Learning to see random-dot stereograms

Alice J O'Toole
School of Human Development, The University of Texas at Dallas, Richardson, TX 75083-0688, USA
Daniel J Kersten
Department of Psychology, University of Minnesota, Minneapolis, MN 55455, USA
Received 27 July 1990, in revised form 6 June 1991

Abstract. In the present study some specific properties of the learning effects reported for random-dot stereograms are examined. In experiment 1 the retinal position-specific learning effect was reproduced and in a follow-up experiment it was shown that the position specificity of learning can be accounted for by selective visual attention. In experiments 2 and 3 evidence was obtained that suggests that observers can learn, to a certain degree, monocular random-dot patterns and that this learning facilitates the depth percept. This result indicates that the traditional belief that random-dot stereograms are devoid of monocularly recognizable or useful forms should be reconsidered. In the second set of experiments the learning of two binocular surface properties of random-dot stereograms, depth edges and internal depth regions, was investigated. It was shown in experiment 4 that the depth edges of random-dot stereograms are not learned, whereas the results of experiment 5 indicate that the internal depth regions are learned. Finally, in experiment 6 it was shown that depth edges are learned when the internal depth regions of the stereogram are ambiguous. The results are discussed in terms of the importance of the particular type of stimulus used in the learning process and in terms of perceptual learning and attention.

1 Introduction
The perception of depth in random-dot stereograms is a purely cyclopean phenomenon that occurs only when the monocular-eye images are combined to form a single unified percept. Since the development of random-dot stereograms by Julesz in 1960, many aspects of the perception of these stimuli have been examined. One phenomenon peculiar to random-dot stimuli is that they seem to be 'learned'. By learning, we mean simply that there are studies and published observations that demonstrate that the time to perceive depth and figure in random-dot stereograms decreases with repeated observation (Julesz 1971; Ramachandran and Braddick 1973; Frisby and Clatworthy 1975; Ramachandran 1976). Although these observations have long been noted, little is known about what is learned and even less is known about the implications of this learning for experiments and theories of stereopsis. Further, the handful of studies in which these issues have been examined have not always been in agreement, about either the what or the why of the learning. Nonetheless, random-dot stereograms are perhaps the most commonly used stimuli for studying human stereopsis and have recently been used for studying the neurophysiological mechanisms of depth-perception in other primates also (Poggio et al 1985).

1.1 Early observations
Julesz (1971) was the first to discuss the learning of random-dot stereograms. His focus was on why learning occurs. He proposed that the learning effects in the perception of random-dot stereograms resulted from the need to learn an appropriate sequence of vergence movements in order to fuse any given stereogram. It would seem that whatever the cue might be for the vergence movement system, it is likely to be binocular and perhaps related to the outputs of disparity detectors as it is generally claimed that random-dot stereograms are devoid of monocular cues. Saye and Frisby (1975),

however, offer support for the idea that, at least in some cases, monocular aspects of the stimuli can cue the vergence system. They looked at the effects of monocularly conspicuous features in facilitating stereopsis and found that when the disparities were large enough to require vergence movements for fusion, the presence of monocularly conspicuous features facilitated stereopsis. For smaller disparities, they found no significant effect of the monocular features. As an additional observation, Julesz (1971) also suggested informally that observers perceive depth in stereograms more quickly when they are told what to expect.

A handful of other researchers in the seventies focused on what was being learned. This work was motivated by the popularity of the spatial frequency hypothesis in vision [see DeValois and DeValois (1980) for a review], and by the very beginnings of interest in computational models of stereopsis (Julesz 1971; Marr and Poggio 1976). Research in both of these areas was strongly concerned with identifying 'primitives' in the visual system. Stereopsis appeared to be quite a useful system for studying these primitives since, clearly, a major part of the computational stereo problem is to establish correspondence between the elements (whatever they may be) of the left-eye and right-eye inputs.

Ramachandran and Braddick (1973) found orientation-specific learning effects in random-dot stereograms. They showed that when the elements comprising a 'learned' stereogram were oriented at 45°, observers did not transfer well to stereograms depicting the same surface with elements that were perpendicular to those learned. They did, however, find some positive transfer between these orthogonal stimuli. Ramachandran and Braddick (1973) suggested that observers may benefit from the experience of seeing any random-dot stereogram. Another possibility is that observers may benefit from seeing stereograms depicting the same surface. In contrast to monocular primitive learning, this kind of surface learning is of a binocular aspect of the stimulus. Ramachandran and Braddick (1973) pointed out that the effects of the experimental manipulation are likely to be short-term effects. They commented, as Julesz (1971) also noted, that the processing of random-dot stereograms requires the interaction of local and global processes, and it is at the global level that learning effects are likely to occur.

The results of Ramachandran and Braddick (1973) are difficult to reconcile with work done by Mayhew and Frisby (1978) that showed that stereopsis masking is not orientationally tuned. Mayhew and Frisby (1978) used a task similar to the one used by Julesz and Miller (1978) but employing orientation-filtered noise rather than spatial-frequency-filtered noise. They showed no difference in the masking effectiveness of noise with orientations similar to the test elements than with noise of different orientations. Further, Mayhew and Frisby (1981), in a thorough examination of some psychophysical and computational studies of stereopsis, offer even stronger evidence that stereopsis is not orientationally tuned.

1.2 Reconsidering the issues
Several other early observations about learning and stereopsis have been challenged more recently as well. First, there is evidence that learning effects in stereopsis are low-level and perceptual, rather than resulting from the expectations of the observers. Second, these effects have been found to be relatively long-lasting, persisting for weeks or longer. Third, the types of learning effects seen may not be accounted for completely by the learning of appropriate sequences of eye movements. Finally, the importance of primitives in the learning process has been reexamined.

