BRIGHTNESS PERCEPTION AND FILLING-IN

MICHAEL A. PARADISO* and KEN NAKAYAMA†

Smith-Kettlewell Eye Research Institute, 2232 Webster St, San Francisco, CA 94115, U.S.A.

(Received 10 July 1990; in revised form 28 November 1990)

Abstract—Three experiments were performed in which a stimulus with homogeneous color and luminance was masked by a second stimulus containing contours. In the first experiment the target was a large white disk and the mask was a white circle concentric with the disk but of smaller radius. We found that the target had a large (up to 2 log unit) suppressive effect on the brightness of the target, but only inside the radius of the mask. With monoptic presentation of target and mask, the greatest suppression was observed with an SOA of 50–100 msec. With dichoptic presentation the strongest suppression was obtained with simultaneous stimuli. The second experiment demonstrated that the latest time at which masking was effective was correlated with the distance between the edge of the target stimulus and the contour in the mask. One possible explanation of the results from these two experiments is that the masking contour is interfering with the propagation of a brightness signal from the target's border. In the third experiment gaps were introduced into the masking circle. Surprisingly, even with rather large gaps there was significant suppression of brightness in the center of the target. We have encountered difficulties attempting to account for these findings with known physiological mechanisms such as lateral inhibition. A qualitative explanation of the results that looks promising is a two-component process involving brightness filling-in and smoothing to satisfy fixed boundary conditions at contours.

Brightness perception  Filling-in  Masking  Scotoma  Stabilized image  Lateral interactions
Metacontrast

INTRODUCTION

For centuries its been known that we all have a perceptual blind spot resulting from the lack of photoreceptors at the optic nerve head. Psychophysical studies have shown that under particular conditions this blind spot disappears and fills-in with the color and brightness of a surrounding stimulus. Such filling-in has also been observed in cases of pathological scotomata and in cleverly arranged stabilized-image experiments (see below). In all of these demonstrations the brightness and color at a border plays a primary role in determining the perception of an enclosed area. These experiments have led to speculation that the perceptual filling-in observed under abnormal experimental conditions may be revealing a filling-in mechanism which is a fundamental component of normal visual processing (Gerrits & Vendrik, 1970; Walls 1954). We have conducted masking experiments to examine this hypothesis.

Previous demonstrations of perceptual filling-in fall into two categories. One set of findings concerns the perception of objects placed near retinal scotomata. As mentioned above, we all have a blind spot which can be demonstrated by putting contours in its part of the visual field. Importantly though, it is not perceived if one views a large uniformly-colored surface. In this situation the color and brightness surrounding the blind spot appear to fill it in. Similar perceptual filling-in occurs with pathological scotomata of various origins, although there is controversy about the types of stimuli which fill-in (Bender & Teuber, 1946; Fuchs, 1921; Gassel & Williams, 1962; Gerrits & Timmerman, 1969; Lashley, 1941; Poppelreuter, 1917). Generally, if a scotoma is surrounded by a uniformly-colored field, the color and brightness of the surround are perceived to fill-in the blind spot. This phenomenon appears to primarily involve brightness and color as there is much less, if any, completion of luminance contours across pathological scotomata.

The second type of demonstration of filling-in comes from experiments with stabilized images. Using various mechanical systems it is possible to create stimuli that are partially or entirely stabilized on the retina. Generally if a stimulus
is entirely stabilized its color and brightness fade until it is no longer seen (Riggs, Ratliff, Cornsweet & Cornsweet, 1953). In close analogy with the studies mentioned above, one can create an artificial scotoma in a normal observer by stabilizing a small spot. As in the clinical studies, if a uniform color surrounds the artificial scotoma, that color perceptually fills-in the blind spot (Gerrits, de Haan & Vendrik, 1966; Yarbus, 1967). In another important experiment a disk of one color is surrounded by an annulus of a different color and the border between the two areas is stabilized on the retina (Krauskopf, 1963, Larimer & Piantanida, 1988; Yarbus, 1967). The observer’s perception is that the color in the surrounding annulus “spills into” the central disk and the stimulus appears to uniformly have the color of the annulus. Another observation that parallels the perception associated with scotomata is that contours do not fill-in across stabilized regions. For instance, Yarbus (1967) arranged a stimulus in which a disk stabilized on the retina is placed over the border between two background areas of different colors, which were not stabilized. He found that the color and brightness of the disk faded but the disk was only partially filled from the sides by the two different colors of the background. In the area of the disk near the junction of the two background regions, the observer saw a black unfilled patch. Again, the implication is that color and brightness can be assigned solely on the basis of edge information whereas contours themselves cannot fill-in in this manner. The experiments with stabilized images demonstrate that perceptual filling-in can occur even in normal observers. This raises the interesting question of whether filling-in is a part of normal (nonstabilized) visual perception. In the percepts of people with retinal scotomata and the experiments with stabilized images the perceived color and brightness of an area are entirely determined by the color and brightness at an edge of the stimulus. In normal vision, does edge information play a major role in determining the color and brightness of a homogeneous area?

Single unit recordings made in primary visual cortex of cats and monkeys suggest that this is not an entirely unreasonable speculation and that there may actually be a need for a filling-in mechanism. Since Hubel and Wiesel's initial studies it has been found that neurons in striate cortex give their largest response to oriented luminance or color discontinuities (Hubel & Wiesel, 1962, 1968). While there are some cells that respond to diffuse illumination (Bartlett & Doty, 1974; Kayama, Riso, Bartlett & Doty, 1979), most cells are not strongly excited if their receptive field is always inside a large spot of light that is shifted about. One point that appears certain is that if one compares cells in V1 with receptive fields inside the stimulus to those with receptive fields on the stimulus’ border there will be a larger number of strongly excited neurons at the border. There are two straightforward ways that the visual system might elicit a homogeneous perception from this inhomogeneous activity in striate cortex. One possibility is that the system is wired in such a manner as to automatically interpret the strong edge response as a filled-in object. In this case, the brightness at the edge is automatically assigned to the objects’ interior and there is a nonisomorphic relationship between perception and the activity in the population of neurons. The alternative to this nonisomorphic assignment of brightness is that at some stage of visual processing, beyond the edge-activated cells in V1, there is a mechanism that uses the edge responses to influence the firing of other cells responsible for the filled-in percept. This process might establish a homogeneous pattern of activity somewhere in the brain which mirrors the homogeneous percept.

