiced an object naming task for multiple days. In Experiment 1, recognition thresholds decreased on average by over 20% over 5 days of training but increased reliably following the transfer to a new set of objects. This suggests that the learning was specific to the practiced objects. Experiment 2 ruled out familiarity with strategies specific to the experimental context, such as stimulus discrimination, as the source of the improvement. Experiments 3 and 4 found that learning transferred across changes in image size. Learning could not be accounted for solely by an improvement in general perceptual abilities, nor by learning of the specific experimental context. Our results indicate that a large amount of learning took place in object-specific mechanisms that are insensitive to image size. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Perceptual learning; Object recognition; Psychophysics

1. Introduction

The human visual system recognizes objects accurately while discounting factors that greatly change the visual image, such as viewing angle and viewing distance. Perceptual psychologists have developed a variety of experimental methods to break recognition into identifiable components. This paper uses perceptual learning[1] to identify the mechanisms that underlie recognition of everyday objects. We use learning to reveal mechanisms that are specific to individual objects, and then show that these mechanisms are insensitive to changes in image size. The methods developed here can be used to study many other aspects of object recognition.

For many tasks, human visual performance improves dramatically with practice. Examining how this practice transfers to different stimuli can reveal properties of visual mechanisms[2]. For example, practicing a Vernier acuity task with a horizontal stimulus leads to improvement mainly for the practiced stimulus; learning does not transfer to unpracticed orientations (e.g. Fahle, 1994). This pattern suggests that Vernier acuity is not produced by one indivisible mechanism. Rather, multiple mechanisms exist, each one responsible for a narrow range of orientations. Most studies adopting this approach have focused on relatively early stages of visual processing, using tasks such as Vernier acuity and pop-out (e.g. Karni & Sagi, 1991; Ahissar & Hochstein, 1993; Fahle & Edelman, 1993). This paper uses a similar logic to identify mechanisms responsible for recognition of common objects.

Prior studies of long-term perceptual learning in recognition have not used common objects. Instead, subjects made fine discriminations on novel shapes

* Corresponding author. Tel.: +1-310-267-2030; fax: +1-310-206-5895; http://rocky.psych.ucla.edu/furmansk.
E-mail address: furmansk@psych.ucla.edu (C.S. Furmanski)

[1] The term perceptual learning has many interpretations and sometimes refers to improvement on a known stimulus dimension. However, we adopt a broader definition, as outlined by Gibson (1969): 'Perceptual learning refers to an increase in the ability to extract information from the environment, as a result of experience and practice …'. A similar definition has been used by several other experimenters (e.g. Karni & Sagi, 1991).

[2] In this paper, mechanism refers to a behaviorally identifiable component of perceptual processing. An example of a perceptual mechanism is a spatial frequency channel (e.g. Graham, 1989).

PII: S0042-6989(99)00134-0

(Tarr & Pinker, 1989; Bulthoff & Edelman, 1992; Edelman & Bulthoff, 1992; Edelman, 1995; Liu, Knill & Kersten, 1995; Tarr, 1995; Gauthier & Tarr, 1997; Gauthier, Williams, Tarr & Tanaka 1998). However, these studies have all examined subordinate-level recognition (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). It is possible that different mechanisms exist for basic-level and subordinate-level recognition (Bulthoff, Edelman & Tarr, 1995; Hummel & Stankiewicz, 1996; but also see Edelman, 1995; Edelman & Duvdevani-Bar 1997; Tarr & Gauthier, 1998). Accordingly, previous studies of perceptual learning in recognition may not generalize to recognition of common objects.

Common objects have been used in some studies measuring viewpoint specificity (e.g. Murray, Jolicoeur, McMullen & Ingleton, 1993; Liu, 1996) and picture naming (Bartram, 1973, 1974). However, these studies measured learning that occurred within a single day. Such rapid learning may reflect 'fast perceptual learning,' whose mechanisms may differ from those responsible for learning that occurs over many days (Karni & Sagi, 1993).

The experiments reported here measure the transfer of perceptual learning in order to reveal mechanisms that perform basic-level recognition. Subjects practiced recognizing objects for multiple days, which resulted in improved performance in a naming task. Subjects were then tested on new stimuli; transfer was measured by comparing performance on the trained and new stimuli. The pattern of transfer provides evidence about the nature of the mechanisms that changed with practice. These results will constrain theories of recognition if learning is due to changes in the same mechanisms that support unpracticed, everyday object recognition.

Perceptual learning in object recognition, however, might result from changes in two alternative types of mechanisms. First, learning could be due to improvement in general mechanisms. Training might improve subjects' general perceptual abilities, such as overall visual sensitivity. Second, learning could reflect development of very specific mechanisms. Subjects might learn to use strategies that take advantage of the experimental context. For example, they could learn to discriminate between the particular stimuli used in the study.

Experiments 1 and 2 rule out both alternatives and validate perceptual learning as a technique for studying basic-level object recognition. Experiment 1 establishes the basic learning method and tests the generality of learning by measuring transfer to new objects. Experiment 2 tests the specificity of learning by measuring transfer to a new experimental context. Finally, Experiments 3 and 4 demonstrate the mechanisms that improve with practice are insensitive to changes in size. Together the results of these experiments establish perceptual learning as a powerful technique for studying the mechanisms of basic-level recognition.

Fig. 1. Two examples of the stimuli and a sample mask image.

2. Experiment 1

Experiment 1 tested whether recognition of common objects improves with practice and measured the generality of the improvement. Performance was measured using an object-naming task. Subjects viewed briefly presented gray-scale images of objects that were followed by a pattern mask. They responded by typing the name of the object and were provided with feedback. Subjects practiced the task on a small number of objects for 5 consecutive days and on the sixth day were tested on a new set of objects.

We expected that recognition would improve with practice and that this improvement would be object specific: Switching to a new set of objects following training should produce a decrease in performance. If learning is object specific, then it cannot be explained as an improvement of general perceptual abilities.