For the first issue, Frisby and Clatworthy (1975) did not offer support for Julesz's (1971) informal observation of the facilitation of depth perception by giving observers information about the surface in a random-dot stereogram. In a very careful study,

they used several conditions which included giving observers a detailed description of the surface, giving information about the depth planes and the object, and finally, showing observers a full scale three-dimensional model of the surface. They found no facilitation of the depth percept in any of their conditions. The study did, however, offer more support for a learning effect in which the time to perceive depth decreased with repeated exposure. Frisby and Clatworthy (1975) also showed that the learning persisted three weeks after the original study. Further testing of the observers was not performed.

Ramachandran (1976) confirmed the learning observations of other researchers and extended them, pointing out two other interesting additions to what is known about the learning of stereograms. These additions offer support for all four of the statements listed above. First, the low-level nature of the learning; Ramachandran (1976) found that learning did not transfer when the stereogram was shifted in position on the retina. He suggested that the learning mechanism was therefore position specific on the retina. Second, the learning in this study, as in Frisby and Clatworthy's (1975) study, persisted for several days. In support of the third issue, Ramachandran (1976) argued that learning of sequences of vergence eye movements, as suggested by Julesz (1971), was not responsible for the learning in his experiment, since observers' asymptotic perception times for the perception of the stereograms were too small to allow for eye movements. He suggested that since the disparities in the stereogram were much greater than Panum's fusional area, it was possible that observers were able to learn "cortical shifts" which were able to replace actual vergence movements. Some interesting neurophysiological hypotheses about cortical shifting have recently been put forth by Anderson and Van Essen (1987). They propose cortical shifter circuits as a possible strategy for solving some computation problems in stereopsis, motion analysis, and directed visual attention.

Ramachandran (1976) also found that learning was not affected by the particular sequence of random dots used in the figure (though this is reported more as an observation than as an experiment). This is somewhat at odds with the results of Ramachandran and Braddick (1973) which demonstrated the existence of orientation specificity for the learning of random-dot stereograms. The Ramachandran (1976) result implies that the forms of surfaces, rather than 'surface markings', may be learned. On the other hand, Ramachandran and Braddick's (1973) study implies that surface markings are learnt. The results of these studies leave open a number of questions about what exactly is learned and how this learning may be interpreted.

2 General methods

The first three experiments offer a replication of the retinal position-specific learning found by Ramachandran (1976) and a reexamination and extension of the issue of random-dot pattern learning. We develop a three-part method for studying learning in random-dot stereograms that incorporates controls for some potential difficulties in previous experiments. In particular, with our method we tried to control for the instability of observer criteria, floor effects that are result from reaction times reaching an asymptote, task-learning effects that may be confused with stimulus learning, and the learning of a general class of stimuli (ie random-dot stereograms) that may be confused with the learning of the specific stimulus tested.

The principles of this method are as follows. First, Ramachandran's (1976) observers were asked simply to indicate when they could clearly see the cyclopean figure. This leaves open the possibility that observers may have used different criteria for "clearly perceiving the figure". Also, these criteria may have changed over time. Since there was no test that depended upon seeing the figure clearly, it is impossible to know if this was the case. We have added a discrimination task so as to be sure

that observers are at least able to see figures clearly enough to discriminate them from an alternative figure.

A second difficulty is that asymptotic perception times in Ramachandran's (1976) study were taken as perception times. Finer control of the perception times can be achieved by varying the stimulus presentation time. We adjusted stimulus presentation times for each observer individually during a preliminary session. Test stimuli could then be presented at this adjusted exposure time. Thus, more subtle learning effects can be detected, even when reaction times are fast enough to cause a floor effect. The procedure is also less likely to be affected by reaction-time delays that only result from the observers noting a difference between the learned and test stimuli and not from a difficulty in perceiving the test stimulus.

Finally, Ramachandran's (1976) argument about the lack of involvement of eye movements in the perception process is critically dependent on observer responses being made in less than the amount of time necessary for eye movements. It is thus important to show that this perception is possible with presentation times less than or equal to 150 ms,[1] because Ramachandran (1976) makes this claim as an observation without presenting data.

2.1 Procedure

The overall procedure consisted of four parts: a practice session, and experimental procedures parts 1, 2, and 3. In the practice session, observers looked at several examples of random-dot stereograms that were not used in the experiments. Part 1 of the experiment was a learning stage in which observers viewed the random-dot stereograms without time constraints. Part 2 was designed to ensure that observers were able to identify the stimuli with perception times of less than 150 ms. This part was also used to adjust stimulus exposure times so that all observers performed at approximately the same level of accuracy (between 70% and 80% correct). In part 3 a procedure similar to part 2 was used, but it was run at the exposure time found for the individual in part 2, and was used to test the learning transfer to the conditions of interest.

The practice session was intended to control for the possibility that initial learning effects were due to the experience of seeing any random-dot stereogram. In the practice session, observers saw a nonius fixation point (shown in figure 1). When fused, this appeared as a single square with a line above and below. When the observers had fused the fixation stimulus, they pressed a button on the response mouse and a random-dot stereogram was presented. The practice stimuli were the capital letters E, H, I, and L. Each appeared with convergent and divergent disparity of 9 min arc. In other respects these figures adhered to the specifications of the experimental stimuli detailed below. Observers were required to identify all letters and to indicate the direction of depth before proceeding in the experiment.

In part 1, observers viewed the fixation point and initiated a trial by pressing a mouse button. One of two random-dot stereograms, presented either in front of or behind the zero disparity plane, appeared on the screen. The observer pressed a mouse button when the figure appeared clearly enough to identify it. The figure disappeared from the screen with this button press, and the observer indicated which of the two figures was presented by pressing the appropriate response button. The second button press (when the observers felt they could identify the figure) was included so that beginning reaction times would be as accurate as possible and would not include task-learning factors such as the observer not remembering which button to press for a given stimulus. The identification of the figure was made on the third

[1] About the time needed to make a vergence eye movement.

button press with no time constraints. This procedure continued for 200 trials, or until the observer was correct on 15 consecutive trials and responded in under 1 s on each.

