Visual learning by cue-dependent and cue-invariant mechanisms

Volodymyr Ivanchenko, Robert A. Jacobs [*]

Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, NY 14627, USA

Abstract

We examined learning at multiple levels of the visual system. Subjects were trained and tested on a same/different slant judgment task or a same/different curvature judgment task using simulated planar surfaces or curved surfaces defined by either stereo or monocular (texture and motion) cues. Taken as a whole, the results of four experiments are consistent with the hypothesis that learning takes place at both cue-dependent and cue-invariant levels, and that learning at these levels can have different generalization properties. If so, then cue-invariant mechanisms may mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient. We claim that learning takes place at multiple levels of the visual system, and that a comprehensive understanding of visual perception requires a good understanding of learning at each of these levels.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Visual learning; Cue-invariance

1. Introduction

Despite decades of research, perceptual learning is a poorly understood phenomenon. Perhaps the most important lesson that research has taught us is that our current theories and experiments are too simple and too narrowly focused. It is likely that perceptual learning takes place at multiple levels of the human perceptual system, and that a comprehensive understanding of perception will require a good understanding of learning at each of these levels. Unfortunately, the study of perceptual learning at multiple levels is nearly unexplored in the scientific literature (see Ahissar & Hochstein, 1997, 2002, for a notable exception). This lack of understanding of learning at multiple levels is, we believe, a major reason why the literature on perceptual learning often contains seemingly confusing (and contradictory) results.

This article reports the results of experiments investigating learning at two levels of the visual system, namely the levels of visual cue-dependent and visual cue-invariant mechanisms (e.g., shape-from-visual-texture or shape-from-visual-motion mechanisms versus a mechanism for perceiving shape that is independent of the visual cue used to define the shape).[1] Within the vision sciences, the study of visual cue-invariant mechanisms is relatively unusual. These mechanisms ought to be of fundamental interest to scientists because visual perception of natural environments must integrate information provided by multiple cues. In this sense, these mechanisms can be regarded as among the "highest level" mechanisms of our visual systems.

---

[1] We hypothesize that visual cue-invariant mechanisms are constructed from cue-dependent mechanisms. For example, a cue-invariant mechanism for representing visual shape might receive inputs from both a mechanism that represents shape-from-visual texture and a mechanism that represents shape-from-visual-motion (and, perhaps, inputs from several other cue-dependent mechanisms for representing shape). If this mechanism's output at any moment in time does not depend on which mechanism provided an input, then its output would be cue-invariant. To our knowledge, the vision sciences literature does not contain any studies directly evaluating this hypothesis.

---

[*] Corresponding author. Fax: +1 585 442 9216.
E-mail addresses: vivanchenko@bcs.rochester.edu (V. Ivanchenko), robbie@bcs.rochester.edu (R.A. Jacobs).

Do visual cue-invariant mechanisms exist? Recent psychophysical data suggests that the human visual system may contain neural mechanisms that represent object shape or depth independent from the visual cue(s) specifying the shape or depth. For example, Poom and Börjesson (1999) reported that prolonged viewing of an adaptation surface caused a test surface to appear to slant in the direction opposite to that of the adaptation surface regardless of whether the two surfaces were defined by the same cue (either motion parallax or binocular disparity) or different cues. Other behavioral studies suggesting visual cue-invariant mechanisms are Bradshaw and Rogers (1996) and Domini et al. (2001).

Related data have been found in neuroscientific studies using monkeys. For example, Sakata et al. (1999) showed that some visually responsive neurons in the macaque anterior intraparietal area encode surface tilt regardless of whether the tilt is specified by disparity alone, monocular cues alone, or both. Other neuroscientific studies indicating visual cue-invariant mechanisms in monkeys are Sary, Vogels, and Orban (1993), Sereno, Trinath, Augath, and Logothetis (2002), Tsutsui, Sakata, Naganuma, and Taira (2002).

Brain-imaging studies using human observers have reported similar data. Grill-Spector, Kushnir, Edelman, Itzchak, and Malach (1998) found that a region located on the lateral aspect of the occipital lobe was preferentially activated during a visual object recognition task relative to control conditions irrespective of whether the object shape was defined by luminance, motion, or texture cues. Kourtzi and Kanwisher (2000) reported overlapping activations in the lateral and ventral occipital cortex for objects depicted by different visual formats (grayscale images and line drawings), and a reduced response when objects were repeated, independent of whether they recurred in the same or a different format.[2] Other relevant brain-imaging studies using human observers are reported in Kourtzi, Betts, Sarkhei, and Welchman (2005) Welchman, Deubelius, Conrad, Bülthoff, and Kourtzi (2005).

Although the studies cited above suggest the existence of visual cue-invariant mechanisms, they did not examine the nature of these mechanisms in a detailed way and, importantly for our purposes, they did not examine the role of these mechanisms in perceptual learning. To date, we are aware of only one study on cue-invariant mechanisms and perceptual learning. Rivest, Boutet, and Intrilligator (1996) trained different sets of observers to visually discrim-

inate the orientations of color-defined bars, of luminance-defined bars, or of motion-defined bars. A similar improvement from pre-test to post-test was found regardless of whether the bars seen after training were defined by the same or by a different cue as the cue seen during training. The authors concluded that training changed the sensitivity of cells that represent visual orientation in a cue-invariant manner.

This article studies the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient. For example, if an observer learns to make more accurate depth-from-visual-texture judgments, then it would be advantageous to the observer to generalize this gained knowledge so that it can be used when estimating depth from cues other than texture, such as when making depth-from-visual-motion judgments. An important goal of the reported experiments is to evaluate this hypothesis. A secondary goal is to compare the generalization properties of visual cue-dependent versus cue-invariant mechanisms. We hypothesize that the "lower level" cue-dependent mechanisms tend to use local representations that lead to stimulus-specific learning (i.e., learning effects are limited to the specific stimulus conditions used during training), whereas the "higher level" cue-invariant mechanisms tend to use global representations that lead to stimulus-general learning (i.e., learning effects generalize to novel stimulus conditions). To our knowledge, there are currently no studies comparing the properties of cue-dependent versus cue-invariant mechanisms.