To clarify the nature of perceptual filling-in we sought to answer two questions. The first is whether, in normal observers, edge information largely determines the perception of brightness and color in homogeneous regions. Assuming that edge information is of primary importance, the second question concerns the dynamics of filling-in. Is it an instantaneous response or a spread of activation over time? Our approach has been to consider a simplistic working model of filling-in and to challenge it experimentally. Specifically, the response properties of neurons in V1 suggest that a large uniform stimulus might initially produce a response predominantly to its edge. Over time the color and brightness at the border could influence the perception at neighboring areas and this might start a chain of interactions which continues until another border is encountered. Presumably we would not normally be aware of this process because of its speed. However, if filling-in involves the spread of activity it might be possible to demonstrate the existence of the filling-in mechanism by interrupting it. In other words, if borders stop filling-in, what will happen if new borders are
introduced before the filling-in process is complete?

**GENERAL METHODS**

All of the results we report in this paper derive from two-stimulus masking experiments. Observers viewed stimuli generated by a Number Nine graphics board in a microcomputer and shown on a CRT display (480 × 640 pixels, 60 Hz refresh rate). A target stimulus, usually consisting of a large uniformly-colored area, was displayed briefly (usually for 16 msec). A second masking stimulus was presented for 16 msec either before (forward masking) or after (backward masking) the target. Experiments were conducted with both monoptic and dichoptic presentation of target and mask. There was a relatively long interval (400–500 msec) before the target and mask were repeated. The subject’s task was to make a brightness match between a specified area of the target and a palette of gray tones ranging from black to white. In the experiments involving both monoptic and dichoptic trials, the comparison palette was 5.4 deg above the center of the target on the CRT. In the binocular-only experiments, the palette was 8.8 deg to the left of the target’s center. A brightness match was made by using a pointer controlled by the computer’s mouse to indicate which element of the comparison palette appeared most similar to the specified region of the target. Subjects were allowed to view unlimited repetitions of the target/mask cycle before responding. The luminance of the stimuli was 160 cd/m² (measured during continuous rather than transient presentation) and the background luminance was less than 0.1 cd/m². The 12 elements in the comparison palette ranged from 0.1 to 160 cd/m² in 0.25 log unit increments of luminance.

**PRELIMINARY OBSERVATIONS**

To establish a context for the quantitative experiments described below, we first present several qualitative observations. The fundamental result of our experiments is that a mask consisting of contours within the boundaries of a uniform target can have a dramatic effect on the brightness of the interior of the target. A demonstration of this effect is illustrated in Fig. 1. In this experiment the target consists of a bright white disk several degrees in diameter and the mask is a grid of thin white lines on a black background. Binocularly we viewed the target followed by the mask and varied the temporal interval between them. When the mask follows the target by 50–100 msec we observe that the brightness of the central area of the disk is greatly reduced. Interestingly, the brightness of the target near its edge appears about the same as when there is no mask. Thus the brightness suppression caused by the mask rapidly increases with distance from the target’s edge even though the lines in the mask are uniformly spaced.

Insight into the nature of the masking caused by contours can be gained by looking at the effect of a single line. If the target is again a disk and the mask is a vertical line contained within the boundaries of the target then there is relatively little masking (Fig. 2, top). The only obvious effect is that the brightness of the disk is reduced just adjacent to the vertical line yielding thin black lines to the sides of the white masking line. Its interesting to compare this result to the case in which the masking line is bent to form a “c” shape. With this alteration one observes quite asymmetric brightness suppression in the vicinity of the vertical and horizontal segments of the mask. The brightness of the disk is considerably darker inside the “c” than outside it (Fig. 2, middle). As with the grid-
like mask, the brightness suppression is maximal with a delay of 50–100 msec. Comparison of the masking effects of a vertical line and a line bent into a “c” shape indicates that the brightness suppression is not simply a function of distance from the masking contours; rather, the configuration of the lines is highly important. The clearest example of the importance of configuration comes from an experiment in which the mask is a thin white circle (smaller than the target). One perceives the outer edge of the disk to be bright extending inwards approximately to the radius of the mask. There is a small area of darkening just outside the masking circle and the entire central area of the target is dark (Fig. 2, bottom). This is consistent with the general finding that there is a great asymmetry of brightness suppression inside vs outside masking contours. We have repeated the masking paradigm and observed brightness suppression with a variety of targets and masks including lines, spots, grids and asymmetric objects. This phenomenon is similar to masking termed “area suppression” by Stoper and Mansfield (1978). The findings are formalized in the three experiments described below.

The suppressive effect of a circular mask on the perceived brightness of a uniform disk is quantified in expt 1. We examine the cases in which the target and mask are presented to the same or different eyes. In expt 2 the effect of varying the distance between the edges of the target and mask is explored. If brightness fills-in over a finite time then brightness suppression might be determined by both the spatial and temporal arrangement of the stimuli. In expt 3 we use a number of different masks consisting of segments of a circle rather than a complete circle. The question is whether continuity is essential for a contour to have a suppressive effect on brightness.

EXPERIMENT 1: TEMPORAL PROPERTIES OF FORWARD AND BACKWARD MASKING

This experiment examines the hypothesis that a mask containing luminance contours can interfere with the perception of uniform brightness in a target stimulus.

