

## Depth perception from pairs of overlapping cues in pictorial displays

BIRGITTA DRESP<sup>1,\*</sup>, SÉVERINE DURAND<sup>2</sup> and STEPHEN GROSSBERG<sup>3</sup>

<sup>1</sup> *Laboratoire des Systèmes BioMécaniques et Cognitifs, Ecole Nationale Supérieure de Physique, IMF, UMR 7507 CNRS-Université Louis Pasteur, 67400 Illkirch/Strasbourg, France*

<sup>2</sup> *Institute of Neuroinformatics, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*

<sup>3</sup> *Department of Cognitive and Neural Systems and Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston MA 02215, USA*

Received 1 February 2001; revised 9 July 2001; accepted 31 July 2001

**Abstract**—The experiments reported herein probe the visual cortical mechanisms that control near–far percepts in response to two-dimensional stimuli. Figural contrast is found to be a principal factor for the emergence of percepts of near *versus* far in pictorial stimuli, especially when stimulus duration is brief. Pictorial factors such as interposition (Experiment 1) and partial occlusion (Experiments 2 and 3) may cooperate, as generally predicted by cue combination models, or compete with contrast factors in the manner predicted by the FACADE model. In particular, if the geometrical configuration of an image favors activation of cortical bipole grouping cells, as at the top of a T-junction, then this advantage can cooperate with the contrast of the configuration to facilitate a near–far percept at a *lower* contrast than at an X-junction. Varying the exposure duration of the stimuli shows that the more balanced bipole competition in the X-junction case takes longer exposures to resolve than the bipole competition in the T-junction case (Experiment 3).

**Keywords:** Perceptual grouping; depth perception; pictorial cues; occlusion; T-junctions; X-junctions; FACADE theory; Boundary Contour System.

### INTRODUCTION

The geometrical characteristics of visual stimuli that determine figure–ground segregation, or how we perceive what appears near to us and what appears further away in two-dimensional images, were described and categorized for the first time by Leonardo da Vinci in the 17th century in his *Trattato della Pittura*. Important cues available to the visual system for the processing of figure and ground, or

\*To whom correspondence should be addressed.

relative depth, in a 'cartoon world' where objects and scenes are represented by two-dimensional drawings, pictures, or computer-generated images, are aerial and linear perspective, relative size, interposition or partial occlusion of parts and wholes, and relative visibility of objects.

The relative visibility of an object in a picture or a scene is partially determined by local variations in luminance, or brightness contrast. Generally, objects with a stronger contrast have been found to attract visual attention away from other objects with a weaker contrast (Yantis and Jones, 1991; Dresch and Grossberg, 1999). How relative visibility correlates with perceived depth is demonstrated by observations showing that the apparent depth of a given region within the visual field is determined by local brightness or hue (Egusa, 1983). In experiments on the *kinetic depth effect*, Schwartz and Sperling (1983) have shown that brightness contrast is used by the visual system to render this depth phenomenon perceptually non-ambiguous, that is, to resolve the problem of what is near and what is far in the stimulus. This observation, which they termed 'proximity-luminance-covariance' in binocular viewing, has motivated other psychophysical studies of contrast as a depth cue. O'Shea *et al.* (1994), for example, have shown that the higher-contrast stimulus of a pair of stimuli appears nearer than the lower-contrast stimulus in monocular viewing. These authors concluded that relative visibility, or contrast, should be sufficient as a pictorial depth cue because it simulates the optical consequences of aerial perspective. Possible interactions of contrast with other cues such as interposition or partial occlusion, which are considered as major determinants of pictorial depth (e.g. Kanizsa, 1979, 1985), were not taken into account in these studies.

Whether different pictorial cues to near and far are used in combination or separately by the visual system is really an open question. The idea that information provided by multiple image cues needs to be combined for the generation of unified estimates and percepts of depth (and shape) has been discussed extensively by Gibson (1950), for example. Within the framework of a computational approach to the perception of apparent depth in pictures and scenes, hypotheses of depth *cue combination* relating to Bayesian theories of cue combination have been proposed (Landy *et al.*, 1995). Therein it is suggested that a single cue from one view of a scene cannot be used to promote itself, and thus interaction between different depth cues is inevitable. It is furthermore stated that such a cooperation of cues, or sharing of information, must occur if two qualitatively different depth cues are to contribute to the depth percept at a given location. A conflict between cues would occur in situations where an unambiguous cue fails to disambiguate an ambiguous one. Most of the studies on cue combination or conflict, however, tend to focus on interactions between stereo disparity cues and other types, such as motion parallax or pictorial cues (e.g. Stevens *et al.*, 1991), rather than on interactions between the different pictorial cues themselves.

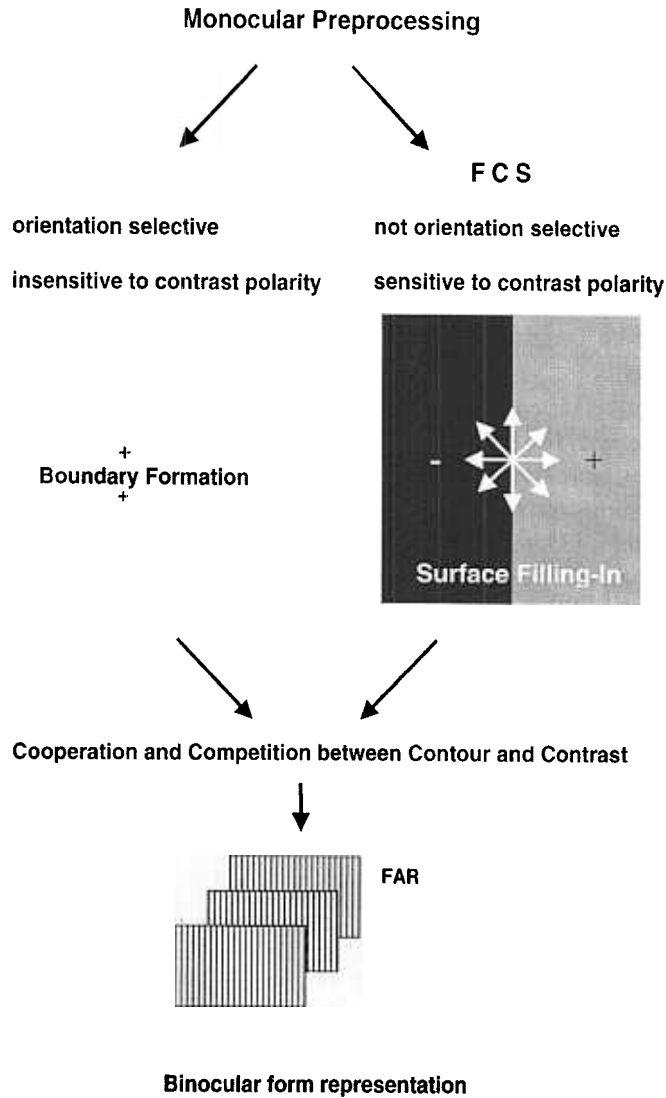
Grossberg (1994, 1997) introduced FACADE (Form-And-Color-And-Depth) theory in order to clarify how the visual cortex gives rise to 3D percepts of objects

separated from their backgrounds. A satisfying consequence of this analysis was the demonstration that the same cortical mechanisms clarify how 2D pictures give rise to percepts of objects separated from, and in front of, their backgrounds. A major theme of the theory is that pictorial cues, such as contrastive and geometrical relationships among contours, can activate several different types of cooperative and competitive processes whose interactions give rise to 3D scenic percepts and 2D pictorial percepts.

