Spatial Vision, Vol. 5, No. 2, pp. 81–100 (1990)

Some results on translation invariance in the human visual system

TATJANA A. NAZIR and J. KEVIN O'REGAN

Groupe Regard, Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes, EPHE, EHESS, 28 rue Serpente, 75006 Paris, France

Received 31 July 1989; revised 19 March 1990; accepted 4 April 1990

Abstract—Four experiments were conducted to study the nature of visual translation invariance in humans. In all the experiments, subjects were trained to discriminate between a previously unknown target and two non-target distractors presented at a fixed retinal location to one side of the fixation point. In a subsequent test phase, this performance was compared with the performance when the patterns were presented either centrally at the fixation point or at a location on the other side of the fixation point, opposite to the location where the patterns were learned, but where acuity was identical to what it was at the learned location. Two different experimental paradigms were used. One used an eye movement control device (Experiment 1) to ensure the eye could not move relative to the patterns to be learned. In the other three experiments, presentation duration of the patterns was restricted to a short enough period to preclude eye movements. During the training period in Experiments 1 and 2, presentation location of the patterns was centered at 2.4 deg in the periphery, whereas in Experiments 3 and 4 presentation eccentricity was reduced to 0.86 and 0.49 deg. In all four experiments performance dropped when the pattern had to be recognized at new test positions. This result suggests that the visual system does not apply a global transposition transformation to the retinal image to compensate for translations. We propose that, instead, it decomposes the image into simple features which themselves are more-or-less translation invariant. If in a given task, patterns can be discriminated using these simple features, then translation invariance will occur. If not, then translation invariance will fail or be incomplete.

INTRODUCTION

How do humans recognize objects despite changes in size, orientation and position? In artificial image recognition systems, the problem is often solved by building global transformation operators which normalize an image so that it is in a standard form before applying recognition algorithms. One way of doing this is to use a preprocessing algorithm which provides an invariant representation of the image. This can be done, for example, by taking a log polar transform, followed by a two-dimensional Fourier transform (e.g., Schwarz, 1977, 1981; Cavanagh, 1978, 1985) or other translation invariant transform (Burkhardt and Muller, 1980; Reitboeck and Altmann, 1984), or by representing the image in terms of an autocorrelation function (Uttal, 1975; Gerrissen, 1982; Kroese, 1985).

At first sight, there does appear to exist behavioural evidence that some kind of renormalization might be done in the human visual system. For example, classic studies on mental rotation (see e.g., Shepard and Metzler, 1971; Cooper and Shepard, 1973; Cooper, 1975) show that the time it takes to compare a rotated pattern (such as a random polygon) to an unrotated version increases linearly as a function of angular disparity from upright. Related results were found while using displays of limited duration (Dearborn, 1899; Arnoult, 1954; Rock, 1973; Foster, 1978; Kahn and Foster, 1981), showing that accuracy in discriminating patterns falls off with rotation angle up to about 90 deg. Similar relations for size transformations were obtained by Bundesen and Larsen (1975). However, while these studies point to the appealing suggestion that the renormalization process depends on the amplitude of the transformation, other studies, using different tasks and stimuli, found there is no penalty as a function of the amplitude of the transformation. Thus, Corballis *et al.* (1978) and Eley (1982) found that using letters or letter-like forms, recognition time is not affected by degree of rotation. Kubovy and Podgorny (1981), using random polygons, found that recognition was unaffected by the extent of a size change. Further confusion arises when invariance to more complex shapes is considered. Cursive handwriting is almost impossible to read upside down (Rock, 1973, 1984). Faces look 'different' when seen upside down, and the ability to recognize facial expressions is poor, as illustrated by Thompson's (1980) famous 'Thatcher illusion'. The horrible expression of the face is only noticed when it is seen rightside up. Difficulties in recognizing complex stimuli arise even for size transformations. Kolers *et al.* (1985) showed that recognition accuracy drops when a previously learned face is shown in a different size (Kolers *et al.*, 1985).

It therefore appears that the behavioural evidence in favour of a global transformation operation done on the retinal image is not at all clear-cut. In some cases transformation of the stimulus has no effect on performance (Corbalis, 1978; Kubovy and Podgorny, 1981; Eley, 1982), in other cases a linear relation was found between the drop of performance and the magnitude of the modification (Dearborn, 1899; Arnoult, 1954; Shepard and Metzler, 1971; Cooper and Shepard, 1973; Rock, 1973; Bundesen and Larsen, 1975; Cooper 1975; Foster, 1978; Kahn and Foster, 1981) and in some cases, when it comes to more complex stimuli like facial expressions (Thompson, 1980), recognition seems to fail completely. The ease of recognition of transformed images appears to depend on the nature of the image and on its complexity: at least as far as rotation and size transformations are concerned the visual system seems only able to rotate or resize simple and overlearned images correctly.

The above evidence has been concerned with the questions of rotation and size invariance. But what is the situation in regard to an even simpler kind of geometric transformation, namely translation? Interestingly, it seems that apart from Kahn and Foster (1981) and Foster and Kahn (1985), no one has seriously considered to what extent human vision is invariant to translational shifts in image position. Most people think that it is obvious that we have no difficulty recognizing an image independently of what part of the image we look at. Eye movements can undoubtedly account for an important part of this insensitivity to position. But in addition, most people would expect that if the eyes were not allowed to move, and we were to learn a new image at one retinal location, we would have no problem recognizing it at other retinal locations, providing of course, that the image was not moved so far into peripheral vision that acuity became insufficient to recognize it. Kahn and Foster (1981) and Foster and Kahn (1985) found that recognition accuracy in discriminating 'same' patterns from 'different' patterns, presented successively with an interval of 1 s, decreases linearly as a function of inter-pattern distance on the retina. This effect was shown not to be an acuity or attention effect. Thus, as was the case for variation in size or rotation, translation seems to affect recognition accuracy, and again a linear relation between performance and amplitude of transformation was found.

The following sequence of experiments was done to investigate the question of translation invariance. Contrary to the experiments done by Kahn and Foster (1981) and Foster and Kahn (1985), where at each trial a new, random stimulus pair was generated, and the subject had only to decide if the two successively presented stimuli were identical or not, in our experiments the subject had to commit to memory a small set of previously unknown stimulus patterns. This is an important difference because when a pattern is learned, it has somehow to be internally represented to allow later access. We were interested in knowing to what extend this access to the internal representation depends on the position on the retina, where the target was learned. Thus, in all the experiments, subjects were taught to recognize a previously unknown pattern, presented at *one* fixed retinal location until recognition was perfect. Then we measured to what extent the information learned at that location could be used afterwards at other locations. Because we were interested in studying both the transferability of information between locations of different acuity as well as between locations of the same acuity, the subjects learned the discrimination task with the patterns presented at a location to one side of the fixation point. In the subsequent test phase, we compared this performance with the performance when the patterns were presented either centrally at the fixation point (higher acuity) or at a location on the other side of the fixation point, opposite to the location where the patterns were learned (same acuity). These two test positions allow us additionally to investigate if the retinal distance between the learned position and the test position has the same effect on performance, as was found by Foster and Kahn. To allow sufficient information processing, subjects were allowed unlimited time to respond. The stimuli to be recognized were dot patterns that resembled Chinese characters (see Fig. 1). These patterns were chosen to be symmetrical about the vertical axis so that the parafoveal retinal acuity gradient would not affect recognition differentially when, in the test phase, the patterns were presented on the other side of the retina relative to the learned location. We used only two non-targets instead of a larger set to avoid the discrimination becoming too difficult to learn.

