PERGAMON

Vision Research 39 (1999) 2399–2410

Vision Research

Behavioral evidence for visual perception of 3-dimensional surface structures in monkeys

Takanori Uka [a], Hiroki Tanaka [c], Makoto Kato [a,b], Ichiro Fujita [a,b,c,*]

[a] *Department of Cognitive Neuroscience, Osaka University Medical School, Yamadaoka 2-2, Suita, Osaka 565-0871, Japan*
[b] *CREST, Japan Science and Technology Corporation, Yamadaoka 2-2, Suita, Osaka 565-0871, Japan*
[c] *Graduate School of Engineering Science, Osaka University, Machikaneyama 1-3, Toyonaka, Osaka 560-8531, Japan*

Received 15 June 1998; received in revised form 3 September 1998

Abstract

Human subjects perceive two crossing bars, one in front of the other, when shown a cross with disparity added to its horizontal limbs, and they also perceive neon-color spreading when shown a stereoscopic Redies–Spillmann figure. It has thus been hypothesized that the human visual system follows the principle of generic image sampling in reconstructing 3-dimensional (3-D) surface structures. Here we examine whether monkeys also perceive these surface structures. The results indicate that monkeys, like humans, perceive two crossing bars and neon-color spreading and suggest that the principle of generic image sampling may also be applied to visual perception in monkeys. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Visual illusion; Neon-color spreading; Binocular disparity; Illusory contour; The principle of generic image sampling

1. Introduction

The addition of horizontal binocular disparity to a 2-dimensional (2-D) shape enables us to perceive a 3-D structure consisting of multiple surfaces at different depths and orientations (Wheatstone, 1838). This makes binocular disparity an important cue by which 3-D surface structures are reconstructed from 2-D retinal images. The task of the human visual system, however, goes far beyond the detection of disparity when it reconstructs 3-D surface structures. One reason for this is that a specific depth cannot be assigned to contours and surfaces lacking a fixed disparity in a retinal image. Consider the horizontal limb of a cross shape, for example (Fig. 1a). If crossed disparity is added to its vertical edges, their depths can be determined as being closer to the observer than the fixation plane, whereas the depths of its horizontal edges are ambiguous. At least two surface structures could be reconstructed from this retinal image; a horizontal bar in front of a vertical bar, or a cross with arms bent to the front (Fig. 1b). Either surface structure is equally consistent with the retinal image, and yet humans almost always perceive the former (Nakayama & Shimojo, 1992). Therefore, the human visual system consistently reconstructs a unique surface structure from a retinal image that could otherwise be interpreted in many ways.

Nakayama and Shimojo (1992) proposed a rule, the principle of generic image sampling, which predicts that "when faced with more than one surface interpretation of an image, the visual system assumes it is viewing the scene from a generic, not an accidental, vantage point" (also see Freeman, 1994; Albert & Hoffman, 1995). Psychophysical evidence indicates that the human visual system follows this rule to reconstruct a 3-D surface structure from a 2-D retinal image (Nakayama & Shimojo, 1992). The perception of two crossing bars in the cross shape described above (Fig. 1a) is one example, as is neon-color spreading, which is the perception of neon-like color leaking outside the physical color boundaries, thus forming illusory contours, in a stereoscopic version of the Redies–Spillmann (s_R–S) figure (Fig. 1d; Redies & Spillmann, 1981; Nakayama & Shimojo, 1992). Considering the cross example, if a horizontal bar lies in front of a vertical bar, the shape

* Corresponding author. Tel.: + 81-6-879-2600; fax: + 81-6-879-2609.
E-mail address: fujita@bpe.es.osaka-u.ac.jp (I. Fujita)