Figure 1 schematically summarizes relevant computational hypotheses of the model. After monocular preprocessing, the visual input is fed in parallel into two subsystems of the cortical network: the BCS (*Boundary Contour System*) and the FCS (*Feature Contour System*). The BCS forms boundary representations of an image. It is orientation-selective, and its outputs become insensitive to the sign of contrast by pooling signals that are derived from opposite image contrasts. This latter property enables the BCS to form boundaries that can completely surround objects in front of textured backgrounds. The FCS forms visible surface representations of an image. It is sensitive to contrast polarity, and uses a combination of filtering and filling-in mechanisms to compensate for variable illumination, and to fill-in surfaces using brightness and color signals from which the illuminant has been discounted.

These complementary BCS and FCS properties are able to generate mutually consistent percepts via interactions between the two systems. Output from the BCS to the FCS are used to define the boundaries within which the FCS filling-in occurs at multiple processing stages. The BCS outputs represent different depths from the observer, and they can 'capture' and fill-in FCS surface properties that are spatially aligned with them at the corresponding depths. Outputs from the FCS to the BCS help to select and strengthen those boundaries which are consistent with successfully filled-in surface representations, and to suppress other boundaries. This feedback loop between BCS to FCS and back to BCS has been predicted to initiate figure-ground separation, and to do so at a cortical processing stage no later than cortical area V2. Other interactions between BCS and FCS complete figure-ground separation at a processing stage that is compared with data from cortical area V4.

As noted above, the model predicts that contrastive and geometrical properties of a 2D picture can influence the BCS and FCS in different ways, and thereby alter the ensuing figure-ground percept (Grossberg, 1997). The present article explores this possibility on the basis of psychophysical data. In three experiments, we have tested interactions of contrast and contour factors with other pictorial depth cues such as interposition and partial occlusion. We expect that the emergence of near-far percepts in briefly presented, two-dimensional images may be strongly influenced by the contrast of a given visual object. However, the contrast cue is shown to cooperate or compete with other pictorial cues, including the geometrical relationships among image contours, in the manner predicted by FACADE theory.

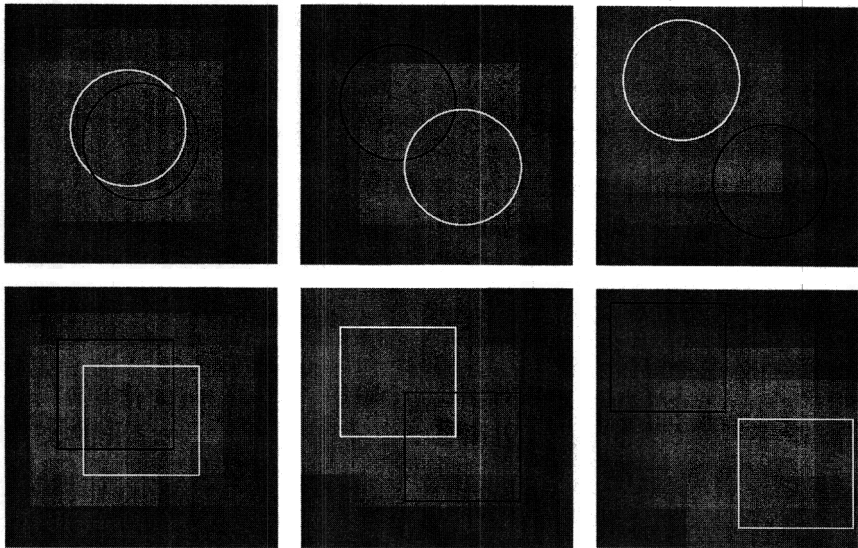


**Figure 1.** FACADE theory (Grossberg, 1994) provides a model for the formation of 3D percepts from 2D images. After monocular preprocessing, the visual input is fed in parallel into two subsystems of the cortical network: the BCS (*Boundary Contour System*) and the FCS (*Feature Contour System*). The BCS is orientation-selective and pools opposite contrast polarities. It generates early representations of contour groupings in the image. The FCS generates visible surface representations that are sensitive to contrast polarity. At an early stage of processing, monocular outputs of both subsystems cooperate and compete to determine which boundaries and surfaces will be selected. A relatively strong grouping signal that coincides with a relatively strong contrast signal stands a better chance to win the competition than weaker coinciding grouping signals. The binocular form representations that emerge after this competition are stored and activate different representations of surface depth within the visual cortex. The stronger groupings that have survived the competition before binocular integration are predicted to be perceived as ‘nearer’ by an observer. The weaker groupings that have lost the competition are predicted to be perceived as ‘further away’.

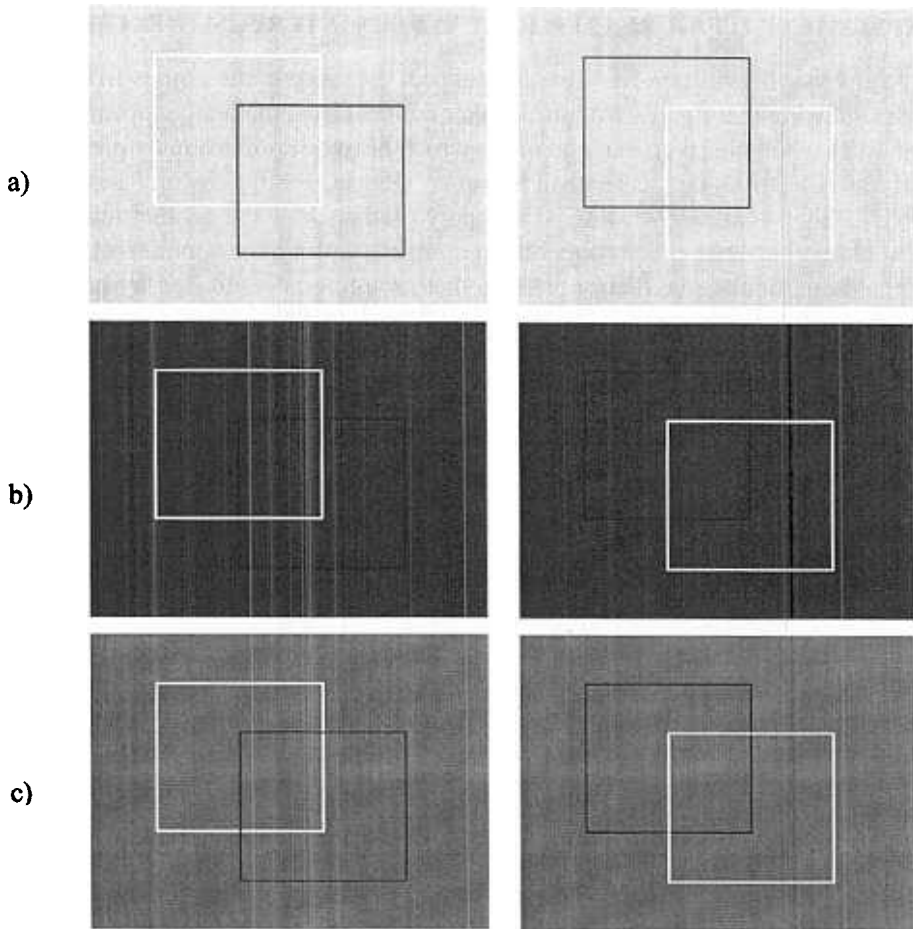
### EXPERIMENT 1: FIGURE CONTRAST VERSUS INTERPOSITION CUES

Psychophysical evidence for stimulus contrast as a depth cue comes from experiments showing that figures with the weaker contrast systematically appear to be further away when observers have to judge which of two simultaneously presented visual forms seems to be 'nearer than the other' (Egusa, 1983; Schwartz and Sperling, 1983; O'Shea *et al.*, 1994). FACADE theory predicts how the geometrical arrangement of contours can either cooperate or compete with image contrasts (Grossberg, 1997). In particular, the theory predicts that grouping is controlled by *bipole cells* which can be activated if there are colinear, or almost colinear, signals on both sides of the cell body. Bipole cells were predicted to exist in the early 1980s (Cohen and Grossberg, 1984; Grossberg, 1984; Grossberg and Mingolla, 1985a, b). Psychophysical data (Field *et al.*, 1993; Polat and Sagi, 1994; Dresch and Grossberg, 1997) and neurophysiological data (e.g. von der Heydt *et al.*, 1984; Peterhans and von der Heydt *et al.*, 1989; Kapadia *et al.*, 1995; Polat *et al.*, 1998) have provided accumulating evidence in support of the prediction that bipole cells control perceptual grouping. The present experiments test their predicted role in figure-ground perception.