2.1. Methods

2.1.1. Participants

Five observers participated in the experiment, including both authors. Observer SAE was very experienced at viewing stimuli under similar speeded conditions. Other subjects had little or no practice on this task. Subjects' visual acuity was normal or corrected-to-normal.

2.1.2. Stimuli

Digitized gray-scale images of 60 common objects (see Appendix A; for examples, see Fig. 1) were presented to subjects using a Macintosh computer. Images

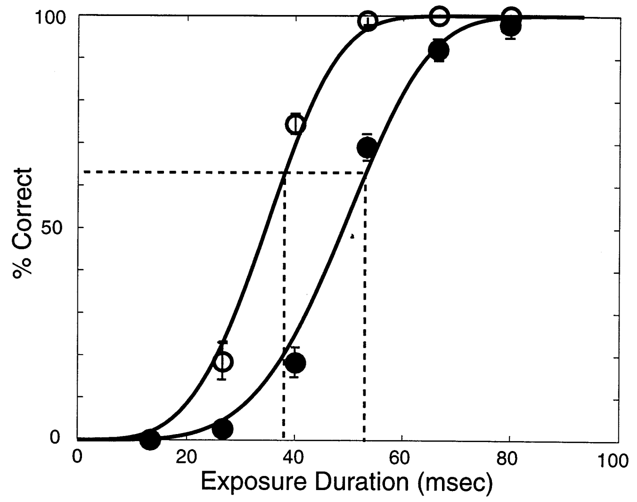

Fig. 2. Typical data and psychometric functions for one subject. Percent correct is plotted as a function of exposure duration. Filled circles show performance on Day 1 and open circles show Day 5. Smooth curves are Weibull functions fit to each day's data (see text for details). Dashed lines show each day's threshold, which is the exposure duration that yields performance of approximately 63% correct.

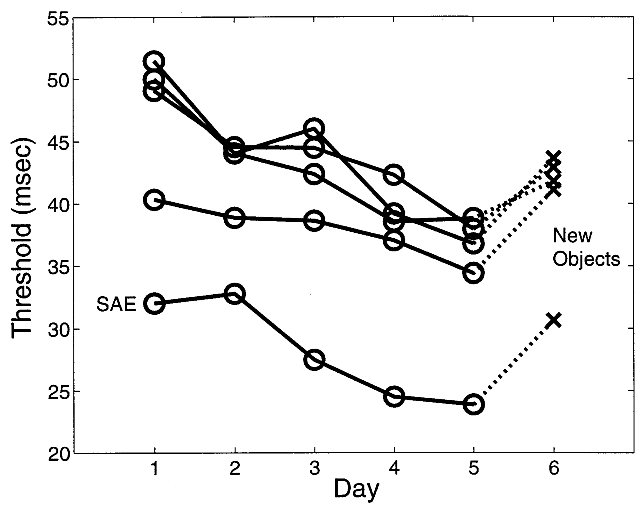

Fig. 3. Experiment 1: individual-subject thresholds are plotted as a function of training day. Horizontal lines show individual subjects. Circles represent data for trained objects and crosses represent data for new objects.

were presented at 12.5% of their original contrast in order to increase task difficulty and avoid ceiling effects. The masking stimulus was a high-contrast, gray-scale image of randomly generated intersecting black and white lines. Images were expanded evenly until each image just fit in a rectangle subtending $11.5 \times 9.3°$ of visual angle.

2.1.3. Procedure

Stimuli were presented on a 75 Hz AppleVision AV display using the Psychophysics toolbox (Brainard, 1997) for MATLAB (Mathworks, Inc.). The display's mean luminance was 37 cd/m^2.

Each daily session began with a series of 2 s displays in which each object was presented along with its name. The remainder of the session consisted of blocks of recognition trials. Each trial began with a 750 ms presentation of a fixation point on a mean field. A stimulus image was then presented for one of eight exposure durations (13, 27, 40, 53, 67, 80, 93, or 107 ms)[3]. The image was immediately followed by a 505 ms presentation of the pattern mask. Subjects responded by typing the first four letters of the object name. Exposure durations were chosen using a two-down, one-up staircase procedure. On error trials, the correct object name was displayed for 1000 ms before the beginning of the next trial.

2.1.4. Design

The experiment used three sets of 20 objects; no object was in more than one set. Subjects trained on one set of objects for five consecutive days and were then tested on a new, transfer set on the 6th day. Each daily session consisted of 10 blocks of 80 trials. In each block, exposure durations were controlled using two randomly interleaved staircases of 40 trials each. Daily sessions lasted approximately 1 h. The three object sets were counterbalanced between conditions (training and transfer) using a modified between-subjects Latin square.

2.2. Results

All the data reported here were analyzed using the same procedure. For each subject, psychometric functions were generated for each daily session by plotting average percent correct as a function of exposure duration[4]. Thresholds were estimated from the psychometric functions by fitting a Weibull function of the form:

$$\hat{p} = 1 - \exp\left(-\frac{d}{\alpha}\right)^{\beta} \tag{1}$$

where $\hat{p}$ is the estimated percent correct, d is exposure duration, the free parameter α is the threshold where performance equals 63.2%, and the fixed parameter β, the slope of the function, was set at 3.5.[5] Parameters were estimated using a maximum-likelihood procedure (Watson, 1979).

[3] Actual exposure durations were multiplies of the 75 Hz monitor refresh rate. Reported values are rounded to the nearest millisecond.

[4] The data were first corrected for guessing using the formula: $p = (x - g)/(1 - g)$, where x is the raw proportion correct, p is the corrected proportion, and g is likelihood of guessing correctly.

[5] Initially, psychometric functions were fit using two free parameters (i.e. α and β). The fixed value for β was the average of those obtained in the initial fits. The quality of the fits for the one and two free parameter models did not differ reliably.

Fig. 2 shows psychometric functions fit to one subject's data for Days 1 and 5. The dashed lines indicate thresholds where performance is 63.2% correct. Training produced a leftward shift of the psychometric function and a corresponding decrease in threshold.