EXPERIMENT 1

In addition to recognition accuracy under changes of location, we were also interested in eye-movement behavior in this task. The tendency to make eye movements when the stimuli were presented in a new retinal location might be a hint that the information extraction at that specific retinal position is not sufficient to recognize the target. Thus in the first experiment we measured recognition performance and eye-movement behavior.

Subjects

Five subjects participated in the study. All had normal or corrected-to-normal vision.

Stimuli

The stimuli were dot configurations (set 1 in Fig. 1) that were symmetric about the vertical axis. The width of the dot patterns corresponded to a visual angle of 0.97 deg and their height was 0.86 deg. The mask was made out of four uppercase Ms (see Fig. 1) covering the same surface as the dot stimuli.

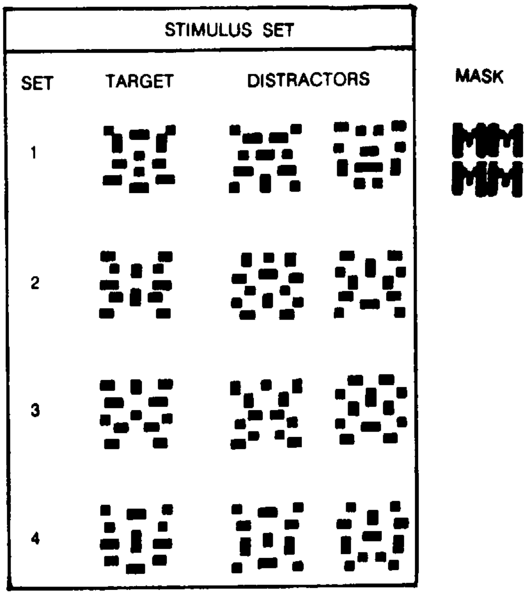

Figure 1. The four different stimulus sets used. In Experiment 1 only set 1 was used; in the other two experiments all the four sets served as stimuli.

Task and procedure

The experiment had three different phases: (1) A presentation phase, where subjects familiarized themselves with the target presented at a fixed eccentricity. (2) A learning phase, where subjects learnt to discriminate between the target and two non-target stimuli. In this phase, targets and non targets were all presented at the same eccentricity as in the presentation phase. (3) A test phase, where subjects had to recognize the target presented at three different retinal eccentricities.

The presentation phase. The subject sat in front of a computer screen wearing a photoelectric device that monitors eye movements. The computer displayed a fixation point at the center of the screen. When the computer detected that the subject was fixating on the point, a dot pattern appeared with its center at an eccentricity of 2.4 deg from the fixation point. For different subjects, the side on which the pattern was presented could be either on the left or the right of the fixation point, but was held constant across all the trials for the subject. The presentation side was counterbalanced across subjects. The subject was instructed to inspect the target dot pattern in peripheral vision without moving his or her gaze away from the central fixation point. When the computer detected that the subject's eye left the fixation point (a criterion of 0.3 deg was used), the computer immediately masked the target dot pattern. In that way, subjects were prevented from seeing the target at other retinal positions. When the subject refixated the fixation point, the mask disappeared and the target reappeared, so the subject could continue inspecting the target. The subject thus had unlimited time for this phase of the experiment.

The learning phase. When the subject was sure he or she knew the target well, the learning part started. The subject's task was now to make a target/non-target decision.

In the same way as indicated above, the computer displayed a fixation point that had to be fixated. Once the subject fixated the point, one of three possible stimuli appeared: either the target dot pattern or one of two non-targets. The subject had to decide whether the stimulus was the target pattern or one of two possible non-targets by pressing the 'target' or 'non-target' key on the computer keyboard. The stimulus remained on the screen until the subject responded. After each response the fixation point and the stimulus disappeared and the next trial started. Every incorrect response was followed by an auditory signal. Again the stimuli were masked whenever the subject's eye attempted to move away from the fixation point. Eye movements were counted as errors and were followed by an auditory signal. This was done in order to train the subject to keep his or her eyes on the fixation point, and allowed the use of eye movements as an index of task difficulty when presentation location was changed. The learning phase continued until the subject reached an accuracy criterion of at least 95%. This required up to 450 trials. After every 90 trials there was a pause, but subjects were free to interrupt the session during the 90 trial blocks by pressing the spacebar of the keyboard. Each of the three stimuli appeared equally often. The order of appearance was chosen randomly by the computer.

The test phase. When a subject reached the accuracy criterion, the test phase started. In this, subjects performed the same target/non-target decision task as in the learning part, by pressing a key on the computer keyboard. However, the test phase contained only three blocks of 30 trials, in which the target appeared 10 times, and each of the non-targets 10 times. In one of the three blocks, the stimuli were presented at the learned location (which could be either +2.4 or −2.4 deg); in another block they were presented at the opposite (−2.4 or +2.4 deg respectively), and in a third block at 0 deg (the fixation point). Contrary to what occurred in the learning phase, in the test phase there was no auditory feedback. Incorrect keypresses and (disallowed) eye movements were recorded. The order of the three test blocks was random across subjects.

Results

In the following analyses we present the proportion of errors made to the target stimulus, computed separately for the three different retinal locations: the position where the stimuli were presented during the learning phase, the LEARNED position (either +2.4 or −2.4 deg); the presentation position at the central fixation point, the CENTRAL position (0 deg); and the position opposite to the learned position, the OPPOSITE position (either −2.4 or +2.4 deg). The results are displayed for each subject individually in Table 1. A wrong response indicates that the subject did not recognize the target as target and thus pressed the non-target key. Except for one subject there were no recognition errors made in the test phase when the target was presented at the learned retinal position. A remarkable increase of, on average, more than 30% in the number of wrong responses appeared when the previously learned stimulus was presented in central vision ($t(4) = 5.59$, $P < 0.005$). But no significant effect was found for changing the presentation position to the opposite side of the learned position. The analysis of eye movement errors (when the subject left the fixation point) given in Table 2 shows a slightly different pattern. Again, all except one subject could do the discrimination task without moving their eyes from the fixation point when the stimuli were presented at the learned position. However, in

Table 1.
Proportion of wrong responses for each individual subject and mean error rate for all 5 subjects, when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position

Ecc. 2.4 deg Subject no.	STIMULUS SET 1 LEARNED	 CENTRAL	 OPPOSITE
1	—	0.7	0.1
2	—	0.4	0.1
3	—	0.2	—
4	0.1	0.3	0.1
5	—	0.1	—
mean	0.02	0.34	0.06
S.D.	0.04	0.23	0.05

Table 2.
Proportion of unallowed eye movements made by each subject when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position

Ecc. 2.4 deg Subject no.	STIMULUS SET 1 LEARNED	 CENTRAL	 OPPOSITE
1	—	0.3	0.1
2	0.4	—	0.5
3	—	—	0.2
4	—	—	—
5	—	0.3	0.1
mean	0.08	0.12	0.18
S.D.	0.17	0.16	0.19

central vision and in particular at the opposite side, eye movement errors occurred more frequently, indicating that the information extracted at a new retinal position was not sufficient to do the task.