At locations where two contours intersect, such as the X-junctions in the intersecting squares and circles of Fig. 2, the geometrical effects of the contours are approximately balanced if the contours are oriented at the intersection points in equally salient orientations, and if they both extend sufficiently far in both direc-



**Figure 2.** Pairs of outlined forms (squares or circles) with varying degrees of interposition were presented to the observers in Experiment 1. The luminance of the background was varied to create noticeable differences in contrast between the left and the right stimulus of a pair. Observers had to decide as quickly as possible which circle or square of a pair appeared to be 'nearer' to them than the other. Exposure duration of the stimuli was 128 ms.



**Figure 3.** Varying background luminance allows manipulation of the relative visibility of dark and bright figures, presented in random order on the right or the left hand side of a given stimulus pair in Experiment 2. This representation does not reproduce the exact luminance values that were used in the experiment. It just roughly shows the principle of the manipulation. The four pairs in panels a and b illustrate how cues of relative contrast and interposition cues cooperate to determine which figure of a stimulus pair is seen as nearer: the dark square of a pair is seen as nearer in the examples given in panel a, the bright square of a pair is seen as nearer in the examples given in panel b. Such a cue combination effect is demonstrated by the results of Experiment 2. The two pairs in panel c show that the near–far percepts remain ambiguous when the interposition cue is provided without the contrast cue.

tions from the intersection points to adequately activate the corresponding bipole cells. Grossberg (1997) predicted how, under such circumstances, contours with a higher relative contrast could facilitate a near–far percept. FACADE theory is generally consistent with models of visual discrimination where events with contrasts of greater magnitude compete with events with contrasts of weaker magnitude.

To highlight the role of the contrast intensity of visual objects in the genesis of near–far percepts in 2D stimuli, we have designed an experiment where the luminance contrast of briefly flashed pairs of forms is varied simultaneously with other figure properties such as interposition, shape, and relative size (see Figs 2 and 3). As in the experiments of O'Shea *et al.* (1994), the observers had to judge which form of a given pair appeared to be nearer than the other, with the difference that, in our tasks, the observers were not given the opportunity to look at the stimuli for as long as they wanted.

### *Subjects*

Four subjects (21 to 25 years old), three of them male and one of them female, all students at the Ecole Nationale Supérieure de Physique de Strasbourg, participated in the experiments. They were all volunteers, had normal vision, and were naive with regard to the purpose of the experiment.

### *Stimuli*

The stimuli (see Fig. 2) were presented binocularly on a high-resolution computer screen (Sony, 60 Hz, non-interlaced). They were generated with an IBM compatible PC (HP 486) equipped with a VGA Trident graphic card. The luminance of the grey levels of the screen was carefully measured with an OPTICAL photometer used in combination with the appropriate software. The length of each side of a square figure was 2.5 degrees of visual angle, the diameter of a circle was 1.5 degrees of visual angle. All line contours were one minute of visual arc thick. In one set of conditions, two figures of a pair had equal size; in another set of conditions, one figure was half the size of the other. In this case, the position of the smaller figure in a given pair (left or right) was randomly generated within a session. The spacing between two figures of a given pair was varied. In one condition, three quarters of the figures were overlapping; in two other conditions, one-half and one-quarter of the figures were overlapping. In a fourth condition, the two figures of a pair were completely separated (no interposition cues) by a gap of about 10 arc min between the nearest contours. These different overlap conditions were also varied randomly within an experimental session.

Figure pairs with different shapes (pairs of squares or pairs of circles) were presented in separate blocks. A dark and a bright figure were presented in each pair. A dark or a bright figure in a given pair appeared as many times to the left as to the right, in random order. While the luminance of the figures was constant at 16 cd/m<sup>2</sup> for bright figures and 2 cd/m<sup>2</sup> for dark figures, the luminance of the background was varied to equal 4, 6, 8, 10, and 12 cd/m<sup>2</sup>. Different luminance combinations were presented in random order within an experimental session. The combination between figure and background luminances led to five different levels of relative visibility for a bright figure, and to five different levels for a dark figure. How varying the background luminance in such a way allows observers to manipulate

the relative visibility, or contrast, of a dark or a white figure in a stimulus pair is illustrated schematically in Fig. 3.

The contrast levels (signed Michelson contrast) of the different figure-ground combinations were calculated as follows:

$$(L_{\min} - L_{\max}) / (L_{\min} + L_{\max}),$$

for contrasts with negative sign (dark figures) and:

$$(L_{\max} - L_{\min}) / (L_{\min} + L_{\max}),$$

for contrasts with positive sign (bright figures). For example, a figure of 2 cd/m<sup>2</sup> on a field of 12 cd/m<sup>2</sup> has a signed Michelson contrast of -0.71.

### *Procedure*

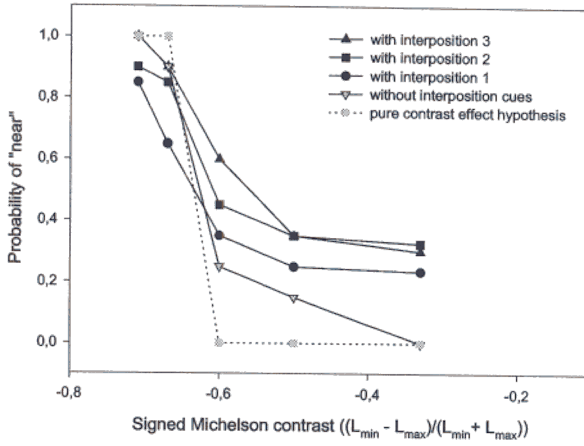
A given pair of figures was flashed for 32 ms (two frames) on the screen, and observers had to decide as quickly as possible, by pressing one of two response keys on the computer keyboard, which figure of the pair (the left or the right one) seemed nearer than the other. The choice of the observer and the response time were recorded. A new trial was initiated about 1000 ms after the keyboard signal. The shape and relative size conditions were presented in separate blocks, the interposition and polarity factors were varied within a given block. The number of observations recorded for each level of each factor tested was perfectly balanced, and each observer was run in a total of 1600 trials.

### *Results*

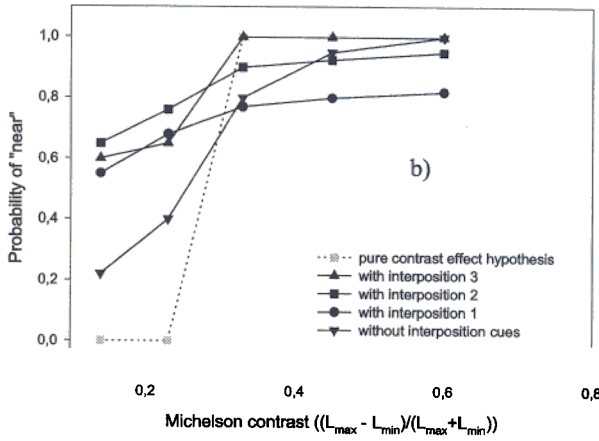
The results from Experiment 1 are represented in Fig. 4a and b. The probability of 'near' responses is plotted as a function of Michelson contrasts of bright and dark figures in pairs of stimuli with interposition cues, and in pairs without interposition cues.