PII: S0042-6989(98)00305-8

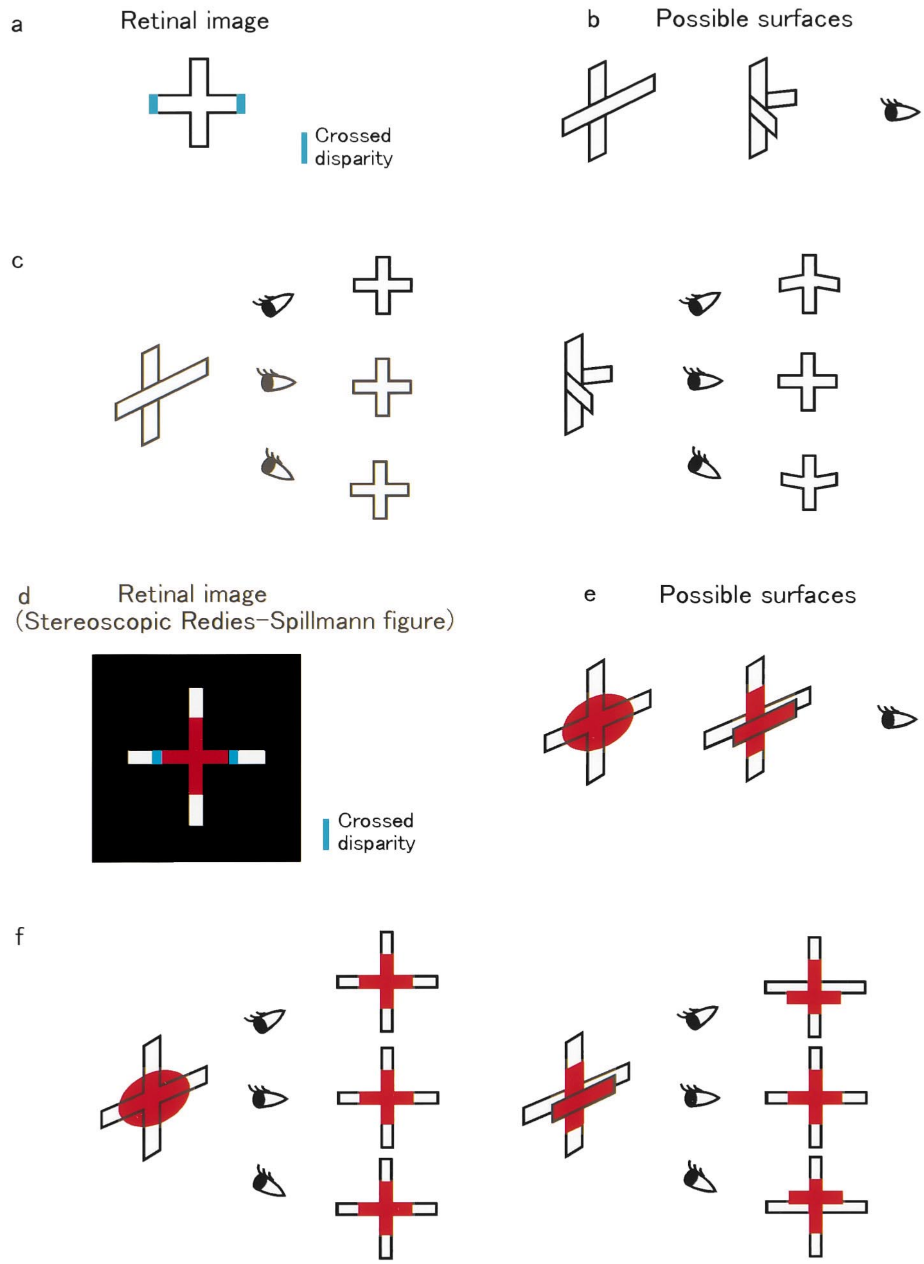

Fig. 1

Fig. 1. Two illusory images to illustrate the principle of generic image sampling of Nakayama and Shimojo (1992). (a) Retinal configuration of a cross with crossed disparity (colored blue) on the vertical edges of its horizontal limbs. (b) Two possible surface structures can be perceived from the retinal image in (a), a horizontal bar in front of a vertical bar (left), or a cross with horizontal arms bent to the front (right). Humans, however, almost always perceive the former. (c) Two crossing bars qualitatively produce the same image as the retinal image when it is projected from different vantage points (left). The cross with bent arms, on the other hand, produces different images depending on the vantage point angle (right). Thus, the retinal image in (a) is the 'generic' view of the two crossing bars, whereas it is an 'accidental' view of the cross with bent arms. The human visual system tends to reconstruct 3-D surface structures which produce generic retinal images. (d) Retinal configuration of a stereoscopic version of the Redies–Spillmann (s_R–S) figure. Crossed disparity (colored blue) is added to the vertical edges of the horizontal limbs of the red cross. (e) Two possible surface structures can be perceived from the retinal image in (d); a translucent red disk hovering over a white cross (left), or a horizontal red bar hovering over a white cross bearing a stationary vertical red bar (right). (f) The image of the first structure does not differ from the retinal image with different vantage point angles (left). The image of the second structure, on the other hand, differs depending on the vantage point angle (right). Thus, the retinal image in (d) is the 'generic' view of the translucent red disk superimposed over a white cross, whereas it is an 'accidental' view of a horizontal bar hovering over a white cross with a fixed vertical red bar.

projected from different vantage points will qualitatively remain the same as the retinal image (Fig. 1c, left). On the other hand, if a cross has arms bent to the front, the shape of the projected image will differ from the retinal image depending on the angle of the vantage point. The horizontal limbs of the cross will appear straight only when the eyes are directly in front of the object (Fig. 1c, right). Therefore, the cross with disparity added to its horizontal limbs is a 'generic' view of the two crossing bars, whereas it is an 'accidental' view of the cross with bent arms. Likewise, if a translucent red disk is hovering over a white cross, the projected image from different vantage points will qualitatively remain the same as the retinal image (Fig. 1f, left). However, if a hovering horizontal red bar is placed over a white cross bearing a vertical red bar, the projected image will differ from the retinal image depending on the vantage point angle. The horizontal limbs of the red cross will no longer be aligned with the horizontal limbs of the white cross unless the observation point is directly in front of the object (Fig. 1f, right). Therefore, the s_R–S figure is a generic view of a translucent red disk hovering over a white cross, whereas it is an accidental view of a horizontal red bar hovering over a white cross bearing a vertical red bar. The human visual system generally reconstructs 3-D surface structures which produces a generic retinal image.

The principle of generic image sampling provides an 'ecological' interpretation of how our brain reconstructs 3-D surface structures from 2-D retinal images. It does not, however, provide an explanation of its underlying neural mechanisms. In order to explore the neuronal components mediating this phenomenon, electrophysiological studies are necessary in a species like the monkey, whose visual system has been extensively characterized. As a first step towards this goal, the present study addresses the question of whether monkeys perceive 3-D surface structures as humans do. The results of the psychophysical experiments presented here indicate that monkeys perceive a cross with disparity added to its horizontal limbs as two crossing bars, as do humans, and also perceive neon-color spreading in the s_R–S figure just as human subjects do. Preliminary data in this study have been reported elsewhere (Uka, Tanaka, Kato & Fujita, 1998).

2. Methods

2.1. General procedures

Two male Japanese monkeys (*Macaca fuscata*, 8 and 11 kg b.w.) were used. The monkeys sat in a primate chair and faced a 15-inch color monitor display (screen size: 260 × 195 mm) placed 57 cm away. They were trained on a fixation task first, and then on a two-alternative forced choice discrimination task. Both tasks were controlled by a computer (PC486FS: Epson, Suwa, Japan). The position of one eye was monitored using the search coil technique, and the position signals were sampled at a rate of 100 Hz. Stimuli were presented using a PC/AT computer (Asus Computer International, San Jose, California; display resolution: 1024 × 768 pixels). Disparity was added using a liquid crystal stereoscopic modulator (SGS610: Tektronix, Beaverton, Oregon; refresh rate: 70 Hz for each eye).

Throughout the experiments, the water supply to the monkeys was restricted at their home cages. They were rewarded for correct responses with a drop of water during the experimental sessions. After each session, they were returned to their cages and given a adequate amount of vegetables or fruits. Monkey chow was always available to them ad libitum. All of the animal care and experimental procedures were in accordance with the NIH Guide for the Care and Use of Laboratory Animals (1996) and were approved by the animal experiment committee of Osaka University Medical School.

2.2. Surgery

A head holder for fixing the monkey's head to the chair was attached to the top of the skull, and a search coil for monitoring eye position was implanted under the conjunctiva, using standard aseptic surgical procedures and sodium pentobarbital anesthesia (35 mg kg^{-1}, i.p.). After the surgery, the monkeys were treated with an antibiotic (piperacillin sodium, 30 mg kg^{-1}, i.m.), an analgesic (ketoprofen, 0.5 mg kg^{-1}, i.m.), and a corticosteroid (dexamethasone sodium phosphate, 0.1 mg kg^{-1}, i.m.) to reduce potential inflammation, and were allowed to recover for at least 3 weeks before the first training session.

2.3. Pretraining

The monkeys were first trained to fixate on a gray spot (0.2 × 0.2°) on a black background (luminance 1.0 cd m^{-2}). The stimulus appeared on the fixation point which was located at the center of the monitor. The monkeys were required to maintain their fixation within a 2.0 × 2.0° window throughout the period of visual stimulus presentation in order to receive a reward. Otherwise, the task was aborted the moment their fixation was broken. After extensive training on the fixation task, the monkeys were next trained on a two-alternative forced choice discrimination task (Fig. 2). Two different shapes were used, one associated with a rightward saccade, the other with a leftward saccade. After a 1 s period of fixation on the gray spot, one of the two stimuli appeared on the fixation point for 2 s.