Taking these two results together, there is evidence that it is harder to recognize a stimulus presented at a retinal location where it was never seen before. This effect seems to be stronger for a positional change to locations with different acuity, as is the case for the CENTRAL position, than to locations with the same acuity (OPPOSITE position). Although the subject had unlimited time to inspect the stimuli, the transfer of information between locations seems not to be perfect. However, the data from the present experiment, where only a single stimulus pattern was used, are insufficient to make a clear statement concerning translation invariance. Unfortunately the time it takes to train a subject to hold the eyes fixed on a fixation point is unreasonably long, and we decided to change the experimental paradigm so as to be able to collect more data. Thus, in Experiment 2, we measured only recognition accuracy and abandoned eye-movement measures. To obtain more generalizable results, we also used a number of different targets and distractors.

EXPERIMENT 2

Instead of training the subject to maintain gaze at a fixation point as was done in Experiment 1, in Experiment 2 the presentation duration of the stimuli was limited to 150 ms. This is too short to make eye movements and thus ensures that the stimulus falls only on the desired retinal location. Additionally we used three other sets of stimuli, each containing a target and two non-targets as before, to see if the effect could be generalized to other stimuli.

Stimuli

Four different stimulus sets were used in this experiment. The first set was identical to the set used in Experiment 1. The other three sets were new dot configurations similar to set 1 (see Fig. 1). The size of the new stimuli as well as the mask were the same as in Experiment 1.

Subjects

Nine subjects contributed data to each stimulus set for this experiment. Most of the subjects participated in two different stimulus-set blocks. All subjects had normal or corrected-to-normal vision.

Task and procedure

The experimental procedure was the same as in Experiment 1. Each stimulus set was tested in a separate session. There was a presentation phase in which subjects familiarized themselves with the target to learn; a learning phase in which subject learned to discriminate between target and two non-targets; and a test phase in which the stimuli were presented at three different retinal locations. To guarantee that during the learning phase the patterns were seen only at one retinal location, the following was done: A fixation point appeared on the screen and the subject was instructed to fixate it. 500 ms after the fixation point appeared, the stimulus appeared for a period of 150 ms either 2.4 deg to the right or to the left (depending on the subject) of the fixation point, followed immediately by a mask. As was the case in Experiment 1, the task was to decide, by pressing the 'target' or 'non-target' key on the computer keyboard, whether the stimulus was the target previously seen in the presentation phase. The fixation point and the mask remained on the screen until the subject responded, and the time to respond was not limited. Wrong responses were followed by an auditory signal. One second after giving the response, the next trial started with the reappearance of the fixation point. The position of the fixation point on the screen changed slightly from trial to trial, while holding the eccentricity of the stimulus relative to the fixation point constant. This was done so that subjects had to slightly move their eyes between the trials, which helped to prevent fatigue. The learning phase continued until the subject reached an accuracy criterion of at least 95%. Depending on the subjects and the stimulus sets, this could require from 60 to several hundred trials. After every 60 trials there was a short break and feedback about performance level was given. Each of the three stimuli (one target and two distractors) appeared equally often and the order of appearance was chosen randomly by the computer. As before, the test phase contained 30 trials, with target and distractors appearing 10 times each.

Results

Error proportions for each of the four targets, computed separately for the three different retinal locations, are given in Table 3. For each of the four stimuli, recognition errors increased in a statistically significant manner for at least one of the two position changes (CENTRAL or OPPOSITE). Stimulus set 1, which is identical to the set used in Experiment 1, showed the same pattern of results as in Experiment 1: a significant increase of error rate from nearly 0 at the learned position to on average 36% when the stimuli were presented in central vision ($t(8) = 4.72$, $P < 0.005$), and a less marked effect of 14% at the position opposite to the learned position ($t(8) = 2.82$, $P < 0.025$). The difference between central position and opposite position was significant ($t(8) = 2.4$ $P < 0.025$). Using stimulus set 2, the results are slightly different. Here a change in position to the side opposite to the learned position produced a stronger effect than a change to central vision. Error rate increased significantly from 6% at the learned position to 32% at the opposite position ($t(8) = 2.66$ $P < 0.025$). The effect of position change to central vision just failed to be significant ($t(8) = 1.62$; $0.05 < P < 0.1$). Although for 3 of the 9 subjects it was extremely difficult to recognize the target when it was presented in central vision, 4 out of 9 had no problems at all and recognized the target in 100% of the trials. The variability of errors made by different subjects was in general very high (see Table 3). Probable reasons for this will be discussed later. Stimulus set 3 was the most difficult set to learn and even at the learned position subjects made more errors than with set 1 or set 2. Nevertheless, errors were still significantly more frequent in both the CENTRAL and OPPOSITE positions. Changing to the opposite side produced a strong increase of error rate, from 12 to nearly 50% ($t(8) = 2.62$ $P < 0.025$). A position change from learned position to central vision increased the error rate on average from 12 to 27% ($t(8) = 2.15$ $P < 0.05$). However, the difference between the CENTRAL and OPPOSITE position was not significant. Stimulus set 4 showed the smallest effect. For CENTRAL and OPPOSITE test positions the error rate increased on average from 1 to 15 and 14%, respectively. The difference between performance at the learned position and the OPPOSITE position was significant ($t(8) = 2.53$ $P < 0.025$). However, the difference between learned position and CENTRAL position did not reach significance.