The results of four experiments are reported. In the first experiment, subjects were trained to judge the 3D orientations of planar surfaces slanted in depth when surfaces were defined by a training cue and when slants were centered near a training slant. Subjects were tested on the same task when surfaces were defined by either the training cue or a novel cue, and when slants were centered either near the training slant or near a novel slant. Because subjects showed improved performance when tested both with the training cue and with the novel cue, the results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms. Furthermore, these two sets of mechanisms seem to have different properties—cue-dependent mechanisms of visual slant are slant-specific whereas cue-invariant mechanisms are not. Experiment 2 was similar to Experiment 1, but it required subjects to judge the slants of cylinders. As in Experiment 1, its results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms, thereby producing transfer of learning from training to novel cue conditions. In addition, this experiment found that both sets of mechanisms either ignored or generalized over an irrelevant shape attribute. Experiment 3 required subjects to judge the curvature-in-depth of cylinders. The results again demonstrate learning by both cue-dependent and cue-invariant mechanisms. Experiment 4 found that learning

---

[2] It is interesting to note that cue-invariance may take place across sensory modalities, not just within the visual modality. Brain-imaging studies with humans have provided evidence for neural mechanisms which are modality-invariant. Amedi, Malach, Hendler, Peled, and Zohary (2001) found preferential activation in the lateral occipital complex when observers viewed objects and also when they grasped the same objects. Pietrini et al. (2004) found that visual and tactile recognition of man-made objects evoked category-related patterns of responses in a ventral extrastriate visual area in the inferior temporal gyrus that were correlated across sensory modality.

was task-specific—training on the curvature judgment task did not produce improved performance on the slant judgment task. This result indicates that learning was not due to adaptations of ''cognitive'' factors, and that observers do not have one set of mechanisms for judging visual depth but rather have different mechanisms for judging curvature-in-depth and slant-in-depth. Taken as a whole, the experiments support the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient.

2. Experiment 1

To motivate Experiment 1, consider an observer viewing a planar surface slanted in depth. Suppose the observer is trained to discriminate the slants of surfaces defined by a stereo cue when the slants are near 45° from vertical (the top of the surface is closer to the observer than the bottom), and the observer improves at this task over time due to training. The observer is then tested with surfaces defined by either the training cue (stereo) or by a novel cue (e.g., visual texture) using slants that are either near the training slant (45°) or far from the training slant (e.g., −45°).

In regard to generalization, at least four possibilities exist:

(i) learning does not generalize to any novel stimulus conditions. Because learning did not transfer across cues in this case, this outcome suggests that learning did not influence cue-invariant mechanisms. Furthermore, because training with the training cue and training slant did not lead to improved performance with the training cue and a novel slant, learning must have influenced representations that can be characterized as *slant-local*. Slant-local representations would occur in a population of mechanisms in which each individual mechanism represents a specific (or small range) of slants, and different mechanisms represent different slants (e.g., consider a neural network that uses a localist representation of surface slant). In the case considered here, training might have influenced an individual mechanism that represents stereo-defined surfaces slanted at about 45°, and not influenced other mechanisms such as cue-invariant mechanisms or a mechanism that represents stereo-defined surfaces slanted at −45°. Hence there was no transfer of learning when discriminating surfaces defined by a novel cue, or surfaces defined by the training cue but at a novel slant;

(ii) transfer of learning occurs when surfaces are defined by a novel visual cue but only when the surfaces are near the training slant (45°). The outcome in this case suggests the existence of both visual cue-dependent and cue-invariant representations for surface slant and, moreover, that both these representations are slant-local;

(iii) transfer of learning occurs when surfaces are at the novel slant but only when they are defined by the training cue. This outcome suggests that learning did not generalize across cues and, thus, did not influence cue-invariant representations. In addition, because training with the training cue and training slant led to improved performance with the training cue and novel slant, learning influenced representations which can be characterized as *slant-global*. Slant-global representations would occur if there exists a population of mechanisms in which all (or at least many) mechanisms are active in representing surface slant for all (or at least many) possible slants (e.g., consider a neural network that uses a fully distributed representation of surface slant). Modification of cue-dependent slant-global mechanisms during training would lead to transfer of learning between training and novel slants when surfaces are defined by the training cue;

(iv) transfer of learning occurs to all novel stimulus conditions. This outcome suggests the existence of both visual cue-dependent and cue-invariant representations that can both be characterized as slant-global.

As this example illustrates, perceptual events can vary along many stimulus dimensions (cue, surface slant, etc.), and generalization of learning might not occur, it might occur along some dimensions but not others, or it might occur along all dimensions. Moreover, experimental studies of observers' patterns of generalizations can inform us about the nature of the underlying perceptual representations modified during training.

2.1. Methods

2.1.1. Apparatus and stimuli

Stimuli simulated perspective views of planar surfaces slanted in depth relative to the frontal image plane. Surface slant varied, but tilt direction was always vertical (i.e., the gradient of surface depth relative to the observer was vertical in the cyclopean projection). Two cue conditions were used in the experiment:

(i) *Stereo-only*: Stimuli were stereoscopic views of a planar surface densely covered with dots. Because individual dots were small (subtending 0.08° of visual angle when a surface was frontoparallel), and because of the way that the placement of dots in each display was randomized (average dot density of 3.89 texels/deg^2), the set of dots in a display did not provide a useful texture cue to surface slant based on gradients of dot area, foreshortening, or density. To evaluate the assertion that the set of dots in a display did not provide a useful texture cue to surface slant, we conducted the following control experiment. Eight subjects first performed 60 practice trials. On each practice trial, they viewed two successive pairs of stereo stimuli, and judged whether the slants of the surfaces depicted in the pairs were the
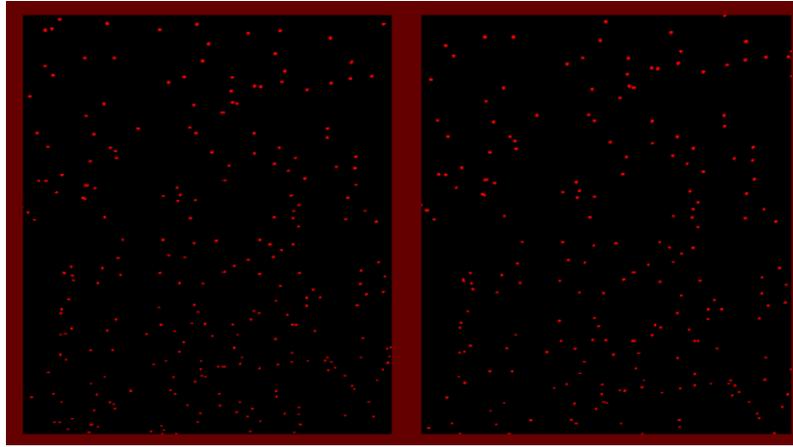
Fig. 1. Two left-eye views from stereo pairs depicting planar surfaces with slants of 41° and 49°, respectively.

same or different. The two stereo pairs were chosen so that either they both depicted planar surfaces slanted at 45°, or one pair depicted a surface slanted at 41° and the other depicted a surface slanted at 49°. After judging whether the surface slants were the same or different, subjects received auditory feedback indicating whether their response was correct. Next, subjects performed 240 test trials designed to evaluate whether they could perform the slant judgment task on the basis of the sets of dots present in individual images from stereo pairs. On each trial, subjects viewed (monocularly) an individual image from a stereo pair of images and then viewed a second individual image from a different stereo pair. For example, the left and right sides of Fig. 1 illustrate individual images from stereo pairs depicting surfaces with slants of 41° and 49°, respectively. After viewing the two individual images, subjects judged whether the depicted surface slants were the same or different. Auditory feedback was not provided. The results are shown in Fig. 2. The height of a bar in this figure

indicates a subject's performance on the test trials in terms of percent correct (the error bars indicate 95% confidence intervals based on 9000 simulated bootstrap trials; Efron & Tibshirani, 1994). Chance performance is 50% correct. The performances of 7 of 8 subjects did not differ significantly from chance performance. Consequently, we conclude that the set of dots in an individual left-eye or right-eye image from a stereo pair did not provide a useful texture cue to surface slant.

