PERGAMON

Vision Research 39 (1999) 4062–4075

Vision Research

www.elsevier.com/locate/visres

Experience-dependent integration of texture and motion cues to depth

Robert A. Jacobs [a,*], I. Fine [b]

[a] *Center for Visual Science, University of Rochester, Rochester, NY 14627, USA*
[b] *Department of Psychology, University of California, San Diego, La Jolla, CA 92093, USA*

Received 22 September 1998; received in revised form 20 January 1999

Abstract

Previous investigators have shown that observers' visual cue combination strategies are remarkably flexible in the sense that these strategies adapt on the basis of the estimated reliabilities of the visual cues. However, these researchers have not addressed how observers' acquire these estimated reliabilities. This article studies observers' abilities to learn cue combination strategies. Subjects made depth judgments about simulated cylinders whose shapes were indicated by motion and texture cues. Because the two cues could indicate different shapes, it was possible to design tasks in which one cue provided useful information for making depth judgments, whereas the other cue was irrelevant. The results of experiment 1 suggest that observers' cue combination strategies are adaptable as a function of training; subjects adjusted their cue combination rules to use a cue more heavily when the cue was informative on a task versus when the cue was irrelevant. Experiment 2 demonstrated that experience-dependent adaptation of cue combination rules is context-sensitive. On trials with presentations of short cylinders, one cue was informative, whereas on trials with presentations of tall cylinders, the other cue was informative. The results suggest that observers can learn multiple cue combination rules, and can learn to apply each rule in the appropriate context. Experiment 3 demonstrated a possible limitation on the context-sensitivity of adaptation of cue combination rules. One cue was informative on trials with presentations of cylinders at a left oblique orientation, whereas the other cue was informative on trials with presentations of cylinders at a right oblique orientation. The results indicate that observers did not learn to use different cue combination rules in different contexts under these circumstances. These results are consistent with the hypothesis that observers' visual systems are biased to learn to perceive in the same way views of bilaterally symmetric objects that differ solely by a symmetry transformation. Taken in conjunction with the results of Experiment 2, this means that the visual learning mechanism underlying cue combination adaptation is biased such that some sets of statistics are more easily learned than others.

Keywords: Cue combination; Motion and texture cues; Perception

1. Introduction

The human visual system obtains information about depth and shape from a large number of cues. For instance, cues to depth and shape result from object rotation (kinetic depth effect), observer motion (motion parallax), binocular vision in which the two eyes receive different patterns of light (stereopsis), texture gradients in retinal images, as well as other features of retinal images arising from the way in which a 3-D world is projected onto a 2-D retina (perspective). No single cue is necessary for depth or shape perception or dominates our perception of depth or shape in all situations (Cutting & Vishton, 1995). In addition, no single cue has been shown to be capable of supporting depth or shape perception with the robustness and accuracy demonstrated by human observers in natural settings. Consequently, there has been a large increase in recent years in the number of studies examining strategies observers use to combine information provided by each of multiple cues in a visual environment (e.g. Dosher, Sperling & Wurst, 1986; Bruno & Cutting, 1988; Bülthoff & Mallot, 1988; Rogers & Collett, 1989; Johnston, Cumming & Parker, 1993; Nawrot & Blake, 1993;

* Corresponding author. Tel.: +1-716-275-0753; fax: +1-716-442-9216.
E-mail address: robbie@bcs.rochester.edu (R.A. Jacobs)

0042-6989/99/$ - see front matter
PII: S0042-6989(99)00120-0

Young, Landy & Maloney, 1993; Landy, Maloney, Johnston & Young, 1995; Tittle, Norman, Perotti & Phillips, 1997; Turner, Braunstein & Anderson, 1997).

An important finding of these studies is that observers visual cue combination strategies are remarkably flexible in that these strategies adapt so as to make greater or lesser use of different cues in different visual environments. Maloney and Landy (1989) argued that the weight assigned to a depth estimate derived from a particular cue should reflect the estimated reliability of that cue in the current scene under the current viewing conditions. Evidence in support of this conjecture was provided by Johnston, Cumming and Landy (1994). They reported that subjects relied about equally on stereo and motion cues when making shape judgments at near viewing distances, whereas they relied more on the motion cue at far viewing distances. They speculated that this strategy is sensible because stereo disparities are small at far viewing distances and, thus, small misestimates of disparity can lead to large errors in calculated depth. These researchers also found that when the motion cue was weakened (only two frames of each motion sequence were presented), subjects relied on stereo more heavily. Related results were reported by Young et al. (1993). When either a texture cue or a motion cue was corrupted by added noise, subjects tended to rely more heavily on the uncontaminated cue when making depth judgments. Some limits on the flexibility of observers' cue combination strategies are suggested by the findings of Turner et al. (1997). When stereo and motion cues specified incompatible depths, subjects' performance on a surface detection task was impaired when motion, but not stereo, indicated a surface. Performance was not as severely degraded when stereo indicated a surface, but motion did not. This heavy reliance on information provided by stereo persisted in the presence of foreknowledge about which cue would be relevant for the task.

Although these experiments reveal the flexibility of observers' cue combination strategies on the basis of estimated visual cue reliabilities, they do not address how observers acquire these estimated reliabilities. That is the topic of the present article. The article studies observers' abilities to learn cue combination strategies. It reports the results of three experiments examining how observers adapt their strategies for combining visual depth information in an experience-dependent manner. On each trial in an experiment, subjects monocularly viewed two sequentially presented stimuli where each stimulus depicted a cylinder defined by texture and motion cues. Subjects then performed a two-alternative forced-choice comparison by judging which of the two depicted cylinders was greater in depth. Because motion and texture cues could indicate different shapes, it was possible to design tasks in which one cue provided useful information for making depth judgments, whereas the other cue was irrelevant. In experiment 1, subjects initially received training in which one cue (e.g. motion) was informative and the other cue (e.g. texture) was irrelevant. Each subject's relative weighting of motion and texture cues was then estimated. Then subjects were re-trained under new experimental conditions; in these new conditions, the previously informative cue was irrelevant, and the previously irrelevant cue was informative. Subjects relative weighting of motion and texture cues was again estimated. The results of experiment 1 suggest that observers cue combination strategies are adaptable as a function of training; subjects adjusted their cue combination rules to use a cue more heavily after training in which the cue was informative versus after training in which the cue was irrelevant.

Because experiment 1 provided direct evidence of learning, it was possible for subsequent experiments to evaluate properties of the underlying learning mechanism. Experiment 2 evaluated whether or not experience-dependent adaptation of cue combination rules is context-sensitive. That is, can observers learn to use one cue combination rule in one context, and a different combination rule in a second context? This experiment was identical to experiment 1 except that there was only one training period. During this period, two sets of stimuli were used, one set depicting short cylinders and the other set depicting tall cylinders. One cue (e.g. motion) was informative in displays of short cylinders, whereas the other cue (e.g. texture) was informative in displays of tall cylinders. The data indicate that subjects weighted each cue more heavily in the context in which that cue was informative (e.g. displays of short cylinders) versus the context in which the cue was irrelevant (e.g. displays of tall cylinders). These results suggest that observers can learn multiple cue combination rules, and can learn to apply each rule in the appropriate context.

Experiment 3 evaluated a possible limitation on the context-sensitivity of adaptation of cue combination rules. Vetter, Poggio, and Bülthoff (1994) hypothesized that observers visual systems are biased to learn to perceive in the same way views of bilaterally symmetric objects that differ solely by a symmetry transformation. Similar to experiment 2, experiment 3 used two sets of stimuli; one cue was informative in displays of cylinders at a left oblique orientation, whereas the other cue was informative in displays of cylinders at a right oblique orientation. The results indicate that observers did not learn to use different cue combination rules in different contexts under these circumstances. Thus, the results are consistent with the hypothesis of Vetter et al. (1994). Taken in conjunction with the results of experiment 2, this suggests that the visual learning mechanism underlying cue combination adaptation is biased such that some sets of statistics are more easily learned than others.