The effect of figure contrast on an observer's judgement of which image in a given pair is seen as being nearer when no other cue is available in the stimulus is reflected by results in the 'no interposition' condition, which follow the predicted contrast effect. Images with higher relative contrast in a picture are likely to be seen as being nearer to the observer. The global effect of figure contrast on observers' judgements is statistically significant ( $F(9, 27) = 24.967$ ;  $p < 0.001$ ). Whether a dark figure overlapped a bright one, or a bright figure overlapped a dark one had no effect on the data. Only contrast intensity determined whether a given figure of a pair was seen as 'nearer' than the other. The effect of interposition cues as a function of figure contrast on observers' judgements is represented in three curves correspond to the three different degrees of interposition used here. The graphs show that figure pairs without interposition cues yield a stronger contrast effect than figure pairs with interposition cues. While the global effect of interposition on observers'





(a)



(b)

**Figure 4.** The probability that the left or the right figure of a given stimulus pair is seen as 'nearer' is plotted as a function of signed Michelson contrasts for dark (4a) and bright (4b) figures of a pair. Each probability is estimated on the basis of a total number of 160 observations per datapoint. The same data were analyzed twice: in Fig. 4a, the data were averaged over the contrasts of the lighter figure, and plotted against the contrast of the darker figure of the pair. In Fig. 4b the data were averaged over the contrasts of the darker figure, and plotted against those of the lighter figure. Four of the curves show data for the four different levels of the 'interposition' factor. A fifth curve (grey square symbols) shows the effect that is predicted under the (strong) assumption that contrast alone determines whether the left or the right figure of a pair is seen as 'nearer'. The 'pure contrast effect' hypothesis is based on the relative visibility of a figure of a stimulus pair with regard to the background (for a schematic illustration, see Fig. 3). When the dark figure of the stimulus pair has a strong contrast, the bright figure of that pair will automatically have the weaker contrast. Models of visual discrimination generally predict that the visual event with the contrast of the greater magnitude will yield a perceptual decision more readily than the event with the weaker contrast. Within a strictly probabilistic framework and a binary choice paradigm like the one used here, the figure with the stronger contrast is assigned a probability of 1 to be seen as nearer, the figure with the weaker contrast a probability of 0. It is shown that the psychophysical observations follow this 'pure contrast effect' hypothesis more closely when pairs of stimuli do not contain interposition cues.

judgements is statistically not significant, the interaction between interposition and figure contrast is found to be highly significant ( $F(27, 81) = 2.468$ ;  $p < 0.001$ ).

As the data suggest, interposition cues may become a more important determinant of the subjects' perceptual judgements when the contrast of a given figure is relatively weak. The relative effect of interposition decreases as the contrast of a given figure increases. In fact, further analyses of variance under conditions with figures of strong contrast only, selected *ad hoc* as Michelson contrasts of  $-0.71$  and  $-0.67$  for dark figures, and Michelson contrasts of  $0.60$  and  $0.45$  for bright figures, show that the effect of interposition is not statistically significant when figure contrast is relatively strong. However, when conditions with figures of the weaker contrasts only are grouped in the analysis, the effect of interposition is found to be statistically significant ( $F(3, 9) = 8.96$ ;  $p < 0.05$ ).

The amount of contrast carried by a given figure, whether bright or dark, was also found to have a significant effect on observers' response times. Mean response time is found to decrease systematically when absolute figure contrast increases. The effect is statistically significant ( $F(9, 27) = 8.043$ ;  $p < 0.001$ ) and reproduces the classic psychophysical observation that response latencies decrease with stimulus intensity in various perceptual tasks (e.g. Pins and Bonnet, 1996).

Neither the shape of the figures (that is, whether they represented squares or circles) nor the relative figure size (that is, whether a pair of figures with equal or with different size was presented) had statistically significant effects on either perceptual judgments or response times. Interactions between these factors, and of each factor with figure contrast, were tested. None of these interactions was found to be statistically significant.

## EXPERIMENT 2: FIGURE CONTRAST VERSUS PARTIAL OCCLUSION CUES

Experiment 2 was designed to further test the FACADE theory prediction that bipole cells are involved in the interaction between contrastive and geometrical properties of an image. In particular, FACADE theory predicts that, other things being equal, geometrical factors, such as the strength of perceptual groupings, can more powerfully compete with contrastive factors at T-junctions than at X-junctions. This is because the bipole cells respond at a T-junction more vigorously to the top of a T than to its stem. This advantage is predicted to initiate the process whereby the surface that is attached to the top of the T appears to occlude the surface that is attached to the stem of the T (see Grossberg (1997) for details). In contrast, at an X-junction, bipole cells can compete well with each other in both orientations, other things (including contrast) being equal. This more balanced situation can inhibit the selection of an occluder, or can elicit a more bistable percept of occluding and occluded surfaces, as one arm of the X-junction gains dominance over the other. FACADE theory also predicts how boundary and contrast effects can cooperate or compete when a prescribed boundary configuration is fixed and contrast is varied.

In this specific sense, FACADE theory predicts how stronger boundary groupings in a 2D image stand a better chance of winning the competition that gives rise to a 3D representation of figure and ground. Other types of local variations in the relative amount of contour at the intersection of a pair of figures can also create a 'boundary advantage' — read 'bipole cell advantage' — which modulates contrast effects in a way that is similar to the interactions between interposition cues and figure contrast found in the previous experiment.

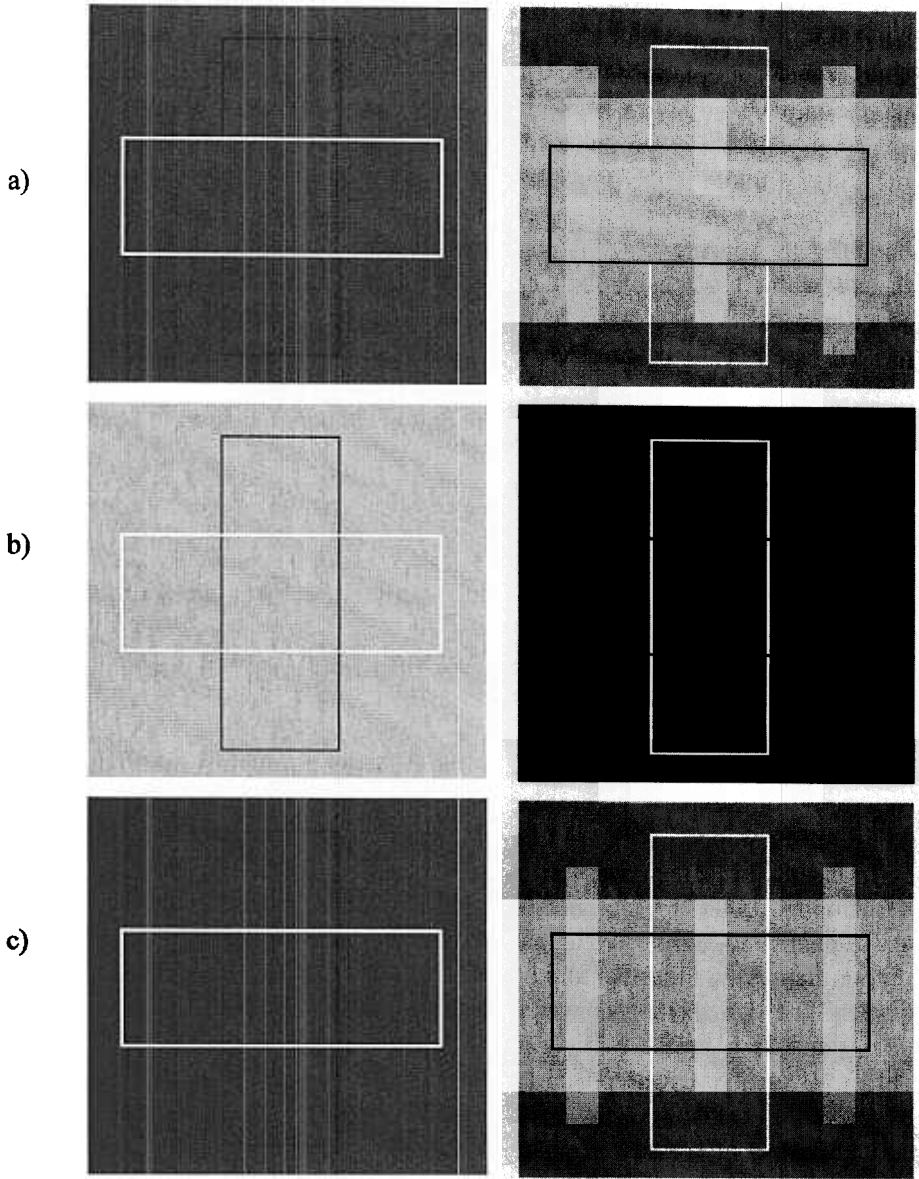
The second experiment was run to test whether the predicted bipole advantage does occur by varying the luminance contrast of briefly flashed pairs of forms with and without a local boundary advantage, that is, with T-junctions *versus* X-junctions. In the pair with boundary advantage, partial occlusion is clearly perceived (see Fig. 5) when observers have time to explore the image. Here, the exposure duration of the stimuli was as brief as in the previous experiment, and the observers again had to judge which figure of a given pair appeared to be nearer than the other.