(ii) *Texture and motion*: Stimuli were monocular views of a planar surface densely covered with a homogeneous texture consisting of square patches. Fig. 3 illustrates a display of a planar surface defined by a texture cue. Each texture element subtended 0.82° of visual angle when a surface was frontoparallel. The placement of texture elements was randomized in each display (average density of 0.33 texels/deg^2). A motion cue to surface slant was added to each display by rotating a surface back and forth around a vertical axis that passed through the center of the surface. The range of rotation was ±15°, and the speed of rotation was 30° per second.
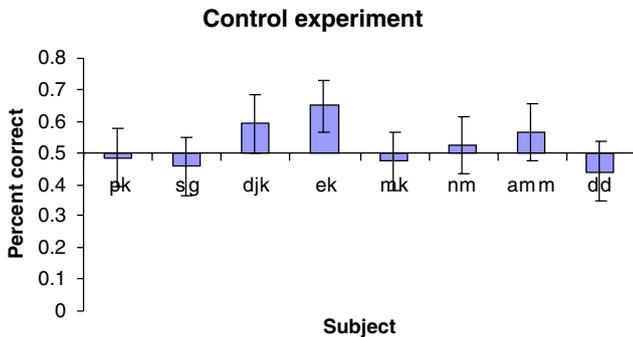
Stimuli were presented on a standard CRT monitor using a resolution of 1024 × 768 pixels and a 100 Hz refresh

Fig. 2. The height of a bar indicates a subject's performance on the control experiment to Experiment 1 in terms of percent correct (the error bars indicate 95% confidence intervals based on simulated bootstrap trials). Chance performance is 50% correct. The performances of 7 of 8 subjects did not differ significantly from chance performance, indicating that the set of dots in an individual left-eye or right-eye display from a stereo pair did not provide a useful texture cue to surface slant.
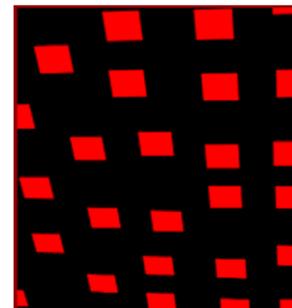

Fig. 3. Illustration of a display of a planar surface defined by a texture cue. The surface slant is 45° (the top is closer than the bottom).

rate. All stimuli were rendered in red because the red phosphor has a relatively fast decay rate. Subjects viewed the monitor from a distance of 75 cm, and displays depicted surfaces whose centers were 26 cm behind the monitor. Surfaces were viewed through a black, rectangular window rendered at the monitor depth. This window provided a view of a surface subtending 9° of visual angle in the horizontal direction and 10° in the vertical direction. The window occluded the edges of a surface, thereby eliminating contour cues to a surface's slant. Subjects viewed stereo stimuli using LCD shutter glasses (CrystalEyes 3 from Stereographics) to present different stereo views to the left and right eyes. They viewed monocular (texture and motion) stimuli while wearing an eye patch over one eye.

2.1.2. Procedure

Subjects performed a two-alternative forced-choice same/different slant judgment task. On each trial, subjects were presented with a successive pair of surfaces, and judged whether the surfaces had the same or different slant. The surfaces were always defined by the same cue(s), either the stereo cue or the monocular cues. Slant was defined to be the angle between the surface normal and the line of sight to a cyclopean eye mid-way between a subject's left and right eyes. For positive slants, the bottoms of surfaces appeared to recede in depth; for negative slants, the tops appeared to recede in depth. On trials in which the surfaces had the same slant, this slant was equal to a value referred to as the "center slant." When the surfaces had different slants, one surface's slant was set to the center slant plus a deviation denoted Δ, and the other surface's slant was set to the center slant minus Δ. The deviation Δ was set to 4°. Each surface was displayed for 1500 ms, and there was a 1000 ms inter-stimulus interval between displays during which the monitor was blank. After both surfaces were displayed, the monitor was blank, and subjects responded "same" or "different" by pressing the right or left mouse buttons, respectively. On practice and training trials, subjects received feedback regarding the correctness of their response—a sound was produced if the subject responded correctly; no sound was produced if the subject responded incorrectly. Feedback was not provided on test trials. The next trial began following a 1000 ms inter-trial interval.

Participants included both experimental and control subjects. Each experimental subject performed practice, pre-test, training, and post-test blocks of trials over 4 days, where each block consisted of 60 trials. The number and order of "same" versus "different" trials was counterbalanced and randomized within each block. On Day 1, experimental subjects performed 4 practice blocks, one block for each combination of cue (stereo or monocular) and center slant (45° or −45°). Practice trials allowed subjects to become comfortable with the experimental environment and task, and allowed us to evaluate a subject's perception of stereo stimuli (subjects with poor stereo vision were dismissed from the study). Experimental subjects also performed 4 pre-test blocks on Day 1—one block for each

combination of cue and center slant. On Days 2 and 3, they performed 4 training blocks. The center slant was 45° for all training trials. Training trials used the stereo cue for half the experimental subjects, and the monocular cues for the remaining experimental subjects. Recall that subjects received feedback regarding the correctness of their responses on training trials. On Day 4, experimental subjects performed 1 training block followed by 4 post-test blocks (which were identical to pre-test blocks). In contrast to experimental subjects, control subjects did not receive training—they performed 4 practice and 4 pre-test blocks on Day 1, and 4 post-test blocks on Day 2.

2.1.3. Subjects

Sixteen undergraduate students at the University of Rochester served as experimental subjects and eight students served as control subjects. All subjects were naïve to the goals of the experiment, and all had normal or corrected-to-normal vision.