2. General methods

2.1. Stimuli and apparatus

The stimuli consisted of elliptical cylinders defined by texture and motion cues. The height of a cylinder is the distance from its topmost point to its bottom-most point; the width of a cylinder is the distance from its leftmost point to its rightmost point (where left and right are defined relative to the observer); the depth of a cylinder is the distance from its point nearest to the observer to its point furthest from the observer (with the restriction that the two points lie in the same horizontal cross-section). The horizontal cross-section of a cylinder could be circular, in which case the cylinder was equally deep as wide, could be elliptical with a principal axis parallel to the observers line of sight (and minor axis parallel to the frontoparallel plane), in which case the cylinder was more deep than wide, or could be elliptical with a principal axis parallel to the frontoparallel plane (and minor axis parallel to the observers line of sight), in which case the cylinder was less deep than wide. Seven cylinder shapes were used in the experiments. The shapes were simulated on a 2-D video display by appropriate texture and motion algorithms. The heights (320 pixels) and widths (160 pixels) of the cylinders were constant (the only exception is the short and tall cylinders used in experiment 2); only the simulated depths of the cylinder shapes varied. The seven shapes had simulated depths of 53, 69, 96, 160, 267, 373 and 480 pixels, respectively.

The texture cue was created by mapping a homogeneous and isotropic texture consisting of circular spots to the surface of each cylinder using a texture mapping algorithm. The details of this algorithm are described in Hearn and Baker (1997). Circular spots were placed on a 2-D sheet whose width was equal to the circumference of a horizontal cross-section of the cylinder, and whose height was equal to the height of the cylinder. The placement of the spots was initially random with the restriction that spots could overlap by only a small amount. Either 45 or 60 spots were placed on the sheet. The radius of each spot was randomly sampled from a uniform distribution ranging from 10 to 16 pixels. The texture mapping algorithm mapped the sheet to the curved surface of the cylinder (the top and bottom of the cylinder were never visible to the observer). When a 3-D curved surface is projected onto a 2-D image, changes in surface orientation result in gradients of texture element size, shape and density in the image. These gradients are texture cues to the shape of a cylinder (see Fig. 1).

The motion cue was created by moving the spots horizontally along the simulated surface of a cylinder in either a clockwise or anticlockwise direction. The motion of a spot may be regarded as analogous to the motion of a train traveling around a track; the shape of the track is given by the horizontal cross-section of a cylinder's surface. The velocity of the spots on the surface was constant within a stimulus presentation; this velocity was varied between presentations. Spots traveled the circumference of a cylinder's horizontal cross-section in either 55 or 75 frames. Note that the cylinder did not rotate; rather, the spots moved along the simulated surface of static cylinders. Thus, the stimuli were different from kinetic depth effect (KDE) stimuli (except when the horizontal cross-section of a cylinder was circular, in which case the stimuli were identical to KDE stimuli). KDE stimuli were not used because they produce artifactual depth cues when the horizontal cross-section of a cylinder is non-circular, such as changes in retinal angle subtended by the cylinder over time. The motion cue in the stimuli used here is an instance of a constant flow field. Constant flow fields produce reliable and robust perceptions of depth (e.g. Perotti, Todd & Norman, 1996; Perotti, Todd, Lappin & Phillips, 1998).

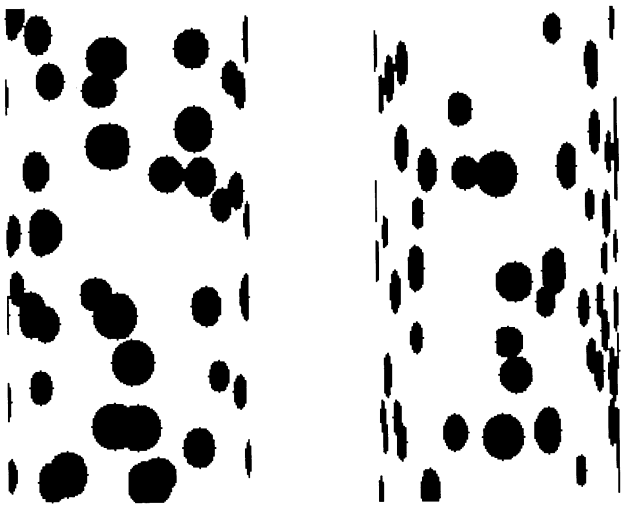

Fig. 1. Two example stimuli. The stimulus on the left is a cylinder whose depth equals its width; the stimulus on the right is a cylinder whose depth is three times its width. Because the cylinders are transparent, spots appearing in a stimulus may either be on the front or back surface of a cylinder.

A computer graphics manipulation was used to independently manipulate the shapes indicated by texture and motion cues[1]. For example, the motion cue might

[1] When a 3-D curved surface is projected onto a 2-D image, changes in surface orientation result in gradients of texture element size, shape (compression) and density in the image. It is not possible to independently manipulate the depths indicated by the motion cue and by all three texture gradients. However, as described in this article, it is possible to independently manipulate the depths indicated by the motion cue and the gradients of texture element compression. This is adequate for our purposes because experimental data indicates that, of the three texture gradients, gradients of texture element compression are the primary (nearly exclusive) determinants of observers perceptions of depth or shape for the types of stimuli used in the experiments reported in this article (cf. Blake, Bülthoff & Sheinberg, 1993; Cumming, Johnston & Parker, 1993; Cutting & Millard, 1984; Knill, 1998).

have indicated a cylinder with a circular horizontal cross-section, whereas the texture cue indicated a cylinder of identical height and width but with an elliptical horizontal cross-section that was more deep than wide. Pilot experiments, as well as debriefings of the subjects used in the experiments reported below, suggest that subjects were never aware that motion and texture cues could indicate different shapes under the experimental conditions considered in this article. The graphics manipulation was nearly identical to the manipulation described in Young et al. (1993).

Two cylinders of identical heights and widths, but different depths, were defined. The cylinders were positioned so that their midpoints lay at the origin of a 3-D coordinate system. The x-axis of this system is parallel to the frontoparallel plane (the width of a cylinder is measured along this axis), the y-axis is parallel to the direction of gravity (a cylinders height is measured along this axis), and the z-axis is parallel to the observers line of sight (a cylinders depth is measured along this axis). Let W denote one-half the width of each cylinder, d_m denote one-half the depth of the first cylinder, also referred to as the motion cylinder, and let d_t denote one-half the depth of the second cylinder, also referred to as the texture cylinder. Initially, texture spots were defined as described above, meaning that values were assigned to the number of spots, the location of the center of each spot on the texture cylinder, and the radius of each spot. The centers of the spots were mapped from the surface of the texture cylinder to the surface of the motion cylinder. Specifically, the center of the ith spot, with coordinates (x_i, y_i, z_i), was mapped to the surface of the motion cylinder by scaling the depth of the center using the transformation $z_i \leftarrow (d_m/d_t)z_i$. The motions of the centers of the spots were determined based on the motion cylinders shape, thereby providing the positions of the centers of the spots at each frame (see above). Next, the centers of the spots were mapped from the surface of the motion cylinder to the surface of the texture cylinder using the transformation $z_i \leftarrow (d_t/d_m)z_i$. The shape of each texture spot was determined based on the texture cylinders shape using a texture mapping algorithm (see above). A stimulus view was rendered from a parallel projection along the observers line of sight of the spots on the texture cylinder. The end result was a projected motion of an inhomogenously and anisotropically textured cylinder with half-depth d_m; the projected texture compression was that of a homogeneously and isotropically textured cylinder with half-depth d_t.

The visual image of a cylinder at a vertical orientation subtended 2.21° of visual angle in the horizontal dimension and 4.6° in the vertical dimension. Stimuli were viewed from a distance of 1.45 m. They were rendered using a PowerComputing 225 computer (a clone of an Apple Macintosh) and a Sony Trinitron Multiscan 20sf II monitor. The video format was 100 Hz, noninterlace. The background luminance was 0.02 cd/m^2 and the luminance of the texture spots was 30 cd/m^2 (this pattern of luminances is the reverse of that illustrated in Fig. 1). Stimuli were viewed monocularly. A black frame matched in color and luminance to the background of the stimuli was placed directly in front of the monitor so as to hide the monitors edges.