Methods

The target was a white homogeneous disk with a luminance of 160 cd/m² and a radius of 3.4 deg presented for 16 msec. The mask was a white circle with luminance of 160 cd/m² on a black background and also presented for 16 msec. Its diameter was 2.0 deg and the line thickness was approx. 0.03 deg. Subjects sat at a distance of 80 cm and viewed the computer display through a stereoscope. The target was presented randomly to either the left or right eye and after a delay the mask was presented to
either eye. This was a "blind" comparison of monoptic and dichoptic masking because subjects generally reported not knowing at any given time which eye was seeing target or mask. Concurrently with varying the eyes to which the stimuli were presented the temporal interval between the target and mask was also varied. The temporal delays used were 0, ±16, ±64, ±96, ±126 msec. Subjects made brightness matches by indicating which of a palette of gray tones appeared most similar in brightness to the center of the target.

Results

Consistent with the qualitative observations mentioned above, the circular mask has a powerful effect on the brightness in the center of the target (Fig. 3). Under optimal conditions the brightness/luminance matches made in the center of the disk are reduced as much as 2 log units relative to the condition in which no masking was observed (i.e. at long target/mask intervals). The magnitude of the suppressive effect depends strongly on the temporal interval between the target and mask. At fairly long target/mask intervals the two stimuli are perceived independently and the brightness matches are high. The delay times which yield maximal masking are determined, in part, by whether the target and mask are seen by the same eye. For two subjects (MP and MB) the strength of the monoptic suppression is between 1 and 2 log units and it is maximal with an interval of about +64 msec (positive intervals indicate backward masking in which the mask follows the target). The third subject reported a maximal suppression of 0.4 log units and the effect was strongest at delays of 16–64 msec (the monoptic effect for this observer is clearer when plotted on a linear scale). For all three subjects there is little or no brightness suppression with forward masking.

The dependence of masking on the target/mask delay is noticeably different when the two stimuli are seen by different eyes. Overall, masking is stronger with dichoptic than with monoptic presentation. A big difference between monoptic and dichoptic masking is that, consistently across observers, maximal dichoptic suppression is obtained with simultaneous masking. The strong suppressive effect extends to delays of 16–64 msec. Also, for all observers there is a considerable effect when the mask precedes the target by 16 msec. For two of the subjects some suppression is seen when the delay is −64 msec. Although monoptic and dichoptic masking are markedly different at short target/mask intervals, the decline of masking at long intervals (64–100 msec) appears similar for all observers.

Discussion

The thin circular mask used in this experiment produces a striking decrease in brightness at the center of the target. Particularly with optimal conditions for dichoptic masking, the central area of the target looks absolutely black.
Importantly, the brightness of the target is relatively unaffected outside the radius of the target even at target/mask intervals which cause a 2 log unit decrease in brightness at the center. This is consistent with the hypothesis that the outside edge of the target plays a role in determining the brightness of the target at internal points and that the masking circle can only interfere with this mechanism inside the radius of the mask. The dichoptic results indicate that the masking contour can interfere with brightness perception if it is presented shortly before or simultaneously with the target. Both the monoptic and dichoptic results show that the mask can have a suppressive effect when it is presented 50–100 msec after the target. This finding bears some resemblance to the masking times observed in metacontrast. In the General Discussion we consider the similarities and differences between metacontrast and the effect we've observed.

There are a number of possible explanations for these results (see General Discussion) but in terms of the filling-in hypothesis they might be explained as follows. The presentation of the target initiates a propagation of brightness away from the border. This process can be interrupted as long as the masking contour is presented before the propagation of brightness has proceeded past it. For this reason the masking circle can be effective if it is within the boundaries of the target and if it is presented after the target. In the case of monoptic masking we assume that there is no suppression with short target/mask delays because of temporal integration. When two stimuli are presented in close temporal succession they become integrated (e.g. yielding Bloch's law, 1885) presumably reflecting a summation of the responses to the individual stimuli. Physiological recordings in striate cortex show that successive stimuli produce reliably independent responses only if they are separated by about 60 msec or more (Watanabe, Gawne, Richmond & Optican, 1989). Thus there may not be a distinct response to the masking contour which can interfere with filling-in of the target.

This is in contrast to the dichoptic results in which powerful brightness suppression is observed even with simultaneous presentation of target and mask. Presumably this occurs because there is incomplete integration when the stimuli are presented to opposite eyes. This may be because not all cortical cells are binocular (Hubel & Wiesel, 1962, 1968). With dichoptic presentation the mask may always produce a response in monocular cortical cells that is independent of the response to the target and this separate edge signal may disrupt filling-in of the target. The other noticeable feature of the dichoptic suppression curves (Fig. 3) is that masking occurs with longer positive than negative inter-stimulus intervals (at least for observers MB and MP). As with monoptic masking the longest positive target/mask delay at which brightness suppression occurs may be a direct indication of the speed of filling-in. The data show that suppression is also obtained when the mask is introduced dichoptically shortly before the target. The independent response to the masking contour presumably lasts for a long enough time that it can interfere with the filling-in of the target even when the target arrives afterwards. However, if the mask is presented too early the response to it may no longer be sufficiently strong to interfere with the target. Since there is strong masking when the target and mask are presented to different eyes the interaction between the stimuli presumably occurs in cortex where there are binocular neurons.

**EXPERIMENT 2: DEPENDENCE OF SUPPRESSION ON THE DISTANCE BETWEEN TARGET AND MASK**

The results of the previous experiment can be interpreted in the framework of brightness filling-in, however, there are other possible explanations of the data. In this experiment we make a more direct test of the hypothesis that brightness propagates inward from the target's border. Assuming that this occurs with some finite velocity there is an immediate prediction that the latest time at which masking is possible should be correlated with the distance between the mask and the edge of the target. One should be able to see masking at a later time if the masking contour is farther from the edge of the target.