2.2. Results

The results are illustrated in Fig. 4. The graphs on the top and bottom plot the data for the experimental and control subjects, respectively. We first consider the top graph.
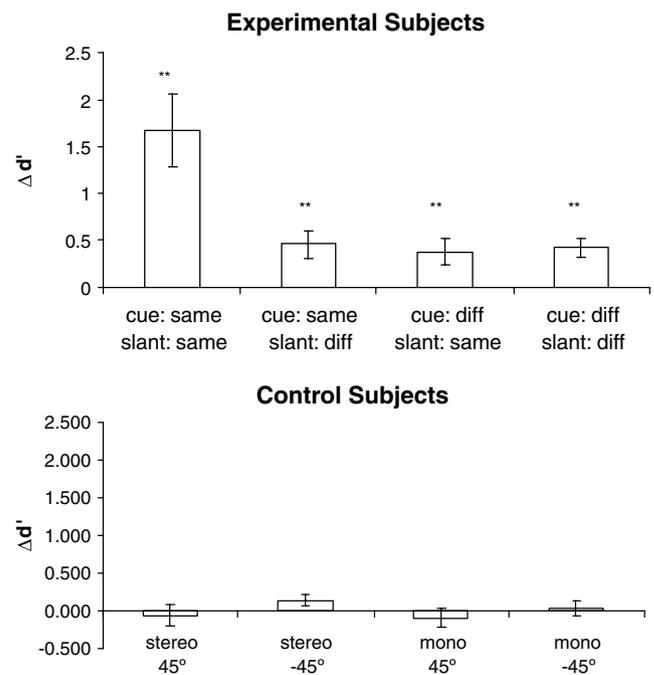

Fig. 4. (Top) The results for the experimental subjects in Experiment 1. The horizontal axis indicates the test condition. For example "cue: same, slant: same" means that subjects were tested using the same cue and center slant as were used during training. The vertical axis plots $\Delta d'$ which is a subject's d' on post-test trials minus this value on pre-test trials averaged over all subjects. Error bars give the standard errors of the means. The two asterisks ("**") above a bar mean that the value indicated by the bar is significantly greater than zero at the $p < .01$ level based on a two-tailed t-test; (Bottom) the results for the control subjects. The horizontal axis indicates the test condition, and the vertical axis plots $\Delta d'$.

The vertical axis plots experimental subjects' average improvement in performance in units of $\Delta d'$ which is the value of a subject's d' on post-test trials minus this value on pre-test trials averaged over all subjects. Error bars give the standard errors of the means. The horizontal axis indicates the experimental condition. For example, condition "cue: same, slant: same" is the set of pre- and post-test trials that used the same cue and center slant as were used during a subject's training trials. Condition "cue: same, slant: diff" is the set of test trials that used the same cue as was used during a subject's training trials but a different (novel) center slant. The graph on the bottom has a similar format, but is not identical because control subjects did not receive training trials. In this case, condition "stereo, 45°" is the set of test trials that used the stereo cue and the 45° center slant, whereas condition "stereo, −45°" is the set of test trials that used the stereo cue and the −45° center slant.

Several observations can be made. First, the experimental subjects showed large learning effects when evaluated

with the same cue and center slant as used during training (top graph of Fig. 4, condition "cue: same, slant: same"). Second, these subjects also showed moderate-sized learning effects when evaluated with novel cues and/or novel slants. In all conditions, $\Delta d'$ values are significantly greater than zero (based on two-tailed *t*-tests with a $p < .01$ significance level). Moreover, it appears that roughly equal amounts of transfer of learning were found to novel cues as to novel slants. In contrast, control subjects never showed improvements from pre- to post-test (bottom graph in Fig. 4). This result was expected as control subjects never received training.

A more detailed view of experimental subjects' data is given in Fig. 5. The first set of eight subjects in each graph were trained with planar surfaces defined by the monocular cues, whereas the second set of eight subjects were trained with surfaces defined by the stereo cue. For each subject, there are two bars showing a subject's performances (in units of d') on the pre-test and post-test trials. For the sake of brevity, this data is only provided for the test trials that used the same cue as a subject's training trials but a novel slant (top graph of Fig. 5), and for the test trials that used a novel cue and a novel slant (bottom graph).
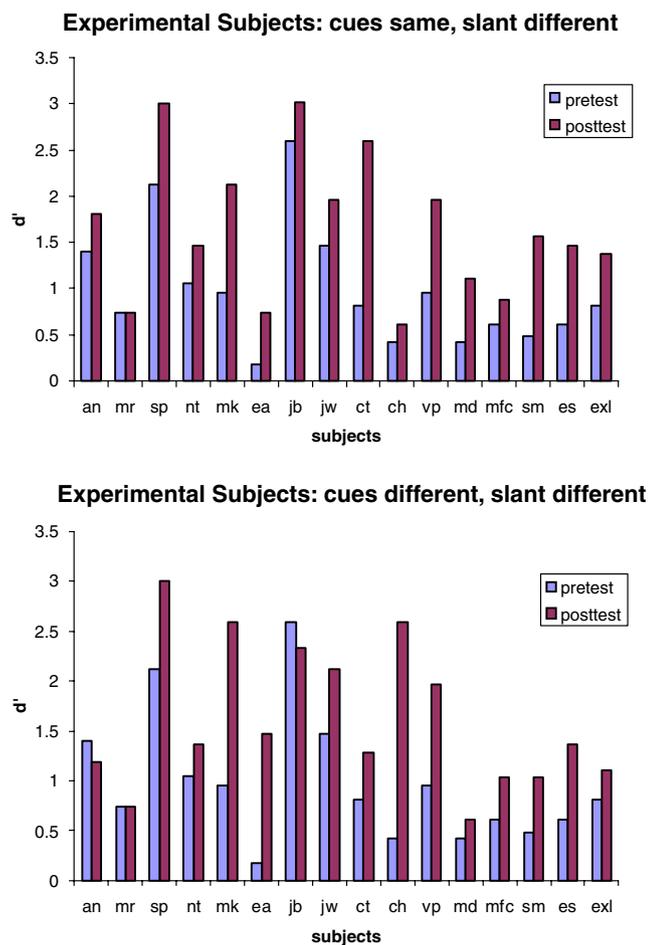
2.3. Discussion

Based on this data, we can conclude the following. First, training produced modifications to experimental subjects' cue-dependent representations of visual slant (e.g., slant-from-stereo, slant-from-texture, slant-from-motion) as evidenced by the large amounts of learning with a subject's training cue and training slant (condition "cue: same, slant: same"). Second, and importantly for our purposes, training also produced modifications to subjects' representations of visual slant which are visual cue-invariant, as demonstrated by subjects' cue-invariant generalizations (conditions "cue: diff, slant: same" and "cue: diff, slant: diff"). Above we hypothesized that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient. The results of Experiment 1 support this hypothesis.

In regard to the issue of whether the modified slant representations are slant-local versus slant-global, these data indicate that the cue-dependent representations are best characterized as slant-local because learning effects with the training slant were much larger than learning effects with the novel slant.[3] The cue-invariant representations, in contrast, are best characterized as slant-global because subjects showed as much transfer of learning to a novel cue and novel slant as to a novel cue but familiar slant. That is, these data indicate that cue-dependent and cue-in-

Experimental Subjects: cues same, slant different

Experimental Subjects: cues different, slant different

Fig. 5. Experimental subjects' values of d' on pre-test and post-test. The first set of eight subjects were trained with planar surfaces defined by the monocular cues, whereas the second set of eight subjects were trained with surfaces defined by the stereo cue. For the sake of brevity, d' values are only given for the test trials that used the same cue as during a subject's training but a different, novel slant (top graph), and for the test trials that used a novel cue and a novel slant (bottom graph).