### *Subjects*

The same four subjects were used as in Experiment 1, plus one additional, naive observer.

### *Stimuli*

The stimuli (see Fig. 5) were presented binocularly on a high-resolution computer screen (Sony, 60 Hz, non-interlaced). They were generated with an IBM compatible PC (HP 486) equipped with a VGA Trident graphic card. The luminance of the grey levels of the screen was measured with an OPTICAL photometer used in combination with the appropriate software. The length of each rectangle in a cross was 2.5 degrees of visual angle, the width was 1.5 degrees of visual angle. As in Experiment 1, all line contours were one minute of visual arc thick. In one condition, the two rectangles were simply superimposed with all their contours visible (transparent crosses with X-junctions), in the other condition, the horizontal rectangle of the cross was given a 'contour advantage' which gave rise to local cues of partial occlusion in the figure (opaque crosses with T-junctions). The two figures (transparent crosses or opaque crosses) were presented in separate blocks. A dark and a bright rectangle were presented in each cross, randomly varying over horizontal and vertical positions. While the luminance of the crosses was constant at 16 cd/m<sup>2</sup> for white rectangles and 2 cd/m<sup>2</sup> for black rectangles, the luminance of the background was varied among 4, 6, 8, 10, and 12 cd/m<sup>2</sup>, as in the previous experiment. Different luminance combinations were presented in random order within an experimental session. The combination between figure and background luminances led to five different levels of contrast, or relative visibility for a bright rectangle, and to five different levels of contrast for a dark rectangle.



**Figure 5.** Perceptually ‘transparent’ and ‘opaque’ crosses were presented in Experiment 2. The local contour advantage of the horizontal rectangle in the so-called ‘opaque’ crosses produces partial occlusion cues (5a). As in Experiment 1, background luminance was varied to create noticeable differences in contrast between horizontal and vertical rectangles of a cross. Observers had to decide as quickly as possible which rectangle of a cross (‘horizontal’ or ‘vertical’) appeared to be ‘nearer’ than the other. Exposure duration of the stimuli was 128 ms.

### Procedure

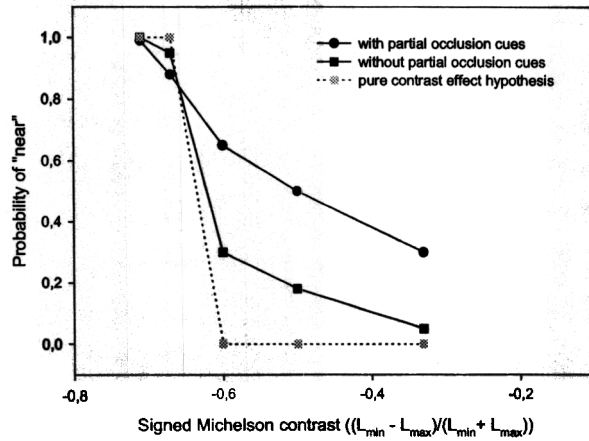
A given pair of rectangles forming a cross was flashed for 32 ms (two frames) on the screen, and observers had to decide as quickly as possible, by pressing one of two response keys on the computer keyboard, which rectangle of the cross (the horizontal or the vertical one) seemed nearer than the other. The choice of the observer and the response time were recorded. A new trial was initiated 1000 ms after the keyboard signal. The two figure conditions (transparent crosses or crosses with partial occlusion) were presented in separate blocks of trials and each observer was run in a total of 400 trials.

### Results

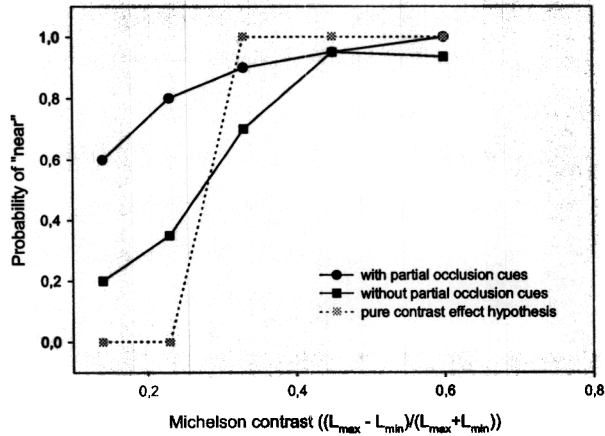
The results from Experiment 2 are represented in Fig. 6a and b. The probability of 'near' responses is plotted as a function of Michelson contrasts of bright and dark bars in crosses with partial occlusion cues, and in crosses without partial occlusion cues. The effect of figure contrast on perceptual judgements is statistically significant ( $F(9, 27) = 28.021$ ;  $p < 0.001$ ). The data curves show similarities with the data reported on contrast effects and interposition cues from the first experiment. When partial occlusion cues are additionally available in the figures, however, observers' judgements tend to deviate from the predicted contrast effect for figures with weaker Michelson contrasts. When the stimuli do not contain partial occlusion cues (transparent crosses), perceptual judgements follow the predicted 'pure contrast' effect quite closely in all conditions. The results hereby support the FACADE prediction that, when bipoles in the vertical and horizontal orientations are geometrically balanced in their activation (transparent crosses), then contrast differences can strengthen the boundary formed by one of the bipoles and thus allow it to win the competition.

The global effect of partial occlusion cues on observers' judgements is statistically significant ( $F(1, 3) = 12.16$ ;  $p < 0.05$ ). The interaction between contrast and partial occlusion is statistically significant ( $F(9, 27) = 26.78$ ;  $p < 0.001$ ). Subjects had a noticeable tendency to respond faster to the figures with partial occlusion; however, this effect is not statistically significant here.

Further analyses of variance, grouping figures with the stronger Michelson contrasts on the one hand, and figures with the weaker contrasts on the other, reveal that partial occlusion is not significant in the case of the strong contrasts, but significant in the case of the weaker contrasts ( $F(1, 3) = 15.68$ ;  $p < 0.05$ ). Similar statistics have been found to describe interactions between contrast and interposition cues in Experiment 1. This pattern of results also supports the FACADE theory prediction of how bipole cells interact with contrast differences to determine figure-ground percepts, since relatively strong contrasts can overwhelm a geometrical advantage by strengthening the weaker geometrical configuration.