Fig. 3 shows individual subject thresholds plotted for each day in Experiment 1. Thresholds decreased during training with the same object set (Days 1–5), but increased following the switch to a new object set (Day 6). Subject SAE had extensive experience with the task, but not with the stimuli. Though his overall thresholds were lower than other subjects', the amount of perceptual learning was comparable to that observed in other subjects.

In order to confirm that the shifts in thresholds corresponded to reliable changes in the raw data for individual subjects, we tested whether performance improved between Days 1 and 5. For each subject, we conducted a χ^2 test to determine if the distribution of correct responses at each exposure duration for Day 1 was independent of the distribution of responses for Day 5. More specifically, χ^2 contingency table entries were the number of correct responses, and the rows and columns represented the day of training and the exposure duration, respectively. Improvement from Day 1 to Day 5 reveals itself in this test as independence in the contingency table, because the entries for the two rows (corresponding to Days 1 and 5) would not be equal. Such an improvement in performance would result in more correct responses at lower exposure durations. All five subjects showed significant learning between Day 1 and Day 5, minimum $\chi^2(6, N = 800) = 191.56$, $P < 0.001$. All five subjects also showed a significant decrement in performance between the trained stimuli set on Day 5 and the new set on Day 6, min $\chi^2(6, N = 800) = 48.08$, $P < 0.001$.

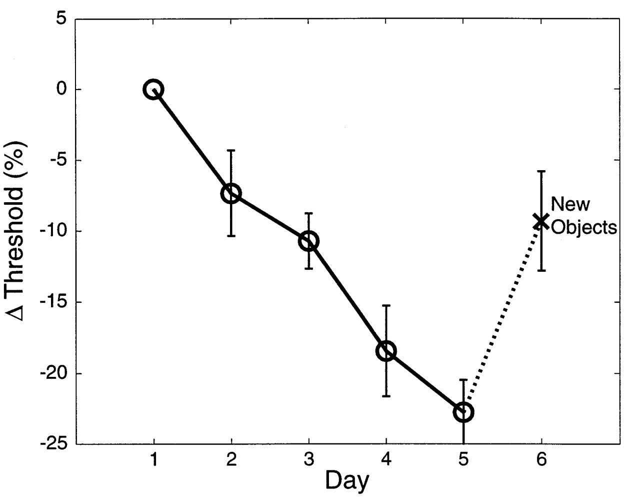

Fig. 4. Experiment 1: percent change in threshold plotted as a function of training day averaged across observers. Circles represent data for trained objects and a cross represents data for new objects. In this and all following graphs, error--bars show one standard error of the mean.

In order to reduce the between-subject variability caused by differences in starting performance, change scores were computed and then normalized by dividing them by Day 1 thresholds, yielding percent-change scores. For example, subject SAE's starting thresholds were lower than the other four subjects perhaps due to extensive practice of the task in other experiments. Percent change scores eliminate this between-subjects variability and allow for more meaningful tests of learning across subjects.

Fig. 4 shows percent change in thresholds averaged across all five subjects. Thresholds decreased reliably between Days 1 and 5 (average decrease, 23.09%), $t(4) = 9.27$, $P < 0.001$. Thresholds on Day 6, the transfer day, reliably increased from Day 5 levels (average increase, 14.02%), $t(4) = 4.50$, $P < 0.005$.

2.3. Discussion

Practice recognizing objects led to steady improvement in performance. Despite different initial thresholds, subjects improved by similar amounts when measured in percent change (ranging from 16 to 29%). On average, thresholds decreased by 23% (SD = 5.0). This effect is actually quite large; for a given exposure duration performance often increased from chance to over 75% correct. The apparent disparity between performance increase and threshold decrease is due to the steepness of the psychometric function (see Fig. 2).

Importantly, switching to unstudied objects caused a decrement in performance. This pattern of results is inconsistent with the hypothesis that all learning was caused by an improvement in general perceptual abilities. If subjects in Experiment 1 were only improving their general perceptual abilities, then performance on Days 5 and 6, for trained and untrained objects, should have been equal. However, performance decreased following the change to untrained objects. But because performance on Day 6 did not return to the untrained levels of Day 1, some of the learning was general. The remainder of the learning, accounting for more than half of the decrement in thresholds, must have occurred in mechanisms that were specific to the studied objects.

3. Experiment 1a

Perceptual learning of simple stimuli can be quite long-lasting; for example, improved thresholds for texture segregation showed almost no changes over 22 months without intervening practice (Karni & Sagi, 1993). A follow-up to Experiment 1 replicated the basic learning effect and tested its longevity.

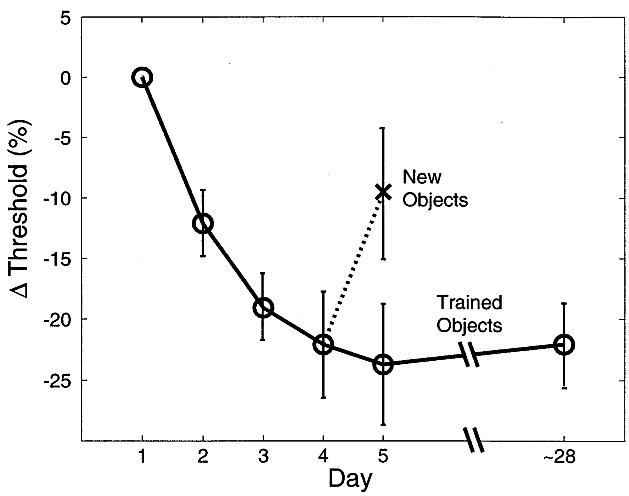

Fig. 5. Experiment 1a: percent change in threshold plotted as a function of training day averaged across observers. Circles represent data for trained objects and a cross represents data for new objects. The broken line indicates a mean delay of approximately 23 days.

3.1. Methods

3.1.1. Participants

Four new observers participated in this experiment.