Part 2 was a two-alternative forced-choice (2AFC) task with both of the stimuli being presented consecutively for 150 ms with an interstimulus interval controlled by the observer. The course of each trial was as follows. The fixation stimulus was presented. The observer initiated the first stimulus presentation by pressing a mouse button. One of the two stimuli was then flashed for 150 ms. A second fixation stimulus appeared. The observer initiated the second stimulus presentation by pressing a mouse button. The second stimulus of the pair was then flashed. A small computer beep prompted for the observer's response. Observers indicated whether the square was the first or second of the two stimuli by pressing the appropriate mouse button.

The percentage of correct responses was calculated after every 10 trials and was adjusted by varying presentation time in 16 ms steps (the smallest controllable unit of time for the apparatus used) to keep observer performance between 70% and 80% correct. The only constraint on the adjustment procedure was that the exposure time was never adjusted to be above 150 ms. The exposure time found after 100 trials was used in part 3 of the study. Observers whose performance was not above chance after 100 trials were excluded from part 3.

The task for part 3 was the same as that for part 2 with two exceptions. First, the exposure time for part 3 remained constant at the adjusted exposure time found for the individual observer in part 2. And second, for testing purposes, half of the stimuli were those learned, and half were new stimuli. The number of test stimuli was limited by the fact that large numbers of presentations might allow for learning these new stimuli, which might have decreased the size of the effect. We used 40 trials in this third part, 20 presentations of the learned stimulus and 20 presentations of the new stimulus.
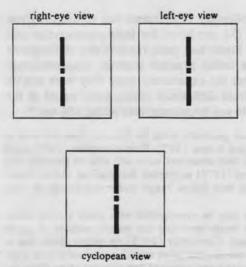

Figure 1. The top part of the figure shows the left-eye and right-eye components of the nonius fixation point presented to observers before each stimulus. The bottom part of the figure shows how the fixation point appears when fused.

2.2 Stimulus design

Except for the practice session with the letter stimuli we used simple stimuli consisting of either a diamond or a square embedded in a background square for which we found the initial perception times varied between 1 and 24 s. Practice-session times

varied to a greater degree so it seems that observers do benefit somewhat from the experience of seeing any random-dot stereogram. The diamond and square were actually identical figures which were rotated versions of each other. These are pictured in figure 2. Care was taken to assure that the area of the two figures was identical. Each background square subtended 3 deg visual angle horizontally and 2 deg vertically, and the individual dots subtended 4.5 min vertically and 3 min horizontally. Unless otherwise noted, the disparity of the figures with respect to the zero-disparity background was ±0.15 deg (9 min arc)[2].

The base random-dot pattern was generated randomly for each observer but was the same for both the diamond and the square pattern for any given observer. This was to assure that discrimination of the diamond and the square could not be based on any recognition of the random-dot pattern. The distance between the left and right random-dot presentations was adjusted individually to the interpupillary distance of the observer.
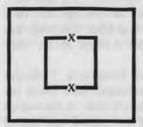

Figure 2. This figure shows the diamond and square random-dot stereograms used as stimuli for the experiments. The Xs show the possible learning/test positions of the center of the nonius fixation point for the stimuli in experiment 1.

2.3 Observers

Unless otherwise indicated, in each of the following experiments ten volunteers from the Brown University community were paid ($4 per hour) for their participation and were naive as to the purpose of the study. Some had seen random-dot stereograms before, but all observers participated in the initial practice session. Approximately one third of the observers were excluded from the experiment since they were unable to perform the task.[3] This resulted either from difficulties seeing some or all of the random-dot stereograms, or from an inability to see the stereograms within 150 ms.[4]

[2] This figure is just above the estimate of 6 min arc generally given for Panum's fusional area in the central degree or so of visual space (Fender and Julesz 1967). Ramachandran (1976) used much larger disparities (up to 30 min) and found that observers were still able to perceive the stereogram in under 150 ms. Thus Ramachandran (1976) suggested that cortical shifts, rather than actual vergence shifts, bring the stereogram into fusion range under conditions of very short exposure.

[3] Although this rejection rate may seem high, it may be comparable with other studies when the task difficulties are taken into account. The study involving the largest number of naive observers cited in this paper is that of Frisby and Clatworthy (1975), in which there was a rejection rate of 24.2%. Given that their task did not involve short stimulus-exposures and large disparities, and that the present experiments were long and required full concentration throughout, the additional rejection of approximately 10% seen here is perhaps not surprising. Further, estimates (Richards 1971) of stereo deficits in the general population indicate that approximately 30% of the population have some stereo deficiency.

[4] Visual screening of the observers showed that the former type of difficulty was generally associated with inability to see depth or with lateral and vertical phorias, and the latter type was associated with lack of good acuity in one or both eyes. Details of the kinds of stereo deficits that have been documented, as well as their frequency of occurrence in the general population, may be found in Richards (1970, 1971).

2.4 Apparatus

Observers viewed the stereograms through a stereoscope-like apparatus in which the left and right visual field were separated. The stimuli appeared on a Tektronix graphics screen, and all experimental trials were controlled by an IBM PC-XT. The viewing distance was approximately 57 cm, and observers viewed the screen through convex lenses to make the screen appear at infinity, thus making the lines of sight parallel. Occasionally additional lens corrections were used for observers with weak or uneven visual acuity.