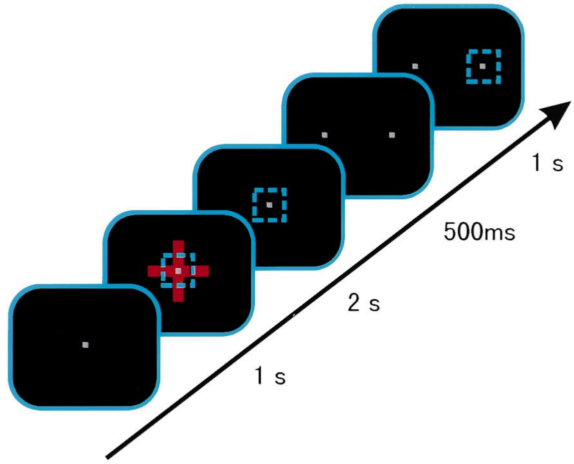

Fig. 2

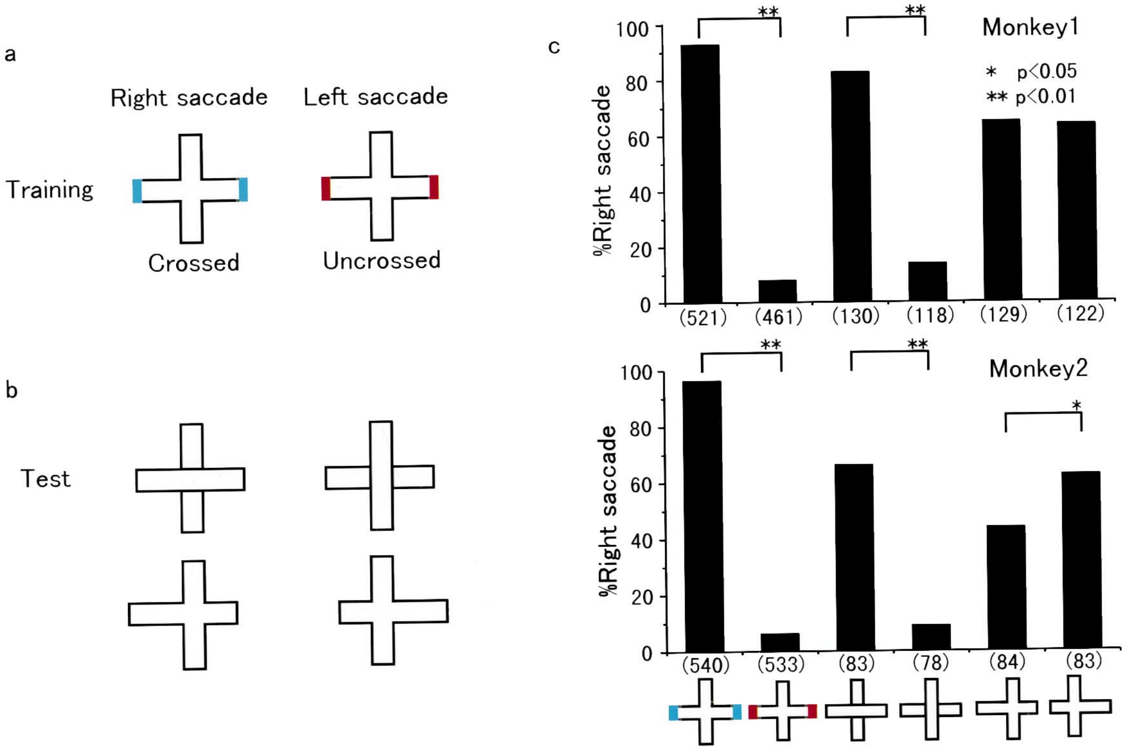

Fig. 3

Fig. 2. The 2-alternative forced-choice discrimination task. The monkeys were trained to fixate on a small gray spot at the center of the display within a 2.0 × 2.0° window, delineated by a dotted blue box here which did not physically appear in the display. After a 1 s fixation period, one of two visual stimuli is presented for 2 s. The red cross here is just one possible stimulus of several used during the pretraining and experimental trials. After the 2 s stimulus presentation, the fixation point was presented for another 500 ms, followed by two target spots located 5° to either side of the fixation point. The monkeys were required to make a saccade within 1 s to one of the two spots, depending on the initial stimulus. This task schedule was the same in all the experiments.

Fig. 3. Design and results of Experiment 1. (a) In the training sessions, the monkeys were required to make a rightward saccade when a cross with crossed disparity (colored blue) on the horizontal limbs was presented, and a leftward saccade when a cross with uncrossed disparity (colored red) was presented. The actual stimuli presented were solid red crosses with disparity. (b) In the test sessions, the monkeys were presented four test shapes: (1) a cross segmented by occluding contours into a horizontal bar in front of a vertical bar; (2) a cross segmented by occluding contours into a vertical bar in front of a horizontal bar; (3) a cross with a short right limb and a long left limb, and (4) a cross with a long right limb and a short left limb. None of the test figures contained disparity cues. During the test session, these figures were randomly interspersed among presentations of the training figures. (c) The percentage of rightward saccades for each shape presented in the test sessions is plotted for each monkey. Note the large differences between the two crosses segmented by occluding contours (the middle two bars). The rightward responses to a horizontal bar in front of a vertical bar were significantly greater than to a vertical bar in front of a horizontal bar, indicating that the discrimination learned in the training session transferred to these two test shapes. There was little (monkey 2) if any (monkey 1) difference between the control crosses with short or long right and left limbs. The number of stimulus presentations is shown in parentheses.

The fixation point disappeared after another 500 ms, immediately followed by a presentation of two gray target spots ($0.2 \times 0.2°$) located 5° to the right and left of the fixation point. The monkeys were required to make a saccade within 1 s to either one of these two spots depending on the stimulus shown, and to maintain their eye position within $2.0 \times 2.0°$ of the selected spot. The pretraining sessions were conducted using stimuli different from those used in the following experiments. Once the monkeys achieved a 90% success rate on the pretraining discrimination task, they were ready for the experiments.

2.4. Data analysis

The trials in which the monkey's fixation was broken were discarded from the analysis, thus leaving the trials in which the monkey made a saccade to either one of the two gray spots for analysis. The percentage of rightward or leftward saccades from these trials was calculated for each figure in the test sessions. For the figures that should produce rightward saccades, the percentage of rightward saccades is indicative of their correct responses, whereas the percentage of leftward saccades is indicative of their false responses. The χ^2 test was used for all the statistical analysis with the Fisher's probability $P < 0.05$ as being significant.