The results of Experiment 2 give evidence that information transfer between different retinal locations is not perfect. For all the four stimulus sets, recognition accuracy dropped when the target had to be recognized at a new retinal location. For some stimuli, like stimulus set 4, target discrimination was overall easier, independently of the retinal location of the stimulus, whereas using other stimuli (set 1 and set 3) the task was much more difficult. This indicates that although all the stimuli used in the experiment were unfamiliar to the subjects, there are nevertheless some kinds of stimulus information which are easier to transfer along the retina than others. Contrary to what would have been supposed from the result of Experiment 1, acuity *per se* did not influence the results: Taking the results of all four stimulus sets together, the effect of positional change was equally strong at both CENTRAL and OPPOSITE test positions (see Fig. 2). Mean error rate at the learned position was 5.5% (S.D. = 9.3%); at the CENTRAL position it was 26.1% (S.D. = 27.5%), and at the OPPOSITE position it was 26.9% (S.D. = 28.7%). The differences between both test positions and the learned position were significant: $t(35) = 4.54$ $P < 0.01$ for the CENTRAL position and $t(35) = 4.76$ $P < 0.01$ for the OPPOSITE position. This equality for the two test

Table 3.
Proportion of wrong responses for each subject and mean for all nine subjects for each stimulus set (1–4) when the target had to be recognized either at the LEARNED, CENTRAL, or OPPOSITE position in Experiment 2

Ecc. 2.4 deg	STIMULUS SET 1			STIMULUS SET 2		
Subject no.	LEARNED	CENTRAL	OPPOSITE	LEARNED	CENTRAL	OPPOSITE
1	—	0.7	0.1	0.3	0.2	0.6
2	—	0.2	—	—	—	0.5
3	—	0.6	—	—	—	0.1
4	—	0.2	**0.2**	0.1	0.1	1.0
5	0.1	0.4	**0.2**	0.1	1.0	0.1
6	—	0.1	**0.2**	—	—	0.1
7	0.1	0.5	**0.5**	—	0.5	0.2
8	—	0.5	—	—	0.5	0.2
9	—	0.1	0.1	0.1	—	0.1
mean	0.2	0.36	0.14	0.06	0.25	0.32
S.D.	0.04	0.22	0.15	0.10	0.34	0.31

Ecc. 2.4 deg	STIMULUS SET 3			STIMULUS SET 4		
Subject no.	LEARNED	CENTRAL	OPPOSITE	LEARNED	CENTRAL	OPPOSITE
1	—	0.2	0.4	—	0.1	0.1
2	—	—	—	—	0.1	0.1
3	0.2	0.5	**0.6**	—	0.1	—
4	0.2	—	—	—	—	—
5	0.1	0.1	**0.3**	—	0.1	0.4
6	—	—	**0.9**	—	0.8	0.1
7	0.3	0.6	**0.4**	—	0.1	0.1
8	0.3	0.7	**0.6**	0.1	—	0.5
9	—	0.3	**1.0**	—	0.1	—
mean	0.12	0.27	0.46	0.01	0.15	0.14
S.D.	0.13	0.27	0.35	0.03	0.24	0.18

positions contradicts the results found by Kahn and Foster (1981) and Foster and Kahn (1985), where performance decreases linearly with the distance between the two dot-patterns. If the distance between the patterns plays a role in our experiment, then the translation effect should have been stronger for the OPPOSITE test position than for the CENTRAL test position, because the distance between the OPPOSITE test position and the LEARNED position was twice as large as the distance between the CENTRAL position and the OPPOSITE position.

To summarize the results of Experiment 2, we found the following:

(1) Although the subjects had unlimited time to respond, the transfer of extracted information from one retinal position to another was not perfect. Nevertheless some information must have been transferred, else subjects could not have recognized the target at a new retinal position at all.

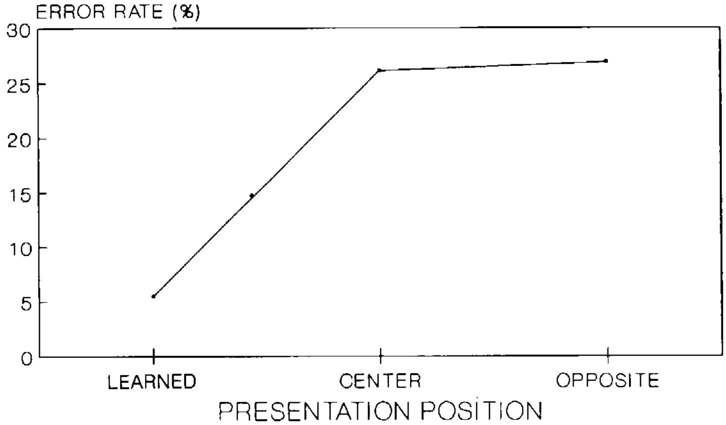

Figure 2. Mean percent errors for all the four stimulus sets together, when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position in Experiment 2.

(2) There are some kinds of information which are easier to transfer than others, and there are subjects who can transfer information better than others. The first claim is supported by the different error rates of the four sets and the second by the high inter-individual variability.

(3) The reduced information transfer along the retina is independent of acuity and independent of the distance between the LEARNED position and the test positions. The transfer was as poor to a location of higher acuity (CENTRAL position) as to a location with the same acuity (OPPOSITE position).

Thus the main factor which seems to influence performance in this experiment is the positional change. However, even though acuity seems not to play a major role in determining the (in)ability to transfer information across the retina, it must be pointed out that the patterns in Experiment 1 and 2 were learned at a location on the retina with somewhat reduced acuity. That is, whatever the nature of the extracted information during the learning phase, this will have been influenced by the eye's acuity at that location. To test the influence of acuity on information transfer, we did a third experiment, where the eccentricity of the learned position was reduced.

EXPERIMENT 3

Experiment 3 was exactly identical to Experiment 2, except that the eccentricity of the peripherally presented patterns was reduced from 2.4 to 0.86 deg. We used the same four stimulus sets as in Experiment 2. Nine subjects participated in each block. None of the subjects had participated in Experiments 1 or 2. Most subjects did two stimulus-set blocks.

Results

Error proportions for each of the four targets, computed separately for the three retinal positions, are given in Table 4. Taking the results of all four stimulus sets together, mean error rate at the learned position was 4.4% (S.D. 6%), at the CENTRAL position it was 9.4% (S.D. 15%), and at the OPPOSITE position 10.3% (S.D. 16%) (see Fig. 3). The differences for both test positions and learned position were significant

Table 4.
Proportion of wrong responses for each subject and mean for all nine subjects for each stimulus set (1–4), when the target had to be recognized either at the LEARNED, CENTRAL, or OPPOSITE position in Experiment 3

Ecc. 0.86 deg	STIMULUS SET 1			STIMULUS SET 2		
Subject no.	LEARNED	CENTRAL	OPPOSITE	LEARNED	CENTRAL	OPPOSITE
1	—	0.4	—	—	—	0.3
2	—	0.7	0.5	—	—	—
3	0.1	0.4	—	—	—	—
4	—	—	—	—	—	—
5	—	0.1	0.4	—	—	—
6	—	0.1	—	—	—	—
7	—	—	—	—	0.1	—
8	—	—	—	0.2	0.1	0.4
9	0.1	—	0.1	—	—	0.1
mean	0.02	0.19	0.11	0.02	0.02	0.08
S.D.	0.04	0.25	0.19	0.06	0.04	0.15

Ecc. 0.86 deg	STIMULUS SET 3			STIMULUS SET 4		
Subject no.	LEARNED	CENTRAL	OPPOSITE	LEARNED	CENTRAL	OPPOSITE
1	—	0.1	—	0.1	0.3	—
2	—	—	0.1	0.1	0.3	0.6
3	0.1	—	0.1	0.1	0.1	0.3
4	0.2	—	0.3	—	—	—
5	—	—	—	—	—	—
6	0.2	0.3	—	—	—	0.1
7	0.1	—	—	0.1	—	—
8	0.1	0.2	0.1	—	—	—
9	—	0.1	0.1	—	0.1	0.2
mean	0.08	0.07	0.07	0.04	0.08	0.13
S.D.	0.08	0.10	0.09	0.05	0.12	0.20

($t(35) = 1.88$ $P < 0.05$ for CENTRAL position and $t(35) = 2.19$ $P < 0.025$ for the OPPOSITE position). The difference between the two test positions was not significant.