3 Experiment 1: Retinal position specificity of learning

In experiment 1 the position-specific learning of the random-dot patterns was tested. Observers learned the random-dot stereograms in one of two positions with respect to the fixation point. In part 3 the observers were tested with half of the stimuli presented in this original position relative to the fixation point (Old) and with half of the stimuli in a new position with respect to the fixation point (New). Note that the fixation point remained constant, and it was the random-dot stereogram that moved. The position of the learning/test stimuli with respect to the fixation point was counterbalanced across observers with half of the observers trained with the stimulus displaced 20 min vertically above the fixation point and half with the stimulus displaced 20 min vertically below the fixation point. These relative positions are marked on figure 2 with Xs. The Old and New positions were vertically aligned so that there was no possibility of the effect being due to the differences in stereo perception with horizontal eccentricity or due to the occasional stereo-anomalous observer who showed deficits in only one hemisphere (Richards 1970, 1971). Further, they were aligned to be in symmetrical positions of the stimuli so that the information content of the area of the figure around the Old and New relative fixation points was identical.

3.1 Results

3.1.1 *Part 1.* As expected, observers showed a decrease in the latency for response over the course of the trials in part 1. This is shown in figure 3, where the latencies for response to the four individual stimuli (diamonds crossed, diamonds uncrossed, squares crossed, squares uncrossed) are averaged over the ten observers in experiment 1 and over blocks of 10 trials. Overall, observers performed in an average of 113.8 trials with the range varying between 17 and 200 trials.

3.1.2 *Part 2.* The exposure time of the stimulus was adjusted individually for observers over the course of the 100 trials in this part of the experiment. Figure 4 shows the average for the ten observers for adjusted exposure time over blocks of 10 trials. The average adjusted exposure time after the 100 trials for the ten observers was 105 ms.

3.1.3 *Part 3.* A reliable difference between performance on the stimuli in the learned retinal position (Old) and those in the shifted (New) position was seen in part 3. The means for the Old and New conditions and the specifics of the hypothesis testing appear in row 1 of table 1. A repeated-measures analysis of variance (ANOVA) was done on the arc sine-transformed (Winer 1971) percentage correct scores from observers in the AFC task of part 3. Percentage correct scores could be used because each observer in part 3 performed the experiment at the exposure time found for that individual in part 2. This analysis was performed in the subsequent experiments as well.

Although in this experiment Ramachandran's (1976) finding that learning in random-dot stereograms is position specific on the retinae was replicated, the basis of this effect is not clear from this experiment. We address this issue in detail in experiment 7 where we will evaluate the role of attention and perceptual learning in the effect.
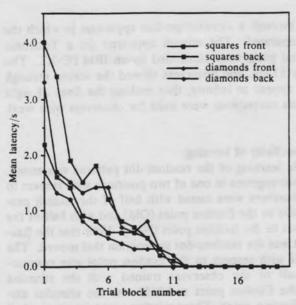
Figure 3. This figure shows the response latencies in s averaged over the ten subjects and blocks of 10 trials in experiment 1 for each learning stimulus separately. The standard deviations in the first few blocks of trials ranged between 3 and 5 s, decreasing in the final trial blocks to a response-time floor, with most of the observers meeting the criterion before the final block of trials.
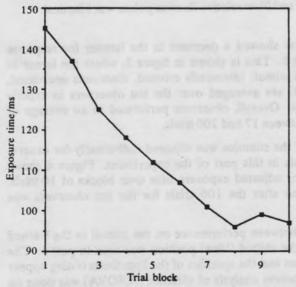

Figure 4. This figure shows the adjusted exposure times for the blocks of 10 trials in part 2, averaged over the ten observers in experiment 1. The standard deviation of the first block of trials across observers was 0 because all observers began part 2 at a fixed exposure time. The standard deviation of exposure times increased across trial blocks to 44 ms by the final block of trials.

Table 1. The percentage correct scores for the Old and New conditions are shown for each experiment. The results of ANOVA for the difference between the Old and New conditions are also shown.

Experiment	Percentage correct		F	p
	Old	New		
1: Position specificity	83.0	64.0	10.8	<0.01
2: Monocular dot learning	81.5	72.5	7.51	<0.025
3: Negative monocular dot learning	81.5	79.0	<1	ns
4: Contour learning	87.0	87.5	<1	ns
5: Disparity learning	81.5	72.5	10.2	<0.01
6: Contour learning with ambiguous depth regions	79.0	68.5	7.2	<0.05

4 Experiment 2: Monocular dot-pattern learning

In experiment 2 we attempted to replicate Ramachandran's (1976) finding that changing the particular pattern of random dots learned does not affect the subject performance. The method was similar to the one used in experiment 1 except that observers learned stimuli with a single pattern of random dots and were tested in part 3 with 50% of stimuli with the pattern learned and 50% with a second randomly generated pattern [5].

4.1 *Results and discussion*

The pattern of results in parts 1 and 2 of this experiment, as well as that of the remaining experiments, was identical and will not be detailed further. As is shown in row 2 of table 1, a significant difference between the performance on the learned stimuli (Old) and the New stimuli was observed in part 3. Thus we did find a pattern-specific effect, though the size of the effect was smaller than the position-specific effect. This finding does not replicate Ramachandran's (1976) claim that changing the pattern of random dots does not affect the performance of observers. Perhaps the difference between Ramachandran's (1976) results and our results is related to the differences in the methods used. Ramachandran used only reaction-time data in his task. It is possible that the observers may have been at a floor for reaction times that did not show this effect. The fact that the size of the effect we found here is smaller (11%) than the size of the effect we found for the change of retinal position (19%) might explain why Ramachandran's method showed the retinal position effect but not the random-dot effect.

The significance of this learning for the stereo process is difficult to discern. That observers are sensitive to the pattern of random dots that portrays a figure in depth suggests that the learning of the random dots can in fact aid in the matching and fusion processes. Thus, our observers were able to match and fuse the stimuli portrayed with the learned pattern of random dots faster than they were able to perceive stimuli otherwise identical with respect to figure, disparity, and position on the retinae. Our results suggest, contrary to previous thought, that random dots (or the patterns they form) can in some cases serve as monocular cues for the stereo process in the sense that they may be learned, and, once learned, may be matched and fused more quickly than unlearned or rather 'unfamiliar' random dots.