3. Experiment 1

3.1. Methods

The aim of this experiment was to determine whether the monkeys perceive two crossing bars or a cross with bent arms when they are shown a cross with disparity added to its horizontal limbs (Fig. 1a, b). In the training sessions which preceded the test sessions, the monkeys were trained to make a rightward saccade when a red cross ($3 \times 3°$, bar width 0.5°, luminance 5.7 cd m^{-2}) with crossed disparity ($-0.2°$) was presented, and a leftward saccade when a red cross with uncrossed disparity (0.2°) was presented (Fig. 3a). When presented with these two stimuli, humans perceived a horizontal bar in front of a vertical bar in the first case, and a vertical bar in front of a horizontal bar in the latter (personal observations), which was consistent with earlier studies (Nakayama, Shimojo & Ramachandran, 1990; Nakayama & Shimojo, 1992). After the monkeys were able to discriminate between the two shapes with an accuracy greater than 90%, discrimination tests were conducted to investigate whether the training effect extended to crosses segmented by occluding contours, that is, an actual horizontal bar in front of a vertical bar or a vertical bar in front of a horizontal bar (Fig. 3b, top). The test images were presented binocularly with no disparity cues. If the monkeys perceived two crossing bars in the training sessions, then they should be able to discriminate between the two test figures. Alternatively, if the monkeys perceived a cross with bent arms, they will not be able to discriminate between the test crosses because whether an arm is bent towards or away from the observer, the vertical illusory contours at the intersection of the crosses appear similar.

We also determined whether the training effect was transferred to crosses with either a short right limb and a long left limb or a cross with a long right limb and a short left limb (Fig. 3b, bottom). These two shapes were also presented binocularly and lacked any disparity cues. In the training sessions, these two figures were simultaneously presented to the right and left eyes, respectively (or vice versa) to induce disparity. Thus this test provided a control by determining whether the monkeys based their response on the fused binocular images rather than the monocular images presented to each eye.

So that we would not bias the monkey's responses during the test sessions, rewards were given randomly for the test shapes. During the test sessions, the test figures were presented randomly, in sequence with the training figures, at a ratio of 1:4 (monkey 1) or 1:6 (monkey 2) presentations. This kept the total percentage of reward receipt at a high level. A total of approximately 100 trials were given for the test figures, conducted over four to five daily sessions, each consisting of approximately 300 trials of discrimination.

3.2. Results

Fig. 3c shows the percentage of rightward saccades made to each shape during the test sessions for each monkey. The monkeys were able to discriminate between crosses with crossed and uncrossed disparity on the horizontal limbs with an accuracy of more than 90%. When the crosses segmented into a two-bar configuration by occluding contours were presented, the monkeys more frequently made a saccade to the right when the cross with horizontal contours was presented, and more frequently to the left when the cross with vertical contours was presented. The difference in the responses to the two crosses segmented by occluding contours was statistically significant ($\chi^2(1) = 119.6$ for monkey 1, 55.7 for monkey 2, $P < 0.0001$ in both cases). Since the monkeys received rewards randomly for the test figures, regardless of the correctness of the response, the effect seen could not have been due to reward contingency. Therefore, the learned discrimination between the training figures generalized to the test figures in both monkeys.

There was no difference between the responses to crosses with either a short right limb and a long left limb or a long right limb and a short left limb in monkey 1 ($\chi^2(1) = 0.04$, $P > 0.8$), and a small but significant difference in monkey 2 ($\chi^2(1) = 5.8$, $P < 0.05$). The percentages of rightward saccades in monkey 2 for the cross with a short right limb and a long left limb were 57.1% (24/42) and 31.0% (13/42), and for the cross with a long right limb and a short left limb were 63.9% (26/41) and 61.9% (26/42), when averaged for the first and latter half of each session, respectively (data not shown). Thus the responses of monkey 2 were similar in the early half of each test session ($\chi^2(1) = 0.3$, $P > 0.6$), and the difference in his response pattern appeared only in the latter half ($\chi^2(1) = 8.1$, $P < 0.01$). The fact that the rightward saccade response to either figure was larger than 50% (57.1%, 63.9%) during the first half of each session suggests that monkey 2 initially had a behavioral bias to making rightward saccades to both control figures, as did monkey 1. Why monkey 2 developed a leftward saccade bias during the test session for the figure with the short right limb and long left limb is currently unknown. Overall, it is unlikely that the monkeys were making discriminations based on the monocular images of the training stimuli.

3.3. Discussion

The learned discrimination between the two crosses with crossed or uncrossed disparity transferred to a discrimination between two crosses segmented by occluding contours with no disparity. This transfer of the trained response clearly demonstrates that the discrimination between the two training figures was not dependent on the local disparity cues, but rather on the perceived depth cues and/or the illusory contour cues of the stimuli. The lack of a difference in responses to the two control stimuli confirmed that the monkeys discriminated between the binocularly fused images and not the monocular images.

This transfer of the learned effect would be expected if, when looking at the training figures, the monkeys perceived two crossing bars rather than a cross with bent arms. In this case, the training stimuli would appear to have horizontal or vertical illusory contours at the intersection of the cross, a perception presented as occluding contours in the test figures. The learned response would then automatically transfer to the crosses segmented by occluding contours because they would be perceived as similar, if not identical to the training configurations. On the other hand, if the monkeys perceived the training stimuli as crosses with forward- or backward-bent arms, both of which have the same vertical illusory contours, the monkeys would have no means to discriminate between the two crosses segmented by occluding contours. The results therefore indicate that the monkeys, like humans, perceive the training images as two crossing bars.

4. Experiment 2

4.1. Methods

This experiment was one of the two experiments determining whether the monkeys perceive neon-color spreading in the s_R–S figure (Fig. 1d, e). In the training sessions which preceded the test sessions, the monkeys were trained to make a rightward saccade when a simple red cross ($3 \times 3°$, bar width 0.5°, luminance 5.7 cd m^{-2}) with crossed disparity ($-0.2°$) on its horizontal limbs was presented, and a leftward saccade when a red cross with the same disparity encircled by a translucent red disk was presented (Fig. 4a). After extensive training of this task, the luminance of the translucent red disk was gradually lowered, and the monkeys were further trained to discriminate between the two stimuli. This procedure was repeated until a luminance level was reached where correct discrimination between the two shapes occurred approximately 70% of the time, even after extensive training. At this threshold level of discrimination (the luminance of the translucent red disk was 1.2 cd m^{-2} for monkey 1 and 1.3 cd m^{-2} for monkey 2), we might expect that if the monkeys perceive neon-color spreading in the s_R–S figure, they should transfer their learned discrimination of the red cross encircled by a translucent red disk to the s_R–S figure. After the monkeys were extensively trained under the threshold condition, test sessions were conducted in which the s_R–S figure (the inner red cross: $3 \times 3°$, luminance 5.7 cd m^{-2}, $-0.2°$ of disparity on the horizontal limbs, the outer white bars: 1° each in length, bar width 0.5°, luminance 15.8 cd m^{-2}) or a control figure with gray bars (luminance 3.9 cd m^{-2}) instead of white bars (otherwise the control figure was identical to the s_R–S figure) was presented (Fig. 4b). Humans perceived neon-color spreading in the former, but not in the