2.2. Procedure

On each training trial, subjects monocularly viewed two sequentially presented stimuli where each stimulus depicted a different cylinder. Subjects then performed a two-alternative forced-choice (2AFC) comparison by judging which of the two depicted cylinders was greater in depth. An auditory signal provided feedback as to whether the response was correct or incorrect. Each stimulus was presented for 750 ms (16 frames). The monitor was blank (black) between the two stimuli of a trial for 920 ms (experiments 1 and 2) or 1780 ms (experiment 3) (these times varied across experiments because the speed of generating a stimulus varied with the orientation of the cylinder).

Define set M to be the collection of displays in which the cylinder shape indicated by the motion cue was one of the seven possible shapes, and in which the shape indicated by the texture cue was circular (the cylinder was equally deep as wide). Define set T to be the collection of displays in which texture indicated one of the seven possible shapes, whereas motion indicated a circular shape. Sets M and T each consisted of 224 displays. The size of these sets was determined as follows. The displays of set M may be denoted by the depth pairs (1, 4), ..., (7, 4) where the numbers in a pair give the depths indicated by the motion and texture cues, respectively (the value 1 indicates the smallest depth, the value 4 indicates a depth that is equal to a cylinders width, and the value 7 indicates the largest depth). For example, the depth pair (7, 4) means that the motion cue indicates a cylinder with the greatest depth, whereas the texture cue indicates a cylinder whose depth equals its width. Similarly, the displays of set T may be denoted by the depth pairs (4, 1), ..., (4, 7). For each possible depth pair, 32 displays of cylinders were created. These 32 displays indicated cylinders of the same depth pair; however, they differed in their appearances (e.g. different numbers of texture elements, different initial locations of the texture elements, different sizes of the texture elements, different velocities of the texture elements, etc.).

Under *motion relevant* training conditions, the two displays of a trial were randomly sampled from set M such that the motion cues in the displays indicated cylinders of different depths; that is, only the motion cue distinguished the shapes of the cylinders in the two displays; the texture cue was identical in the two displays. In this case, only the motion cue provided information useful for performing the task. Under *texture relevant* training conditions, the two displays were sampled from set T such that the texture cues in the displays indicated cylinders of different depths; that is, only the texture cue distinguished the shapes of the cylinders in the two displays; the motion cue was identical in the two displays. Only the texture cue provided information useful for performing the task under texture relevant training conditions.

Test trials were identical to training trials with the following two exceptions. First, subjects did not receive an auditory signal indicating whether or not their response was correct. Instead, subjects heard a 'click' indicating that they had made a response. Second, one of the displays in a test trial was sampled from set T whereas the other display was sampled from set M. As discussed below, test trials evaluated the extent to which subjects made their depth judgments on the basis of texture information versus motion information.

Subjects participated in each experiment for 5 days. The 5 days generally occurred within a 2-week period. Four blocks of 125 trials (experiment 1) or 120 trials (experiments 2–3) were conducted on each day (a block of trials required about 8 min to complete). Subjects rested between blocks of trials. A block of trials contained either all training trials or a mixture of training and test trials. If the block contained a mixture, then 98 of the trials were test trials (seven possible cylinder shapes from set M, seven possible cylinder shapes from set T, and two presentations of each possible pair of shapes). The use of training trials randomly intermixed with test trials was intended to force subjects to remain mentally focused on the task (recall that subjects did not receive feedback during test trials).

In order to measure the adaptation in an observers cue combination strategy as a function of training it was necessary to specify a parametric model of this strategy. It was assumed that observers combine visual depth information based on the texture and motion cues using a linear cue combination rule:

$$d(t,m) = w_T d(t) + w_M d(m) \quad (1)$$

where t denotes the texture cue, m denotes the motion cue, $d(t,m)$ is the composite percept of visual depth, $d(t)$ is the depth percept based on the texture cue, $d(m)$ is the depth percept based on the motion cue, and w_T and w_M are the linear coefficients corresponding to the texture and motion cues, respectively. It was also assumed that w_T and w_M are non-negative and sum to one. Linear cue combination rules are often assumed in the visual perception literature, and they have received a considerable degree of empirical support (e.g. Dosher et al., 1986; Bruno & Cutting, 1988; Landy et al., 1995). Evidence is presented below indicating that a linear combination rule provides a good fit to the experimental data reported in this article.

To complete the specification of Eq. (1), it is necessary to specify observers depth perceptions based on the texture cue and based on the motion cue, the functions $d(t)$ and $d(m)$, respectively. However, the values of these functions are not known, and there is no uncontroversial method for estimating these values. We assumed that the depth estimates based on the texture cue and the motion cue are each veridical. For example, when viewing a display in which the texture cue indicates a cylinder whose depth is 69 pixels and the motion cue indicates a cylinder whose depth is 160 pixels, then $d(t) = 69$ and $d(m) = 160$ (in the calculations presented below, the functions $d(t)$ and $d(m)$ were then linearly scaled to lie between -1 and 1; this scaling made the calculations more numerically stable). The veridical assumption is approximately correct, and is commonly made by researchers studying cue combination rules (e.g. Tittle et al., 1997).

On the basis of the results of a set of test trials, it is possible to estimate an observer's linear coefficients w_T and w_M using a statistical model that relates the visual stimuli to an observer's responses. In particular, a model known as a *logistic* model is suitable for these purposes. Mathematically, the logistic model is an instance of a generalized linear model that is appropriate for binary response data, such as subjects responses in a 2AFC experiment (McCullagh & Nelder, 1989; Dobson, 1990). Because the model has a probabilistic interpretation, it is possible to characterize its performance using a likelihood function, and to estimate the values of its parameters using a maximum likelihood estimation procedure. The mathematical details of the logistic model are presented in the appendix; here we present an intuitive description of the model.

Recall that one of the displays in a test trial is sampled from set T and the other display is sampled from set M, and that the subject judged which display depicted a cylinder with greater depth. The logistic model is a function with four independent variables and one dependent variable. The independent variables are observer's depth perceptions based on the motion and texture cues in the displays from sets M and T. Using the superscript M or T to indicate the set, and the argument m or t to indicate the cue, the independent variables may be denoted $d^M(m)$, $d^M(t)$, $d^T(m)$ and $d^T(t)$, where $d^M(m)$ is the depth percept based on the motion cue in the display from set M, and the other variables follow the same notational convention. The dependent variable of the logistic model is a prediction

of the probability that the subject chose the display from set M as depicting the deeper cylinder, denoted $P(\text{response} = M)$. The model may be summarized as the mapping

$$P(\text{response} = M) \leftarrow d^M(m), d^M(t), d^T(m), d^T(t) \quad (2)$$

The output of the model is a probability distribution that is a monotonic, differentiable function whose shape resembles a multidimensional 'S'. As we now discuss, the exact shape of this function is determined by two parameters.

We might expect that if a subject's composite depth percept based on the display from set M was greater than his or her composite depth percept based on the display from set T, then the subject was more likely to select the display from set M as depicting the deeper cylinder. On the other hand, if the composite depth percept based on the set M display was less than the percept based on the set T display, then the subject was less likely to select the display from set M as depicting the deeper cylinder. If the two depth percepts were equal, then the subject would choose the set M display about half the time.

According to this reasoning, the probability that a subject chose the set M display as depicting the deeper cylinder depends on the difference between the value of the composite depth percept based on the set M display and the value of the composite depth percept based on the set T display. This difference is denoted $d^M(t,m) - d^T(t,m)$. Using the fact that each composite depth percept depends upon the values of the linear coefficients w_M and w_T (see Eq. (1)), and using the fact that $w_T = 1 - w_M$, we arrive at the conclusion that the probability that a subject chose the set M display on a test trial is dependent on the value of the motion coefficient w_M.