(a)



(b)

**Figure 6.** The probability that the vertical or the horizontal stimulus part of a given cross is seen as 'nearer' is plotted as a function of signed Michelson contrasts for dark (6a) and bright (6b) stimulus parts. Each probability is estimated on the basis of a total number of 100 observations per datapoint. Two of the curves show data for the two levels of the 'partial occlusion' factor. The third curve (grey square symbols) shows the effect that is predicted when contrast alone ('pure contrast effect' hypothesis) determines whether the vertical or the horizontal part of a cross is seen as 'nearer'. The psychophysical observations follow the 'pure contrast effect' hypothesis more closely when the crosses do not contain partial occlusion cues.

### EXPERIMENT 3: INTERPOSITION, PARTIAL OCCLUSION, AND FILLING-IN DOMAIN

Experiment 3 provides a more direct test of the predicted interaction between figural geometry and contrast. Three conditions were tested. Two of the conditions are variants of those used in Experiment 2. They use either X-junctions (transparent) or T-junctions (opaque) in line drawings of overlapping surfaces. The third condition supplements the T-junctions with a uniformly luminant surface. The goal was to test

how much additional luminance is needed in each case to the target to look nearer. If X-junctions can, in fact, compete more effectively due to their ability to strongly activate bipole cells in both orientations, then more contrast should be needed in that case than the T-junction case to make the target look nearer. The addition of a uniformly illuminant surface should, if anything, tend to lower the amount of contrast needed in the T-junction case, since it might strengthen feedback from surfaces to boundaries at a later processing stage (see Grossberg (1997) for details).

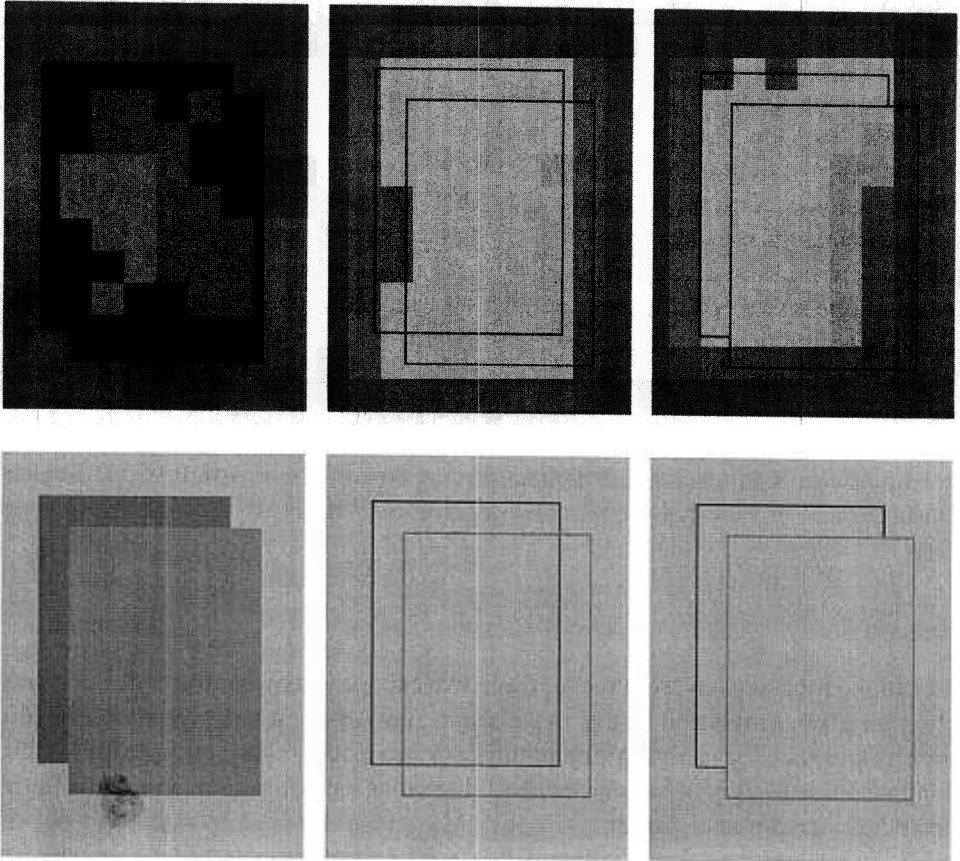
In Experiment 3, we used a procedure where the subjects had to adjust the luminance of one figure of a given pair in each experimental condition until the modified figure appeared unambiguously as being 'nearer' than the other. To investigate the influence of temporal factors on the emergence of these depth percepts, we varied the exposure duration of the stimuli in the following way. In particular, a longer exposure duration was needed to generate an equilibrated figure-ground separation in response to X-junctions than T-junctions, again consistent with FACADE mechanisms which predict that the bipole competition is harder to resolve in the former case.

### *Subjects*

Two of the four subjects from the previous two experiments were used. A naive third observer also participated in this experiment. Two of them were psychophysically trained and familiar with the psychophysical procedure. The third observer, also a volunteer with normal vision like the other subjects, was made familiar with the procedure in a pre-test session.

### *Stimuli*

The stimuli (see Fig. 7) were presented binocularly on a high-resolution computer screen (Mitsubishi, 60 Hz for observer SD and BD, TAXAN for observer CT). They were generated with an IBM compatible PC (Pentium II) equipped with a VGA graphic card. The luminance of the grey levels of the Mitsubishi screen was measured with a PRITCHARD photometer used in combination with the appropriate software. Luminance output of the TAXAN screen was calibrated with an OPTICAL photometer and software. Stimuli were presented in pairs, and their exposure duration was varied (16, 32, 64, 128, 256, and 512 ms). The length of each rectangle of a pair was 2.5 degrees of visual angle, with a width of 1.5 degrees. In one condition, the surfaces of the two rectangles were filled, in the other condition, only the contours of the rectangles were presented. Two rectangles of a pair always had the same contrast polarity. One of the rectangles, either the left or the right rectangle of a pair, had constant luminance ( $0 \text{ cd/m}^2$  for black rectangles, and  $50 \text{ cd/m}^2$  for white rectangles). The location (left or right) of the rectangle with constant luminance varied randomly. Background luminance was constant at  $8 \text{ cd/m}^2$ .



**Figure 7.** Pairs of rectangles defined by a figure–ground contrast extending over the whole rectangular surface in one experimental condition, and rectangles defined by a figure–ground contrast at their boundaries only in the other experimental condition were presented in Experiment 3. The stimuli in condition one, filled-in in the physical domain, give rise to the perception of two opaque surfaces and massive partial occlusion. The figures in the other conditions either formed two transparent surfaces with local interposition cues (X-junctions), or two opaque surfaces with cues of partial occlusion (T-junctions). Observers had to adjust the contrast of either the left or the right figure of a pair until that test figure unambiguously appeared to stand in front of the other figure. At the beginning of each adjustment session, the test figure was set at background luminance. Adjustments were made by using luminance increments (adjustments towards ‘brighter’) and decrements (adjustments towards ‘darker’) in separate sessions.