*With simultaneous presentation of a disk-shaped target to one eye and a circular mask to the other eye the central area of the target is greatly suppressed in brightness. This is not rivalry because stimulus durations considerably longer than 16 msec are required to obtain rivalry (Wolfe, 1983). In fact, if one presents typical rivalrous stimuli, such as orthogonal lines to opposite eyes, for 16 msec they are fused rather than having a suppressive interaction. Another reason why we don't believe the dichoptic masking is rivalry is because it does not differ qualitatively from monoptic masking with the same stimuli (at least for observers MB and MP).
Methods

The shape and luminance of the stimuli in this experiment are identical to those used in expt 1. Subjects again sat at a distance of 80 cm viewing the computer display through a stereoscope. However, in this experiment trials were interleaved with different size targets and masks. Two masks were used with radii of 1.4 and 2.0 deg. Targets with radii ranging from 0.55 to 3.4 deg were used. As a control we used masks larger than the target, in addition to smaller ones, to see if there is significant brightness suppression when the two stimuli have nearby edges but the mask is not within the target. Because dichoptic masking is stronger than monoptic masking, this experiment was conducted with the target and mask always presented to opposite eyes. Trials with the target presented to the left or right eye were randomly interleaved. Only backward masking was used in this experiment since this is where the strong effects were obtained in expt 1.

Results

The graphs in Fig. 4 illustrate how the suppressive effect of the mask depends on the temporal delay between the target and mask. The data in these graphs were obtained with a mask having a radius of 2.0 deg and targets with radii from 1.2 to 3.4 deg. The different curves are for different size targets. Consistent with the dichoptic results in expt 1, strongest masking is obtained with simultaneous presentation of the target and mask. Masking decreases as the interval between the stimuli is increased until there is no masking with a delay of about 100–120 msec. Although there are inter-subject differences, there are two consistent effects of changing the target radius. The first is that the suppressive effect of the mask is greater as the targets increase in size. For example, for observer M.B. the average brightness matches made with simultaneous presentation of target and mask progress from 81 cd/m² (target radius = 1.2 deg) to 13 cd/m² (target radius = 2.0 deg) to 0.1 cd/m² (target radius = 3.4 deg). The other change is that as the targets increase in size the curves rise back up to the baseline brightness level at later target/mask delay times. To clarify this point, the data from Fig. 4 have been transformed and re plotted with additional data (obtained with a 1.4 deg radius mask) in Fig. 5. The abscissa is the distance between the edges of the target and mask. The ordinate is a criterion level defined as the target/mask delay time at which the brightness at the center of the target recovers to half of its unmasked level. These plots indicate that masking remains effective at later times as the distance between the edges of the two stimuli increases.

Discussion

An important finding in expt 2 is that there is a tradeoff between the target/mask distance
and the latest time at which the mask has a suppressive effect on the target's brightness.

This is consistent with the hypothesis that the mask is interfering with a process begun at the target's outer edge. One interpretation of the results is that some internal correlate of brightness is spreading inward from the edge of the target. The contour in the mask can block the spread but only if it is presented at a given location before the spreading brightness gets there. Thus, the farther the mask is from the target the later the time at which it can interfere in the spreading process.

One can use the data in Fig. 5 to estimate the speed at which this spreading would have to occur. By making least-squares fits to the data the slopes obtained for the different observers are in the range 6.7–9.2 msec/deg (110–150 deg/sec). There are some differences between observers, the largest being that, with a 2 deg mask, observer M.P. saw suppression at longer delays than the other subjects. Importantly though, the slopes of the curves are similar for different size stimuli and for different observers. To try and understand the hypothetical mechanism of brightness spreading in terms of neural activity it is useful to express the results in terms of the rate of spread across cortex rather than the angular subtense in the visual field. This requires that the data be replotted using a cortical magnification factor to transform degrees of arc into mm of cortex. In Fig. 6 the data are plotted using Levi, Klein and Aiisebaomo's (1985) estimate of the magnification factor for human striate cortex. The slopes of the curves give an estimate of the rate of spread of brightness that would be required if this process were occurring in area V1. For the three observers these rates are in the range of 2.5–6.7 msec/mm (0.15–0.40 m/sec). Physiologically, this isn't a totally unrealistic number given that lateral excitation in visual cortex spreads at the rate of 10–20 msec/mm (Chervin, Pierce & Connors, 1988). One way to account for the discrepancy between the rates calculated from our data and measured speeds of lateral excitation in cortex is that filling-in does not occur in striate cortex. If the cortical area underlying filling-in had a lesser magnification factor (as do all extrastriate areas), then the calculated rates of filling-in would be closer to 10–20 msec/mm.

One curious finding that can't be ignored in the data is that there was some masking even when the target disk was smaller than the masking circle. Its not obvious that this finding can
be explained as a blocking effect on the inward flow of brightness. However, masking with the target smaller than the mask only occurs when the edges of the two stimuli are quite close and the strength of the brightness suppression is considerably less than when the target is larger than the mask. Resolution of this point may involve a consideration of the spatial scale of brightness mechanisms. We deal with this issue below in the General Discussion.

