The role of visual field position in pattern-discrimination learning

MARCUS DILL AND MANFRED FAHLE

University Eye Clinic Tübingen, Section of Visual Science, Waldhoernlestr. 22, D-72072 Tübingen, Germany

SUMMARY

Invariance of object recognition to translation in the visual field is a fundamental property of human pattern vision. In three experiments we investigated this capability by training subjects to distinguish between random checkerboard stimuli. We show that the improvement of discrimination performance does not transfer across the visual field if learning is restricted to a particular location in the retinal image. Accuracy after retinal translation shows no sign of decay over time and remains at the same level it had at the beginning of the training. It is suggested that in two-dimensional translation invariance—as in three-dimensional rotation invariance—the human visual system is relying on memory-intensive rather than computation-intensive processes. Multiple position- and stimulus-specific learning events may be required before recognition is independent of retinal location.

1. INTRODUCTION

Visual constancy—our ability to recognize an object even if its projection on the retina varies considerably—is among the most fascinating properties of human perception. Despite considerable efforts during the last decades (for reviews see, for example, Shepard & Cooper 1982; Walsh & Kulikowski 1997) the mechanisms underlying this ability remain largely unknown. This holds true even for such apparently simple cases as the displacement of an object in the visual field. Unlike lower animals like flies (Dill *et al.* 1993; Dill & Heisenberg 1995), human observers are able to tolerate retinal translations (Biederman & Cooper 1991), at least to a certain extent.

Several methods have been suggested by which the brain might accomplish translation invariance. Many of these accounts assume that visual comparison processes depend on supplementary operations aligning visual input and stored memory by some kind of shifting process (e.g. Foster & Kahn 1985; Foster 1991; Olshausen *et al.* 1993; Van Essen *et al.* 1994). Psychophysical evidence for such accounts comes from studies showing that response latencies and error rates increase monotonically with the degree of transformation applied to the stimulus (Foster & Kahn 1985).

Alternatively, tolerance to retinal translation may be explained as the result of repeated stimulus exposure at various locations in the visual field: by associating different input images produced by the same visual object at different positions one may gradually acquire a position-invariant representation of this object. For simple or familiar objects the learning process leading to translation invariance may already be completed in our brains. Only for novel visual patterns or in a highly demanding discrimination task can one hope to find evidence for positional specificity, since none of the available memory modules may be sensitive enough to allow discrimination. However, learning may tune the modules more precisely or create new ones so that differences between the patterns can be better detected. If training is spatially restricted, this acquired improvement should remain spatially specific too. This is, indeed, what has been found in several perceptual learning studies, with tasks ranging from vernier discrimination to texture segmentation (Ramachandran 1976; Fiorentini & Berardi 1981; Karni & Sagi 1991; Shiu & Pashler 1992; Fahle 1994; Fahle *et al.* 1995). All tasks have in common that they used relatively simple stimuli that were either very similar to each other or were presented under severely limited viewing conditions involving masking and short presentation times.

Nazir & O'Regan (Nazir & O'Regan 1990; O'Regan 1992) showed that learning of more complex visual patterns is also not translation invariant. They trained subjects to discriminate a defined target dot pattern from two distractor stimuli. Once they had reached a criterion of 95% correct responses, observers were tested both at the retinal location where the patterns had been presented during training and at new locations. In most tests, performance decreased significantly after displacement, indicating that the improvement during training was at least partially specific to the retinal location. Similar results have been reported for a classification task employing Gabor patches (Rentschler *et al.* 1994; Jüttner & Rentschler 1996).

In both kinds of studies, however, accuracy at untrained locations was clearly above chance, indicating that recognition of novel patterns can be at least

Printed in Great Britain

TEX Paper

somewhat position invariant. A possible explanation for this finding is offered by introspective reports: some observers in the Nazir & O'Regan (1990) study described their learning strategy to solve the task as looking for a distinctive feature that discriminates one pattern from the others. This feature could be an abstract association (like, for example, a 'bizarre telephone', p. 99) as well as a subgroup of two or three dots of a particular configuration or orientation. Stimuli in that study were different in all their parts and only two distractors were used with each target, allowing such an approach to be successful. In the experiments presented below we tried to reduce the influence of this strategy by using a large set of very similar distractors for each target.