The motion coefficient is not the only parameter that determines the prediction of the probability of a subjects response. This prediction is also determined by a parameter known as temperature, denoted τ. The value of τ scales the difference in the values of the composite depth percepts. The exact form of the logistic model is

$$P(\text{response} = M) = \frac{1}{1 + \exp\{-[d^M(t,m) - d^T(t,m)]/\tau\}} \quad (3)$$

Instances of this model are illustrated in Fig. 2.

Fig. 2 illustrates how different values of w_M and τ influence the estimate of the probability distribution $P(\text{response} = M)$. As noted in Eq. (2), this estimate changes according to the motion and texture cues in the displays from sets M and T. A five-dimensional plot is therefore needed to visualize the entire estimated distribution. Fortunately, the texture cue in the displays from set M and the motion cue in the displays from set T were constant (they always indicated a cylinder whose horizontal cross-section was circular), and so the probability distribution $P(\text{response} = M)$ can be plotted in a 3-D graph that omits these constant values. The four graphs in Fig. 2 show the predicted probability distributions corresponding to four different sets of values of w_M and τ. In each graph, the axis labeled 'motion shape' gives the cylinder shape indicated by the motion cue in the display from set M (recall that seven shapes were used in the experiments; 1 is the shape with the smallest depth, 7 is the shape with the greatest depth); the 'texture shape' axis gives the cylinder shape indicated by the texture cue in the display from set T.

The four graphs have sensible shapes. As the motion cue in the display from set M indicates a deeper cylinder (that is, as the value along the motion shape axis increases), the logistic model predicts that the probability that a subject judges the display from set M as depicting a deeper cylinder increases. Similarly, as the texture cue in the display from set T indicates a deeper cylinder (as the value along the texture shape axis increases), the model predicts that $P(\text{response} = M)$ decreases. The degree to which predictions of $P(\text{response} = M)$ rise along the motion shape axis and decline along the texture shape axis is altered by the value of the motion coefficient w_M. As illustrated in the left column of Fig. 2, if w_M is large (and thus w_T is small), then the prediction of $P(\text{response} = M)$ rises quickly along the motion shape axis, and declines slowly along the texture shape axis. In contrast, if w_M is small (w_T is large), then the prediction of $P(\text{response} = M)$ rises slowly along the motion shape axis, and declines quickly along the texture shape axis (illustrated in the graphs in the right column). This general effect is scaled by the temperature parameter τ. A comparison of the top and bottom rows of Fig. 2 indicates that the predicted probability distribution is relatively flat when τ is large (top row), and more closely approximates a step function as τ decreases (bottom row).

The estimated values of τ are not discussed in the presentation of the experiments below; this parameter is useful for enabling the logistic model to accurately fit a subjects response data on the test trials, but it is not directly relevant to the research questions addressed in this article. The estimates of w_M (but not w_T because w_T is simply $1 - w_M$) are extensively discussed below because they can be used to quantify how much a subject relied on the motion cue versus the texture cue under different experimental conditions. By comparing the values of w_M across experimental conditions, it is possible to measure the degree to which a subject adapted his or her cue combination strategy in an experience-dependent manner.

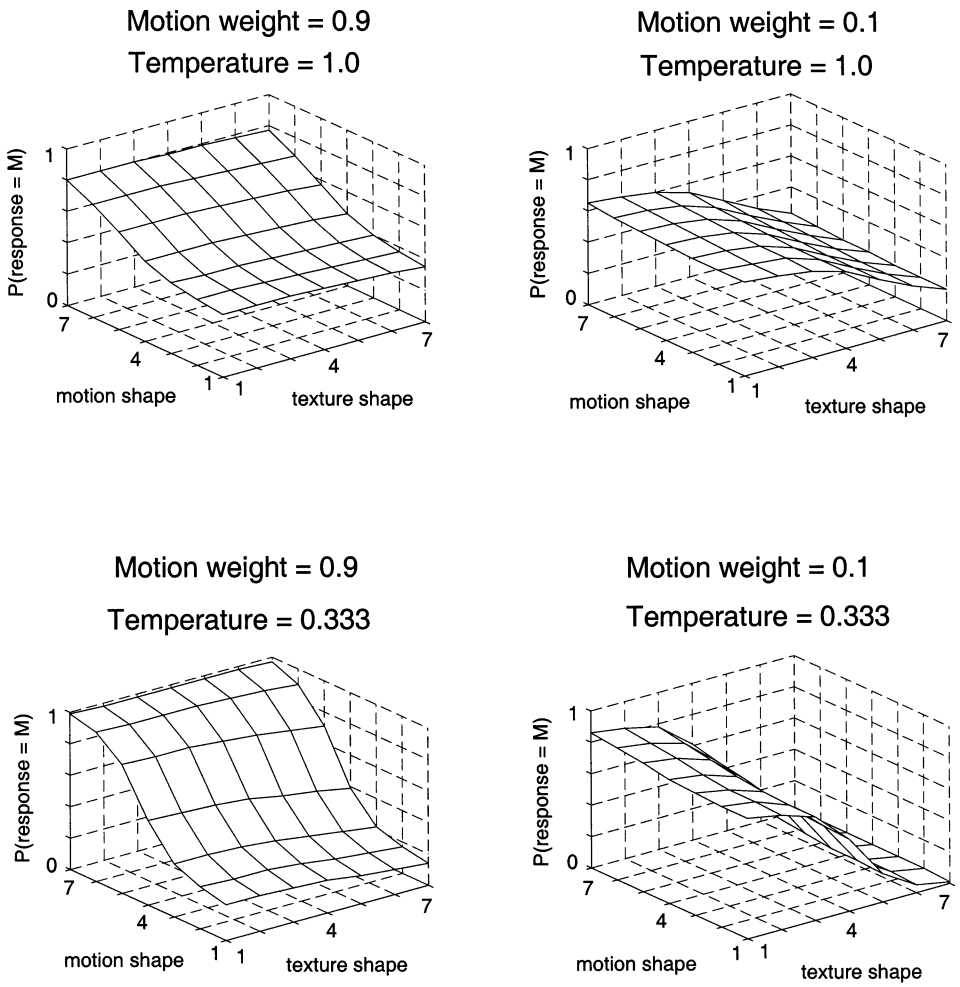

Fig. 2. The probability distributions $P(\text{response} = M)$ corresponding to four different sets of values of the motion coefficient w_M and the temperature τ.

2.3. Subjects

Subjects were undergraduate students at the University of Rochester. They had normal or corrected-to-normal vision. They were naive to the purposes of the experiments.

3. Experiment 1

Experiment 1 studied differences in observers cue combination rules after prolonged training under the motion relevant condition (the motion cue distinguished the shapes of the cylinders in the two displays of a trial; the texture cue was identical in the two displays) versus after prolonged training under the texture relevant condition. On the first day of participation in the experiment, all subjects received training trials in which texture and motion cues were consistent; that is, the two cues indicated cylinders with the same shape. This allowed subjects to become comfortable with the experimental situation. On day 2 and the first half of day 3, half the subjects were trained under the motion relevant training condition. They received four blocks of 125 trials on day 2; on day 3 they received two blocks of trials. At the end of the third day, subjects were given two blocks of trials consisting of a mixture of motion relevant training trials and test trials. The test trials were included so that estimates of the subjects linear coefficients w_M and w_T could be obtained. On day 4 and the first half of day 5 the subjects were trained under the texture relevant training condition (four blocks and two blocks of trials on the two days, respectively). At the end of day 5 they received two blocks of trials consisting of a mixture of texture relevant training trials and test trials. The order of training conditions was counterbalanced across subjects (the other half of the subjects were trained and tested in the reverse order: first texture relevant training and testing, then motion relevant training and testing). Our prediction was that subjects would adapt their cue combination strategies so that they relied more on the motion