### *Procedure*

A luminance adjustment procedure was used, and observers were asked to change the contrast of one of the two rectangles in a pair (the test figure) by means of a key on the computer keyboard until this rectangle appeared to be nearer than the other rectangle with the constant luminance contrast (the comparison figure). In particular, the luminance adjustments were done in small increments, trial by trial. Each hit on the 1 keyboard incremented the figure by a small luminance amount;



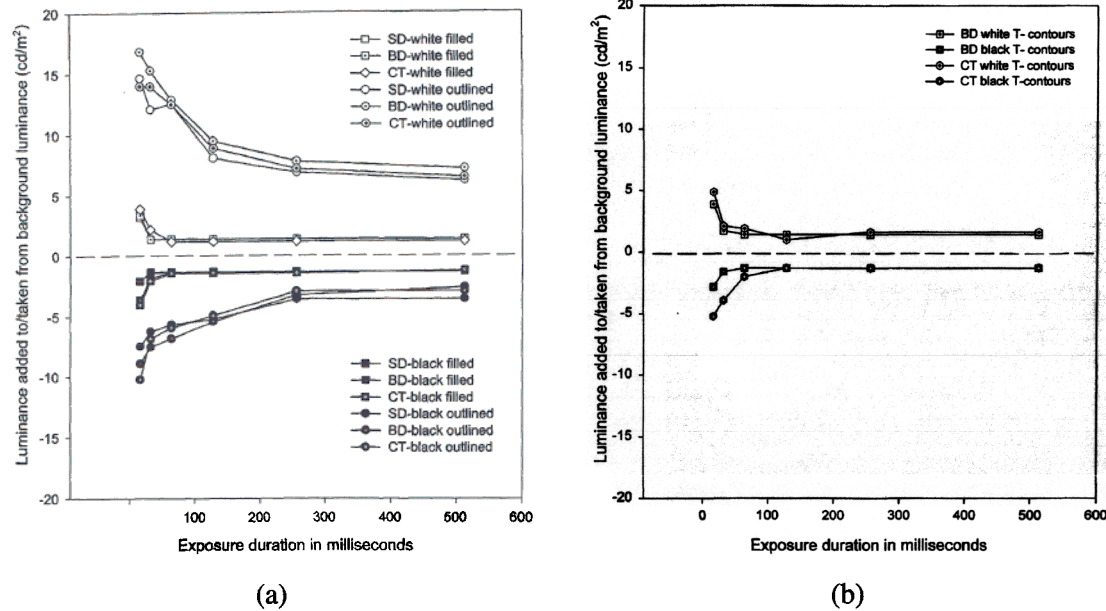
each hit on the 2 keyboard decremented it by a small luminance amount. The figure pair came on for a few milliseconds, limiting the time the subject had to explore it. Then it disappears and the subject chooses one key to either increment or decrement the target figure. This decision takes about 500 ms (average) on each trial. Then comes the next trial, the figure is flashed again. If the subject thinks that the target figure needs to be incremented/decremented further to stand out as 'nearer', one of the keys is hit again. When the subject thinks that the target figure stands out clearly as 'nearer' on a given trial, the 3 key is pressed to end the procedure. The final contrast level of the test figure was recorded. The initial contrast level of the test figure was constant at background luminance for both white and black figures. As in the previous experiments, the stimuli were flashed in pairs, and the different exposure durations varied between sessions. For each observer, figure type (filled or outlined rectangles), polarity (black or white rectangles), and exposure duration, two or three sessions were run. The different experimental conditions were presented in separate blocks of trials. Observers SD and BD were run in 48 blocks each, observer CT was run in 72 blocks.

### Results

The data of each observer from Experiment 3 are represented in Fig. 8. The final contrast levels of the test figure after luminance adjustment by the observers are plotted in  $\text{cd}/\text{m}^2$  differences from the no-contrast level, which means that the luminance of the test figures before adjustment was always equal to background luminance. The starting contrast is therefore represented by the number zero on the  $x$ -axis. Negative values on the  $y$ -axis indicate contrast decrements (adjustments towards 'darker'), positive values indicate contrast increments (adjustments towards 'lighter'). The adjusted luminance levels are plotted for each observer as a function of the exposure duration of a given pair of test and comparison figures. The results show that figure pairs already filled-in in the physical domain and containing strong cues of partial occlusion give rise to a near-far percept after minimal contrast adjustments only, at even the shortest exposure durations. Figure pairs represented by their contours only (outlined rectangles with interposition or occlusion cues) require noticeably stronger differences in contrast to generate unambiguous percepts of relative depth. The shorter the exposure duration of the figures, the greater is the contrast difference needed to produce these percepts. The individual results of the three subjects are shown to be very similar, and coherent in every respect. Coefficients of intra-individual variability ( $w$ ) were computed, but too small ( $w < 1$ ) to make the plotting of error bars necessary.

### GENERAL DISCUSSION

The data reported herein support the hypothesis that figural contrast is an important pictorial cue for the emergence of percepts of near *versus* far in two-dimensional



**Figure 8.** The final luminance level of a test figure after adjustment by the observers (Experiment 3) is plotted in  $\text{cd/m}^2$  added or taken away from the starting level where all figures were set equal to background luminance. This level is represented by the number zero on the x-axis. Negative values on the y-axis indicate contrast decrements (adjustments towards 'darker'), positive values indicate contrast increments (adjustments towards 'lighter'). The adjusted contrast levels (means) are plotted as a function of the exposure duration of a given pair of test and comparison figures. Figure pairs already filled-in in the physical domain and containing strong cues of partial occlusion are shown to give rise to a near-far percept after minimal contrast adjustments only, at even the shortest exposure durations (Fig. 8a). Figure pairs represented by their contours only (referred to as 'outlined' here) require noticeably stronger differences in contrast to generate unambiguous percepts of relative depth when they generate interposition cues, but no cues of partial occlusion (Fig. 8a). However, when the rectangles generate cues of partial occlusion via T-junctions (referred to as 'T-contours' here), near-far percepts again emerge after only a few contrast adjustments (Fig. 8b). Generally, we observe that shorter exposure duration of the figures yield more contrast adjustments to produce near-far percepts. This effect of exposure duration on the amount of contrast added to a figure of a given pair (i.e. on the number of adjustments made) becomes asymptotic at exposure durations between 120 and 250 ms for all figure types.

stimuli (O'Shea *et al.*, 1994). Our results show that this hypothesis is valid in situations where the stimulus duration is brief.

Other pictorial factors such as interposition (Experiment 1) and partial occlusion (Experiment 2), often considered as major determinants of pictorial depth (Kanzisa, 1979, 1985), were tested. Pictorial cues are found to interact with contrast cues. Interposition and partial occlusion contribute to generate perceived depth when combined with a cooperative contrast cue. This conclusion is generally consistent with cue combination models (e.g. Gibson, 1950). Interposition and partial occlusion on their own are not strong enough in the stimuli presented here to compete with a strong, conflicting contrast cue. The results furthermore support the hypothesis that image geometries which effectively activate bipole cells can compete better with conflicting contrast cues than geometries that do not.

#### *Cooperation and/or competition between contrast and other pictorial cues?*

The interactions between contrast and interposition and between contrast and partial occlusion found here suggest that weaker contrasts do not cooperate with the two other pictorial cues but tend to compete. This would seem to indicate, as suggested by Landy *et al.* (1995), that unambiguous cues, which are interposition and partial occlusion here, fail to disambiguate the ambiguous one, which is a weak contrast here. From this observation, it is tempting to conclude that pictorial cues do not have equal status in determining the perception of near and far, which would support the O'Shea *et al.* (1994) claim that contrast alone, as the optical consequence of aerial perspective, is an absolute and self-sufficient pictorial depth cue.