Nazir & O'Regan (1990) concluded from their experiments that visual recognition in general is specific to the location in the visual field and that more or less position-invariant performance can only be achieved by detecting particular features. However, the reported data are not sufficient to exclude explanations based on supplementary shift mechanisms. Given that these could be imperfect they may lead to partial, but not complete, tolerance for retinal shifts. It is even imaginable that the observed positional specificity of performance does not reflect a lack of invariance of the stimulus representation at all, but is instead the result of a narrowing or sharper tuning of the focus of attention leading to a general improvement of visual capabilities at the trained part of the visual field. Additionally, there is the problem that Nazir & O'Regan (1990) restricted data analysis to the one-third of trials presenting the target and discarded the remaining two-thirds displaying a distractor. This procedure may lead to an overestimation of positional specificity because part of the difference between control and transfer tests may be due to differential response biases.

The purpose of the present study is twofold: first, to exclude the aforementioned explanations for the observed positional specificity that are independent of pattern memory itself; and, second, to characterize the invariant part of discrimination performance. Of special interest is the question whether above-chance performance in tests at new locations indicates learning transfer. At least for some patterns, training may start at a baseline that is different from chance because subjects immediately recognize certain features that are either simple or already familiar from earlier encounters, e.g. as part of similar visual stimuli. These features may allow partial fulfilment of the discrimination task right after the first (few) trial(s).

2. METHODS

Except for one of the authors (M.D.), observers were undergraduate students from Tübingen University being paid for their participation. Each observer had normal or corrected-to-be-normal visual acuity (at least 20/20).

Stimuli were produced on a black and white CRT monitor (refresh rate 75 Hz) controlled by a Macintosh Power PC. The display was viewed binocularly at a viewing distance of 1 m. Patterns were presented for *ca.* 100 ms. This time period is below the normal onset latency for saccades (about 200 ms; Saslow 1967) and was chosen in order to prevent directed eye movements towards the stimuli. Patterns appeared on a uniform white background (mean luminance *ca.* 100 cd m^{-2}; total visual angle 12° width, 16° height). Their centre was located at 2.4° horizontal eccentricity left or right of a fixation spot of 0.13° diameter. Each pattern was composed of a matrix of elements filling a rectangular field of edge length 48 pixels (corresponding to about 0.85°). In experiments 1–3, a 6×6 matrix of elements was used. Each element measured six pixel in width and height and was separated from its neighbours by two-pixel wide white (i.e. background) lines (see figure 1). Each matrix element could be either black or white randomly chosen by the computer. Randomness was restricted only by bilateral symmetry between the left and right three columns of fields. Gradients in retinal acuity, therefore, should not affect recognition of the patterns differentially when presented in the left and right visual field, respectively (cf. Nazir & O'Regan 1990). For each target a set of all possible distractors was formed by changing, one at a time, all symmetrical pairs of target elements from black to white or vice versa. Figure 1 illustrates an example of a target and its 18 possible distractors. In the pilot experiments slightly different matrix types were employed (see insets in figure 2). Decisions were communicated by hand-held press buttons. Except during test periods in experiments 1 and 2, a computer tone provided error feedback immediately after incorrect responses.

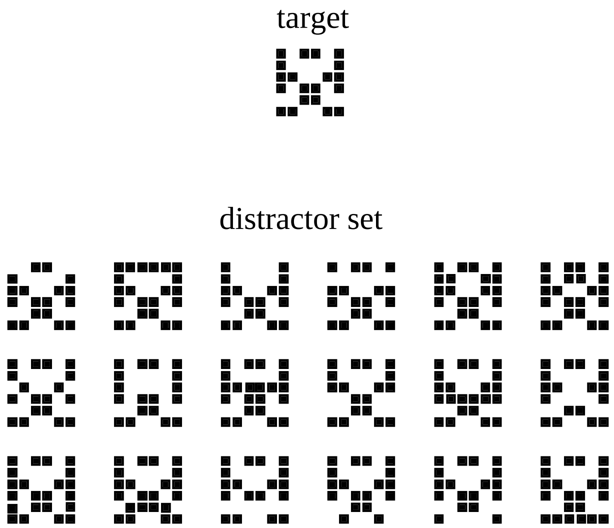

Figure 1. Example for a target dot-matrix pattern (top) and its 18 possible distractors.