EXPERIMENT 3: MASKING WITH DISCONTINUOUS CONTOURS

In the discussion above it is proposed that the suppressive effect of the mask on target brightness might be due to a blockade of spreading brightness. This is an intuitively appealing idea because it accounts for the great reduction in target brightness inside the radius of the mask. The hypothesis that the mask blocks the inward flow of brightness also can explain the dependence of masking time on the distance between the two stimuli. A physical analogy for this hypothesis is that brightness flows in from the outer edge of the target filling up the central area like water running to lower ground. In this analogy the mask is a barrier which stops the inward flow which represents spreading brightness. An alternative analogy using a neural network is that brightness fills-in as excitation spreads laterally through a network of cells (Gerrits & Vendrik, 1970; Cohen & Grossberg, 1985; Grossberg & Todorovic, 1988). The mask would somehow serve to stop the chain of lateral excitation through the network. Both of these analogies have the property that the effectiveness of the mask relies on its continuity. If there is a gap in the mask the spreading brightness could get through. We conducted this experiment to see if, in fact, there is a sharp decline in mask efficacy as gaps are introduced in the circular mask.

Methods

As in the previous experiments, the target was a uniformly white disk, in this case with a radius of 5.4 deg and a luminance of 160 cd/m². Circular masks with radii of 2.7 deg were used. They were constructed from 5 deg segments of a circle. The masks differed in the number of segments and correspondingly the subtense of the gaps in their perimeter. The gap sizes were 0, 35, 55, 85, 115, 175 and 360 deg. The mask with 0 deg gaps is the usual complete circle and the mask with 360 deg gaps is simply a control case where there is no mask at all. Trials with the different masks were randomly interleaved and subjects always made brightness matches with the central area of the target disk. The temporal interval between target and mask was fixed at 96 msec. This experiment was conducted with normal binocular viewing.

As a test of the generality of the results we also conducted the experiment under dichoptic viewing conditions using a stereoscope. The masks consisted of 5 deg segments of circles with the same gap sizes as in the binocular experiment. The main difference was that the target disk had a radius of 3.4 deg and the masks had radii of 2.0 deg. In the dichoptic experiment the target and mask were simultaneously presented.

Results

Because of the great similarity of the binocular and dichoptic results we present them together in Fig. 7. In these graphs the brightness matches made by the subjects are plotted as functions of the gap size in the masks. In all the graphs for both dichoptic and binocular viewing, there is a gradual trend that as the gap size increases the brightness matches increase eventually reaching the unmasked levels at the far right of the plots (gap size = 360 deg).

Discussion

The results in Fig. 7 are reasonable in the sense that we expected the suppressive effect of the masks to decrease as increasingly larger gaps are made in their perimeter. However, what is striking is how gradually the strength of masking decreases as rather large gaps are introduced. To our surprise, even masks consisting of only 2–4, 5 deg segments had significant effects on the brightness seen at the center of the target. Figure 8 gives a rough idea of the percept in these conditions. One observer describes the percept in the following way:

"There is a small area of darkening in the target just to the outside of the segments in the mask. There is a considerably larger area of darkening to the inside of the masking lines as if the line segment in the mask were casting a shadow. The white extends from the border of the target inward toward the center but the brightness decreases inside the radius of the mask."
GENERAL DISCUSSION

When a large homogeneous stimulus is briefly presented and followed by a second masking stimulus, there is a dramatic reduction in the brightness of the first stimulus in the area central to the contours in the mask. When the two stimuli are presented to the same eye the strongest brightness suppression is obtained when the mask follows the target by 50–100 msec. On the other hand, if the target and mask are dichoptically presented the optimal masking interval extends from simultaneity to a delay of 50–100 msec. Experiment 2 showed that the latest times at which masking is obtained increase as the distance between the edges of the mask and target increases. Experiment 3 demonstrated that significant brightness suppression is observed even when the mask consists of a few isolated line segments rather than a continuous and closed contour.

In the following discussion we present the considerations that have led us to the tentative conclusion that our experiments reveal a brightness filling-in process. We begin by considering several possible explanations for the masking we’ve observed. As it currently seems that filling-in is the most parsimonious explanation for the experimental results, detailed mechanisms of filling-in are discussed. Those mechanisms best able to account for the results serve to delimit the properties of the underlying neural process.

Possible explanations for brightness suppression produced by contours

**Mechanism 1: lateral inhibition.** A simple explanation for the suppressive effect of the masks might be that there is an interaction between the neural responses to the target and mask stimuli in which the mask contours laterally inhibit parts of the target. This could explain why the brightness in the target is always lowest adjacent to the masking contours. However, there are two reasons why lateral inhibition isn’t a viable explanation for our findings. First, our qualitative perception of the suppressive effect is inconsistent with lateral inhibition. In expts 1 and 2 there was always brightness suppression extending much farther toward the inside of masking contours than to the outside. The configuration of the mask is critical and suppression is not observed in a fixed range extending from the masking line as one might expect with lateral inhibition. One could

---

**Fig. 7.** Brightness suppression in the center of the target is inversely correlated with the size of gaps in the circular mask. Effects with binocular (△) and dichoptic (○) viewing are similar. The target was a disk and the masks were constructed from 5 deg segments of a circle. The masks differed in the number of segments and correspondingly in the subtense of the gaps in their perimenter (see Fig. 8). The gap sizes were 0, 35, 55, 85, 115, 175 and 360 deg. The mask with 0 deg gaps was a complete circle and the mask with 360 deg gaps was a no-mask control.

It is worth noting that the transient presentation of the masks does not produce the perception of subjective contours between the line segments. Thus, it cannot be argued that subjective contours are interfering with filling-in as do real contours.
argue that inhibition with a fixed range has a greater suppressive effect on brightness inside than outside a circular mask because the neural response to the target disk is weaker inside the disk. However, this explanation is at odds with the observation that a single masking line near the center of a disk (where the response might be lowest) has very little effect on brightness whereas the same line formed into a circle has a much larger effect even when it is nearer the target's edge (where the response to the disk might be higher).

The second difficulty with lateral inhibition is that it can't account for the distance/time tradeoff seen in the results of exp 2. In that experiment the radius of the mask was constant and the radius of the disk-shaped target was varied. As the outer radius of the target was increased it was possible to get masking at later times. If we were observing simple lateral inhibition with a fixed timecourse, there shouldn't be any dependence on target size. Whatever kind of inhibition we've observed, it clearly depends on both the timing and the distance between edges of the stimuli.