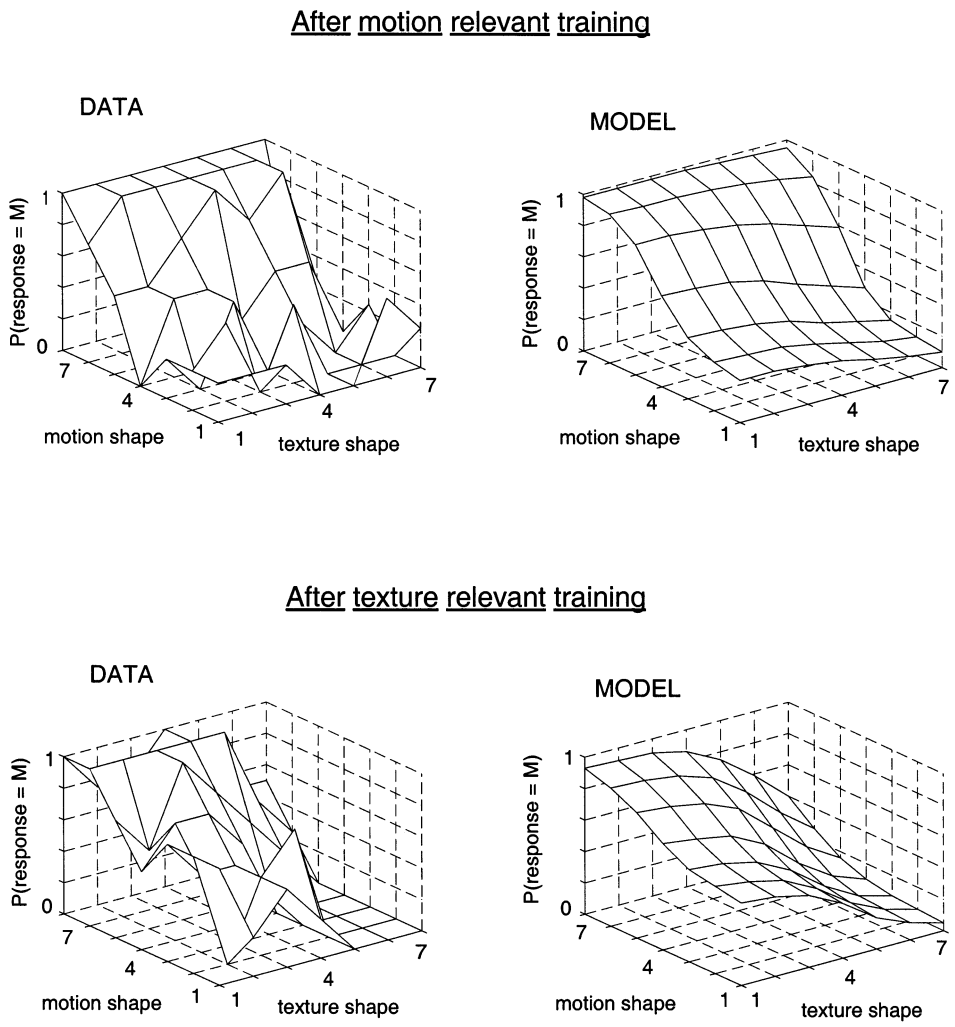

Fig. 3. The response data of subject DA on test trials following motion relevant training (top-left graph) and texture relevant training (bottom-left graph). The logistic model was used to fit surfaces to these two datasets. These surfaces are shown in the top-right and bottom-right graphs, respectively.

cue after motion relevant training than after texture relevant training. If so, then estimates of the linear coefficient w_M should be larger after motion relevant training.

The response data of one subject, subject DA, on the test trials are shown in Fig. 3. This subject was initially trained under the motion relevant training condition; this training was then followed by texture relevant training. The top-left graph of Fig. 3 gives the subjects response data on the test trials following motion relevant training. The top-right graph shows the surface that was fit to the subject's data by the logistic model described above. The shape of the data is sensible. As the motion cue in the display from set M indicated a deeper cylinder (that is, as the value along the motion shape axis increased), the probability that the subject judged the display from set M as depicting a deeper cylinder increased. Similarly, as the texture cue in the display from set T indicated a deeper cylinder (as the value along the texture shape axis increased), the probability that the subject judged the display from set M as depicting a deeper cylinder decreased. Analogous graphs for the test trials following texture relevant training are shown in the bottom row of Fig. 3. The bottom-left graph shows the subjects response data; the bottom-right graph shows the surface that was fit to the subjects data by the logistic model.

A comparison of the graphs in the top and bottom rows of Fig. 3 reveals experience-dependent adaptation. The subject responded to the same set of test trials in different ways following training under motion relevant and texture relevant conditions. Moreover, the experience-dependent adaptation occurred in a logical way. Following motion relevant training, the subject was highly sensitive to the motion cue and relatively insensitive to the texture cue. This is evidenced by the fact that the graph of the subject's data rises sharply along the motion shape axis, but declines gradually along the

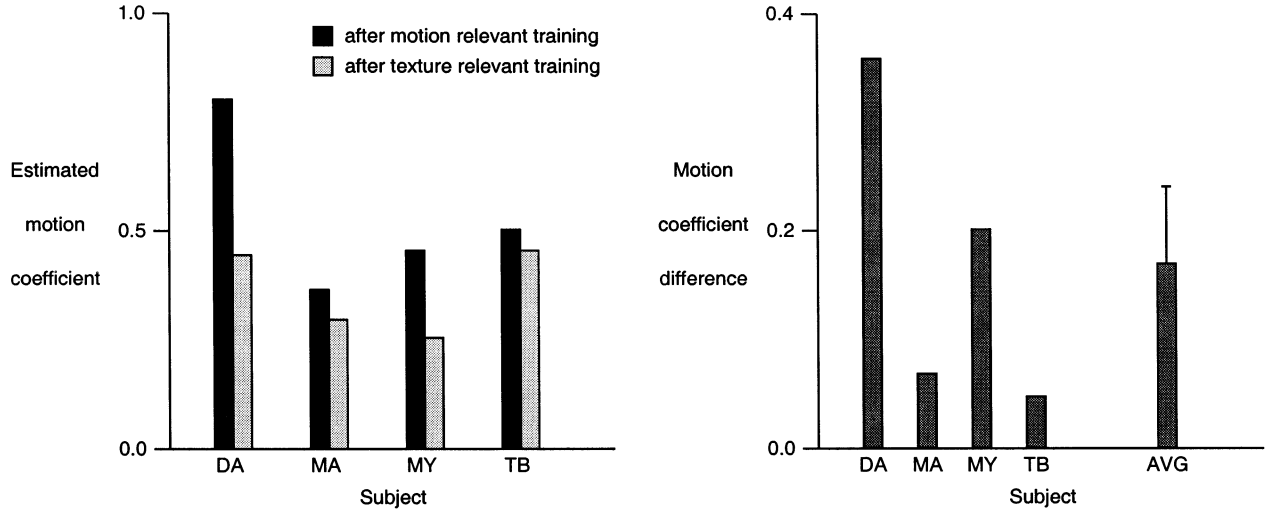

Fig. 4. The results of experiment 1. The graph on the left shows the estimated values of the motion coefficient w_M following motion relevant training and following texture relevant training for the four subjects. The graph on the right shows the motion coefficient difference for each subject. The rightmost bar in this graph gives the average motion coefficient difference; the error bar gives the standard error of the mean.

texture shape axis (this is most easily seen in the top-right graph of Fig. 3). Examining the subjects performance after texture relevant training compared with her performance following motion relevant training, the subject was relatively less sensitive to the motion cue and more sensitive to the texture cue. The graph of the data rises less steeply along the motion shape axis following texture relevant training (bottom row of Fig. 3), and rises more steeply along the texture shape axis.