For each session a new target pattern was randomly chosen by the computer. Training comprised a presentation and a learning stage. During the presentation stage, the target appeared 10 times at the training location with an interstimulus interval of 1 s to allow observers to get acquainted with the target pattern. The fixation spot was continuously visible during this period. No response was required. During the following learning stage, each trial started with the appearance of the fixation spot, followed 1 s later by the brief display of either a target or a distractor at the training location. The probability for display of a target was 50%, in the remaining trials the computer randomly chose one stimulus from the distractor set. In the pilot experiments target frequency was 33%. After stimulus disappearance the fixation spot was replaced by

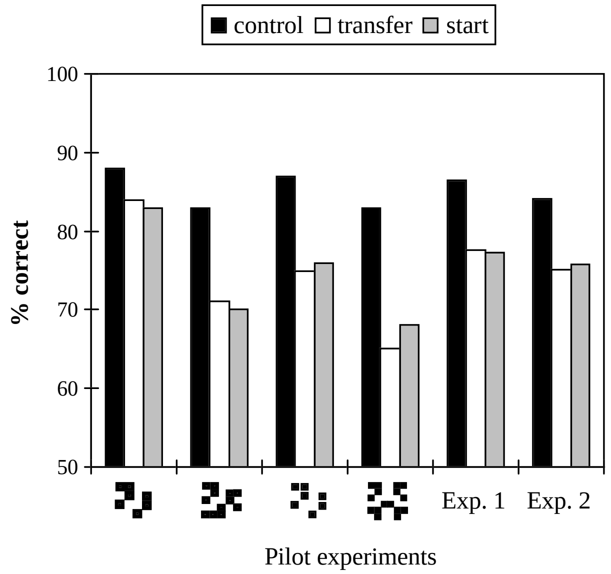

Figure 2. Learning of target identification is position specific. Bars represent the mean percentage of correct responses in four pilot studies and experiments 1 and 2. Control and transfer refer to the two test conditions, start represents the first 30 trials of training. From left to right, the number of training sessions (observers) was: 16 (1), 16 (1), 18 (1), 10 (1), 24 (6) and 24 (6).

a question mark requesting the subject's decision 'target' versus 'no target'. In the pilot experiments (not in experiments 1–3) stimuli were followed immediately by a regular checkerboard mask of the same size (mask duration 350 ms). Subsequent trials were separated by a one-second break. The computer continuously monitored performance, until subjects achieved a percentage of correct responses of more than 90% (28 correct responses in 30 successive trials).

Except for experiment 3, memory was assessed by a test using the same stimuli but without error feedback at both the training (*control*) and the contralateral (*transfer*) location (30 trials each). In the pilot studies and in experiment 1 control and transfer were tested in sequential blocks: in half of the sessions testing started with control trials followed by the transfer block, in the remaining sessions the order was reversed, i.e. transfer was tested first. There was no break between control and transfer test blocks. Both blocks were balanced for target versus distractors. In experiment 2, control and transfer test trials were presented in randomized order, such that, on a given trial, observers did not know in advance at which of two possible locations the test stimulus would appear. Each observer passed four sessions in experiments 1 and 2, respectively.

In experiment 3, memory was tested by a second round of training, again comprising a presentation and learning period and involving either the same or a novel target at either the same (control) or a transfer location. Each observer was trained twice for each of the four conditions.

For statistical analysis, data were pooled for individual subjects. Unless stated otherwise, significance statements refer to paired-comparison t-tests. Because individual test sessions comprised only 15 target and distractor trials, respectively, we had to pool data across all subjects in order to determine discriminability values $d' \pm$ s.e.m (for statistical details see Macmillan & Creelman (1991)).

3. RESULTS

(*a*) *Pattern-discrimination learning is position specific*

In all experiments observers were trained to discriminate a target from a set of very similar distractors (figure 1) until they reached a criterion level of 90%. The number of trials required to achieve criterion performance varied considerably (e.g. in experiment 1 from 31–180 trials, mean 83 trials; experiment 2 from 28–319 trials, mean 98 trials) both between observers and for individual subjects (cf. Nazir & O'Regan 1990). After successful training subjects were tested at the location of training (control) and at the alternative position in the opposite visual half field (transfer). In experiment 1 (and in several pilot experiments) control and transfer were tested in blocks of 30 trials each. In experiment 2 both trial types were randomly interleaved. The results (figure 2) confirm the finding of Nazir & O'Regan (1990) that learning to discriminate complex patterns is not translation invariant: in 20 of the 24 training sessions of experiment 1 (18/24 in experiment 2) the percentage of correct responses is higher at the control than at the transfer location (experiment 1: $t[5] = 3.25$, $p < 0.05$; experiment 2: $t[5] = 3.07$; $p < 0.05$).