Several psychophysical studies have suggested that there is propagating lateral inhibition in the visual system (Smith & Richards, 1969; Meijer, van der Wildt & van den Brink 1978; van der Wildt & Vrolijk, 1981). This is manifest as a time- and distance-dependent interaction between nearby lines or spots of light. Also, Bridgeman (1971) has argued that propagating inhibition might account for metacorsontrast. These observations are probably not related to the masking we've observed because they were seen only over a fraction of a degree whereas we saw target/mask interactions over several degrees. Additionally, propagating inhibition cannot account for the greater suppressed area to the inside of masking contours.

Mechanism 2: processing-time dependent on retinal eccentricity. Our data might be accounted for by simple inhibition if one additional assumption is made. Specifically, consider that the speed of visual processing is dependent on eccentricity. If the response to the outer edge of the target stimulus is faster than the response to the central area of the target then this may explain why a delayed mask can interfere with brightness perceived in the center but not at the border. There is experimental evidence that peripheral processing is faster than that in the fovea. Evoked potentials in the retina have a shorter latency at a few degrees of eccentricity than they do centrally (Erich Sutter, personal communication). Also, with stimuli scaled in size for retinal eccentricity it is found that critical flicker frequency increases with eccentricity (Tyler, 1985) suggesting that peripheral processing is faster.

While some of the data could be explained by faster peripheral vision, this cannot account for all of our findings. The most serious difficulty is encountered in trying to explain the results of exp 2. There we found that with a fixed size mask suppression is observed at later times if the target is larger. The filling-in hypothesis suggests that this is because the distance between the target and mask increases with increasing target size and it takes much longer for the filling-in process to reach the mask. An alternative is that the larger target is masked later because it...
evokes a response at a later time than a smaller target. Thus a long target/mask interval for a large target is equivalent to a shorter target/mask interval for a smaller target. Unfortunately this suggestion that the larger (and thus more peripheral) target gives a slower response is opposite to the evidence cited above which indicates that peripheral processing is actually faster rather than slower. For this reason it doesn't seem possible to explain all the data by eccentricity-dependent differences in the speed of visual processing.

**Mechanism 3: metacontrast.** There are aspects of our data that are reminiscent of metacontrast and one must consider the possibility that we have observed some variant of this phenomenon. Metacontrast is defined as a decrease in brightness of a briefly presented stimulus when it is followed by a second spatially adjacent stimulus. In assessing the relationship between metacontrast and the masking we’ve observed, one must deal with the fact that the term metacontrast comprises a variety of experimental observations and it is not known how they relate to normal visual processing. Given the vast number of experimental studies of metacontrast, what is the common thread that relates them? Based on the definition above, a reasonable answer is that metacontrast has two key elements: first, it involves backward masking and second it occurs when the target and mask have spatially adjacent edges.

Our experimental paradigm incorporates one key element of metacontrast in that we are studying a strong form of backward masking. The optimal masking times in metacontrast and our experiments are similar because in both phenomena, under appropriate conditions, optimal monoptic masking is observed with a target/mask interval of 50–100 msec. However, unlike metacontrast, in our experiments the strength of masking is not determined by the proximity of edges in the target and mask. In fact, the data in expt 2 clearly show that we obtain much better masking when the edges of the stimuli are not near. That the adjacency of edges is fundamental to metacontrast is evidenced by the models that have been constructed to explain it. Most models of metacontrast are based on the idea of parallel visual pathways with different transmission speeds (reviewed by Breitmeyer, 1984). For instance, Breitmeyer assumes that visual stimuli elicit a slow excitatory response and a fast inhibitory response. Metacontrast masking results when the slow excitatory response to one stimulus collides at some location in the brain with the fast inhibitory response to a second stimulus. Implicit in this explanation is the idea that the borders of the two stimuli are adjacent as in the masking of a central disk by an annular surround. It isn’t that the second stimulus produces global visual inhibition. Rather, the edges of the second stimulus produce inhibition in their immediate vicinity. This is a reasonable model for metacontrast given the experimental data showing that the masking of a disk by a delayed annulus decreases rapidly as the inner diameter of the annulus is made larger than the outer diameter of the disk (see Breitmeyer, 1984). However, these models based on local edge inhibition can’t explain our results. The masking we’ve observed has a particular dependence on the spatial and temporal distance between the target and mask which is unlike metacontrast. Also, in metacontrast it is found that the degree of masking depends on the relative “energy” in the target and mask (Weisstein, 1972). The models would probably not predict that a very sparse mask such as the small line segments in expt 3, would yield significant masking. Another interesting point is that while there is considerable controversy about whether metacontrast occurs dichoptically (Breitmeyer, 1984), the masking we’ve observed is strongest under dichoptic viewing conditions.

In summary, we believe it is incorrect to call the masking we’ve observed metacontrast. However, this does not mean that the phenomena are unrelated. Both effects can occur over similar time scales and they both involve the suppression of brightness. Stopfer and Mansfield (1978) describe a metacontrast effect called area suppression which is similar, in part, to the brightness suppression we have studied. Interestingly, they suggest that this effect might result from interference with filling-in. Ultimately it is important to relate both metacontrast and the brightness suppression we’ve observed to normal visual processing. We hope that a model able to explain our findings will suggest a new way to consider metacontrast.