3.1.2. Stimuli, Procedure, and Design

These were identical to Experiment 1, except that subjects only trained for 4 days. On the fifth day, subjects performed 200 trials on the trained set of objects followed by 200 trials on a new set. After an average of 22.8 (SD = 2.5) days with no training, subjects were retested on the trained set.

3.2. Results and discussion

Fig. 5 plots average change in thresholds for the trained and untrained sets of objects. The data for the first 5 days show the same effects that were seen in Experiment 1. Thresholds for the trained set decreased reliably during training (average decrease, 23.2%), $t(4) = 4.22$, $P < 0.01$. On the transfer day, thresholds for the untrained set reliably increased (average difference, 12.30%), $t(4) = 3.83$, $P < 0.02$. After 23 days, thresholds did not increase reliably from Day 5 levels (average difference, 2.68%), $t(4) = 0.499$, $P > 0.1$. The fact that thresholds remained stable, despite 3 weeks without practice, suggests the learning may result from long-term neural changes rather than from short-term shifts in strategy (Karni & Sagi, 1993; Karni & Bertini, 1997).

4. Experiment 2

The two previous experiments demonstrated that recognition of common objects improves dramatically with training. This learning could not be accounted for solely by an improvement in general perceptual abilities. Rather, the data suggest that a substantial amount of learning took place in mechanisms specific to the studied objects.

However, subjects may not have improved at recognition of individual objects per se, but instead, may have learned to better discriminate between the objects within the trained set. For example, subjects may have learned to identify an object using some basic attribute that no other object in the set possessed, such as its global orientation. Such learning would be context dependent in the sense that this improvement is tied to an aspect of the specific experiment and might not reflect more general improvements in object-specific recognition. Such context-dependent learning could have accounted for the basic learning results reported in Experiment 1.

Experiment 2 aimed to rule out context-dependent learning as a possible account of perceptual improvement by testing whether learning generalized to a new training context. Following training, subjects were tested on a mixed set, of both learned and new objects. If learning is specific to the trained context, then performance on practiced objects should decrease when tested in this mixed set because the other objects in the set have changed. We expected that performance would remain the same, adding support to the hypothesis that learning takes place in mechanisms specific for recognition of the studied objects.

4.1. Methods

4.1.1. Participants

Five new observers participated in the experiment. All subjects were naive to the purpose of the study, and had no previous psychophysics experience.

4.1.2. Stimuli and procedure

The stimuli and procedure were identical to those used in Experiment 1, except that the stimuli subtended $16.4 \times 12.7°$ visual angle.

4.1.3. Design

Experiment 2 used two overlapping sets of ten objects each; a subset of four objects was the same in each set. Subjects performed 300 trials on the first day of training followed by 3 consecutive days of 600 trials each on one object set. On Day 5, subjects were tested for 300 trials on the second object set. The new set contained four of the objects from the trained set and six new objects. Object sets were counterbalanced between subjects.

4.2. Results

For each subject on each day, recognition thresholds were calculated as described in Experiment 1. Fig. 6 shows average changes in recognition thresholds for Experiment 2. Thresholds on Day 5 were computed separately for both new and studied objects.

Thresholds for the trained objects decreased reliably over the first 4 days (average decrease, 14.30%), $t(4) = 2.816$, $P < 0.02$. For the trained subset of objects, thresholds between Day 4 and Day 5 were not increased by a change in context, (average difference, 6.00%), $t(4) = 1.82$, $P > 0.05$. Thresholds for the new objects on Day 5 did not differ reliably from untrained levels (average difference, 0.40%), $t(4) = 0.064$, $P > 0.1$.

4.3. Discussion

If subjects had only learned to discriminate objects in the trained set, then recognition for the trained objects should have been more difficult in the mixed set. However, performance for the trained subset of objects remained unchanged or even improved in the new context. Hence, our results suggest that subjects are doing more than simply learning to discriminate between objects in the training set. Instead, subjects are likely to be improving at recognition for objects independent of the training context.

This improvement remains specific to the studied objects, replicating the basic findings of Experiments 1 and 1a. Performance for new objects on the transfer day returned to near untrained levels.

Together, Experiments 1 and 2 suggest that learning can occur in recognition mechanisms that are selective for individual objects. Two alternative hypotheses have been ruled out; learning is not solely due to improvement in general perceptual abilities, nor is it specific to the training context. In short, the learning shown here most likely occurred in the same mechanisms that underlie everyday object recognition. Experiment 3 further investigates the nature of these mechanisms.

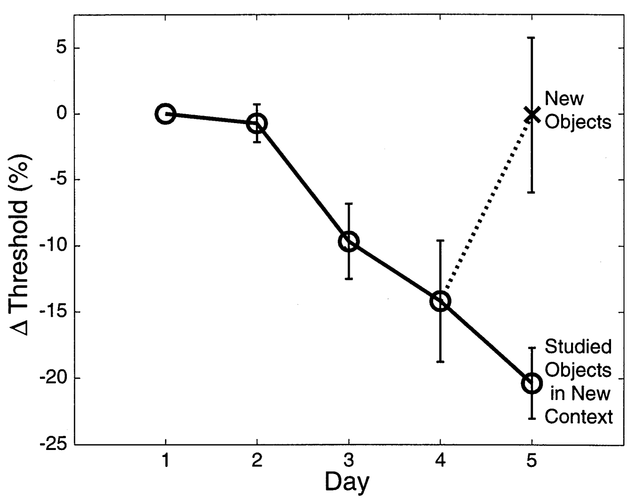

Fig. 6. Experiment 2: percent change in threshold plotted as a function of training day averaged across observers. Circles represent data for trained objects and a cross represents data for new objects. On Day 6 the objects were presented in a mixed set of new and learned objects.