From the beginning of training, accuracy in target trials is at about 90% correct indicating a general tendency of the subjects to press the 'target' button. Learning is, therefore, only evident for distractor trials. These bias effects, however, cannot explain positional specificity, since calculation of d'—a bias-free measure of discriminability—yields a clear difference between control (experiment 1: $d' = 2.21 \pm 0.12$; experiment 2: $d' = 1.95 \pm 0.11$) and transfer tests (experiment 1: $d' = 1.64 \pm 0.11$; experiment 2: $d' = 1.34 \pm 0.10$).

Though the effect of displacement is highly significant, performance in the transfer test also clearly exceeds chance level (50%) indicating that some information about target identity is available after displacement. As discussed above, this might be explained either by an imperfect mental-transformation mechanism or by some immediate ability to recognize at least part of the stimulus right from the beginning and independent of its locations. In order to distinguish between these two alternatives we looked at the first 30 trials of each training session (figure 2, *start*). Also included in the figure are results from a series of pilot studies with a single observer (author M.D.), performed under only slightly different experimental conditions. In these experiments with four different stimulus types (see inset), we had noticed that accuracy at the beginning of the training was always on about the same level as performance in the transfer test. This preliminary observation is confirmed by experiments 1 and 2: performance in transfer tests is not better than the start level of the training ($t[5] = 0.11$ and 0.16, $p > 0.1$, respectively; control versus start: $t[5] = 3.65$ and 2.76; $p < 0.05$, respectively). This finding strongly argues against any transfer of learning. It indicates, instead, that right from the start accuracy is already above the

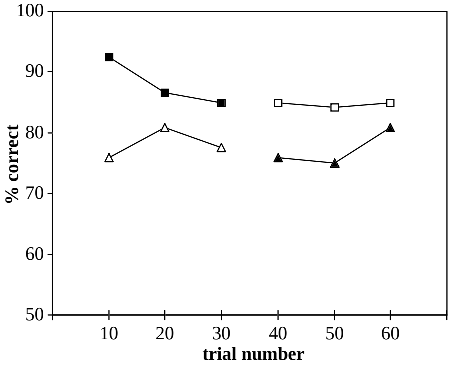

Figure 3. Time course of test performance in experiment 1. Percentage of correct responses in the test presented in separate blocks of ten trials each. In half of the experimental sessions (filled symbols) control (squares) preceded transfer (triangles) tests, in the other half transfer was tested first (open symbols). Mean of 12 sessions (6 observers), respectively.

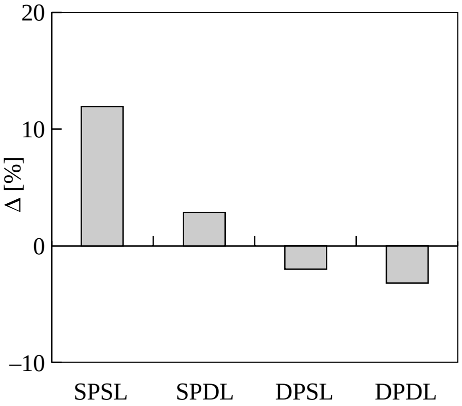

Figure 4. Improvement, Δ, of performance from first to second training period for the four different types of sessions of experiment 3: SPSL same target at control (same) location; SPDL same target at transfer location; DPSL different target at control location; DPDL different target at transfer location. Mean of 16 training sessions (eight observers), respectively.

50% chance level, and that this immediate recognition is 'translation invariant'.

The tests at control and transfer locations had been performed in separate blocks in experiment 1. As can be seen from the time course of test performance in figure 3 the order of the two test blocks had only minor effects: regardless of whether transfer is tested first (open triangles) or second (filled triangles) performance is at about the same level, despite some small internal variation. Even in the first ten trials after training there is no indication of memory decay in transfer tests that would reveal at least partial transfer of learning. An analysis of variance reveals that performance at the control location decays slightly ($F[2, 22] = 3.89$, $p < 0.05$) during the first test block (filled squares), but otherwise remains remarkably constant.