Fig. 4. (Opposite, top) design and results of Experiment 2. (a) In the training sessions, the monkeys were required to make a rightward saccade when a simple red cross (right) with crossed disparity (colored blue) was presented, and a leftward saccade when a red cross with the same disparity encircled by a translucent red disk (left) was presented. (b) In the test sessions, the monkeys were presented two test figures, the s_R–S figure (left) and a control figure (right), interspersed randomly among training figure presentations. (c) The percentage of leftward saccades for each shape presented in the test sessions is plotted for each monkey. Note the difference in the percentage of leftward saccades for the s_R–S figure compared to the control figure. This indicates that the learned discrimination transferred from the training stimuli to these two figures, and suggests that monkeys can perceive neon-color spreading in the s_R–S figure. The number of trials is shown in parentheses.

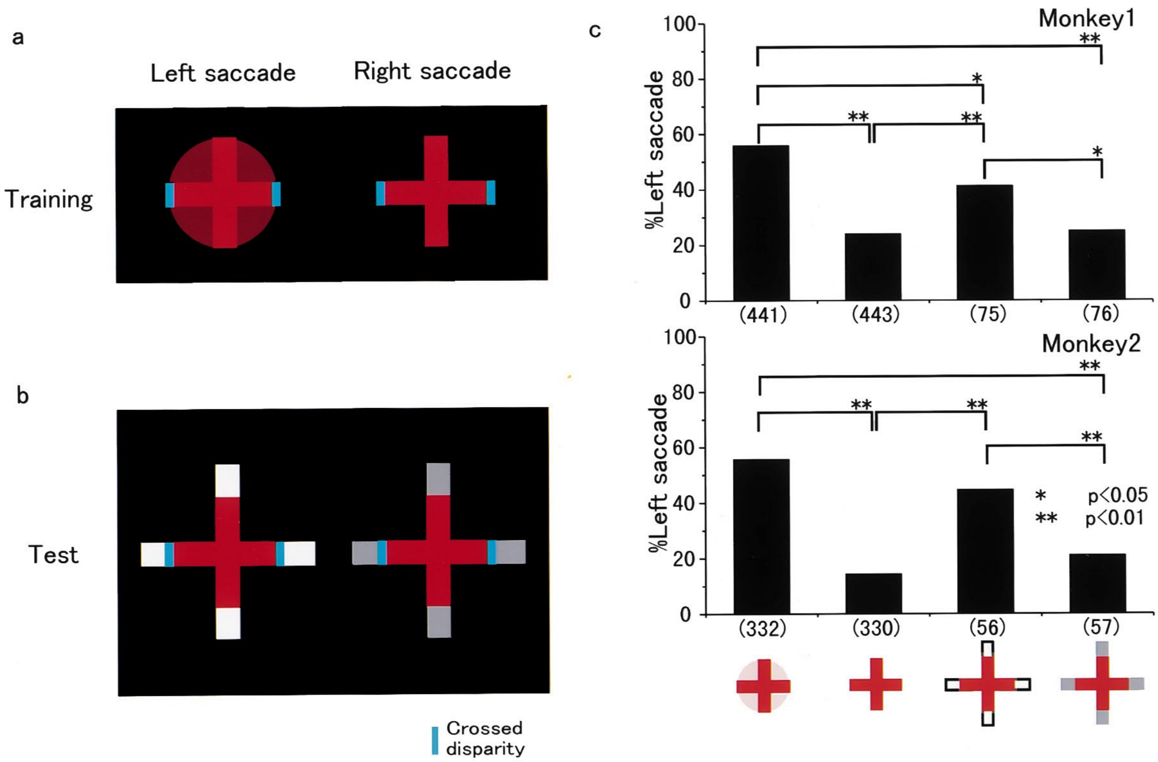

Fig. 4

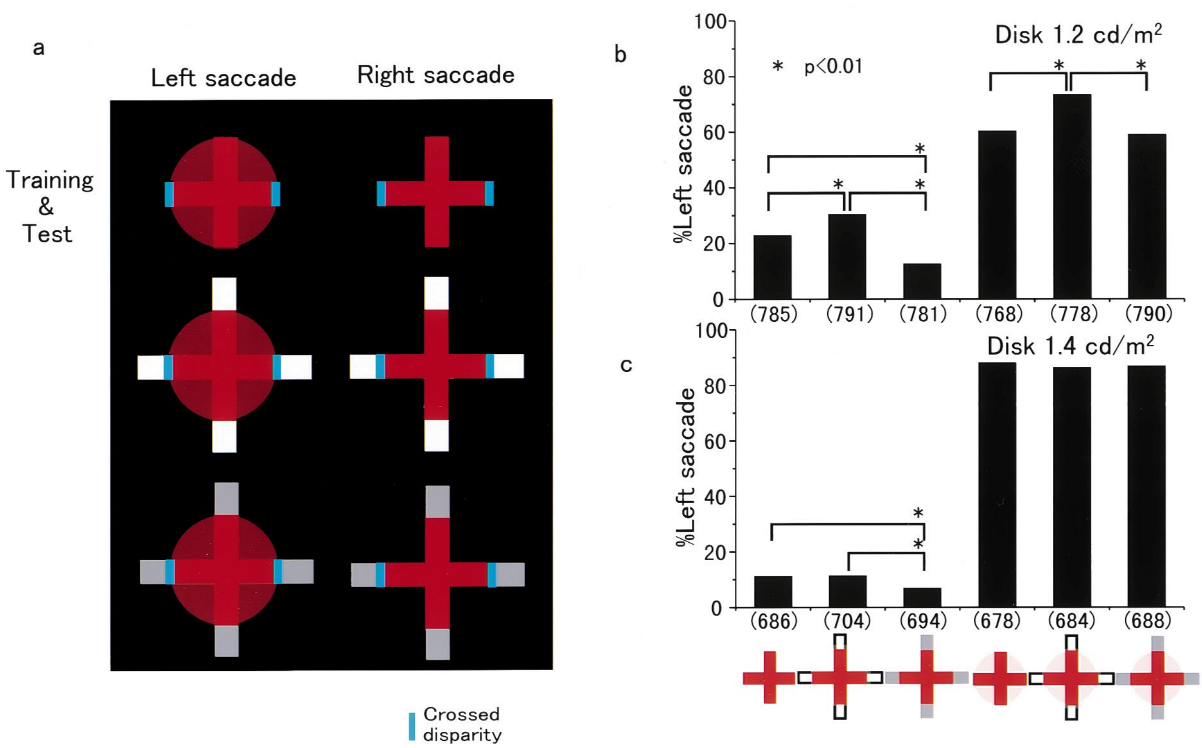

Fig. 5

Fig. 5. Design and results of Experiment 3. (a) A monkey was trained to make a rightward saccade when either a simple red cross, the s_R–S figure, or the control figure was presented, and a leftward saccade when one of these figures was presented encircled by a translucent red disk. (b) The percentage of leftward saccades for each shape with a low luminance disk (1.2 c deg^{-1} m^{-2}). The monkey's decision was biased towards leftward saccades in response to the s_R–S figure, but not the control figure or the red cross. (c) The percentage of leftward saccades for each shape with a high luminance disk (1.4 c deg^{-1} m^{-2}). Note that the leftward saccade bias for the s_R–S figure disappears at the higher luminance condition. The number of trials is shown in parentheses.