**Mechanisms of filling-in**

As mentioned in the Introduction, there is good evidence from studies of pathological and artificial scotomata that perceptual filling of a blind spot occurs if it is surrounded by a region of uniform brightness and color. Similar perceptual filling-in is seen by normal observers in
experiments with stabilized images. Color and brightness can perceptually spread across an edge and fill into an area stabilized on the retina. In cases of both scotomata and stabilized images, information at a border is sufficient to give the perception of a uniformly filled-in area even though there may be no retinal activity corresponding to the filled-in part of the stimulus. While we cannot prove that our masking results are based on the same mechanism as the demonstrations of perceptual filling-in it appears that a single process can account for both phenomena. One possible mechanism is a process of assignment in which, in the lack of conflicting information, the central area of a uniform stimulus is automatically assigned the same brightness as the edge. This implies that there is a non-isomorphic relationship between the perceived brightness pattern and the underlying neural activity. It is conceivable that the masking in expt 1 results from interference with such a process before the central area is assigned the brightness at the target’s border. If this is the case the process must take 50–100 msec to occur since masking was observed with this temporal interval between target and mask. However, if brightness away from borders is assigned without the need for any spreading activation, it is difficult to understand why the timecourse depends on stimulus size. In expt 2, we found that the larger the target is, the longer it is susceptible to masking. This clearly implies that there is a processing time dependent on size. Even more problematic for the assignment explanation is that with a fixed size target disk, the degree of masking depends on the distance of the mask circle from the edge of the target (expt 2). If the disk-shaped stimulus is the same, why should the timecourse of its assignment of brightness depend on the size of the masking stimulus?

The alternative to brightness assignment based on edge information is an active filling-in mechanism involving propagation of brightness. Although the final percept is still determined by the brightness at edges, the process differs from assignment in that it takes a finite time for the signal to propagate. The neural activity underlying a filled-in percept might be isomorphic with perceived brightness or alternatively, there might be brightness filling-in only in a special subpopulation of cells. As discussed after the individual experiments, brightness propagation can account for the suppressive effect of a delayed mask and the relationship between masking times and the sizes of target and mask. The results in expt 2 can be directly interpreted in terms of a propagation speed because they indicate that the latest time at which masking is effective increases as the distance between target and mask increases. This may be because brightness must go from the target’s border to the mask’s border before any masking occurs. If one makes a least squares fit to the data for the change in masking time as the target/mask distance changes, the propagation speed is calculated to be roughly 110–150 deg/sec (Fig. 5). Of course this ignores the possibility that speed might change with eccentricity. For instance, if filling-in involves the spread of activity in visual cortex then the cortical distance that a signal must travel depends on the cortical magnification factor. Our best guess is that filling-in would be a cortical phenomenon because there is strong dichoptic masking. Using the magnification factor estimated for V1 by Levi et al. (1985) we calculate that the propagation speed of filling-in would be roughly 0.15–0.40 m/sec (Fig. 6). This is roughly the same order of magnitude as physiological measurements of lateral excitation in cortex which are 0.06–0.09 m/sec (Chervin et al., 1988).

The results obtained in expt 3 were somewhat unexpected and pose a challenge to the simplest conception of filling-in. The data clearly show that even masks consisting of a few isolated line segments can have a significant effect on the perceived brightness of the central area of the target. Because the gaps between the line segments were in some cases fairly large, it is difficult to imagine that the masking lines simply serve as barriers to block the spread of brightness. The key to understanding the results in expt 3 may lie in the details of the perceived pattern of brightness. There are two features of this pattern that may be particularly significant. The first is that even though the strong suppressive effect of each mask occurs inside its radius, there is a small darkened area just outside the mask. Secondly, the brightness gradually increases moving inward from the masking segments. These features and the quantitative results in expt 3 may be understood by considering the role of brightness boundary conditions.

**Filling-in and brightness boundary conditions**

The features of the brightness profiles observed in expts 1–3 can be understood in terms of the constraints imposed by the target and mask stimuli. To understand what this means, its helpful to consider how our experiments differ.
from normal visual experience. For instance, suppose ones looking at a large white disk on a black background. Proceeding from left to right across the stimulus there is first a black/white edge then an area of uniform white and then a white/black edge. Under normal circumstances filling-in might proceed from the black/white edge until a white/black edge is encountered. Our masking paradigm interferes with this process by presenting a white masking line on a black background thus introducing a black/white/black edge. What may happen when the white disk-shaped target is quickly flashed and followed by the mask is that filling-in starts from the disk’s black/white edge and it encounters the mask’s black/white/black edge rather than the “expected” white/black border at the other side of the target. Based on the qualitative observations above, it appears that the final brightness pattern might result from a combination of filling-in and smoothing. Filling-in starts at the edges of the target and proceeds until the other edges are encountered. If the filling-in level and the newly encountered edge differ in brightness then there is smoothing between the two brightness levels. In this way information at all explicit edges establishes boundary conditions that must be satisfied by the final brightness landscape. A physical analogy that we’ve found useful is to imagine that the final pattern of brightness is a stretchable fabric and the degree of brightness corresponds to the height of this fabric off the floor. Presentation of the target corresponds to suddenly lifting up the edge of the fabric. Because of the elasticity of the material, the central portion of the fabric is gradually pulled up corresponding to the process of brightness filling-in. If a mask is suddenly presented in the center of the target before filling-in has completed then this introduces new boundary conditions specifying that the masking line is bright with dark edges to its sides. In the fabric analogy the dark/light boundary conditions serve to nail down the fabric at the edges of the mask. Even as filling-in proceeds the height of the material must smoothly fall off toward these nailed-down areas. Its easy to see with this analogy how the boundary conditions established by lines can affect the global brightness terrain. For instance, suppose the target is a white disk and the mask is a single line segment contained within the boundaries of the target. This is a relatively ineffectual mask because filling-in can proceed nearly up to the line before there must be smoothing to meet the black/white edge on the side of the masking line. The same thing would happen outside of the masking circle in expts 1–3 above. However, inside the masking circle the only boundary condition may be that the mask has a black edge on its inside. Thus, as we observed, there can be an extensive area of black perceived inside the radius of the mask even though the target is uniformly white. One caveat is that in some cases we perceived a somewhat brighter area (a gray spot) at the very center and darkening at larger radii nearer the circular mask. This might be explicable if the mask is presented at a time at which some filling-in of the target has already occurred inside the radius of the mask. The center of the target disk can no longer be black but there is still the boundary condition that the brightness profile must be black at the mask’s inner edge. A way to satisfy both constraints is that the center of the disk is somewhat filled-in and the brightness rolls off toward the masking circle. This is at least qualitatively consistent with our observations. The perceived pattern of brightness in expt 3 (Fig. 8) can be understood by imagining that the small line segments with their explicit light/dark edges hold down the brightness adjacent to the segments. The resultant brightness landscape is a relatively complex interaction between the process of filling-in and smoothing near the masking contours.