5. Experiment 3

Experiment 3 tested whether the mechanisms underlying object recognition are sensitive to image size. Tests of size sensitivity have been important for both cognitive and neural theories of object recognition. For perceptual theorists, sensitivity to image size is a form of viewpoint specificity; this property may help distinguish between two rival classes of models (Biederman & Cooper, 1992). For neural theorists, sensitivity to image size can help localize learning in the visual system. Receptive fields in early cortical areas, such as V1 and V2, are more sensitive to stimulus size than are receptive fields in later visual areas such as inferior temporal cortex (Gross & Mishkin, 1977; Schwartz, Desimone, Albright & Gross, 1983).

This experiment measured how learning transferred across changes in image size. Subjects studied a set of objects for 3 days, and were then tested on images of different sizes. Image size was either doubled or halved. If learning is size specific, then thresholds should increase when the image size changes. Alternatively, if learning is size invariant, then thresholds should remain the same following changes in image size.

5.1. Methods

5.1.1. Participants

Four new observers participated in the experiment. All subjects were naive to the purpose of the study and had no previous psychophysical experience.

5.1.2. Stimuli and procedure

The procedure was identical to that used in previous experiments with the following exceptions. Two sets of five objects each were used. Objects were displayed at one of two sizes, either $16.4 \times 12.7°$ (large) or $8.2 \times 6.4°$ (small). In addition, a set of 20 mask images were used instead of the single mask image used previously. Each of the mask images was similar to the single high-contrast mask used in Experiments 1 and 2, and consisted of a series of randomly placed black and white lines on a mean field. One image from this set was randomly selected as the mask for each trial. Each mask subtended $16.4 \times 12.7°$ of visual angle for both large and small targets.

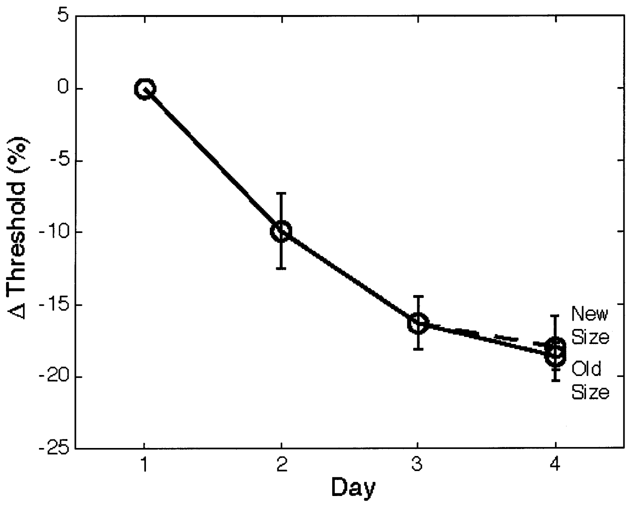

Fig. 7. Experiment 3: percent change in threshold plotted as a function of training day averaged across observers. Circles represent data for trained objects. A dashed line indicates a change in image size.

5.1.3. Design

Subjects trained on both sets of objects with one set displayed at each size. Trials for the two sets were intermixed. Stimulus set and training size were counterbalanced across subjects. Subjects trained for 640 trials per day for 3 consecutive days. On the fourth day, subjects were tested on 640 intermixed trials of both sets of objects displayed at both image sizes.

5.2. Results

Thresholds were estimated for each subject on each day as in previous experiments. Fig. 7 plots the average changes in thresholds for Experiment 3. Because separate analyses did not reveal any systematic differences between small and large training stimuli, data were averaged across trained image size. As in the previous experiments, thresholds decreased reliably during training, (average decrease, 18.78%), $t(3) = 12.82$, $P < 0.003$. On day four, performance did not differ reliably between trained and untrained sizes (mean difference, 0.57%), $t(3) = 0.207$, $P > 0.1$.

5.3. Discussion

Subjects' performance remained unchanged despite a two-fold increase or decrease in image size. These results suggest that learning is largely size invariant. Experiments 1 and 2 found that learning mainly takes place in object-specific mechanisms that support object recognition. Experiment 3 revealed that these mechanisms are not specific to image size. We will elaborate on the implications of these results for cognitive and neural theories in Section 7. Experiment 4 will attempt to rule out an alternative account of the transfer of learning across image size.

6. Experiment 4

Throughout this paper we have argued that learning occurs in the same mechanisms used to recognize everyday objects under normal circumstances. In these experiments, however, subjects received hundreds of trials with the same stimuli. It is possible that training conditions such as these may lead subjects to use specialized mechanisms for recognition (Gauthier & Tarr, 1997 Schyns & Rodet, 1997; Gauthier et al., 1998; Schyns, Goldstone & Thilbaut, 1998). Hence, the size invariance measured in Experiment 3 may only apply when subjects have become 'experts' for specific stimuli.

Experiment 4 was designed to rule out expertise as a source of size invariance. Subjects were trained for just a single day, and then tested on images of different sizes. To further limit expertise, subjects were trained on twice as many objects as in Experiment 3. Because of the larger number of objects, this experiment used a between-subjects design.

6.1. Methods

6.1.1. Participants

Sixteen subjects participated in this experiment. All participants were naive to the purpose of the study.

6.1.2. Stimuli and procedure

The stimuli and procedure were the same as those used in Experiment 3, except that the two stimulus sets contained 20 objects each.

6.1.3. Design

Subjects trained on one set of objects viewed at one size only. Trained object set and size were counterbalanced between subjects. On Day 2, subjects were tested on the trained set and either the trained set displayed at a different size or a new set of objects displayed at the trained size. Subjects performed 640 trials on each day.

6.2. Results

Thresholds were calculated as in Experiment 1. Average change in thresholds for Experiment 4 are presented in Fig. 8. Thresholds decreased reliably even after a single day of training (average decrease, 16.14%), $t(7) = 8.52$, $P < 0.001$. As in Experiment 3, changing object size did not reliably increase thresholds (average difference, 4.4%), $t(7) = 1.56$, $P > 0.1$. Thresholds for new objects did not differ reliably from untrained levels (average difference, 2.0%), $t(7) = 0.545$, $P > 0.1$.

6.3. Discussion

Despite only one day of training, learning transferred to images of different sizes but did not transfer to new objects. This pattern of results replicates Experiment 3. Because subjects only trained for a single day on a relatively large set of objects, expertise is unlikely to account for the observed size invariance.