latter (personal observations), which was consistent with earlier studies (Metelli, 1974; Van Tuijl & De Weert, 1979; Nakayama et al., 1990). If the monkeys perceive neon-color spreading only in the s_R–S figure, then the discrimination learned between the training stimuli should generalize to the two test figures.

As in Experiment 1, rewards were given randomly for the test figures, which were presented randomly interspersed between the training figures at a 1:6 ratio. Approximately 60 presentations of the test figures were conducted over three to four daily sessions, each consisting of approximately 300 trials of discrimination.

4.2. Results

Fig. 4c shows the percentage of leftward saccades made to each figure during the test sessions for each monkey. The luminance of the red disk was lowered to a level where the monkeys discriminated between the simple red cross and the red cross encircled by a translucent red disk (Fig. 4a) with approximately 70% accuracy (66.4% for monkey 1, 70.7% for monkey 2). At this luminance level, it was also difficult for humans to discriminate between the two stimuli (personal observations). Both monkeys made a leftward saccade in slightly more than 50% of the trials when the red cross encircled by a translucent red disk was presented, and approximately 20% of the trials when the simple red cross was presented. The luminance of the translucent red disk was so low that its detectability was diminished, causing this figure to be confused with the simple red cross in nearly half of the trials. Thus, the average level of correct responses to the simple red cross and the cross encircled by a translucent red disk was approximately 70%. At this point, the test sessions were performed. The monkeys more frequently made a leftward saccade when the s_R–S figure was presented than when the control figure was presented, despite the fact that they received rewards randomly for the test figures. The differences between responses to the s_R–S figure and the control figure were statistically significant ($\chi^2(1) = 4.5$, $P < 0.05$ for monkey 1, $\chi^2(1) = 7.1$, $P < 0.01$ for monkey 2). This indicates that the learned discrimination to the training stimuli generalized to the s_R–S figure and the control figure.

Since the leftward saccade responses to the control figure were as low as they were for the simple red cross ($\chi^2(1) = 0.07$, $P > 0.7$ for monkey 1, $\chi^2(1) = 1.6$, $P > 0.2$ for monkey 2), it is likely that the monkeys did not perceive neon-color spreading in the control figure. Moreover, monkey 1 made significantly fewer leftward saccades for the s_R–S figure than for the cross encircled by a translucent red disk ($\chi^2(1) = 5.9$, $P < 0.05$), suggesting that the perception of neon-color spreading in the s_R–S figure may have been weaker in this monkey than the perceived brightness of the translucent red disk.

4.3. Discussion

The discrimination between a red cross encircled by a translucent red disk and a simple red cross transferred to a discrimination between the s_R–S figure and the control figure. Since the only cue they could use to discriminate between the training figures was the translucent red disk superimposed on the cross, the generalized response could be interpreted as evidence that monkeys perceive neon-color spreading in the s_R–S figure. This is further supported by the lack of a response to the control figure, which according to human studies is perceived as having little or no neon-color spreading (Metelli, 1974; Van Tuijl & De Weert, 1979; Nakayama et al., 1990). In fact, the percentage of leftward saccades made to the control figure was at a level comparable to that produced by the simple red cross.

5. Experiment 3

5.1. Methods

We conducted an additional experiment in one monkey to determine whether he could perceive neon-color spreading in the s_R–S figure. This experiment differed from the previous experiments in that the same figures were used in both the training and test sessions. The monkey was trained to make a rightward saccade when the simple red cross, the s_R–S figure, or the control figure with gray bars was presented, and a leftward saccade when any one of these shapes encircled by a translucent red disk was presented (Fig. 5a). After extensive training of this task, the luminance of the translucent red disk was gradually lowered as in Experiment 2, and the monkey was trained until he could discriminate the two sets of figures only with an accuracy of approximately 70%. At this threshold, humans perceived the brightness of the neon-color spreading in the s_R–S figure as nearly the same as the brightness of the translucent red disk encircling the red cross (personal observations). If the monkey perceives neon-color spreading in the s_R–S figure, he should confuse the s_R–S figure with the encircled cross at the low luminance level, and make a false response (leftward saccade) to it. The percentage of incorrect leftward saccades made to the s_R–S figure would thus be larger than that for the simple red cross or the control figure, though the task requirements for the three figures is the same. Therefore, the perception of neon-color spreading in the s_R–S figure would bias the monkey's decision independent of the rule of the task. The monkey's responses were scored with a low luminance disk (1.2 cd m^{-2}), which served as test sessions, as well as with a higher luminance disk (1.4 cd m^{-2}) at which humans perceived the translucent red disk as being brighter than the neon-color spreading in the s_R–S figure. Approximately 700 trials were conducted at each luminance level, with tests with the higher luminance disk recorded first, followed by tests with the lower luminance disk. In this experiment, the monkey was given rewards during the test sessions only after correct responses were made.

5.2. Results

Fig. 5b shows the results. The percentage of leftward saccades for the three figures without a red disk was lower than that for the three figures with a red disk ($\chi^2(1) = 856.9$, $P < 0.0001$), indicating that the monkey had learned to discriminate between the two sets of figures. The average level of correct responses to all six figures under low luminance was low (71.2%), as expected. Interestingly, under the low luminance but not the high luminance condition, a greater tendency towards leftward saccades was seen for the s_R–S figure, with or without the red disk, than for the other figures within each set. This difference was seen despite identical task requirements for each of the three figures in each set. The differences between responses to the s_R–S figure and those made for the simple red cross or the control figure were statistically significant ($\chi^2(1) = 11.8$, $P < 0.001$ and $\chi^2(1) = 73.4$, $P < 0.0001$, respectively), as were the differences between responses to the red-disk s_R–S figure and those for the red cross and control figures with a red disk ($\chi^2(1) = 30.0$ and $\chi^2(1) = 36.4$, respectively, $P < 0.0001$ for both).