This view cannot be fully supported, however, when one acknowledges that contrast also controls the strength of the geometrical cues that influence perceptual grouping. In particular, contrasts in an image activate, via parallel pathways, amodal boundary groupings within the Boundary Contour System *and* (potentially) visible surface features within the Feature Contour System (see Fig. 1). Other things being equal, increasing the contrast in an image will increase both boundary and surface responses. As a result, each contrast cue necessarily cooperates with the geometrical grouping cue that it defines within the BCS. Even if geometrical factors remain fixed, such as the length of the contour that activates BCS bipole cells, increasing the contrast of this contour can increase the strength of the inputs that activate bipole cells at every position along this length. Thus when one pits an oriented linear contrast cue against a differently oriented and weaker linear contrast cue in an X-junction, one is really competitively pitting a *pair* of cooperating contrast-plus-geometrical cues against one another. That is why, in the X-junction, sufficiently high contrast always wins, as shown in Experiment 2: the geometrical cues at the X-junction within the BCS are balanced in strength, so the greater contrast can always tip the balance by activating each position along one branch of the X more than along the other branch. This is also why more contrast is needed to win in an X-junction than a T-junction, as shown in Experiment 3: since the stem of the T-junction is a weaker geometrical cue for activating bipole cells than the top,

less contrast is needed to give the top of the T the advantage that is needed to initiate figure-ground separation.

Taken together, these results supply supportive evidence for the FACADE model hypothesis which predicts that bipole cells underlie perceptual grouping, and that non-colinear bipole cells compete for dominance. The present experiments illustrate how contrast variations can be used to illuminate the relative strengths of these groupings. Furthermore, as predicted by the FACADE model (Grossberg, 1994, 1997), when one set of bipole cells wins out over another, non-colinear set, it initiates the process whereby a figure-ground percept is generated, with the winning boundaries supporting the percept of a nearer surface. Also, as predicted by the FACADE model, contrast and geometrical factors may cooperate or compete to determine the winner and, hence, which surface appears nearer.

#### *Contrast cues, exposure duration, and possible shape effects*

The fact that we do not find an effect of shape on near-far judgments in Experiments 1 and 2 may be a consequence of the short exposure duration (32 ms) of the stimuli. The brief presentation may have masked a possible shape effect, either because there was not enough time for higher-level effects to express themselves, or because the interposition cues in Experiment 1, for example, were not visible enough. However, such a partial cue-masking effect due to the brief stimulus exposure is unlikely. There is, after all, a conditionally significant effect of interposition in Experiment 1 with stimuli of weaker contrast. Furthermore, the data from Experiment 3 which used stimuli with interposition cues (the 'outlined' conditions) as fine as in Experiments 1 and 2 show that the contrast needed to attain a just perceptible depth separation was, for example, about 31% with the bright outlined stimuli, a contrast level similar to the one that yielded a significant effect of interposition in Experiment 1.

O'Shea *et al.* (1994) used unlimited exposure durations and found that size had no effect on near-far judgments. Even when the size cue opposed a contrast cue, the contrast cue took over. We think that something similar might happen to shape cues when contrast is introduced as a depth cue, regardless of the exposure duration of the stimuli. O'Shea's suggestion that contrast is equivalent to aerial perspective in generating perceived depth implies that contrast, under some conditions, acquires the status of an absolute depth cue. It could be that shape information becomes totally irrelevant when such a strong depth cue enters the game. In this regard, the FACADE model suggests that the balance between geometrical and contrastive factors at T-junctions and X-junctions of sufficiently simple images is a key factor in determining the final depth percept. Shape factors can indirectly influence this balance by altering the number of junctions, their relative spacing, and the relative strength of the junction branches. It remains to be seen if shape changes that do not alter these factors can ever have a major effect on depth percepts of images such as those which we have studied herein.

### Acknowledgements

The authors wish to thank Diana Meyers and Robin Amos for their valuable assistance in the preparation of the manuscript and figures. B. Dresch was supported by a fellowship from the Human Frontier Science Program Organization (HFSPO, SF9/98). S. Grossberg was supported in part by the Air Force Office of Scientific Research (F49620-01-1-0397), the Defense Research Projects Agency and the Office of Naval Research (ONR N00014-92-J-1309, ONR N00014-95-1-0494, and ONR N00014-95-1-0657).

### REFERENCES

- Cohen, M. A. and Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance, *Perception and Psychophysics* **36**, 428–456.
- Da Vinci, L. (1651). *Trattato della Pittura di Leonardo da Vinci*. Scritta da Raffaele du Fresne, Langlois, Paris.
- Dresch, B. and Grossberg, S. (1997). Contour integration across polarities and spatial gaps: From local contrast filtering to global grouping, *Vision Research* **37**, 913–924.
- Dresch, B. and Grossberg, S. (1999). Spatial facilitation by color and luminance edges: boundary, surface, and attentional factors, *Vision Research* **37**, 3431–3443.
- Egusa, H. (1983). Effects of brightness, hue, and saturation on perceived depth between adjacent regions in the visual field, *Perception* **12**, 167–175.
- Field, D. J., Hayes, A. and Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local 'association field', *Vision Research* **33**, 173–193.
- Gibson, J. J. (1950). *Perception of the Visual World*. Houghton-Mifflin, Boston, MA.
- Grossberg, S. (1984). Outline of a theory of brightness, color, and form perception, in: *Trends in Mathematical Psychology*, E. Degreef and J. van Buggenhaut (Eds), pp. 5559–5586. Elsevier, Amsterdam.
- Grossberg, S. (1994). 3D vision and figure-ground separation by visual cortex, *Perception and Psychophysics* **55**, 48–120.
- Grossberg, S. (1997). Cortical dynamics of 3D figure-ground perception of 2D pictures, *Psychol. Rev.* **104**, 618–658.
- Grossberg, S. and Mingolla, E. (1985a). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations, *Perception and Psychophysics* **38**, 141–171.
- Grossberg, S. and Mingolla, E. (1985b). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading, *Psychol. Rev.* **92**, 173–211.
- Grossberg, S., Boardman, I. and Cohen, M. (1997). Neural dynamics of variable-rate speech categorization, *J. Exper. Psychol.: Human Perception and Performance* **23**, 481–503.
- Kanzisa, G. (1979). *Organization in vision: Essays in Gestalt Perception*. Praeger Press, New York.
- Kanzisa, G. (1985). Seeing and thinking, *Acta Psychologica* **59**, 23–33.
- Kapadia, M. K., Ito, M., Gilbert, C. D. and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys, *Neuron* **15**, 843–856.
- Landy, M. S., Maloney, L. T. and Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion, *Vision Research* **35**, 389–412.
- Magnussen, S. and Glad, A. (1975). Brightness and darkness enhancement during flicker: perceptual correlates of neuronal B- and D- systems in human vision, *Experimental Brain Research* **22**, 399–413.

- O'Shea, R. P., Blackburn, S. G. and Ono, H. (1994). Contrast as a depth cue, *Vision Research* **34**, 1595–1604.
- Peterhans, E. and von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex, II: Contours bridging gaps, *J. Neurosci.* **9**, 1749–1763.
- Pins, D. and Bonnet, C. (1996). On the relation between stimulus intensity and processing time: Piéron's law and choice reaction time, *Perception and Psychophysics* **58**, 390–400.
- Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T. and Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold, *Nature* **391**, 580–584.
- Polat, U. and Sagi, D. (1994). The architecture of perceptual spatial interactions, *Vision Research* **34**, 73–78.
- Schwartz, B. J. and Sperling, G. (1983). Luminance controls the perceived 3D structure of dynamic 2D displays, *Bull. Psychonomic Soc.* **21**, 456–458.
- Stevens, K. A., Lees, M. and Brookes, A. (1991). Combining binocular and monocular curvature features, *Perception* **20**, 425–440.
- Von der Heydt, R., Peterhans, E. and Baumgartner, G. (1984). Illusory contours and cortical neuron responses, *Science* **224**, 1260–1262.
- Yantis, S. and Jones, E. (1991). Mechanisms of attentional selection: temporally modulated priority tags, *Perception and Psychophysics* **50**, 166–178.