Performance at the control location is superior even if only the second test blocks of experiment 1 are considered. Given that these trials were separated from the training by a delay of more than 1 min (i.e. the time required for the first half of the test), positional specificity cannot be attributed to sensory after-effects. Finally, in contrast to the blocked test of experiment 1, the location of the test pattern in experiment 2 is generally unpredictable, yet both results are almost identical. Obviously, the degree to which spatial attention can be focused to a point in visual space does not influence positional specificity of the results. These findings contradict a specific model for invariant object recognition proposing positional normalization by shifts of a window of attention (Olshausen *et al.* 1993; Van Essen *et al.* 1994). Recent experiments investigating translation invariance in same–different tasks led to similar conclusions (Dill & Fahle 1997).

(*b*) *Testing with continuous reinforcement*

In the above experiments, training success and positional specificity were assessed by tests without feedback. Learning can, however, also be observed under conditions of training: if a subject performs better in a second round of training the improvement relative to the first training can be considered to be the result of a learning process. This way of testing learning and memory has the advantage that, under continued reinforcement, problems like reduced motivation or fast extinction of memory play a minor role. We wondered whether in a second round of training with the same target pattern, but at a new location, subjects would perform better than in the first learning period. In experiment 3 subjects were trained twice at either the same or at different locations. The target pattern in both rounds of training was either identical or not.

In the following we analyse only the start level, i.e. the first 30 trials of the training periods: data are presented as the difference Δ = (percentage correct in training 2) − (percentage correct in training 1). If, immediately after successful training, the same pattern is used for a second round of training at the same location, then subjects reached the criterion very fast and were significantly better than at the beginning of the first training ($t[7] = 2.56$, $p < 0.05$; figure 4 SPSL). If the learned pattern is used to train the subject at a transfer location (SPDL) the improvement compared to the first training is only marginal ($t[7] = 2.01$; $p > 0.05$). Hence, even under conditions of continued reinforcement there is no evidence for significant learning transfer to new locations.

The double-training results with different target patterns show that the memory acquired during training is specific not only to the position in the visual field, but also to the particular target pattern. Pattern-discrimination learning does not profit from a preceding training with a different target stimulus, regardless of whether the location was changed between the two training rounds (DPDL; $t[7] = 0.89$; $p > 0.1$) or not (DPSL; $t[7] = 0.87$; $p > 0.1$). Translation-sensitive improvement of training performance, therefore, does not arise from generally increased efficiency of information processing at that part of the visual field.

4. DISCUSSION

The results from our training experiments confirm the earlier finding (Nazir & O'Regan 1990) that memory for a learned discrimination between novel visual patterns is not invariant to translation in the visual field. They further indicate that positional specificity cannot be attributed to processes independent of pattern recognition and memory: the effect of translation is too robust to be explained by sensory after-images, it is not evoked by response biases and there is no evidence for attentional contributions. It should also be noted that these findings are not specific to bilaterally symmetric stimuli (cf. figure 2) nor to transfer between locations that are symmetric relative to the fovea (Nazir & O'Regan 1990).

Since performance after transfer is clearly above chance level, some information must be available irrespective of location. Our analysis reveals that this invariant information does not result from transfer of discrimination learning, but is based on some immediate ability. The three main findings supporting this view are: first, accuracy is at the same level in transfer tests as it was at the beginning of the training; second, transfer test performance is constant and gives no indication of memory decay—the latter can only be observed at the trained position; and third, performance under retraining with the same pattern at a new location is no better than the initial conditioning.

The acquired memory is, therefore, largely—if not completely—specific to the location of training. Accuracy in transfer tests does not profit from the training period, but can be explained by immediate abilities that the subjects possess right from the start of the experiment: at the beginning of the training the visual system is not a tabula rasa. Inbuilt detectors for relatively simple features and memories of numerous visual stimuli experienced during life may provide the brain with some hints to the structure of the target, allowing partial identification just after the first or a few trials. While performance achieved by this immediate input interpretation can be significantly above chance, it may not be sufficient to discriminate distractors from the target at the required criterion level. A novel learning process has to be initiated in order to establish a more precise definition of the target. Because training is restricted to a particular location in the visual field, the newly acquired memory is no longer available for comparison with the visual input during transfer tests at a different location. The immediate interpretation, on the other hand, could well be translation invariant, for example, because the representation to which the system refers has been built up after repeated experience at different locations in the visual field. For simple visual features, like oriented lines, invariant detectors may even be built into the brain, much in the same way as complex cells pool information from several position-specific simple cells in early visual processing stages (Hubel & Wiesel 1962; Földiák 1991).