In order to limit the amount of learning in Experiment 4, subjects were presented with many objects, fewer total trials, and fewer training days than in previous experiments. However, despite these impediments, subjects' performance in Experiment 4 improved as much or more than the first day of performance in Experiments 1–3. One reason for this might be that the actual thresholds for Day 1 were higher in this experiment than in other experiments reported here. Presenting the data in terms of percent change obscured this fact. The mean threshold on Day 1 for Experiment 4 was 55.3 ms (SD = 0.30), while the mean threshold on Day 1 for Experiment 1 was 42.95 ms (SD = 0.63). The high initial thresholds in Experiment 4 may allow more room for improvement than in previous experiments.

7. General discussion

In the experiments described above, subjects improved their ability to recognize common objects. Subjects did not simply learn to discriminate among a particular set of stimuli, nor did they just improve their general perceptual abilities. The improvement in recognition was long lasting and transferred almost completely to images of different sizes.

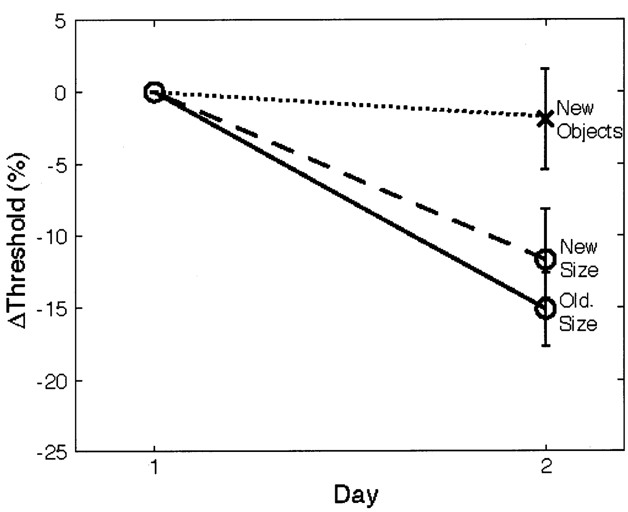

Fig. 8. Experiment 4: percent change in threshold plotted as a function of training day averaged across observers. Circles represent data for trained objects and a cross represents data for new objects. A dashed line indicates a change in image size.

7.1. Implications for theories of perceptual learning

The learning measured in our experiments resembles perceptual learning measured in many other studies. Most of this prior work used tasks designed to probe the early stages of the visual system, including Vernier hyperacuity (e.g. McKee & Westheimer, 1978; Fahle, 1994), line-segment popout (e.g. Karni & Sagi, 1991; Ahissar & Hochstein, 1993), and complex grating discrimination (Fiorentini & Berardi, 1981). Practicing these tasks, like practicing object recognition, produces long-lasting improvement (Karni & Sagi, 1993; Sireteanu & Rettenbach, 1995), stimuli specificity (e.g., Ahissar & Hochstein, 1993; Fahle, Edelman, & Poggio, 1995), between subject variability (McKee & Westheimer, 1978; Beard, Levi, & Reich, 1995; Fahle & Henke-Fahle, 1996), and in some cases, size invariance (Ahissar & Hochstein, 1996). These similarities suggest that the principles of learning uncovered in prior studies may apply to more natural tasks such as object recognition.

While most studies of perceptual learning share these general characteristics, differences arise in the stimulus specificity shown for each task. These differences are important because stimulus specificity helps localize learning within the visual stream. For example, in some studies using Vernier acuity and texture segregation, learning does not transfer across eyes (Fahle et al., 1995; Karni & Sagi, 1993). Although these results remain controversial (Beard et al., 1995; Schoups & Orban, 1996), they would suggest that the cortical locus of learning is at or before primary visual cortex (V1), where information from the two eyes converge. Conversely, for tasks such as visual search, where learning transfers well between eyes, the likely cortical locus is in or after V1 (Ahissar & Hochstein, 1996).

Similarly, the size invariance shown here implies a cortical locus of learning that is relatively late in the visual stream. Single unit studies show that individual neurons in early visual cortex (i.e., areas V1 and V2) only respond well to stimuli of a fairly limited range of sizes (Hubel & Wiesel, 1968). Such neurons could not provide the basis for learning that transfers across size. Later visual areas, such as V4, and IT, do show some size invariance (Gross & Mishkin, 1977; Schwartz et al., 1983), and hence are more plausible cortical loci for the learning observed here. Related arguments have been made about size invariance in visual search (Ahissar & Hochstein, 1993) and priming (Biederman & Cooper, 1992).

Size invariance implies that learning may be happening later in cortex than is required by the task alone. For example, improved recognition could in principle result from an increased signal-to-noise ratio in V1 neurons that encode a specific image. Such learning would not show size invariance, however, because large and small images of the same objects are encoded by different populations of V1 neurons. Hence, our results do not support the idea that learning always happens as early as possible in the visual stream (Ahissar, Laiwand, Kozminsky & Hochstein, 1998).

One important difference between our study and prior work is the use of a basic-level categorization task. In other perceptual learning studies of recognition, subjects made difficult subordinate-level discriminations. For example, Gauthier and Tarr (1997) presented subjects with complex, novel objects ('Greebles'). In this paradigm, subjects were required to make fine discriminations among very similar exemplars of a specific object class; subjects became Greeble experts. In other studies, subjects trained on similar discriminations among complex, novel objects (Tarr & Pinker, 1989; Edelman & Bulthoff, 1992; Bulthoff & Edelman, 1992; Tarr, 1995; Liu et al., 1995; Gauthier & Tarr, 1997; Gauthier et al., 1998; Tarr & Gauthier, 1998). Our study demonstrates that robust perceptual learning occurs for more coarse, basic-level categorization.