The estimated values of the motion coefficient w_M following motion relevant training and following texture relevant training for the four subjects that participated in experiment 1 are shown in the graph on the left of Fig. 4. The horizontal axis gives the subject; the vertical axis gives the estimated value of the motion coefficient. All four subjects had larger motion coefficients following motion relevant training than following texture relevant training. If we define the motion coefficient *difference* to be the estimate of a subjects motion coefficient following motion relevant training minus the value of this estimate following texture relevant training, then a positive difference indicates that a subject weighted the motion cue more heavily in his or her cue combination rule following motion relevant training than following texture relevant training. The motion coefficient differences for the four subjects are shown in the graph on the right of Fig. 4. The horizontal axis gives the subject; the vertical axis gives the motion coefficient difference. The rightmost bar in the graph is the average motion coefficient difference; the error bar gives the standard error of the mean. Using a one-tailed t-test, the average motion coefficient difference is significantly greater than zero ($t = 2.357$, $P < 0.05$); in addition, the 95% confidence interval for the average motion coefficient difference contains only positive values. These results indicate that subjects have a larger motion coefficient after motion relevant training than after texture relevant training. We conclude, therefore, that observers cue combination strategies are adaptable as a function of training; subjects adjusted their cue combination rules to more heavily use a cue when that cue was informative on a given task versus when the cue was irrelevant.

The above conclusion relies on the assumption that the logistic model provides a reasonably good fit to a subjects response data on the test trials. Visual inspection of the left and right columns of Fig. 3 suggests that this is indeed the case for subject DA. A quantitative comparison was made by correlating the actual probability that a subject judged the display from set M as depicting a deeper cylinder on a test trial with the probability predicted by the logistic model. The correlation coefficient is 0.8. A linear regression in which the probability predicted by the logistic model is the independent variable and the actual probability of a subjects response is the dependent variable yields a slope of 1.09 and an intercept of -0.06. We conclude that the logistic model provides a good fit to the experimental data. In fact, the data in all the experiments reported in this article are well fit by the logistic model (the corresponding values of the correlation coefficient, slope, and intercept for experiment 2 are 0.77, 1.08 and -0.08, respectively; for experiment 3 the values are 0.8, 1.01 and 0.0, respectively). Because the logistic model provides a good fit, an additional conclusion that can be drawn from the experimental data is that observers cue combination strategies under the conditions studied here are well-approximated by linear cue combination rules (cf. Landy et al., 1995).

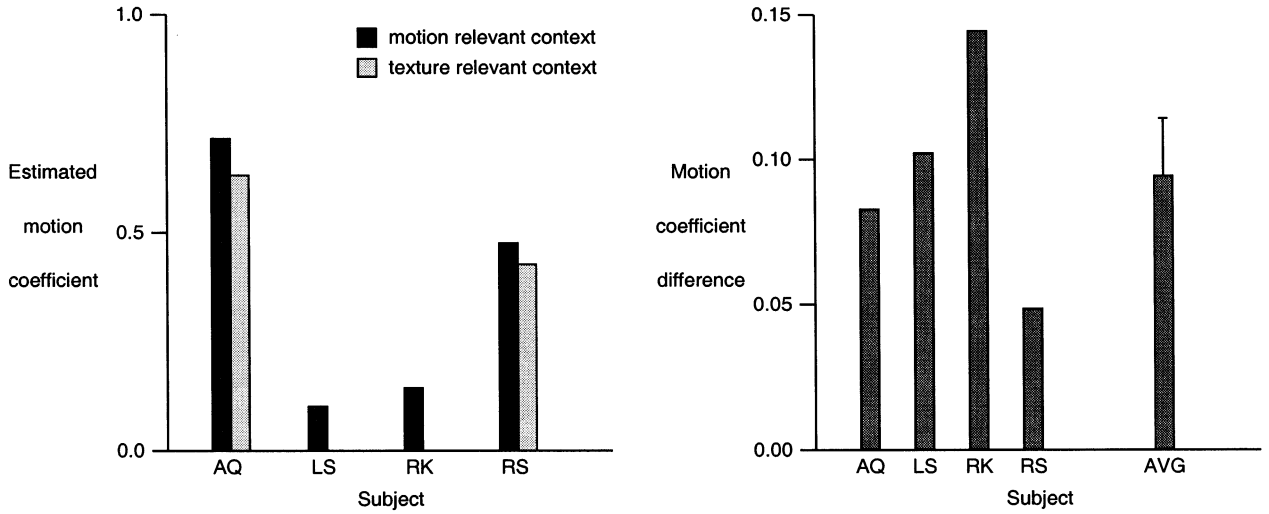

Fig. 5. The results of experiment 2. The graph on the left shows the estimated values of the motion coefficient w_M in the motion relevant context and in the texture relevant context for the four subjects. The graph on the right shows the motion coefficient difference for each subject. The rightmost bar in this graph gives the average motion coefficient difference; the error bar gives the standard error of the mean.

4. Experiment 2

Experiment 2 evaluated whether or not adaptation of cue combination rules is context-sensitive. That is, can observers learn to use one cue combination rule in one context, and a different combination rule in a second context? This experiment used displays of cylinders of different heights. For half the subjects, training trials containing displays of short cylinders (height = 200 pixels) used displays belonging to set M (the motion cue distinguished the shapes of the cylinders depicted in the two displays of a trial; the texture cue was identical in the two displays), whereas trials containing displays of tall cylinders (height = 320 pixels) used displays belonging to set T. In other words, the motion cue was relevant on trials using displays of short cylinders; the texture cue was relevant on trials using displays of tall cylinders. For the other half of the subjects, the relationship between cylinder height and relevant visual cue was reversed.

Subjects received training trials on all 5 days in which they participated in the study. On days 1–3, they received four blocks of 120 training trials; on days 4–5, they received two blocks. On days 4–5 they also received two blocks of trials consisting of a mixture of training trials and test trials. The two displays of a test trial depicted either short or tall cylinders. As before, one display was sampled from set M and the other display was sampled from set T. Test trials were included so that estimates of the subject's linear coefficients w_M and w_T could be obtained in the context of short cylinders and in the context of tall cylinders. During training blocks, trials were organized into groups of 20 trials; on trials in even-numbered groups subjects were required to judge the depths of two short cylinders; on trials in odd-numbered groups subjects were required to judge the depths of two tall cylinders. Importantly, however, during blocks containing test trials, trials with short or tall cylinders were randomly intermixed.

The estimated values of the motion coefficient w_M in the motion relevant context (e.g. short cylinders) and in the texture relevant context (e.g. tall cylinders) for the four subjects that participated in experiment 2 are shown in the graph on the left of Fig. 5. All four subjects had larger motion coefficients in the motion relevant context than in the texture relevant context. Define the motion coefficient difference to be the estimate of a subjects motion coefficient w_M in the motion relevant context minus the value of this estimate in the texture relevant context. We predicted that this difference would be positive, thereby indicating that a subject weighted the motion cue more heavily in his or her cue combination rule in the motion relevant context than in the texture relevant context. The motion coefficient differences for the four subjects are shown in the graph on the right of Fig. 5. Using a one-tailed t-test, the average motion coefficient difference is significantly greater than zero ($t = 4.742$, $P < 0.01$); the 95% confidence interval for the average motion coefficient difference contains only positive values. These results indicate that subjects weighted the motion cue more heavily in the motion relevant context than in the texture relevant context. We conclude that adaptation of observers cue combination rules is context-sensitive. Observers can learn to weight a cue more or less heavily depending on the context in which that cue appears.

As noted above, previous investigators have shown that observers visual cue combination strategies are flexible in the sense that these strategies adapt so as to

make greater or lesser use of different cues in different visual environments (e.g. Johnston et al., 1994; Turner et al., 1997). Maloney and Landy (1989) argued that the weight assigned to a depth estimate derived from a particular cue should reflect the estimated reliability of that cue in the current scene under the current viewing conditions. In experiments 1 and 2, we studied how observers acquire these estimated reliabilities. These experiments reveal that the reliabilities can be learned by a mechanism that is sensitive to the statistical nature of the tasks performed by the observer.