The observation that is the most difficult to account for is the finding in expt 2 that brightness suppression is sometimes observed even when the mask has a larger radius than the target. Importantly, this effect is weak compared to the cases in which the mask is contained within the target and it only occurs when the mask and target are quite similar in size. The cause of this effect might be related to the spatial scale at which edge signals and filling-in occur. For example, if filling-in occurs at a rather coarse scale then masking stimuli near but outside the target could affect the target’s brightness.

At any rate, a mechanism incorporating both filling-in and smoothing presently appears best able to account for the effects of a variety of masks including arrays of small spots, randomly arranged line segments, and generally any stimulus with interior contours. The effectiveness of these discontinuous masks is inconsistent with a simplistic scheme in which the masking contours block the spread of brightness. However, the resulting patterns of brightness suppression are consistent with the idea that brightness fills-in and smooths to satisfy the boundary conditions.
imposed by the masking contours.* It is worth noting that filling-in or smoothing processes have been proposed in other visual modalities such as motion processing (Horn & Schunk, 1981; Hildreth, 1984; Nakayama & Silverman, 1988) and may be an important component of the visual system's solution to ill-posed problems (Poggio, Torre & Koch, 1985). In the case of brightness suppression, both filling-in and smoothing are required to explain the spatio-temporal properties of the masking as well as the details of the brightness percepts near contours. Identifying plausible physiological mechanisms which might underlie these processes is a difficult question deserving further investigation.

**Models of filling-in**

There have been several attempts to model the process of filling-in. Walls (1954) described a hypothetical system in which the response to stimulus borders initiates the filling-in of color and brightness which stops when other borders are encountered. Gerrits and Vendrik (1970) presented a somewhat more formal model in which the responses of on-center retinal cells initiate the spread of brightness in some "higher center". This filling-in propagates until it is stopped by a darkness signal produced by retinal off-center cells. Conversely, when a stimulus is extinguished, off-center cells respond and initiate the spread of darkness. Darkness filling-in then propagates until a brightness signal is encountered. These complementary mechanisms account for the spread of brightness across stabilized borders by assuming that, in the absence of stimulus motion on the retina, the "higher center" adapts to the signals that normally stop filling-in. While this is a workable model for qualitatively explaining the stabilization experiments, its lack of explicit mechanisms and time constants makes it difficult to apply to our experiments.

A more detailed mathematical model has been developed by Grossberg and coworkers (Cohen & Grossberg, 1985; Grossberg & Mingolla, 1985; Grossberg & Todorovic, 1988). One layer in their neural network is a cellular synecytium which supports the rapid spread of excitation. The activity in this layer corresponds to the visible percept. The cell-to-cell diffusion of activity which occurs in the synecytium is started by a signal resulting from convolving the stimulus with difference-of-Gaussians filters. The spreading activity is contained by shunting inhibition produced by edge signals. Within the synecytium there is constant interaction between nearest-neighbor cells until a state of equilibrium is reached and all cells have the same level of activity. This model is able to successfully account for the results of a variety of experiments on brightness perception. However, we have had two difficulties trying to apply Grossberg et al.'s model to our experiments. The first problem is that diffusion continues until it is stopped by shunting inhibition. This suggests that there should generally be sharp edges to filled-in areas corresponding to the sites of shunting inhibition. This is inconsistent with our observations indicating that in many cases brightness gradually rolls off near masking lines. A second difficulty results from the feature that diffusion continues until an equilibrium state is reached. It isn't obvious how such a process could account for the significant masking observed with the line segments in exp 3. The model seems to incorrectly predict that brightness would diffuse around the edges of the masking segments and fill-in the center of the target. Nonetheless, given the ability of this model to explain other aspects of brightness perception it is important to determine if alterations to it will provide an explanation for the suppressive effects of contours on brightness. In developing our own neural model to account for the experimental results we have borrowed important ideas from the models of Gerrits and Vendrik and Grossberg and coworkers (Paradiso & Nakayama, 1989). This modeling work will be presented in detail elsewhere but suffice it to say that it formalizes the idea of combining filling-in (brightness propagation) and smoothing at boundary conditions.

*An important question which remains open is whether filling-in occurs in a color-coded manner or alternatively that brightness fills-in and color is more or less "tacked on". While our results are consistent with a process of brightness filling-in, in the literature the concept of filling-in is often discussed in relation to color perception. The reason for this probably derives from the nature of earlier experiments. In the studies of people with retinal scotomata, both color and brightness perceptually fill-in. Thus there is no particular reason to preferentially call the phenomenon color or brightness filling. In a typical stabilized image experiment the border between areas of different color is stabilized and the color from one area spreads into the other, suggesting that color does fill-in. In our masking experiments we have used targets and masks of various colors and obtain the same effect irrespective of the hue of the stimuli. What is required in the future is a carefully controlled study to look for evidence that the suppression produced by contours is color specific.
Acknowledgement—This work was supported by AFOSR grant 83-0320.

REFERENCES