Our results agree with other studies that measured learning in category discrimination tasks. In one study, subjects practiced discriminating between exemplars from three classes of relatively similar, novel stimuli (Tarr & Gauthier, 1998). Learning transferred more to other objects sharing the same basic shape than to objects from a different shape class. Another study also demonstrated perceptual learning in a discrimination among a few (two) basic-level objects (Edelman, 1995). Our results extend these findings to more natural objects and a larger stimulus set. We explicitly rule out mechanisms specific for discrimination as the source of improvement in recognition, and test for long-term learning and size-invariance. Our data are also in general agreement with a large literature demonstrating perceptual learning occurs during categorization tasks other than object recognition (for a review, see Goldstone, 1998).

Because a single mask was used in Experiments 1 and 2, it is possible that subjects learned to discount the mask or learned some particular interaction of the mask and the stimulus. This seems unlikely, however, because Experiment 3 used 20 random masks and learning was as robust as in the previous experiments. Another possible explanation of our results is that subjects improved the temporal resolution of general visual mechanisms (Wolford, Marchak & Hughes, 1988). It is possible that such an improvement may account for some of the general learning which kept thresholds for new objects slightly below untrained levels. But increases in general temporal resolution can not account for the remainder of the observed learning which is object specific.

7.2. Implications for theories of object recognition

Because learning did not transfer to the untrained objects, it most likely occurred within neural mechanisms that are specific for the recognition of individual objects. The logic supporting this inference is identical to that used in prior studies of perceptual learning. For example, the orientation specificity of learning in a Vernier task suggests that it is performed by mechanisms tuned for a specific orientation (McKee & Westheimer, 1978; Kapadia, Gilbert & Westheimer, 1994; Sireteanu & Rettenbach, 1995). Similarly, the object specificity observed in our experiments suggests that separate neural mechanisms must signal the identity of different objects. On the other hand, the learning measured here is not specific to image size. This suggests that for any given object, the same mechanism signals the object's identity for both of the tested image sizes.

The simplest explanation of our results is that learning occurs in object specific mechanisms that receive size invariant input. Learning in such mechanisms must generalize across image size. This explanation is consistent with models that use size invariant volumetric features (e.g. Biederman, 1987; Hummel & Biederman, 1992). It is also consistent with models that normalize views with respect to size before recognition (Tarr & Pinker, 1989; Ullman, 1989; Tarr, 1995). A speculative neural basis of the observed learning is increased output of IT neurons as a result of strengthened inputs from neurons with size-invariant receptive fields. This input might arise either from neurons that encode particular views of objects, particular parts of objects, or both, but the inputs do not vary with image size.

Our data are more difficult to explain using object specific mechanisms that receive input that varies with image size. Learning in such mechanisms may be specific to the practiced image size (Poggio & Edelman, 1990; Edelman & Weinshall, 1991; Ullman & Basri, 1991; Edelman, 1995; Edelman & Duvdevani-Bar, 1997; Ullman, 1998). In one model of this type, for example, recognition results from interpolation between stored views of different sizes[6]. In this account, learning would be modeled most naturally as improved storage of the trained view, causing better recognition mainly for views of close to the practiced size. Our data exclude this basic class of explanations, however more elaborate versions of these models may be able to explain size invariant learning[7].

Prior studies report conflicting results about size invariance in processing of complex stimuli. One study

[6] It is of course possible that this type of model could use inputs that have already been normalized for size. This normalization would make the inputs to recognition size invariant, making the model a subtype of the ones discussed in the previous paragraph.

[7] For example, training could cause improved storage of all views. Or, training could lead to improvement later in processing, for example at a decision stage, where information from many object specific mechanisms is combined. While such accounts are certainly less parsimonious than size invariant input, our data can not rule them out.

measured priming in an object naming task and found large amount of transfer across image size (Biederman & Cooper, 1992). Our results are consistent with this finding, and extend it in several ways. While priming may be seen as an initial small amount of learning, we show that size invariance holds for much larger amounts of learning. For many of our subjects, recognition rates at a given exposure duration climb from near chance levels to over 75% correct. While it is difficult to compare these figures to the reaction time differences reported in priming studies, one measure of the size of effects is the number of subjects needed to obtain statistical reliability. Typical priming studies use tens of subjects. The current experiments report learning that is reliable in individual subjects. In addition, the neural mechanisms of priming may differ from those supporting perceptual learning. For example, priming may correspond to a temporary multiplicative gain change (Reinitz, Wright & Loftus, 1989). The incremental nature and the longevity of perceptual learning suggest that it is more likely to be due to changes in synaptic weighting (Karni & Bertini, 1997).

Same-different and old-new judgments, however, are affected by changes in image size (Bundesen & Larsen, 1975; Besner, 1983; Larsen, 1985; Jolicoeur 1987; Jolicoeur & Besner, 1987; Ellis, Allport, Humphreys & Collis, 1989). It seems likely that these tasks tap into different mechanisms than those used for recognition. For example, same-different judgments may be performed using mechanisms that encode individual features, rather than entire objects. These mechanisms could be size dependent, and may be supported by neurons at a relatively early stage of the visual stream. Other possible explanations of size dependency include attentional shifting between scales (Biederman & Cooper, 1992). Understanding the details of these task differences is an important direction for future investigation.

In summary, perceptual learning in object recognition is object specific, independent of the experimental context, and size invariant. Perceptual learning can also be used to address many other issues in basic-level recognition. Future work will examine whether learning of objects transfers across changes in viewing angle, illumination and retinal position. Perceptual learning should continue to be an important tool for studying the mechanisms underlying basic-level object recognition.

Acknowledgements

The authors are grateful to John Hummel, Jim Thomas, and two anonymous reviewers for their many helpful comments.

Appendix A. List of objects used in Experiments 1–4.

Experiment 1 used Sets 1a, 1b, and 1c. Experiment 2 used Sets 2a and 2b. Experiment 3 used Sets 3a and 3b at different sizes. Experiment 4 used Sets 1a and 1b at difference sizes.