5 Experiment 3: Negative monocular dot-pattern learning

Before we consider the implications of this random-dot-pattern learning, we wished to see if it would be possible to manipulate the size of the effect by changing the pattern of dots to different degrees. The reason for this follow-up study was that although it

[5] Dot patterns were created randomly for each observer.

seemed clear that the pattern of the random dots had an effect, it was equally clear that it did not completely disrupt observers' performance on the task. One reason for this might have been that as the dot patterns in this experiment were randomly-created patterns in which 50% of the dots were dark and 50% light, the New patterns would have had approximately 50% of the dots in common with the Old pattern. This 50% overlap might have been sufficient to carry over some of the learning. We wanted to see if the size of the effect would change contingent on the degree of overlap between the Old and the New dot patterns. Following parts 1 and 2, which were identical to the first two experiments, we tested two conditions in part 3. In the first condition, 5% of the dots in the random-dot stereogram were changed at random to create the second pattern. For the second condition, 95% of the dots were changed at random to make the New pattern. (Note this is nearly a negative of the Old pattern.)

5.1 Results and discussion
We found no difference between the learning transfer to the 95% common stimuli and the 5% common stimuli. The details of the analysis appear in row 3 of table 1. The fact that near-negatives of the original pattern did not affect observer performance suggests even more strongly than the results of the previous experiment that the monocular forms of the dot patterns were learned and were used for fusing the stimuli to obtain the cyclopean percepts. These monocular forms would be nearly completely preserved in the 95% reversed patterns, despite the fact that the intensity values at nearly all locations would be changed. The results also indicate that individual dots themselves were not learned in that a 50% change of dots produced less learning transfer than did a 95% change.

6 Experiment 4: Contour learning
There is an alternative to the conclusion that the monocular forms present in the dot patterns are used for fusing binocular stimuli. It is possible that in addition to (or instead of) the surface markings, some aspects of the surface properties are also learned. The results of experiments 2 and 3 could be explained if some aspects of the surface properties are preserved with a negative transformaton of the dot pattern but not with a random switch of 50% of the dots. The difference between surface properties and the kinds of monocular cues we have been looking at is that surface properties are primarily binocular cues. For example, depth edges and internal depth regions are binocular properties in that they are only visible in random-dot stereograms when the left and right images are fused.

The results of experiments 2 and 3 could be explained by assuming the great importance in the stereo process of perceiving the shapes of depth edges. Since the depth edges of random-dot stereograms are formed with random dots, they are, in general, slightly jagged (unless the stereogram is of very high resolution). If these depth edges are learned, it would explain why observers did as well on the near-negatives as on the near-positives, but failed to perform as well when the dot patterns were changed by 50%. Specifically, the particular shape of the presumably-learned depth discontinuity was mostly preserved in the near-positive and near-negative cases but not in the 50% case.

Some recent papers have stressed the relative importance of depth discontinuities for stereopsis. Mitchison and McKee (1985) showed that a stereoscopic percept does not always depend on a discrete matching of the features of the left and right eyes. Rather, special importance may be given to the end points or to discontinuities with small disparities (less than 5 – 7 min arc). They used repeating-dot patterns that were ambiguous with respect to depth. At the end points, the leftmost dot in the left-eye image and rightmost dot in the right-eye image were displaced inward a fraction of

the interdot spacing. Mitchison and McKee (1985) found that when the interdot spacing was less than 6 min arc, the depth percept could not be accounted for by any pattern of discrete matches. Observers saw a plane tilted in depth. This percept can be explained by the discrete matching of the end points of the pattern, and by the interpolation of all of the intervening points between the end point matches. Mitchison and McKee (1985) argue that their results support a model of stereo in which disparity discontinuities are detected.

Mitchison and McKee (1987) extended their findings to show that the interpolated states are transient. The change from interpolated percepts to discrete matches takes 2 s or more. The finding that the interpolation effect is transient suggests even more strongly that locating depth discontinuities is a priority for the stereo system.

In experiment 4 the possibility was tested that the reason for observers' ability to transfer learning to new patterns that were near-positives and near-negatives of the originals, despite their reduced ability to transfer to patterns with 50% of the dots in common, resulted from learning the shape of the depth discontinuity. The procedures for parts 1 and 2 were identical to those in the previous experiments. In part 3, however, the Old patterns were identical to those learned and the New patterns consisted of jagged-edged versions of the diamond and square. These patterns are shown in figure 5. It should be noted that these jagged versions are base patterns within which depth-edge dots are randomly perturbed. Thus, the depth edges should be much more jagged than the learned 'straight' diamond and square patterns.
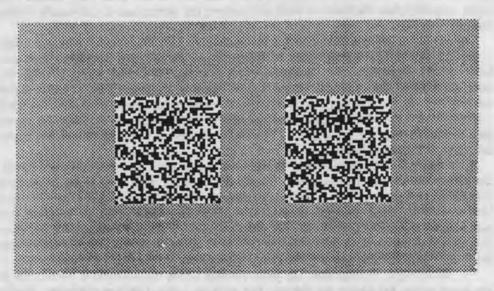

Figure 5. This figure shows an example of the jagged square random-dot stereograms used as stimuli for experiment 4.

6.1 *Results*

As shown in row 4 of table 1, the results of this experiment showed no difference between the Old (straight) and New (jagged) conditions. Thus, learning of the binocular depth contours can not explain the results of experiments 2 and 3.