---

[3] For the sake of simplicity, we assume that only cue-dependent mechanisms are involved in the tests with the training cue and only cue-invariant mechanisms are involved in the tests with a novel cue. Whether this simplifying assumption is correct is a topic of future research.

variant mechanisms have different properties—the "lower level" cue-dependent mechanisms appear to use a localist representation of slant, whereas the "higher level" cue-invariant mechanisms appear to use a global, perhaps distributed, representation of slant.

A reader might wonder if factors other than modifications of cue-dependent and cue-invariant mechanisms might underlie subjects' improvements in performance. For example, could it be the case that the learning effects are due to the fact that subjects learned to ignore the task irrelevant flatness cues in a display (e.g., from accommodation, blur, etc.)? We believe that this is not a viable hypothesis because it does not provide an adequate account of the experimental data. For instance, it does not explain why subjects showed a larger learning effect when tested with a training cue and training slant than when tested with a training cue and novel slant, and it does not explain why subjects showed one pattern of performances across different slants when tested with a training cue but a different pattern when tested with a novel cue. This latter point, in particular, strongly suggests that performances when tested with a training cue versus a novel cue are based on sets of mechanisms which are not identical.

3. Experiment 2

Experiment 2, like Experiment 1, examined whether training on a visual slant judgment task produces modifications to both cue-dependent and cue-invariant mechanisms. Whereas Experiment 1 studied generalization across slants to ask whether the cue-dependent and cue-invariant representations of visual slant are slant-local versus slant-global, Experiment 2 studied generalization across shapes to ask whether these representations are *shape-local* versus *shape-global*. Shape-local representations of visual slant would occur in a population of mechanisms in which each individual mechanism represents the slant of a specific (or small range) of shapes, and different mechanisms represent the slants of different shapes (e.g., consider a neural network that uses a representation of slant that is local along the dimensions coding shape—one set of units might represent the slant of planar surfaces whereas another set represents the slant of curved surfaces). Shape-global representations of visual slant would occur in a population of mechanisms in which all mechanisms are active in representing slant for all possible shapes (e.g., consider a neural network that uses a representation of slant that is fully distributed along the shape dimensions—the same units are active in representing an object's slant regardless of the object's shape). Experiment 2 examined whether cue-dependent and cue-invariant representations of slant are shape-local or shape-global by training subjects on a slant with one cue and shape, and by testing them with a novel cue and shape.

The question of whether subjects' representations of visual slant are shape-specific is motivated by recent findings on perceptual learning in which observers uncon-

sciously learned about stimulus dimensions that were irrelevant for the task that they were performing (e.g., Watanabe, Nanez, & Sasaki, 2001). If observers learn about task-irrelevant stimulus properties, are their representations of these properties local or global? Experiment 2 aimed to replicate the findings of Experiment 1 in terms of providing evidence for learning by cue-invariant mechanisms, but Experiment 1 used the slant judgment task to study generalization along a task-relevant stimulus dimension (slant), whereas Experiment 2 used this task to investigate generalization along an irrelevant dimension (shape).

3.1. Methods

3.1.1. Apparatus and stimuli

Stimuli simulated perspective views of either planar or curved surfaces slanted in depth. The planar surfaces were defined by either stereo or monocular (texture and motion) cues, and were identical to those used in Experiment 1. The curved surfaces were elliptical cylinders (i.e., cylinders whose horizontal cross-sections are ellipses when a cylinder has a slant of 0°). The depth-to-width ratio of cylinders' horizontal cross-sections were 1.0, meaning that cylinders (with a slant of 0°) had an object depth (defined as the distance from the point on the cylinder closest to the observer to the point furthest away) equal to their width. Cylinders were defined by a stereo cue in this experiment (Fig. 7 illustrates a display of an elliptical cylinder defined by a texture cue). The surface of a cylinder was colored green, and small red dots were placed on the surface. Stereoscopic views of cylinders yielded percepts of surface curvature due to the binocular disparities of the dots. Dots were placed on the surface in such a way that the set of dots did not provide a useful texture cue to surface curvature based on gradients of dot area, foreshortening, or density. Displays of cylinders contained red borders at the top and bottom rendered at the monitor depth. These borders occluded the top and bottom edges of a cylinder, thereby eliminating contour cues to a cylinder's curvature. The visible portion of a cylinder subtended 11.9° of visual angle in the horizontal direction and 15.4° in the vertical direction.

3.1.2. Procedure

Similar to Experiment 1, subjects performed a two-alternative forced-choice same/different slant judgment task. On each trial, subjects were presented with a successive pair of planar or curved surfaces, and judged whether the surfaces had the same or different slant. Curved surfaces were displayed for 1000 ms, and inter-stimulus and inter-trial intervals were each 700 ms.

Subjects performed practice, pre-test, training, and post-test blocks of trials over 4 days. On Day 1, they performed 2 practice blocks. One block used planar surfaces defined by the monocular cues, and the other block used curved surfaces defined by a stereo cue. The center slant was 45° on all practice trials. These were followed by 4 pre-test blocks using planar surfaces. There was one pre-test block

for each combination of cue (stereo or monocular) and center slant (45° or −45°). On Days 2 and 3, subjects performed 4 training blocks using the curved surface defined by a stereo cue and a center slant of 45°. On Day 4, they performed 1 training block followed by 4 post-test blocks (which were identical to pre-test blocks).

3.1.3. Subjects

Eight undergraduate students at the University of Rochester served as experimental subjects. All subjects were naïve to the goals of the experiment, and all had normal or corrected-to-normal vision.

3.2. Results

The results are shown in the graph in Fig. 6. The horizontal axis indicates the experimental condition, and the vertical axis plots the subjects' average improvement in performance in units of $\Delta d'$. Error bars give the standard errors of the means. The four leftmost bars show the average performance improvement in the four test conditions (post-test d' minus pre-test d'), whereas the rightmost bar shows the improvement during training (d' on last training block on Day 3 minus d' on first training block on Day 2).

Subjects showed significant improvement on the slant judgment task during training with curved surfaces and a center slant of 45° ($p = .015$; rightmost bar of Fig. 6). In addition, they showed significant improvement from pre- to post-test in three of four test conditions with planar surfaces ($p < .05$ based on a two-tailed t-test; the improvement in the remaining condition—test trials using planar surfaces defined by the monocular cues and a center slant of 45°—is not statistically significant).