Experiments in a *same–different* paradigm have shown that, even without any obvious learning, performance with novel visual stimuli (such as the ones in figure 1) is not entirely invariant to translation (Kahn & Foster 1981; Foster & Kahn 1985; Dill & Fahle 1997). The two tasks involve quite different decision processes. Yet they may have in common that the translation-invariant knowledge that the visual system provides about a stimulus is not sufficient for error-free recognition. By referring to additional position-specific information the brain may at least improve performance at a particular location, though this improvement does not transfer across the visual field. Hence, both kinds of tasks uncover limitations of translation invariance in the human visual system.

M.D. was supported by Boehringer Ingelheim Fonds. M.F. was supported by the von Humboldt society (Max Planck Prize) and by Deutsche Forschungsgemeinschaft (SFB 307). We thank H. Bülthoff, P. Dayan, S. Edelman, G. Geiger, T. Nazir, K. O'Regan and T. Poggio for valuable discussions and two anonymous referees for helpful comments.

REFERENCES

Biederman, I. & Cooper, E. E. 1991 Evidence for complete translational and reflectional invariance in visual object priming. *Perception* **20**, 585–593.

Dill, M., Wolf, R. & Heisenberg, M. 1993 Visual pattern recognition in Drosophila involves retinotopic matching. *Nature* **365**, 751–753.

Dill, M. & Heisenberg, M. 1995 Visual pattern memory without shape recognition. *Phil. Trans. R. Soc. Lond.* B **349**, 143–152.

Dill, M. & Fahle, M. 1997 Limited translation invariance in human visual pattern recognition. *Percept. Psychophys.* (In the press.)

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

Fahle, M., Edelman, S. & Poggio, T. 1995 Fast perceptual learning in hyperacuity. *Vis. Res.* **35**, 3003–3013.

Fiorentini, A. & Berardi, N. 1981 Learning of grating waveform discrimination: specificity for orientation and spatial frequency. *Vis. Res.* **21**, 1149–1158.

Földiák, P. 1991 Learning invariance from transformation sequences. *Neural Computation* **3**, 194–200.

Foster, D. H. 1991 Operating on spatial relations. In *Pattern recognition by man and machine* (ed. R. J. Watt), pp. 50–68. Boca Raton: CRC.

Foster, D. H. & 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. Cybern.* **51**, 305–312.

Hubel, D. H. & Wiesel, T. N. 1962 Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160**, 106–154.

Jüttner, M. & Rentschler, I. 1996 Reduced perceptual dimensionality in extrafoveal vision. *Vis. Res.* **36**, 1007–1022.

Kahn, J. I. & 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.* A **33**, 155–166.

Karni, A. & Sagi, D. 1991 Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. *Proc. Natn. Acad. Sci. USA* **88**, 4966–4970.

Macmillan, N. A. & Creelman, C. D. 1991 *Detection theory: a user's guide.* Cambridge University Press.

Nazir, T. A. & O'Regan, J. K. 1990 Some results on translation invariance in the human visual system. *Spat. Vis.* **5**, 81–100.

Olshausen, B. A., Anderson, C. H. & Van Essen, D. C. 1993 A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J. Neurosci.* **13**, 4700–4719.

O'Regan, J. K. 1992 Solving the 'real' mysteries of visual perception: the world as an outside memory. *Can. J. Psychol.* **46**, 461–488.

Ramachandran, V. S. 1976 Learning-like phenomena in stereopsis. *Nature* **262**, 382–384.

Rentschler, I., Jüttner, M. & Caelli, T. 1994 Probabilistic analysis of human supervised learning and classification. *Vis. Res.* **34**, 669–687.

Saslow, M. G. 1967 Latency for saccadic eye movement. *J. Opt. Soc. Am.* **57**, 1030–1036.

Shepard, R. N. & Cooper, L. A. 1982 *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shiu, L.-P. & Pashler, H. 1992 Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept. Psychophys.* **52**, 582–588.

Van Essen, D. C., Anderson, C. H. & Olshausen, B. A. 1994 Dynamic routing strategies in sensory, motor, and cognitive processing. In *Large-scale neuronal theories of the brain* (ed. C. Koch & J. L. Davis), pp. 271–299. Cambridge, MA: MIT Press.

Walsh, V. & Kulikowski, J. (eds) 1997 *Visual constancies: why things look as they do.* Cambridge University Press.

Received 17 February 1997; accepted 27 March 1997