It is clear that the effects are weaker in this experiment than in Experiment 2. In fact, if analyses are done on stimulus sets individually, only stimulus set 1 gives a significant effect, with average error rate at the central position being significantly higher than at the learned position ($t(8) = 1.96$ $P < 0.05$). The different strengths of the effects in Experiments 2 and 3 suggest that information extraction differs at different retinal locations. The information that was used to perform the discrimination task at an eccentricity of 2.4 deg in Experiment 2 was different from the information used in Experiment 3 at an eccentricity of 0.86 deg. Though both kinds of information allowed correct recognition of the target at the learned position, the information used at 2.4 deg was quite hard to translate across the retina, whereas that used at 0.86 deg

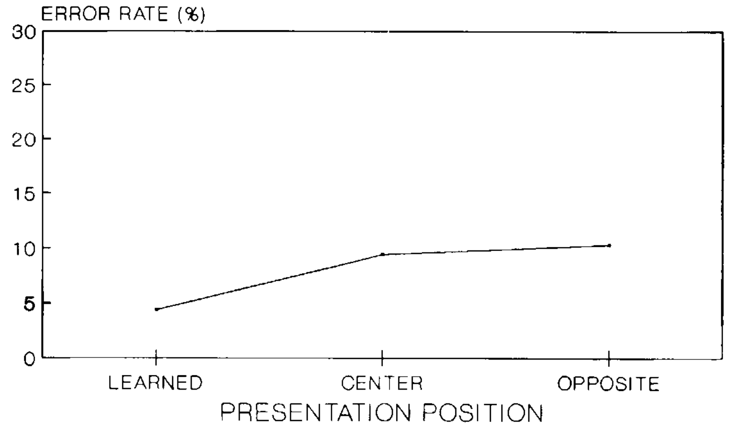

Figure 3. Mean percent errors for all the four stimulus sets together, when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position in Experiment 3.

was easier to translate. The lack of a difference between the two test positions again excludes the possibility that the difference was due to the distance through which the information had effectively to be translated.

A reason for finding a decrement of performance under translation in the three experiments reported above might be related to the type of stimuli used. When presenting a pattern parafoveally, the dots of the pattern which are nearest the fovea are most clearly seen. If these dots are sufficient to discriminate the target and the distractors, it is possible that the subjects never need have learned the rest of the stimulus pattern. When in the test phase the stimuli are presented centrally, it is the central dots that are most clearly seen. This may account for the increase in error rate at the CENTRAL position. For the test on the OPPOSITE position, the inner edge of the stimulus is most clearly seen again, but because the stimulus pattern is symmetrical, the inner edge is now a mirror image of the learned edge. Though it is known that human's recognition of mirror images is quite good (e.g., Kahn and Foster, 1986), this might nevertheless have reduced performance at the OPPOSITE position. Additionally, the change of the presentation position from the LEARNED position to the OPPOSITE position involved a change in the cerebral hemisphere where the information impinged. At least for the OPPOSITE test position, this might be another reason why performance dropped. To exclude these possibilities an additional experiment using a new type of stimulus was done.

EXPERIMENT 4

So that an explanation of the translation effects in terms of the difficulty of extracting mirror-reflected features could be set aside, instead of using symmetrical dot-patterns, in Experiment 4 thin vertical columns, made out of 16 squares of randomly chosen grey levels, were used as stimuli. These new kinds of stimuli are much easier to learn than the stimuli used in the other three experiments, especially when the target has to be distinguished from two distractors only. This is because in this case the grey level of just one square at a given position often suffices to identify the target (e.g., the target contains a black square at the top of the column, whereas the two distractors do not). This kind of information will be easily translatable. To prevent the subjects from adopting such a strategy, the number of distractors was chosen to be unlimited. This increases the probability of distractors and targets sharing squares of the same grey level at given positions. Thus, increasing the size of the distractor set forces the subject to process more than only small parts of the target pattern. The inconvenience of using an unlimited distractor set is that, during the test phase, each distractor appears only once, but the target appears many times. This may allow the subject to identify and so re-learn the target at the new location. Thus, measuring recognition accuracy over a certain number of trials will underestimate the effect of positional change. For this reason, we chose to take recognition time as a possibly more sensitive indicator for the effect of positional change.

To see if the effect we obtained in the other experiments was not due to inter-hemispheric transfer, the present experiment contained an additional condition in which the stimuli were thin horizontal rows of 16 squares presented above, center or below the fixation point (see Fig. 4).

Stimuli

The stimuli were thin lines of width 0.23 and length 3.1 deg. Each line was made of 16 small squares. Each square could be drawn in one of 15 possible grey levels, ranging from white to black. The grey level of the squares was chosen randomly by the computer, with the restriction that two adjacent squares should never have the same grey level. Thus, the difference between target and distractors was defined by the grey level combinations of the squares. For each subject, and in both the vertical and horizontal display sub-experiments, a new target line was defined. The distractor lines were generated at random at each trial. The mask was a white column of the same size as the stimuli.

Subjects

Twelve subjects participated in the study. All had normal or corrected-to-normal vision. Each subject participated in both sub-experiments.

Task and procedure

As was the case in the other three experiments, each sub-experiment in Experiment 4 contained a presentation phase, a learning phase and a test phase. In the left/right sub-experiment (Fig. 4a) the stimuli were presented to the left, at the center, or to the right of the fixation point, as was done before. In the above/below sub-experiment (Fig. 4b) the stimuli were presented above, at the center or below the fixation point. In the left/right sub-experiment the stimulus lines were presented vertically, with the center of the line at the same height as the fixation point. The stimulus was presented with its center at an eccentricity of 0.49 deg. At this eccentricity the distance between the fixation point and the inner edge of the stimulus was the same as in Experiment 3. In sub-experiment above/below, the stimulus lines were presented horizontally, with the center of the line vertically aligned with the fixation point. The presentation eccentricity was 0.49 deg. To be able to compare the reaction-time data of the different subjects, the number of training trials in the learning phase was fixed for every subject to be three blocks of 60 trials, in which the target appeared 20 times in each block (giving a probability of 1/3 for target appearance as was the case in the other experiments). In a pilot study we found that generally after 180 trials target