3.3. Discussion

Training with curved surfaces was effective as evidenced by the improvement in performance on the slant judgment task from the start to the end of training. In addition, train-
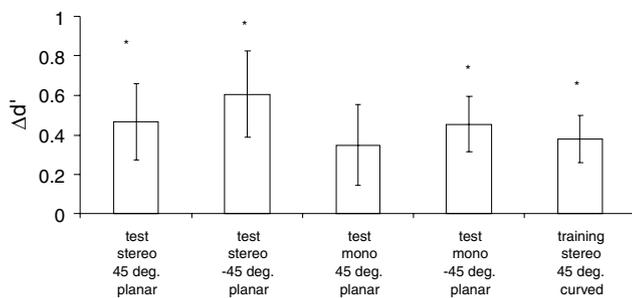

Fig. 6. The first four bars show subjects' performance improvements in the test conditions of Experiment 2, whereas the rightmost bar shows subjects' improvements during training. The horizontal axis indicates the test or training condition, and the vertical axis plots the improvement in units of $\Delta d'$. Error bars give the standard errors of the means. An asterisk ("*") above a bar means that the value indicated by the bar is significantly greater than zero at the $p < .05$ level based on a two-tailed t-test.

ing produced modifications to cue-dependent mechanisms as evidenced by the improved performance on test trials using the training (stereo) cue, as well as to cue-invariant mechanisms as evidenced by the improved performance in one test condition with a novel (monocular) cue and the nearly significant improvement in the other test condition with a novel cue. These results are consistent with the results of Experiment 1 in the sense that both experiments show visual learning by both cue-dependent and cue-invariant mechanisms.

The results also show learning effects of similar sizes in all cases, thereby indicating that both cue-dependent and cue-invariant representations of visual slant are best characterized as shape-global. We first consider the cue-dependent representation. To evaluate whether it is shape-local or shape-global, it is necessary to compare conditions that used the same cue, but different shapes. Compare the performance improvement during training (stereo cue, curved surface; see rightmost bar of Fig. 6) with the improvements in the first and second test conditions (stereo cue, planar surface; see first and second bars of Fig. 6). If the former improvement is larger than the latter improvements, we would conclude that subjects' cue-dependent representations of visual slant are shape-local because a larger improvement is found with the training shape than with a novel shape. However, the data do not show that the former improvement is larger. Instead, the data show that these two improvements are about the same size, meaning that subjects showed complete transfer of learning from displays of curved surfaces to displays of planar surfaces. This result suggests that training produced modifications of subjects' cue-dependent mechanisms that applied equally to all shapes, consistent with a shape-global representation.

In regard to whether subjects' cue-invariant representations of visual slant are best characterized as shape-local or shape-global, it is necessary to compare conditions that used both different cues and different shapes. Compare the performance improvements during training (stereo cue, curved surface; see rightmost bar of Fig. 6) and during the third and fourth test conditions (monocular cues, planar surface; see third and fourth bars of Fig. 6). These performance improvements are about the same size, meaning that subjects showed complete transfer of learning between displays of curved surfaces defined by a stereo cue to displays of planar surfaces defined by the monocular cues. We conclude that training produced modifications of subjects' cue-invariant mechanisms that applied equally to all shapes and, thus, the cue-invariant representations are also best characterized as shape-global.

Experiment 2 produced results which might be regarded as inconsistent with those of Experiment 1 in two ways. First, in the introductory section of this article, we hypothesized that "lower level" cue-dependent mechanisms tend to use local representations that lead to stimulus-specific learning (i.e., learning effects are limited to the specific stimulus conditions used during training), whereas the

"higher level" cue-invariant mechanisms tend to use global representations that lead to stimulus-general learning (i.e., learning effects generalize to novel stimulus conditions). Experiment 1 found evidence supporting this hypothesis—subjects' cue-dependent representations of visual slant were slant-local whereas their cue-invariant representations were slant-global. However, Experiment 2 did not—subjects' cue-dependent representations of visual slant were shape-global, not shape-local. A possible explanation is that the stimuli used during training in Experiment 2 contained an attribute (shape) which was irrelevant and possibly difficult to interpret for the purpose of performing the experimental task (slant judgment task). It may be that subjects' cue-dependent mechanisms either ignored the irrelevant shape information or, perhaps equivalently, generalized across the irrelevant shape dimensions (thereby producing complete transfer of learning from curved to planar surfaces) because shape was an irrelevant attribute. Future research will need to explore this possibility.

Second, Experiment 1 found that subjects' cue-dependent mechanisms showed greater performance improvement with the training slant than with a novel slant (first and second bars of the top graph of Fig. 4), whereas Experiment 2 found that subjects' cue-dependent mechanisms showed equal performance improvement with training and novel slants (first and second bars of Fig. 6). A possible explanation is that Experiment 2 not only used a novel slant, but also a novel shape. Observers' cue-dependent mechanisms may generalize differently to novel slants than to conjunctions of novel slants and novel (irrelevant) shapes. Again, future research will need to study this issue.

4. Experiment 3

An important goal of this research project was to evaluate the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions. Experiments 1 and 2 evaluated subjects' performances in familiar and novel cue conditions on a slant judgment task. The results of these experiments provide compelling evidence supporting this hypothesis. Experiment 3 is a control experiment designed to make sure that the results reported above are not due to the use of a particular type of task (i.e., a slant judgment task) but, rather, that the basic findings can be replicated with at least one other perceptual task. This experiment used a curvature judgment task.

4.1. Methods

4.1.1. Apparatus and stimuli

Stimuli simulated perspective views of curved surfaces. These surfaces were vertically oriented elliptical cylinders. All cylinders had the same width, though different cylinders had different object depths. Cylinders were defined by either a stereo or texture cue. When defined by a stereo cue, displays were identical to those used in Experiment 2.

When defined by a texture cue, an isotropic texture consisting of red circles was mapped to the green surface of a cylinder. The gradient of texture element foreshortening, size, and density provided a useful cue to a cylinder's shape. Fig. 7 illustrates a display of a cylinder defined by a texture cue. As was the case in Experiment 2, displays of cylinders contained red borders at the top and bottom rendered at the monitor depth. These borders occluded the top and bottom edges of a cylinder, thereby eliminating contour cues to a cylinder's shape. The visible portion of a cylinder subtended 11.9° of visual angle in the horizontal direction and 15.4° in the vertical direction.

4.1.2. Procedure

Subjects performed a two-alternative forced-choice same/different curvature judgment task. On each trial, subjects were presented with a successive pair of cylinders, and judged whether the cylinders had the same or different curvature. (Note that this is identical to judging whether the cylinders had the same or different shape, or the same or different object depth.) Cylinders were displayed for 1000 ms, and inter-stimulus and inter-trial intervals were 700 ms.