Set 1a	Set 1b	Set 1c	Set 2a	Set 2b	Set 3a	Set 3b
Clock	Apple	Abacus	Brush	Chair	Bone	Bullet
Balloon	Bolt	Banana	Chair	Clock	Duck	Cookie
Calculator	Box	Cherry	Clock	Cup	Forklift	Gear
Chair	Brush	Clip	Computer	Dart	Horn	Skull
Hat	Computer	Dart	Cup	Glasses	Tack	Truck
Lamp	Copier	Desk	Fish	Key		
Lock	Eraser	Dolly	Stapler	Outlet		
Megaphone	Fish	Glasses	Scissors	Stapler		
Cup	Magnet	Key	Toaster	Taxi		
Projector	Meter	Knife	Tricycle	Watch		
Punch	Papercutter	Mallet				
Rocket	Ribbon	Mirror				
Saw	Rolodex	Outlet				
Scale	Scissors	Phone				
Stamp	Shaker	Register				
Stapler	Sharpener	Roller				
Suitcase	Toaster	Taxi				
Tape	Tricycle	Television				
Videocamera	Videotape	Wagon				
Wateringcan	Wrench	Wrench				

References

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 5718–5722.

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: specificities to stimulus characteristics. *Vision Research*, *36*, 3487–3500.

Ahissar, M., Laiwand, R., Kozminsky, G., & Hochstein, S. (1998). Learning pop-out detection: building representations for conflicting target-distractor relationships. *Vision Research*, *38*, 3095–3107.

Bartram, D. J. (1973). The effects of familiarity and practice on naming pictures of objects. *Memory and Cognition*, *1*, 101–105.

Bartram, D. J. (1974). The role of visual and semantic codes in object naming. *Cognitive Psychology*, *6*, 325–356.

Beard, B., Levi, D., & Reich, L. (1995). Perceptual learning in parafoveal vision. *Vision Research*, *35*, 1679–1690.

Besner, D. (1983). Visual pattern recognition: size preprocessing re-examined. *Quarterly Journal of Experimental Psychology*: *Human Experimental Psychology*, *35A*, 209–216.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–117.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology*: *Human Perception & Performance*, *18*, 121–133.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 60–64.

Bulthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*, 247–260.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology*: *Human Perception & Performance*, *104*, 214–220.

Edelman, S. (1995). Class similarity and viewpoint invariance in the recognition of 3D objects. *Biological Cybernetics*, *72*, 207–220.

Edelman, S., & Duvdevani-Bar, S. (1997). A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London Series B*: *Biological Sciences*, *352*, 1191–1202.

Edelman, S., & Bulthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research*, *32*, 2385–2400.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple–view representation of 3D objects. *Biological Cybernetics*, *64*, 209–219.

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology*: *Human Experimental Psychology*, *41*, 775–796.

Fahle, M. (1994). Human pattern recognition: parallel processing and perceptual learning. *Perception*, *23*, 411–427.

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: effects of stimulus orientation, range and of feedback. *Vision Research*, *33*, 397–412.

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*, 3003–3013.

Fahle, M., & Henke-Fahle, S. (1996). Interobserver variance in perceptual performance and learning. *Investigative Ophthalmology and Visual Science*, *37*, 869–877.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Research*, *21*, 1149–1158.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'greeble' expert: exploring mechanisms for face recognition. *Vision Research*, *37*, 1673–1682.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision Research*, *38*, 2401–2428.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Gross, C., & Mishkin, M. (1977). In S. Harnad, R. Doty, J. Jaynes, L. Goldstein, & G. Krauthamer, *The neural basis of stimulus equivalence across retinal translation*. New York: Academic Press.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.

Hummel, J., & Stankiewicz, B. (1996). Categorical relations in shape perception. *Spatial Vision*, *10*, 201–236.

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory and Cognition*, *15*, 531–543.

Jolicoeur, P., & Besner, D. (1987). Additivity and interaction between size ratio and response category in the comparison of size-discrepant shapes. *Journal of Experimental Psychology*: *Human Perception & Performance*, *13*, 478–487.

Kapadia, M., Gilbert, C., & Westheimer, G. (1994). A quantitative measure for short-term cortical plasticity in human vision. *Journal of Neuroscience*, *14*, 451–457.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 4966–4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*, 250–252.

Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, *7*, 530–535.

Larsen, A. (1985). Pattern matching: effects of size ratio, angular difference in orientation, and familiarity. *Perception & Psychophysics*, *38*, 63–68.

Liu, Z. (1996). Viewpoint dependency in object representation and recognition. *Spatial Vision*, *9*(4), 491–521.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, *35*, 549–568.

McKee, S., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception and Psychophysics*, *24*, 258–262.

Murray, J. E., Jolicoeur, P., McMullen, P. A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory and Cognition*, *21*(5), 601–610.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.

Reinitz, M. T., Wright, E., & Loftus, G. R. (1989). Effects of semantic priming on visual encoding of pictures. *Journal of Experimental Psychology*: *General*, *118*, 280–297.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Schoups, A., & Orban, G. (1996). Interoccular transfer in perceptual learning of a pop-out discrimination task. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 7358–7362.

Schwartz, E., Desimone, R., Albright, T., & Gross, C. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 5776–5778.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning Memory and Cognition*, *23*, 681–696.

Schyns, P., Goldstone, R. L., & Thilbaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Sireteanu, R., & Rettenbach, R. (1995). Perceptual learning in visual search: fast, enduring, but non-specific. *Vision Research*, *35*, 2037–2043.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*, 233–282.

Tarr, M. J. (1995). Rotating objects to recognize them: a case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review*, *2*, 55–82.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition*, *67*, 73–110.

Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition. *Cognition*, *32*, 193–254.

Ullman, S. (1998). Three-dimensional object recognition based on the combination of views. *Cognition*, *67*, 21–44.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 992–1005.

Watson, A. B. (1979). Probability summation over time. *Vision Research*, *19*, 515–522.

Wolford, G., Marchak, F., & Hughes, H. (1988). Practice effects in backward masking. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 101–112.