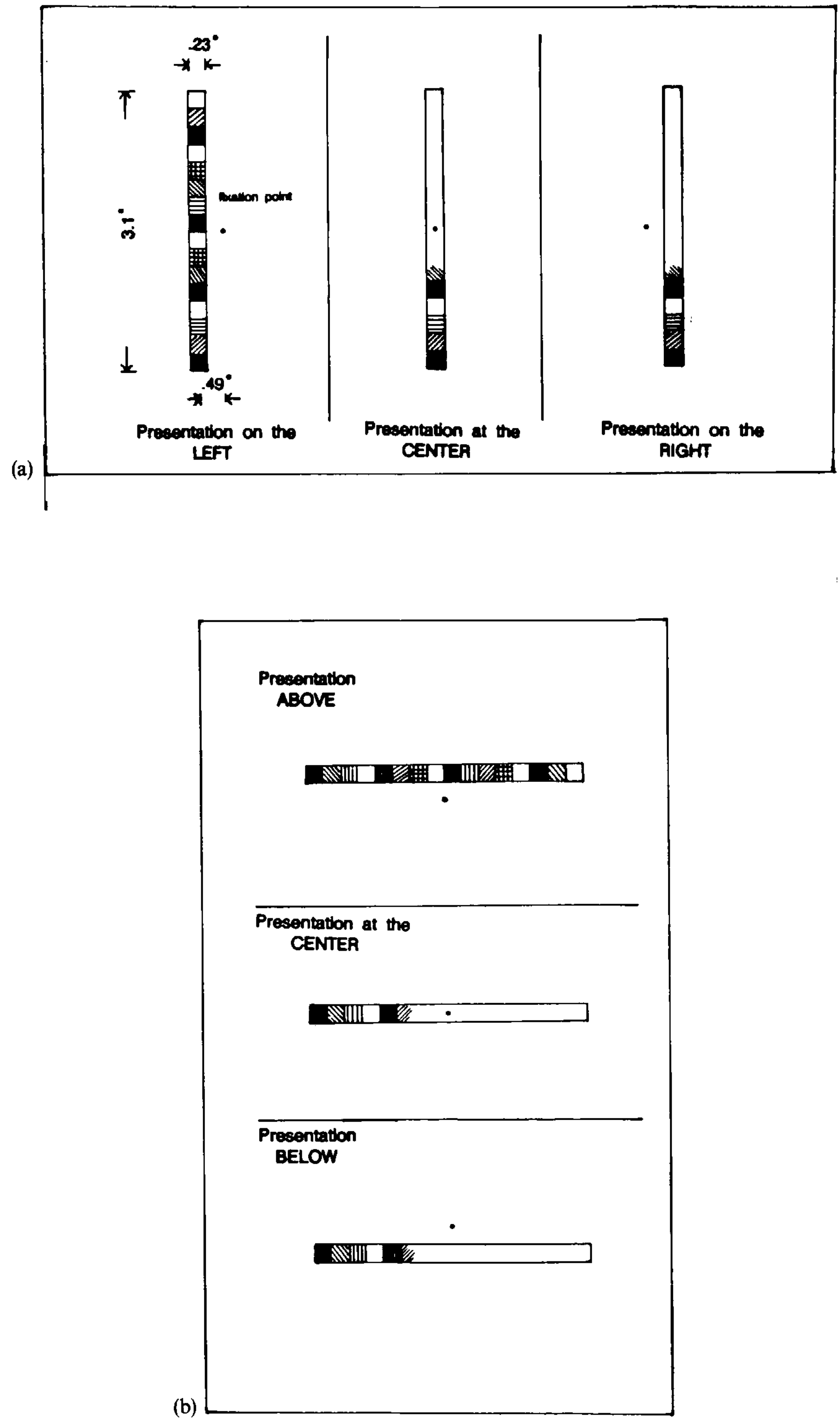

Figure 4(a) and (b). The presentation positions of the stimulus in the two sub-experimental conditions.

Table 5.
Proportion of wrong responses for each individual subject and mean error rate for all 12 subjects for the two sub-experimental conditions (left/right vs. above/below), when the target had to be recognized at either the LEARNED, CENTRAL or OPPOSITE position in Experiment 4

Ecc. 0.49 deg	LEFT/RIGHT			ABOVE/BELOW		
Subject no.	LEARNED	CENTRAL	OPPOSITE	LEARNED	CENTRAL	OPPOSITE
1	0.1	—	—	—	—	—
2	—	—	0.1		0.1	0.1
3	0.2	—	0.1	—	0.9	—
4	—	0.2	0.2	—	—	—
5	—	0.3	—	0.1	0.1	0.1
6	0.1	0.1	0.2	0.1	—	—
7	0.2	0.5	0.1	0.1	0.1	—
8	—	—	0.3	—	0.1	0.1
9	—	0.3	0.1	0.2	0.3	0.4
10	—	0.1	0.2	—	0.1	0.1
11	—	—	0.2	—	—	—
12	0.1	0.1	0.1	—	—	0.1
mean	0.06	0.13	0.13	0.04	0.14	0.08
S.D.	0.08	0.16	0.09	0.06	0.25	0.11

identification was perfect. The test phase was identical to the test phase in the previous experiments. The order of the sub-experiments was counter-balanced across subjects.

Results

Except for two subjects, who did not reach an accuracy level of more than 60% (and thus were replaced by two new subjects), in all cases response accuracy was at least

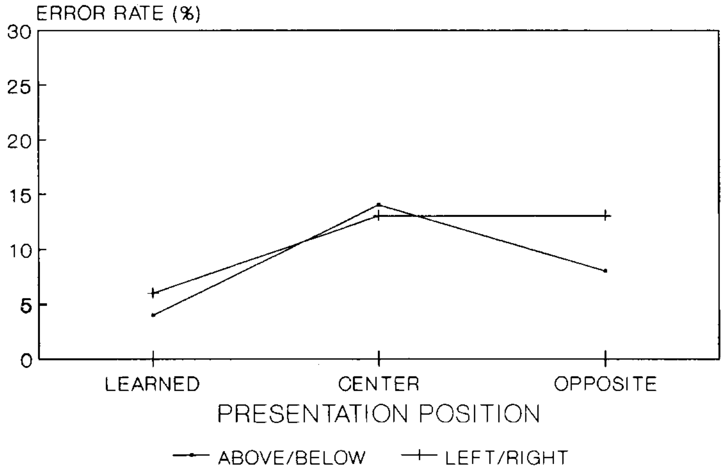

Figure 5. Mean percent errors for the two sub-experimental conditions (left/right vs. above/below), when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position in Experiment 4.