Both experimental and control subjects participated in this experiment. Experimental subjects performed practice, pre-test, training, and post-test blocks of trials over 4 days. On Day 1, they performed 2 practice blocks, one block used the stereo cue and the other block used the texture cue. When cylinders had the same curvature, the depth-to-width ratio of their horizontal cross-sections was 1.0 (the cylinders were equally deep as wide). When their curvatures were different, one cylinder had a depth-to-width ratio of 0.5 (the cylinder's width was twice its depth) and the other had a ratio of 1.5 (the cylinder's depth was 1.5 times its width). Practice blocks were followed by 2 pre-test blocks. All test trials used cylinders defined by the texture cue. When cylinders had the same curvature, the depth-to-width ratio of their horizontal cross-sections was 1.0.
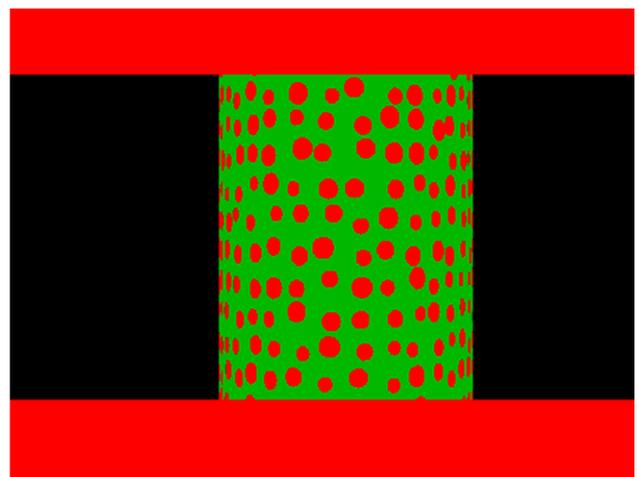

Fig. 7. Illustration of a display of a cylinder defined by a texture cue.

When their curvatures were different, one cylinder had a depth-to-width ratio of 0.8 and the other had a ratio of 1.2. On Days 2 and 3, subjects performed 6 blocks of training trials. Training trials used cylinders defined by the stereo cue. They used the same depth-to-width ratios as test trials. On Day 4, subjects performed 2 training blocks followed by 2 post-test blocks. Post-test blocks were identical to pre-test blocks. In contrast to experimental subjects, control subjects did not receive training—they performed 2 practice and 2 pre-test blocks on Day 1, and 2 post-test blocks on Day 2.

4.1.3. Subjects

Six undergraduate students at the University of Rochester served as experimental subjects, and six students served as control subjects. All subjects were naïve to the goals of the experiment, and all had normal or corrected-to-normal vision.

4.2. Results and discussion

The results are shown in Fig. 8. For each subject, the two bars show a subject's performances (in units of d') on pre-test and post-test trials. Control subjects' performances on post-test trials did not differ significantly from their performances on pre-test trials (bottom graph of Fig. 8). This was expected as control subjects did not receive training. Experimental subjects, in contrast, showed significantly better performance on post-test trials than
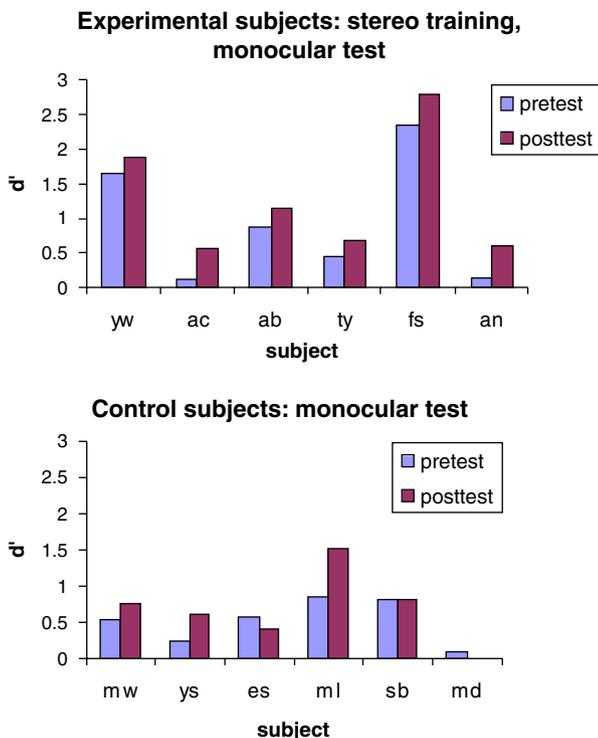

Fig. 8. Results of Experiment 3. Experimental subjects showed significantly better performance on post-test trials than pre-test trials (top graph), whereas control subjects' performances on pre-test and post-test trials did not differ significantly (bottom graph).

pre-test trials (top graph)—their average improvement in units of $\Delta d'$ was 0.352 (standard error of the mean = 0.046; average improvement is significantly greater than zero at the $p < .001$ level). This result suggests that training with a stereo cue on the curvature judgment task led to modifications of subjects' cue-invariant mechanisms, thereby producing improved performance when cylinders were defined by a texture cue. The result is consistent with the findings of Experiments 1 and 2 and, thus, indicates that these findings were not due to the use of a specific experimental task. Overall, the results of Experiments 1–3 support the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions.

5. Experiment 4

Experiment 4 had two goals. The first goal was to serve as a control experiment that would rule out an unlikely, but not impossible, interpretation of the earlier experiments. In Experiment 1, for example, experimental subjects showed improvements in performance from pre-test to post-test (to a large degree in one test condition, and to moderate degrees in other test conditions), whereas control subjects did not. A possible interpretation is that experimental subjects showed improved performance on post-test trials because they learned during training to better control general attentional and other cognitive factors. If so, their learning might be better classified as "cognitive" learning than as "perceptual" learning. According to this interpretation, because control subjects did not receive training, they did not learn to better control attentional and other cognitive factors and, thus, they did not show improved performance. (We regard this interpretation as unlikely, at least in part, because it does not explain why experimental subjects showed different amounts of improvements in performance in different test conditions.) Experiment 4 evaluated this hypothesis by training subjects on one perceptual task, but testing them on a different task.

A second goal of Experiment 4 was to evaluate whether observers have a set of mechanisms for judging all types of visual depth or whether they have different sets of mechanisms for different types of depth judgments, such as judging slant-in-depth and judging curvature-in-depth. If observers have a set of mechanisms for judging all types of visual depth, then we would expect that training with one type of depth task would result in improved performance on another type of depth task. If, on the other hand, observers have different sets of mechanisms for different types of depth judgments, then we would not expect transfer of learning across different depth tasks.

5.1. Methods

5.1.1. Apparatus and stimuli

Stimuli simulated perspective views of either planar or curved surfaces. The planar surfaces were defined by either

stereo or monocular (texture and motion) cues, and were identical to those used in Experiment 1. The curved surfaces were vertically oriented elliptical cylinders defined by a stereo cue, as were those used in Experiment 3.