7 Experiment 5: Disparity learning

Experiment 5 was carried out to see if the internal depth region of random-dot stereograms, another binocular surface property, is learned. Parts 1 and 2 of the study proceeded as in previous experiments. In part 3, however, observers were

presented with half of the stimuli at the depth planes that they had learned (both crossed and uncrossed) and half at a new depth plane (again both crossed and uncrossed). To control for past findings of differences in the detectability of different disparities, twenty observers were divided randomly into two conditions, a larger and a smaller disparity condition. Stimuli appeared at plus and minus 13.5 min arc for the larger disparity condition, and at plus and minus 4.5 min arc for the smaller disparity condition. For both conditions, Old stimuli were trained at plus and minus 9 min arc as before.[6]

7.1 Results and discussion

The results of this study showed a significant difference between the performance on the Old stimuli and on the New stimuli (as shown in row 5 of table 1). The difference between the larger and smaller disparity conditions was not significant ($F_{1,18} = 1.31$), nor was the interaction between the Old/New factor and the size of the disparity factor ($F_{1,18} = 1.67$). Despite the lack of a significant difference between the large-disparity and small-disparity condition, there was a noticeable difference between the percentage correct for the Old stimuli for the two conditions (87% for the small-disparity condition and 77.5% for the large-disparity condition). Since the Old conditions are, in principle, equivalent for the two conditions, these results suggest either a subject-sampling fluke for the conditions or some disruptive effect of the large-disparity stimuli on fusing the already-learned Old stimuli. Erkelens (1988) provides evidence for such a disruptive effect. He showed that presenting stimuli with larger disparity than the fusional limit can hamper the perception of previously fusible stimuli.

The results of experiments 4 and 5 are consistent with the well-documented effect of greater sensitivity to depth than to figure for random-dot stereograms (Over and Long 1973; Harwerth and Rawlings 1975, 1977). Over and Long (1973), for example, showed that depth is visible before figure in the perception of random-dot stereograms. In their study, observers viewed random-dot stereo pairs and reported whether or not the stimulus contained any figural information and whether the stereo pair contained near or far depth. The results showed a large advantage for depth detection over figure detection.

Since figure is defined primarily by the contours or edge discontinuities of objects, the interpolation and stereocapture results are at odds with the results suggesting depth-first perception. This is especially true in view of the fact that Mitchison and McKee (1987) found that the interpolation effect is transient, which suggests that matching of edges is a priority for the stereo system. From a computational perspective, models such as Marr and Poggio's (1976) cooperative model would predict that depth continuities emerge earlier and with greater certainty than depth discontinuities.

For random-dot stereograms, the most reliable depth information is found, not at the depth edges, but within the figures. For surfaces without surface markings, assuming that the intensity of edges are primitives, the reverse is the case. Mitchison and McKee's (1985) stimuli are extreme cases of stimuli for which the depth edges provide the most unambiguous information. Thus, a major difference between these stimuli and random-dot stereograms that might account for the opposing depth-first/figure-first results is the fact that Mitchison and McKee's stimuli are inherently ambiguous in the internal depth regions. That is, there is no real global solution to the stereogram in the cooperative sense. In contrast, random-dot stereograms are most ambiguous at the depth edges.

---

[6] The values of 4.5, 9.0, and 13.5 min were simply plus and minus one, two, and three pixels, respectively.

8 Experiment 6: Contour learning with ambiguous internal depth regions

Given the different kinds of stimuli that were used in the depth-first and figure-first studies, we wondered whether we could create a random-dot stimulus that might be ambiguous enough in the central depth regions to show learning of the depth edges. If learning of the depth discontinuities occurs here, it would indicate that the stimulus is somewhat able to dictate the procedure that the stereo system uses. The background of the figure was created as usual, but the random-dot pattern for the part of the figure in depth was uncorrelated. Figure 6 shows an example. Observers readily saw this figure in depth despite the fact that point-by-point matching was impossible in the internal depth regions. The depth percept was a separation between a background of perfectly correlated left-eye and right-eye dots and a disparity-adjusted region of uncorrelated dots. The depth edges, and hence the figure, were determined by this separation.

Observers reported that they saw these stimuli clearly in depth, with relatively sharp diamond and square outlines and with the internal depth region somewhat blurred. All observers who were able to see regular random-dot stereograms were able to see these stimuli. Experiment 6 was a repeat of experiment 4 but with these depth-region-ambiguous figures. Observers learned depth-region-ambiguous diamonds and squares in parts 1 and 2, and were tested in part 3 with these same stimuli on half of the trials and with jagged depth-region-ambiguous stimuli on the other half of the trials.

Contrary to the results of experiment 4, the results of this experiment revealed a significant difference between performance on the Old stimuli and the New stimuli. Detailed results are shown in row 6 of table 1. Thus, observers learned something of the contours in this case. The results of this experiment indicate simply that the type of stimulus used dictates to a certain degree what observers learn.
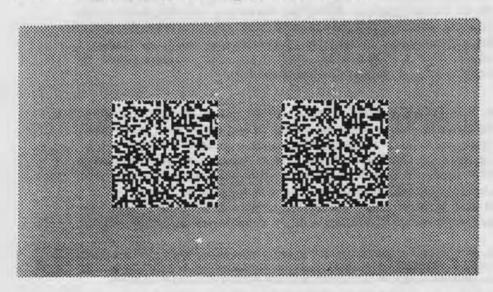

Figure 6. This figure shows an example of the depth-region-ambiguous, random-dot stereograms use as stimuli in experiment 6.

9 Experiment 7: Spatial attention and position specificity

In the handful of papers on the learning effects for random-dot stereograms, the role of simple perceptual processes, such as attention and perceptual learning, has not been explored. Attention and perceptual learning are defined broadly in the

literature, and there is some overlap between the two definitions. Frequently, the methods for inducing perceptual learning and attention are very similar, both involving repeated presentations of perceptual stimuli and measuring improvements in performance.[7] Thus, exclusive attribution of a given effect to one or the other of these categories may not always be appropriate.