This consistent difference, however, disappeared when the luminance of the red disk was increased (Fig. 5c). The percentage of leftward saccades for the three figures without a red disk was lower than that for the three figures with a red disk ($\chi^2(1) = 2472.4$, $P < 0.0001$). As expected, the average performance accuracy for all six figures (88.7%) was higher than that in the low luminance condition. More importantly, under high luminance, there was no difference between the responses to the s_R–S figure and to the simple red cross ($\chi^2(1) = 0.007$, $P > 0.9$). There was also no difference between the responses to the red disk s_R–S figure and to the red cross or the control figure with a red disk ($\chi^2(1) = 0.8$, $P > 0.3$ or $\chi^2(1) = 0.08$, $P > 0.8$, respectively). These results indicate that it is not merely the presence of white outer bars that can explain the biased responses to the s_R–S figure and the red disk s_R–S figure which is observed under conditions of low luminance.

Under the high luminance condition, the monkey made fewer incorrect leftward saccades in the presence of the control figure lacking the disk compared to the simple red cross and the s_R–S figure ($\chi^2(1) = 7.8$, $P < 0.01$ and $\chi^2(1) = 8.3$, $P < 0.005$, respectively). This trend was also seen under the low luminance condition. Fewer leftward saccades were made in the presence of the control figure than in response to the simple red cross and the s_R–S figure ($\chi^2(1) = 27.5$ and $\chi^2(1) = 73.4$, $P < 0.0001$ for both).

5.3. Discussion

The results provide additional evidence that monkeys perceive neon-color spreading. Under low luminance conditions, the monkey more frequently responded with leftward saccades to the s_R–S figure compared to the other two figures, regardless of whether the red disk was present or not. This indicates that the presence of neon-color spreading in the s_R–S figure made it easier to confuse the s_R–S figure with the figures with the red disk, thus more likely resulting in leftward saccades. It is possible that the monkey had a behavioral bias unrelated to neon-color spreading, and preferred to make a leftward saccade when he saw a cross with white outer bars. This is unlikely, however, since the effect disappeared when the luminance of the red disk was increased (Fig. 5c). This disappearance of the response bias under the high luminance condition is not due to a ceiling effect, because even under the high luminance condition, the monkey was not performing at maximal performance, and because there were still differences between the simple red cross and the control figure. More likely, the bias disappears under the high luminance condition because the red disk is more apparent, and confusion due to neon-color spreading in the s_R–S figure is less prone to occur. Since the response bias was also observed under the low luminance condition for the figures with a red disk, the brightness of the red disk may have been enhanced by the presence of the white outer bars as observed in human subjects (personal observations), an enhancement that is not apparent under the high luminance condition.

Under both the high and low luminance conditions, the monkey less frequently made leftward saccades in the presence of the control figure compared to the other two disk-less figures. That is, he responded correctly more often to the control figure. Perhaps the monkey had a behavioral bias to making a rightward saccade when presented this figure, or perhaps the characteristics of the control figure made it more distinguishable from the three figures with the red disk.

6. General discussion

Studies on visual illusions, which have long provided and continue to provide an important contribution to the field of human psychophysics (Coren & Girgus, 1978; Petry & Meyer, 1987), have given us valuable insight into how our visual system processes images projected onto the two retinas. In particular, such studies allow us to probe the constraints used by the visual system in processing visual inputs (Poggio, Torre & Koch, 1985).

Relatively popular and successful approaches have related the psychological responses to illusory phenomena in humans to the functional and organizational aspects of the monkey visual system (e.g. Peterhans & von der Heydt, 1989; von der Heydt & Peterhans, 1989). Clearly, however, the most direct evidence would come from psychophysical and physiological studies conducted in the same species. In furthering our efforts to identify the neural basis of the reconstruction of 3-D surface structures, we investigated whether macaque monkeys perceive 3-D surface structures as humans do. Two of the figures used in Nakayama and Shimojo (1992), a cross with disparity added to its horizontal limbs and the s_R–S figure, were presented to the monkeys in a discrimination task which determined whether learned responses to the training stimuli generalized to test figures, thereby revealing the perceptive capabilities of the monkeys. As for humans, the monkeys perceived two crossing bars in the first figure, and neon-color spreading in the latter. A possible interpretation of these results is that the monkey visual system obeys the principle of generic image sampling in reconstructing 3-D surface structures from retinal images.

6.1. Computational complexity of the perceptual phenomena

Visual illusions have been widely studied in many species, including fish, birds, rodents, cats, and monkeys. Evidence has been described for perceptions of illusory contours in cats (Bravo, Blake & Morrison, 1988), and of 'filling-in' a blind spot (Komatsu & Murakami, 1994) or a retinal scotoma in monkeys (Murakami, Komatsu & Kinoshita, 1997). While contour completion and filling-in represent fundamental properties of the visual system required for segmenting and smoothing images, the visual system must identify what objects to complete and fill-in in order to experience the two perceptual phenomena examined in the present study. This requires a much more complex interpretation of visual illusions than has been previously studied in the monkey. For instance, considering the cross with crossed disparity added to its horizontal limbs (Fig. 1a), modal completion of its contours at the intersection of the cross occurs horizontally, but not vertically. The depth of the horizontal limbs are assigned the same depth as that of its vertical contours, and not interpolated between the vertical edges of the horizontal and vertical limbs. The s_R–S figure (Fig. 1d) contains the same shape and disparity as does the cross with crossed disparity on the vertical edges of its horizontal limbs. Contour completion of this figure, however, connects the ends of the horizontal and vertical limbs of the red cross, and does not occur horizontally at the intersection of the cross. The black background area between the horizontal and vertical limbs of the red cross is assigned the same depth as that of the vertical contours of the horizontal limbs of the red cross. Thus, these two figures are uniquely perceived as two crossing bars for the first, and having neon-color spreading for the second. Each of these perceptual processes likely require a neuronal circuitry to specify the way contour completion and filling-in are to be combined. The present study shows that these processes operate in the visual system of the monkey.

6.2. Perception of 3-D surface structures in stereograms in monkeys

Bough (1970) demonstrated that monkeys can discriminate depth in random dot stereograms. In his study, the monkeys were trained to discriminate between patterns of random dots with or without a luminance-defined square, a discrimination that was transferred to patterns of random dots with or without a disparity-defined square. This study did not determine, however, whether the monkeys based their responses on perceived surfaces, or merely on the uniformity of the luminance or the disparity in the dot patterns. Similarly, Harwerth and Boltz (1979) later demonstrated that monkeys can discriminate whether there was one bar or two bars formed in random dot stereograms using binocular disparity cues. They also did not rule out whether the monkeys were merely using the location of the disparity cues within the stereogram rather than perceiving surfaces.