The results of experiments 1 and 2 suggest the existence of learning mechanisms that adapt observers visual cue combination strategies in flexible ways on the basis of prior training, at least under the circumstances studied here. Given the existence of these mechanisms, it is important to evaluate how powerful they are. For example, is the context-sensitivity of these learning mechanisms unbiased in the sense that observers can learn equally well to use different cue combination strategies in different contexts no matter how the contexts differ from each other? Or is the context-sensitivity of these mechanisms biased such that it is easier to learn to use different combination strategies when the contexts differ in some ways and harder to learn to use different combination strategies when the contexts differ in other ways? Experiment 3 addresses these questions.

5. Experiment 3

Vetter et al. (1994) noted that learning devices frequently need to observe an object from a large number of viewpoints before they can successfully recognize the object from a novel viewpoint. They proposed that more efficient learning can occur if a device exploits invariant properties of an object. For example, if an object is bilaterally symmetric, and if a device observes a single 2-D view of the object, then it is possible for the device to generate a new 'virtual view' of this object using an appropriate symmetry transformation[2]. The device can then learn on the basis of the observed view and the virtual view. Taken to its logical extreme, this hypothesis might predict that devices that use this learning strategy would treat observed views and virtual views equivalently. A device would perceive, for instance, a bilaterally symmetric object's visual properties, such as its depth, in the same way for all viewpoints that are equivalent except for a symmetry transformation. In particular, the hypothesis predicts that an observer using this strategy in order to learn to perceive the depth of cylinders cannot learn to use different cue combination rules in different contexts when one context is characterized by views of a cylinder at one orientation (e.g. cylinder at a left oblique orientation) and the other context is characterized by views of the cylinder that result from a symmetry transformation (e.g. cylinder at a right oblique orientation).

Experiment 3 was identical to experiment 2 except that it used displays of cylinders at different orientations. For half of the subjects, training trials containing displays of cylinders at a left oblique orientation used displays belonging to set M, whereas trials containing displays of cylinders at a right oblique orientation used displays belonging to set T. That is, the motion cue was relevant on trials using displays of left oblique cylinders; the texture cue was relevant on trials using displays of right oblique cylinders. The relationship between cylinder orientation and relevant visual cue was reversed for the other subjects.

The estimated values of the motion coefficient w_M in the motion relevant context (e.g. cylinders at a left oblique orientation) and in the texture relevant context (e.g. cylinders at a right oblique orientation) for the six subjects that participated in experiment 3 are shown in the graph on the left of Fig. 6. Define the motion coefficient difference to be the estimate of a subject's linear coefficient w_M in the motion relevant context minus the value of this estimate in the texture relevant context. The motion coefficient differences for the six subjects are shown in the graph on the right of Fig. 6. There is no systematic pattern among the differences, and the average motion coefficient difference is not significantly different than zero. Because subjects failed to learn to use different cue combination strategies in contexts that differed by a symmetry transformation, this result is consistent with the strong version of the hypothesis by Vetter et al. (1994) outlined above.

Rather than supporting the hypothesis of Vetter et al., another plausible interpretation of the data is that it supports the conjecture that observers visual learning systems are strongly biased to produce orientation invariance (Biederman, 1987). If this conjecture is true, then it ought to be the case that observers find it difficult to learn to use different combination rules in different contexts when the contexts are identical except for a change in orientation. That would explain why subjects failed to learn to use different cue combination rules on trials with cylinders at a left oblique orientation versus a right oblique orientation. We do not favor this conjecture, however, because pilot data has shown that many subjects, though not all, successfully learned to use different combination rules when one cue was informative on trials with vertical cylinders, and the other cue was informative on trials with horizontal cylinders. Thus, these data suggest that the visual learn-

[2] By definition, this transformation exchanges the coordinates of bilaterally symmetric pairs of features, possibly with a change of sign, thereby generating a new view that is not necessarily a simple rotation in the image plane.

ing system is not strongly biased to produce orientation invariance (Bülthoff, Edelman & Tarr, 1995). The results of experiment 3 are more consistent with the hypothesis of Vetter et al. We conclude that observers visual learning systems may be biased such that observers tend to perceive in the same way views of bilaterally symmetric objects that differ solely by a symmetry transformation.

When studying a learning mechanism, it is important to understand both what the mechanism can do and also what it cannot do. The results of experiments 1 and 2 suggested that subjects adapt their cue combination rules in a way that is consistent with the statistics of the tasks. For example, subjects adapt their cue combination rules to more heavily weight a cue in the context in which the cue is informative, and to weight it less heavily in a context in which it is irrelevant. The results of these experiments tell us about what the visual learning mechanism underlying cue combination adaptation can do, but reveal nothing about the limitations of this mechanism. Can the mechanism learn the statistics of all visual tasks equally well, or is it biased such that some sets of statistics are more easily learned than others? The results of experiment 3 suggest that the mechanism may be biased. When contexts differ solely by a symmetry transformation, the mechanism is impaired in its ability to adapt an observers combination strategy so that different combination rules are used in different contexts.

6. Summary and conclusions

Previous investigators have shown that observer's visual cue combination strategies are remarkably flexible in the sense that these strategies adapt on the basis of the estimated reliabilities of the visual cues (e.g. Young et al., 1993; Turner et al., 1997). However, these researchers have not addressed how observers acquire these estimated reliabilities. This article has studied observers abilities to learn cue combination strategies. It reported the results of three experiments examining how observers adapt their strategies for combining visual depth information in an experience-dependent manner. The results of experiment 1 suggest that observers cue combination strategies are adaptable as a function of training; subjects adjusted their cue combination rules to more heavily use a cue when that cue was informative on a given task versus when the cue was irrelevant.

Because experiment 1 provided direct evidence of learning, it was possible for subsequent experiments to evaluate properties of the underlying learning mechanism. Experiment 2 evaluated whether or not experience-dependent adaptation of cue combination rules is context-sensitive. That is, can observers learn to use one cue combination rule in one context, and a different combination rule in a second context? The results suggest that observers can learn multiple cue combination rules, and can learn to apply each rule in the appropriate context.

Experiment 3 evaluated a possible limitation on the context-sensitivity of adaptation of cue combination rules. It used two sets of displays; one cue was informative in displays of cylinders at a left oblique orientation, whereas the other cue was informative in displays of cylinders at a right oblique orientation. The results indicate that observers did not learn to use different cue combination rules in different con-

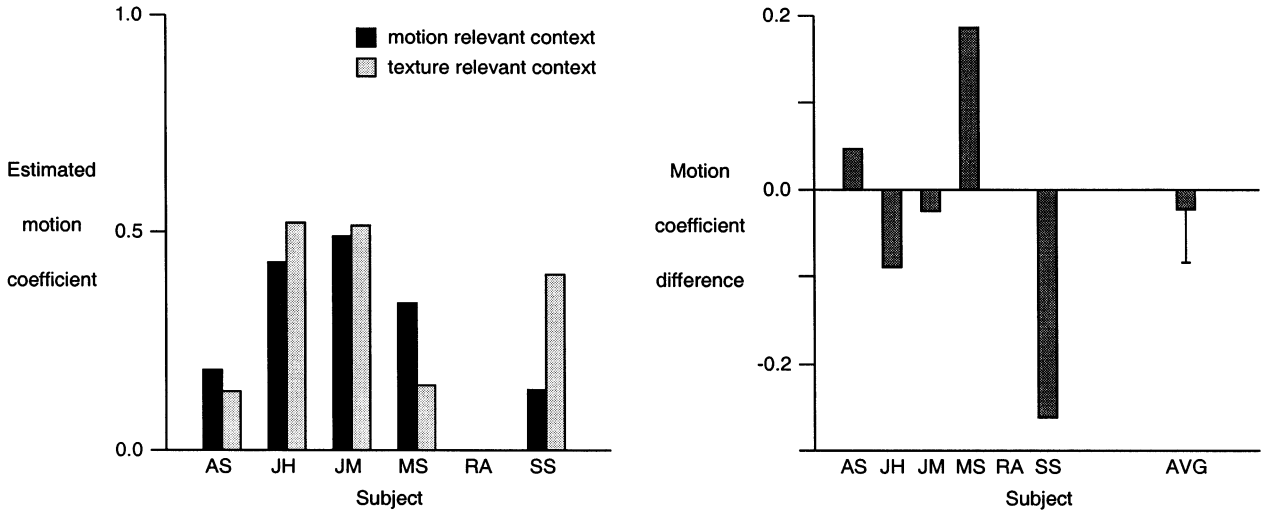

Fig. 6. The results of experiment 3. The graph on the left shows the estimated values of the motion coefficient w_M in the motion relevant context and in the texture relevant context for the six subjects. The graph on the right shows the motion coefficient difference for each subject. The rightmost bar gives the average motion coefficient difference; the error bar gives the standard error of the mean.

texts under these circumstances. These results suggest that observers visual systems may be biased such that observers tend to learn to perceive in the same way views of bilaterally symmetric objects that differ solely by a symmetry transformation. Taken in conjunction with the results of experiment 2, this suggests that the visual learning mechanism underlying cue combination adaptation is biased such that some sets of statistics are more easily learned than others.