Of the typically documented attention effects, selective visual attention is most relevant. Within this category two subcategories are distinguished. If attention is invoked to account for the results of the experiments in which advantages were seen for the Old over the New stimuli (ie position specificity in experiment 1, monocular-dot learning in experiment 2, disparity learning in experiment 5, and contour learning with ambiguous depth regions in experiment 6), it is evident that the object of attention is qualitatively different for the different experiments. For experiment 1, an attention-based account of the results involves attention to the probable spatial location of the stimulus, rather than attention to any aspect of the stimulus itself. In contrast, an attention-based account of experiments 2, 5, and 6 would need to rely on attention to a (some) specific aspect(s) of the stimulus itself. Increased sensitivity to some nongeneral aspect(s) of a stimulus after repeated trials is perhaps best described as perceptual learning.[8] We consider experiment 1 first and then discuss experiments 2, 4, and 6 as a group.

9.1 An attention-based account of retinal position specificity

The results of experiment 1 are consistent with selective spatial attention of the kind demonstrated in Bashinski and Bacharach (1980). They show increased perceptual sensitivity to an area of the retina, based on signal probability. Likewise, it is possible that observers in experiment 1 began to attend an area of the retina selectively during the learning trials due to the high signal probability associated with that area. After the learning trials, therefore, observers should be better at detecting any stimulus presented on the retina in the learned position. Perceptual learning or attention involving an (some) aspect(s) of the stimulus itself, on the other hand, would indicate that better performance should be observed for only the stimulus learned. Since the data presented in experiment 1 do not address this question explicitly, a simple follow up experiment was performed to decide between these hypotheses.

9.2 Method

9.2.1 *Procedure*. We presented three different stimuli during the learning phase. Stimulus 1, a left-pointing arrow, appeared slightly above the fixation point (position 1) on all trials. Stimulus 2, a right-pointing arrow, appeared slightly below the fixation point (position 2) on all trials.[9] The control stimulus, an upward-pointing arrow, appeared in position 1 on half of the trials and in position 2 on the other half of the trials. In test trials, all three stimuli were tested in both positions.

If perceptual learning of an (some) aspect(s) of the individual stimuli can account for the position-specific effects of learning, then we would expect to see a detection

---

[7] Oddly enough, the same type of method is used to induce adaptation effects. The difference is simply that adaptation produces a decrease in sensitivity to some aspect of the repeatedly presented stimulus, whereas perceptual learning and attention produce increases in sensitivity.

[8] Some caution in categorizing the effect as perceptual learning, as opposed to attention, is in order. Very similar effects have been reported in the literature as either perceptual learning or as attention. Sorting out the uses of these terms in the literature is beyond the scope of this paper. We try simply to be clear about the type of effect we mean without forcing a definitive categorization. The interested reader is referred to the paper of LaBerge (1973), in which attention and perceptual learning processes and their interaction with the familiarity of the stimulus are considered.

[9] Unless otherwise noted, stimuli adhere to the specifications given for the other experiments.

advantage for stimulus 1 in position 1, a detection advantage for stimulus 2 in position 2, and no difference for the control stimulus, which appeared equally often in positions 1 and 2. On the other hand, a spatial-attention account would predict no stimulus-by-position interaction. This is because positions 1 and 2 were used equally often across all learning trials.

The experiment involved three parts, as before, with the following change. A 2AFC method was not possible in this experiment because it was impossible to create pairs of stimuli without compromising the position-specific nature of the stimulus presentation. Thus, an identification task was used. Observers identified the stimulus on each trial by pressing the appropriate mouse button. Signal detection theory analysis was used to control for any response bias that might have resulted in an advantage for the stimulus learned in a given position.

9.2.2 *Observers*. Seven volunteer observers from the University of Texas at Dallas participated in the experiment for a credit in a psychology course.

9.3 *Results*
The results of the study were clear and showed no trends in the direction predicted by perceptual learning. A repeated-measures ANOVA performed on the d' scores for within-group factors of position and stimulus, showed no significant effects. Thus, observers were equally good at detecting all stimuli in both of the practiced positions. Selective spatial attention on the basis of signal probability can therefore account for the results in experiment 1. This explanation would also seem to be consistent with Ramachandran's (1976) results that demonstrate position specificity of learning on the retina. Though, Ramachandran (1976) used fewer practice trials, an attention-based account of his results cannot be ruled out as he tested only one position on the retina across all of the learning trials.

10 Attention to, versus perceptual learning of, stimulus-specific properties
The results of the other experiments that resulted in an advantage for the learned stimulus would seem to rely on attention to or perceptual learning of a (some) specific aspect(s) of the stimulus itself rather than a general increase in the sensitivity of one area of the retina. In experiment 2 (monocular dot learning) and in experiment 6 (contour learning with ambiguous depth regions), attention to spatial locations where distinguishing features for the stimuli might be located could not explain the results because these did not change between learning and test trials. The case for experiment 5, disparity learning, is more complicated and may involve both learning of some binocular aspects of some binocular aspects of the stimulus itself, as well as learning of the monocular offset locations in the two eyes for different disparities. This explanation, though, would need to posit simultaneous spatial attention to the convergent and divergent offsets of the learned disparity which varied from trial to trial. Further, more general attention-based effects would have been expected to operate in all experiments showing an advantage for Old compared with New stimuli. We did not find any advantage in the results of experiment 3 (the negative monocular dot manipulation) or of experiment 4 (the contour study).

11 General summary and discussion
The results of experiment 1 and its follow-up show that the position-specific 'learning' of random-dot stereograms can be accounted for by selective spatial attention based on the probable location of the stimulus. The results of experiments 2 and 3 indicate that some, though not all, aspects of the random-dot markings can be learned. Relating this finding to the previous literature helps clarify some of the earlier findings and shows that there may be less disagreement than originally thought. First, the results

of Ramachandran and Braddick's (1973) study, which found that the learning of oriented primitives did not transfer completely to orthogonally oriented primitives, indicate that some aspects of the primitives are learned. However, this does not necessarily mean that the orientations per se were learned. Data from Mayhew and Frisby (1978, 1981) show that stereopsis is not bound at any basic level of analysis by the orientations of the surface markings. We tested the pattern of random dots and found that complete learning transfer did not occur when these were changed. Thus, the results of Ramachandran and Braddick's (1973) study may indicate that learning of some general properties of the surface markings occurs, rather than learning of orientation specifically.