Our demonstration that the monkeys perceived two crossing bars in a cross with disparity added to its horizontal limbs indicates that monkeys perceive multiple surfaces at different depths in untextured stereograms. The monkeys' responses in Experiment 1 were based on the perceived surface depths and/or the illusory contour cues of the stimuli, and not on the disparity cues themselves. Such perceptions rely on the reconstruction of 3-D surface structures, suggesting that the monkey visual system is capable of reconstructing 3-D surface structures from rather sparse local disparity cues in 2-D images. However, it is still possible, though unlikely (e.g. Anderson & Julesz, 1995), that the monkeys perceived illusory contours at the intersection of the cross without perceiving depth of the two crossing bars (or vice versa).

6.3. The neural basis of the perception of 3-D surface structures

One mechanism by which neurons may relay information on surface structures is topographically, by which perceived features are represented by activated neurons in a particular spatial location. This hypothesis assumes that neurons in areas that preserve retinotopy, such as

V1 or V2 (Tootell, Switkes, Silverman & Hamilton, 1988; Tootell & Hamilton, 1989), relay on all of the perceived surface properties of a particular spot in space, including its brightness, color, and depth. Another way surface information may be represented in the brain is via single neurons that encode individual surface segments or structures. Visual areas that contain shape-specific neurons, such as V4 or the inferior temporal cortex (IT) (Gross, Bender & Rocha-Miranda, 1972; Desimone, Albright, Gross & Bruce, 1984; Tanaka, Saito, Fukada & Moriya, 1991; Fujita, Tanaka, Ito & Cheng, 1992; Gallant, Braun & Van Essen, 1993; Kobatake & Tanaka, 1994), are candidate regions for such representation.

Previously, Cowey and Porter (1979) found that lesions of the IT in the monkey impair the ability to discriminate depth in random dot stereograms. Interestingly, we recently found a neural correlate in the IT for depth perception of the two crossing bars (Uka, Tanaka, Kato & Fujita, 1997). A subpopulation of IT neurons that fired in response to a cross shape changed their response magnitude depending on whether the cross was segmented into a horizontal bar in front of a vertical bar or a vertical bar in front of a horizontal bar. This altered response occurred regardless of whether the disparity cues on the cross were crossed or uncrossed. This strongly suggests that the ventral visual pathway of the monkey in which the IT is the final structure (Mishkin, Ungerleider & Macko, 1983) plays an important role in reconstructing 3-D surface structures from 2-D retinal images. The present results provide a basis for relating perceptual responses in the monkey with neuronal responses in the IT. Future studies will determine how individual IT neurons become selectively sensitive to a particular surface structure, and may help reveal the algorithm underlying the reconstruction of 3-D surface structures from 2-D retinal images.

Acknowledgements

We thank Hiroshige Takeichi for his valuable comments on the manuscript. This work was supported by grants to I.F. from CREST of the Japan Science and Technology Corporation, and the Ministry of Education, Science, Sports, and Culture (09268222). T. U. is a recipient of the JSPS Research Fellowship for Young Scientists.

References

Albert, M. K., & Hoffman, D. D. (1995). Genericity in spatial vision. In R. D. Luce, M. D'Zmura, D. D. Hoffman, G. J. Iverson, A. K. Remmey, *Geometric representation of perceptual phenomena: papers in honor of Tarow Indow on his 70th birthday* (pp. 95–112). Mahwah: Lawrence Erlbaum Associates.

Anderson, B. L., & Julesz, B. (1995). A theoretical analysis of illusory contour formation in stereopsis. *Psychological Review*, *102*, 705–743.

Bough, E. W. (1970). Stereoscopic vision in the macaque monkey: a behavioral demonstration. *Nature*, *255*, 42–44.

Bravo, M., Blake, R., & Morrison, S. (1988). Cats see subjective contours. *Vision Research*, *28*, 861–865.

Coren, S., & Girgus, J. S. (1978). *Seeing is deceiving: the psychology of visual illusions*. Mahwah: Lawrence Erlbaum Associates.

Cowey, A., & Porter, J. (1979). Brain damage and global stereopsis. *Proceedings of the Royal Society of London B*, *204*, 399–407.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature*, *368*, 542–545.

Fujita, I., Tanaka, K., Ito, M., & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360*, 343–346.

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science*, *259*, 100–103.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.

Harwerth, R. S., & Boltz, R. L. (1979). Behavioral measures of stereopsis in monkeys using random dot stereograms. *Physiology & Behavior*, *22*, 229–234.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71*, 856–867.

Komatsu, H., & Murakami, I. (1994). Behavioral evidence of filling-in at the blind spot of the monkey. *Visual Neuroscience*, *11*, 1103–1113.

Metelli, F. (1974). The perception of transparency. *Scientific American*, *230*, 90–98.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417.

Murakami, I., Komatsu, H., & Kinoshita, M. (1997). Perceptual filling-in at the scotoma following a monocular retinal lesion in the monkey. *Visual Neuroscience*, *14*, 89–101.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, *257*, 1357–1363.

Nakayama, K., Shimojo, S., & Ramachandran, V. S. (1990). Transparency: relation to depth, subjective contours, luminance, and neon color spreading. *Perception*, *19*, 497–513.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *Journal of Neuroscience*, *9*, 1749–1763.

Petry, S., & Meyer, G. E. (1987). *The perception of illusory contours*. New York: Springer-Verlag.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, *317*, 314–319.

Redies, C., & Spillmann, L. (1981). The neon color effect in the Ehrenstein illusion. *Perception*, *10*, 667–681.

Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex. *Journal of Neurophysiology*, *66*, 170–189.

Tootell, R. B. H., Switkes, E., Silverman, M. S., & Hamilton, S. L. (1988). Functional anatomy of macaque striate cortex. II. Retinotopic organization. *Journal of Neuroscience*, *8*, 1531–1568.

Tootell, R. B. H., & Hamilton, S. L. (1989). Functional anatomy of the second visual area (V2) in the macaque. *Journal of Neuroscience*, *9*, 2620–2644.

Uka, T., Tanaka, H., Kato, M., & Fujita, I. (1997). Disparity sensitivity and representation of 3-dimensional surface structure in monkey inferior temporal cortex. *Society for Neuroscience Abstract*, *27*, 2063.

Uka, T., Tanaka, H., Kato, M., & Fujita, I. (1998). Perception of surface structure obey the principle of generic image sampling in monkeys. *Investigative Ophthalmology & Visual Science Supplement*, *39*, 854.

Van Tuijl, H. F. J. M., & De Weert, C. M. M. (1979). Sensory conditions for the occurrence of the neon spreading illusion. *Perception*, *8*, 211–215.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *Journal of Neuroscience*, *9*, 1731–1748.

Wheatstone, C. (1838). Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transaction of the Royal Society of London*, *2*, 371–393.