90% in the last learning block. Thus contrary to the stimuli used in the previous experiments, where some of the subjects needed about 450 trials to learn to discriminate one out of three possible patterns, the stimulus lines we used in this experiment were easy to learn. As in Experiment 3, recognition accuracy dropped only slightly when presentation position was changed. Percent correct for each of the two sub-experiments, computed separately for the three different retinal locations, are given in Table 5. For the right/left sub-experiment, mean error rate at the LEARNED position was 6% (S.D. 8%), at the CENTRAL position 13% (S.D. 16%), and at the OPPOSITE position 13% (S.D. 9%). Similar effects were obtained for the above/below sub-experiment, where mean error rate for the LEARNED position was 4% (S.D. 6%), at the CENTRAL position 14% (S.D. 25%), and at the OPPOSITE position 8% (S.D. 11%) (see Fig. 5). Recognition errors increased in a statistically significant manner only in the right/left sub-experiment for the OPPOSITE test position ($t(11) = 1.91$ $P < 0.05$).

The response time data are stronger. For each of the subjects we computed the response time for correct responses separately, rejecting those latencies which were further than two standard deviations from the subject's mean. The final means of response time were obtained by taking the average across subjects, and are presented for each of the two sub-experiments separately in Table 6. For the right/left sub-experiment the average correct response time for the target at the LEARNED position was 429 ms. It increased by 136 ms when the target was presented at the CENTRAL test position and by 87 ms when the target was presented at the OPPOSITE test position. The effect was in both cases statistically significant ($t(11) = 3.45$ $P < 0.005$ and $t(11) = 1.78$ $P < 0.05$, respectively). For the above/below sub-experiment the effect was not so strong. The average response time at the LEARNED position was about 365 ms; it increased by 34 ms at the CENTRAL test position and by 66 ms at the OPPOSITE test position (see Fig. 6). The difference between performance at the LEARNED position and the OPPOSITE test position was significant ($t(11) = 3.41$ $P < 0.005$). However, the difference between LEARNED position and CENTRAL test position did not reach significance ($t(11) = 1.2$, n.s.). The differences, in both sub-experimental conditions, between the two test positions (CENTRAL vs. OPPOSITE) were not significant.

The results of Experiment 4 are in line with the results obtained in the previous three experiments. Though recognition accuracy was only slightly affected when the target had to be recognized at a new position, response latencies increased by a value

Table 6.
Mean recognition time (ms) (+ S.D.) for all 12 subjects for the two sub-experimental conditions (left/right vs. above/below), when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position in Experiment 4

Condition	LEARNED	CENTRAL	OPPOSITE
Left/right	429 (133)	565 (223)	516 (205)
Above/below	365 (93)	399 (61)	431 (75)

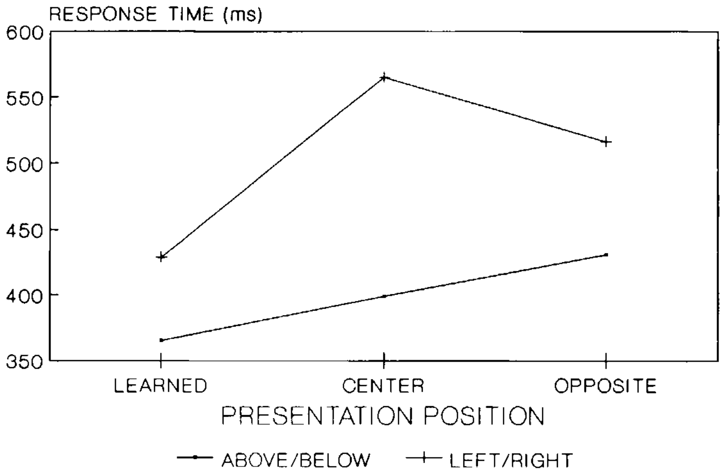

Figure 6. Mean recognition time for the two sub-experimental conditions (left/right vs. above/below), when the target had to be recognized either at the LEARNED, CENTRAL or OPPOSITE position in Experiment 4.

of 20–30% in the right/left sub-experimental condition and by a value of about 10–20% in the above/below sub-experiment. Thus, even with a more simple stimulus type, a small change of retinal position has an effect on recognition. This result excludes the 'mirror-image' argument as a possible explanation of our results. Though the effect is weaker in the above/below experiment than in the right/left sub-experiment, it still exists. Thus, the hemisphere argument can be rejected also. We do not know exactly what the reason for the difference in the two sub-experiments is, but most of the subjects reported that it was much easier to learn the horizontally presented line than the vertically presented column. In fact when comparing the response latencies of both experimental conditions for the LEARNED position (429 ms in the right/left sub-experiment vs. 365 ms in the above/below sub-experiment) we can see that on average the subject recognized the target about 60 ms faster when it was presented horizontally than vertically. The difference is statistically significant ($t(11) = 2.46$ $P < 0.025$). Thus, even before the presentation position of the target is changed, the two sub-experiments are not completely comparable.

DISCUSSION

All the experiments reported here showed that translation invariance is not perfect. To summarize the obtained results the following can be said: (1) Translation caused a decrease of recognition performance. (2) The effect of translation on performance depends on the stimulus set used. (3) The effect of translation on performance varies for the same stimulus between subjects. (4) The information transfer to new positions is independent of the acuity at the new position: it is on average equally difficult to recognize the pattern presented at the CENTRAL test position than at the OPPOSITE test position, though for some patterns recognition is more difficult at the OPPOSITE test position whereas for others recognition is more difficult at the CENTRAL test position. (5) The drop of performance is stronger when the pattern is learned at a retinal position with a reduced acuity (Experiments 1 and 2) than when it is learned at a retinal position with higher acuity (Experiments 3 and 4). (6) The equality of the two test positions suggests additionally, that the drop of performance is not related to the distance over which the information has to be transferred.

During the learning phase in the experiments, the subject must have built an internal representation of the target which persisted over a certain time and which could be accessed in order to identify the target later in the test phase. The performance on changing the presentation position dropped, but it was far better than chance level. Thus, some information in the internal representation can be accessed in a way which is position independent, whereas some additional information is only accessible when the target is presented at the learned position. What is the nature of the information which can be transferred and what is the additional information, not transferable, allowing perfect recognition only at the learned position? The variability of the strength of the translation effect while using different stimulus sets, as well as the differences of the effect for the individual stimuli at the two test positions (some stimuli are easier to recognize at the OPPOSITE test position than at the CENTRAL test position and vice versa) make an interpretation of the data in line with a global renormalization transformation of the retinal image difficult, because in this case, the stimuli should have been either translatable to both test positions equally well (or poorly), or for all stimuli the advantage for one of the two test positions should have been in the same direction, which was not the case. The effect seems highly related to the nature of the stimulus set used but not to the amount of positional change.

When an observer is confronted with a new, unfamiliar pattern, there will always exist some sub-patterns or features which are already familiar to him or her. Thus, though most of our subjects expressed their astonishment when first seeing the stimuli at a new position, none of them reported that they did not recognize that the pattern at the new position was made out of the same kind of dots (or squares) as at the learned position. The 'dotness' of the pattern was transferable, because 'dots' are familiar to us. However, in our experiment the 'dotness' of the stimuli does not allow correct discrimination of the target. But there is no doubt that the stimulus patterns we used contained other familiar features, which might be distinctive for the target. If for example by accident some dots in the target are grouped forming a line or other simple geometrical form, the patterns would be easy to translate (it is known that to identify an object, the visual system does not need to process the whole stimulus; Treisman and Gelade, 1980; Biederman, 1987).