The results of experiment 5 show that the binocular property of internal depth regions in random-dot stereograms was learned. It is likely that some of the transfer observed by Ramachandran and Braddick (1973) on their orthogonally oriented primitives and by ourselves in experiment 2 occurred because the surface portrayed by the random-dot patterns was held constant throughout the experiment.

In contrast, in experiment 4 it was shown that for the binocular property of depth edges, observers transferred completely between straight-edged and jagged-edged stimuli. This was not the case for the depth-region-ambiguous stimuli of experiment 6, in which a deficit of figure discrimination accuracy was observed when the depth edges were made jagged. This deficit can be explained by considering the differences between the stimuli. For standard random-dot stereograms, the most reliable depth information is found in the internal depth regions, whereas for our depth-region-ambiguous stereograms, the most reliable depth information is found at the depth edges.

12 Clarifying the issues
The major claim made about random-dot stereograms is that since they do not contain monocularly conspicuous or recognizable forms to be matched, they provide a phenomenological proof that monocular form-recognition is not necessary for stereopsis. The argument is often further extended to imply that monocular form-recognition does not precede stereopsis. The present results indicate that observers can become familiar with the dot patterns through repeated exposure and can then perceive depth in these stimuli faster than in otherwise-identical stimuli with a different pattern of random dots. If we accept that random-dot stereograms do not contain monocular forms, we are hard-pressed to explain why changing the pattern of random dots disrupts performance on the task.

Second, the assumption that stereopsis precedes form recognition is also problematic. One implication of the present study is that form recognition in the monocular stimuli may occur with, or prior to, stereopsis, and this may be used to facilitate the matching process. The function of the learning, then, might be to make the dot patterns as familiar as possible so that a more normal course of stereo processing can be employed. This might be equivalent to making the clusters of random dots familiar enough to be easily recognized and matched As learning does not seem to occur for natural depth stimuli (perhaps perception is already so fast that learning could not be noted) it is likely that the reason for its occurrence (or at least its exaggerated nature) in these studies has to do with the artificial nature of random-dot stereograms. These stimuli present challenges to the stereo system that must be dealt with by using procedures that may not follow the course of normal stereo processing. Therefore, there is a need to exercise caution in generalizing the results of experiments with random-dot stereograms to normal stereo processes.

References

Anderson C H, Van Essen D C, 1987 "Shifter circuits: A computational strategy for dynamic aspects of visual processing" *Proceedings of the National Academy of Sciences* **84** 6297–6301

Bashinski H S, Bacharach V B, 1980 "Enhancement of perceptual sensitivity as the result of selectivity attending to spatial locations" *Perception & Psychophysics* **28** 241–248

DeValois R L, DeValois K K, 1980 "Spatial vision" *Annual Review of Psychology* **31** 309–341

Erkelens C J, 1988 "Fusional limits for a large random-dot stereogram" *Vision Research* **28** 345–353

Fender D, Julesz B, 1967 "Extension of Panum's fusional area in binocularly stabilized vision" *Journal of the Optical Society of America* **57** 819–830

Frisby J P, Clatworthy J L, 1975 "Learning to see complex random-dot stereograms" *Perception* **4** 173–178

Harwerth R S, Rawlings S C, 1975 "Pattern and depth discrimination from random dot stereograms" *American Journal of Optometry and Physiological Optics* **52** 248–257

Harwerth R S, Rawlings S C, 1977 "Viewing time and stereoscopic threshold with random-dot" *American Journal of Optometry Physiological Optics* **54** 452–457

Julesz B, 1971 *Foundations of Cyclopean Perception* (Chicago, IL: University of Chicago Press)

Julesz B, Miller J E, 1978 "Independent spatial-frequency-tuned channels in binocular fusion and rivalry" *Perception* **4** 125–143

LaBerge D, 1973 "Attention and the measurement of perceptual learning" *Memory and Cognition* **1** 268–276

Marr D, Poggio T, 1976 "Cooperative computation of stereo disparity" *Science* **194** 283–287

Mayhew J E, Frisby J P, 1978 "Stereopsis masking in humans is not orientationally tuned" *Perception* **7** 431–436

Mayhew J E, Frisby J P, 1981 "Psychophysical and computational studies towards a theory of human stereopsis" *Artificial Intelligence* **17** 349–385

Mitchison G J, McKee S P, 1985 "Interpolation in stereoscopic matching" *Nature (London)* **315** 402–404

Mitchison G J, McKee S P, 1987 "Interpolation and the detection of fine structure in stereopsis" *Vision Research* **27** 285–294

Over R, Long N, 1973 "Depth is visible before figure in stereoscopic perception of random-dot patterns" *Vision Research* **13** 1207–1209

Poggio G F, Motter B C, Squatrito S, Trotter Y, 1985 "Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms" *Vision Research* **25** 397–406

Ramachandran V S, 1976 "Learning-like phenomena in stereopsis" *Nature (London)* **262** 392–394

Ramachandran V S, Braddick O, 1973 "Orientation-specific learning in stereopsis" *Perception* **2** 371–376

Richards W, 1970 "Stereopsis and stereoblindness" *Experimental Brain Research* **10** 380–388

Richards W, 1971 "Anomalous stereoscopic depth perception" *Journal of the Optical Society of America* **61** 410–414

Saye A, Frisby J P, 1975 "The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms" *Perception* **4** 159–171

Winer B J, 1971 *Statistical Principles in Experimental Design* (New York: McGraw-Hill)