5.1.2. Procedure

Subjects performed practice, pre-test, training, and post-test blocks of trials over 4 days. On Day 1, they performed 3 practice blocks. Two blocks used the slant judgment task—one block used planar surfaces defined by the monocular cues, and the other block used planar surfaces defined by the stereo cue. The remaining block used the curvature judgment task. When cylinders had the same curvature, the depth-to-width ratio of their horizontal cross-sections was 1.0. When their curvatures were different, one cylinder had a depth-to-width ratio of 0.5 and the other had a ratio of 1.5. Practice blocks were followed by 2 pre-test blocks. All test blocks used the slant judgment task. The two pre-test blocks used planar surfaces defined by the stereo cue and the monocular cues, respectively. On Days 2 and 3, subjects performed 5 training blocks. All training trials used the curvature judgment task. When cylinders had the same curvature, their depth-to-width ratio was 1.0. When their curvatures were different, one cylinder had a depth-to-width ratio of 0.8 and the other had a ratio of 1.2. Subjects performed 2 training blocks and 2 post-test blocks on Day 4.

5.1.3. Subjects

Eight undergraduate subjects at the University of Rochester served as subjects. All subjects were naïve to the goals of the experiment, and all had normal or corrected-to-normal vision.

5.2. Results and discussion

The results are shown in Fig. 9. The first (leftmost) bar shows subjects' average performance improvement during training in units of $\Delta d'$ (d' on the last training block of

Fig. 9. Results for Experiment 4. The first (leftmost) bar shows subjects' average performance improvement during training in units of $\Delta d'$ (d' on the last training block of Day 3 minus d' on the first training block of Day 2). The remaining bars show subjects' average performance improvements on the post-test versus pre-test trials when planar surfaces were defined by the stereo cue or the monocular cues, respectively. Error bars give the standard errors of the means. The two asterisks ("**") above a bar mean that the value indicated by the bar is significantly greater than zero at the $p < .01$ level based on a two-tailed t-test.

Day 3 minus d' on the first training block of Day 2). The remaining bars show subjects' average performance improvements on the post-test versus pre-test trials when planar surfaces were defined by the stereo cue or the monocular cues, respectively. Subjects showed large improvements in performance during training on the curvature judgment task (average improvement during training is 1.11 in units of $\Delta d'$ [standard error = 0.17; improvement is significantly greater than zero at the $p < .01$ level based on a two-tailed t-test]). Despite this, their post-test performances on the slant judgment task did not significantly differ from their pre-test performances in either test condition (planar surfaces defined by the stereo cue, or defined by the monocular cues). In other words, performance improvements were task-specific—there was no transfer of learning from curvature to slant judgment tasks.

As discussed above, Experiment 4 had two goals. The first goal was to evaluate whether performance improvements in test conditions could be due to "cognitive" learning during training. Because subjects showed significant improvement during training but no improvements during testing, the "cognitive" learning interpretation can be rejected. The second goal was to evaluate whether observers have a set of mechanisms for judging all types of visual depth or whether they have different sets of mechanisms for different types of depth judgments. Because subjects showed improvements on judging curvature-in-depth but not on judging slant-in-depth, we conclude that observers have different sets of mechanisms for different types of depth judgments.

6. Conclusions

In summary, the results of four experiments were reported. In the first experiment, subjects were trained to discriminate the 3D orientations of planar surfaces slanted in depth when surfaces were defined by a training cue and when slants were centered near a training slant. Subjects were tested on the same task when surfaces were defined by either the training cue or a novel cue, and when slants were centered either near the training slant or near a novel slant. Because subjects showed improved performance both with the training cue and with the novel cue, the results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms. Furthermore, these two sets of mechanisms seem to have different properties—cue-dependent mechanisms of visual slant are slant-specific, whereas cue-invariant mechanisms are not. Experiment 2 was similar to Experiment 1, but it required subjects to judge the slants of cylinders. As in Experiment 1, its results suggest that training produced modifications to both cue-dependent and cue-invariant mechanisms, thereby producing transfer of learning from training to novel cue conditions. In addition, this experiment found that both sets of mechanisms either ignored or generalized over an irrelevant shape attribute. Experiment 3 required subjects to judge the curvature-in-depth of cylinders. The results again demon-
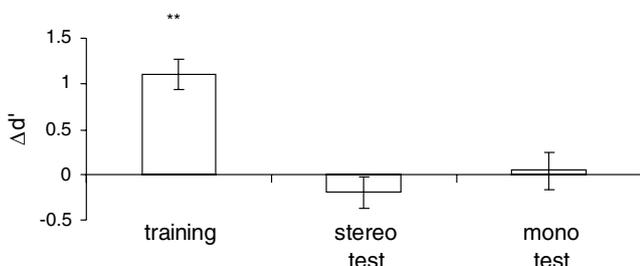
strate learning by cue-invariant mechanisms. Experiment 4 found that learning was task-specific—training on the curvature judgment task did not produce improved performance on the slant judgment task. This result indicates that learning was not due to adaptations of "cognitive" factors; it also shows that observers do not have one set of mechanisms for judging visual depth but rather have different mechanisms for judging curvature-in-depth and slant-in-depth. Taken as a whole, the experiments support the hypothesis that cue-invariant mechanisms mediate the transfer of learning from familiar cue conditions to novel cue conditions, thereby allowing perceptual learning to be robust and efficient.

Our results suggest that visual learning takes place at multiple levels of the human visual system, and that a comprehensive understanding of visual perception will require a good understanding of learning at each of these levels. Unfortunately, the study of visual learning at multiple levels is nearly unexplored in the scientific literature. This lack of understanding of learning at multiple levels is, we believe, a major reason why the literature on visual learning often contains seemingly confusing (and contradictory) results. Our work represents an early step toward an examination of learning at multiple levels of the visual system. We hope that the study of visual learning at multiple levels becomes a common practice in the field.

Acknowledgments

References

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature, 387*, 401–406.

Ahissar, M., & Hochstein, S. (2002). The role of attention in learning simple visual tasks. In M. Fahle & T. Poggio (Eds.), *Perceptual learning*. Cambridge, MA: MIT Press.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*, 324–330.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research, 36*, 3457–3468.

Domini, F., Adams, W., & Banks, M. S. (2001). 3D after-effects are due to shape and not disparity adaptation. *Vision Research, 41*, 2733–2739.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Boca Raton, FL: CRC Press.

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

Kourtzi, Z., Betts, L. R., Sarkhei, P., & Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLOS Biology, 3*, e204.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *The Journal of Neuroscience, 20*, 3310–3318.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W.-H. C., Cohen, L., et al. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences USA, 101*, 5658–5663.

Poom, L., & Börjesson, E. (1999). Perceptual depth synthesis in the visual system as revealed by selective adaptation. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 504–517.

Rivest, J., Boutet, I., & Intrilligator, J. (1996). Perceptual learning of orientation discrimination by more than one attribute. *Vision Research, 37*, 273–281.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tsutsui, K., Tanaka, Y., et al. (1999). Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research, 128*, 160–169.

Sary, G., Vogels, R., & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science, 260*, 995–997.

Sereno, M. E., Trinath, T., Augath, M., & Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron, 33*, 635–652.

Tsutsui, K.-I., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298*, 409–412.

Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844–848.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in the human visual cortex. *Nature Neuroscience, 8*, 820–827.