Context-sensitivity is one property of the learning mechanism that produces experience-dependent adaptation of observer's cue combination strategies. Future work will consider additional properties. One important issue concerns the asymptotes of learning. The changes in cue combination strategies reported in this article are modest in size. It is likely that longer training periods would have produced larger changes. What other factors influence the amount of learning? In particular, under what conditions is learning maximal? A second issue, related to the first, concerns the influence of the nature of the feedback on the learning process. We speculate that rapid learning, and large amounts of learning, will occur if subjects can interact with the cylinders. For example, rapid learning may occur if motor feedback when grasping a cylinder is consistent with one cue, and inconsistent with another cue. We are currently using a virtual reality environment to test this hypothesis. Another important issue concerns observers' cue combination generalization gradients. If, for example, subjects are trained to believe that one cue is relevant to depth perception when viewing short cylinders, and another cue is relevant when viewing tall cylinders (as in experiment 2), what cue combination strategy will they use when viewing cylinders of intermediate heights? A fourth issue is the degree of similarity between the learning mechanism underlying the adaptation of cue combination strategies and learning mechanisms evidenced in other perceptual and cognitive domains. For example, is the learning mechanism underlying the adaptation of cue combination strategies similar to other learning mechanisms in the sense that its primary operation is to compute statistics about stimuli and responses? Or is it different in the sense that it is primarily concerned with other sorts of variables and/or non-statistical relationships? The results reported here give a mixed answer. On the one hand, the results suggest that the learning mechanism underlying the adaptation of cue combination strategies is similar to other learning mechanisms in that it is sensitive to statistical relationships among stimuli and responses. On the other hand, the learning mechanism is dissimilar in that it contains a domain-specific bias such that it is sensitive to symmetry transformations.

Acknowledgements

We thank R. Aslin, M. Banks, and two anonymous reviewers for commenting on an earlier draft of this manuscript, and N. Rubin for suggestions regarding the design of experiment 3. We also thank E. Bero, L. O'Brien, A. Pauls and M. Saran for help in conducting the experiments. This work was supported by NIH grant R29-MH54770.

Appendix A

The main body of the text provides an intuitive description of the logistic model. This appendix provides the mathematical details of this model.

A common method in the statistics literature of modeling response data that depends on a set of predictor variables is to use a generalized linear model (McCullagh & Nelder, 1989; Dobson, 1990). Generalized linear models are particularly useful for modeling response data whose distribution is a member of the exponential family of distributions. They consist of a linear transformation followed by a monotonic and differentiable nonlinear transformation such that the output of the model is the expected value of the response variable (for example, which of two displays depicted a cylinder of greater depth) conditioned on a set of predictor variables (for instance, variables characterizing the current stimulus conditions).

Because the results of the test trials are binary response data which can be closely modeled by a Bernoulli distribution (Bernoulli distributions characterize the distribution of events with two possible outcomes, such as the probability that a flipped coin will land heads-up or tails-up), the generalized linear model takes the form of a logistic function (the inverse of the logistic function is the canonical link function for a Bernoulli distribution; see McCullagh & Nelder, 1989, for details). In general, the logistic function has the form

$$y = \frac{1}{1 + \exp\{-f(\mathbf{x})/\tau\}} \tag{4}$$

where y is the probability that the response variable takes the value associated with one outcome ($1 - y$ is the probability that the variable takes the value associated with the other possible outcome), $f(\mathbf{x})$ is a linear function of the vector of predictor variables $\mathbf{x}$, and τ is referred to as a 'temperature' parameter. Note that y always lies between 0 and 1 which is sensible because probabilities must always lie within this range. Also note that the temperature τ scales the logistic function; this function rises more quickly as τ decreases, and less quickly as τ increases.

The logistic model used in this article has the form

$$P(response = M) = \frac{1}{1 + \exp\{-[d^M(t,m) - d^T(t,m)]/\tau\}} \quad (5)$$

where $d^M(t,m)$ and $d^T(t,m)$ are the composite depth percepts based on the texture and motion cues in the displays from sets M and T, respectively. Equation 5 has two free parameters, namely the temperature τ and either the linear coefficient associated with the motion cue, w_M, or the coefficient associated with the texture cue, w_T (recall that $w_T = 1 - w_M$). Maximum likelihood estimates of these parameters were found using gradient ascent on an appropriate Bernoulli likelihood function. This likelihood function gives the joint probability of the response data for a set of test trials:

$$\prod_{i=1}^{n} P(r_i = M)^{r_i}[1 - P(r_i = M)^{1 - r_i}] \quad (6)$$

where $P(r_i = M)$ *is* the probability that the subject selected the set M display on trial i (this quantity is obtained from Eq. (5)), r_i is a binary variable indicating whether the subject selected the set M display ($r_i = 1$) or the set T display ($r_i = 0$) on trial i, and n is the number of test trials.

References

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115–147.

Blake, A., Bülthoff, H. H., & Sheinberg, D. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research*, *33*, 1723–1737.

Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology*, *117*, 161–170.

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

Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America*, *5*, 1749–1758.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1993). Effects of different texture cues on curved surfaces viewed stereoscopically. *Vision Research*, *33*, 827–838.

Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198–216.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth. In W. Epstein, & S. Rogers, *Perception of space and motion*. San Diego: Academic Press.

Dobson, A. J. (1990). *An introduction to generalized linear models*. London: Chapman and Hall.

Dosher, B. A., Sperling, G., & Wurst, S. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, *26*, 973–990.

Hearn, D., & Baker, M. P. (1997). *Computer graphics (C Version)*. Upper Saddle River, NJ: Prentice Hall.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of motion and stereopsis cues. *Vision Research*, *34*, 2259–2275.

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: stereopsis and texture. *Vision Research*, *33*, 813–826.

Knill, D. C. (1998). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, *38*, 2635–2656.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research*, *35*, 389–412.

Maloney, L. T., & Landy, M. S. (1989). *A statistical framework for robust fusion of depth information*. *Visual Communications and Image Processing IV: Proceedings of the SPIE*, *1199*, 1154–1163.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. London: Chapman and Hall.

Nawrot, M., & Blake, R. (1993). On the perceptual identity of dynamic stereopsis and kinetic depth. *Vision Research*, *33*, 1561–1571.

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception and Psychophysics*, *60*, 377–388.

Perotti, V. J., Todd, J. T., & Norman, J. F. (1996). The visual perception of rigid motion from constant flow fields. *Perception and Psychophysics*, *58*, 666–679.

Rogers, B. J., & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *The Quarterly Journal of Experimental Psychology*, *41*, 697–717.

Tittle, J. S., Norman, J. F., Perotti, V. J., & Phillips, F. (1997). The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture and shading. *Perception*, *26*, 147–166.

Turner, J., Braunstein, M. L., & Anderson, G. J. (1997). The relationship between binocular disparity and motion parallax in surface detection. *Perception and Psychophysics*, *59*, 370–380.

Vetter, T., Poggio, T., & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology*, *4*, 18–23.

Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685–2696.