Although it is premature, from the results we have obtained, to make strong hypotheses about the underlying mechanism of human translation performance, a possible explanation of our results might be the following. Suppose that when learning a pattern at a given retinal location, the pattern is coded by an ensemble of different feature 'units', which are associated with the pattern of activation resulting from the retinal stimulation. If the stimulus is shifted across the retina, the retinal pattern of activation will change. As a consequence the same stimulus will be coded by a different set of feature units. As long as no connection is made between the two sets of feature units, the stimulus cannot be recognized as the same. To be able to recognize a stimulus independently of retinal location, the visual system must learn to associate the different retinal activations produced by the same stimulus with the same set of feature units. This is only possible when enough experience with different visual objects presented at different retinal positions has accumulated. Now, if an unfamiliar pattern has to be learned at one specific retinal location, as in our experiment, the pattern will also be coded in terms of feature units. Due to earlier experience, some of these features are 'invariant', but some of them not (else the pattern would not be unfamiliar). The more the unfamiliar pattern contains familiar, 'invariant' features, the easier or faster the target will be to recognize at different retinal locations (providing the 'invariant' features are distinctive). The advantage in recognizing the target at the learned position is due to the fact that only at the learned position do all the activated feature units serve as information, whereas at the test position only the 'invariant' feature units can be used for target identification.

Seen from this point of view the variability of performance while using different stimulus sets, as well as the variability between different subjects can be explained. The variability is due to the differing degrees of familiarity of the features extracted by the subjects. We do not know which distinctive features our subjects used to solve the discrimination task, because there are always many different possible choices of feature, all allowing identification of an object. But apparently some subjects used more 'efficient' features than others. For example, when asking the subjects afterwards what the nature of the target was, some of them gave a global description of the target as being like a chessman having something round on its head, or a bizarre telephone, whereas, other subjects simply said the target had three dots at the bottom, or they described a specific arrangement of two or three dots contained in the target. This latter kind of description was particularly common in the experiments which employed the small eccentricity, where local arrangements of a few dots were easier to see.

Thus, according to this explanation, translation invariance is the result of learning and experience. This idea is not new, since it was already proposed by Hebb (1949), who, after prolonged studies with rats reared in darkness, and after studying apprehension of figure perception by cataract patients (after operation), claimed that during the continuous, intensive, and prolonged visual training of infancy and childhood, we learn to recognize visual stimuli separately for each separate part of the visual field. Unfortunately, this kind of explanation does not seem consistent with the results obtained by Kahn and Foster (1981) and Foster and Kahn (1985), who found a linear relation between translation distance and performance decrement. We think that much work must be done before a more coherent explanation of the underlying mechanism of translation invariance can be given.

Acknowledgements

This research was supported by the B.R.A.I.N. (Basic Research in Adaptive Intelligence and Neurocomputing) initiative of the European Community. We wish to thank A. Gorea, O. Heller, J. Juola and two anonymous reviewers for valuable suggestions.

REFERENCES

Arnoult, M. D. (1954). Shape discrimination as a function of the angular orientation of the stimuli. *J. Exp. Psychol.* **47**, 323–328.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychol. Rev.* **94**, 115–147.

Bundesen, C. and Larsen, A. (1975). Visual transformation of size. *J. Exp. Psychol. Human Percept. Perform.* **3**, 214–220.

Burkhardt, H. and Muller, X. (1980). On invariant sets of a certain class of fast translation–invariant transforms. *IEEE Trans. ASSP* **28**, 517–523.

Cavanagh, P. (1978). Size and position invariance in the visual system. *Perception*, **7**, 167–177.

Cavanagh, P. (1985). Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance. In: *Models of the Visual Cortex.* Rose, D. and Dobson, V. (Eds.) Wiley, pp. 85–95.

Cooper, L. A. and Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In: *Visual Information Processing.* Chase, G. (Ed.). Academic Press, New York.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychol.* **7**, 20–43.

Corballis, M.C., Zbrodoff, N. J., Shetzer, L. I. and Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory Cognition* **6**, 98–107.

Dearborn, G. V. N. (1899). Recognition under objective reversal. *Psychol. Rev.* **6**, 395–406.

Eley, M. G. (1982). Identifying rotated letter-like symbols. *Memory Cognition* **10**, 25–32.

Foster, D. H. (1978). Visual comparison of random-dot patterns: evidence concerning a fixed visual association between features and feature-relations. *Q. J. Exp. Psychol.* **30**, 637–654.

Foster, D. H. and Kahn, J. I. (1985). Internal representations and operations in the visual comparison of transformed patterns: effects of pattern point-inversion, positional symmetry, and separation. *Biol. Cybernet.* **51**, 305–312.

Gerrissen, J. F. (1982). Theory and model of the human global analysis of visual structure. *IEEE Trans. Systems, Man, Cybernet.* **12**, 805–817.

Hebb, D. O. (1949). *The Organization of Behavior.* Wiley, New York.

Kahn, J. I. and Foster, D. H. (1981). Visual comparison of rotated and reflected random-dot patterns as a function of their positional symmetry and separation in the field. *Q. J. Exp. Psychol.* **33A**, 155–166.

Kahn, J. I. and Foster, D. H. (1986). Horizontal–vertical structure in the visual comparison of rigidly transformed patterns. *J. Exp. Psychol. Human Percep. Perform.* **12**, 422–433.

Kolers, P. A., Duchnicky, R. L. and Sundstroem, G. (1985). Size in visual processing of faces and words. *J. Exp. Psychol. Human Percept. Perform.* **11**, 726–751.

Kröse, B. J. A. (1985). A structure description of visual information. *Pattern Recognition Lett.* **3**, 41–50.

Kubovy, M. and Podgorny, P. (1981). Does pattern matching require the normalization of size and orientation? *Percept. Psychophys.* **30**, 24–28.

Reitboeck, H. J. and Altman, J. (1984). A model for size- and rotation-invariant pattern processing in the visual system. *Biol. Cybernet.* **51**, 113–121.

Rock, I. (1973). *Orientation and Form.* Academic Press, New York.

Rock, I. (1984). *Perception.* Scientific American Library, New York.

Schwartz, E. L. (1977). Spatial mapping in primate visual context: analytic structure and relevance to perception. *Biol. Cybernet.* **25**, 181–194.

Schwartz, E. L. (1981). Cortical anatomy, size invariance and spatial frequency analysis. *Perception* **10**, 455–468.

Shepard, R. N. and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* **3**, 701–703.

Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception* **9**, 483–484.

Treisman, A. and Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychol.* **12**, 97–136.

Uttal, W. R. (1975). An autocorrelation theory of visual form detection. Erlbaum, Hillsdale, New Jersey.