

# Competition for Consciousness Among Visual Events: The Psychophysics of Reentrant Visual Processes

Vincent Di Lollo and James T. Enns  
University of British Columbia

Ronald A. Rensink  
Nissan Research & Development, Inc.,  
Cambridge, Massachusetts

Advances in neuroscience implicate reentrant signaling as the predominant form of communication between brain areas. This principle was used in a series of masking experiments that defy explanation by feed-forward theories. The masking occurs when a brief display of target plus mask is continued with the mask alone. Two masking processes were found: an early process affected by physical factors such as adapting luminance and a later process affected by attentional factors such as set size. This later process is called *masking by object substitution*, because it occurs whenever there is a mismatch between the reentrant visual representation and the ongoing lower level activity. Iterative reentrant processing was formalized in a computational model that provides an excellent fit to the data. The model provides a more comprehensive account of all forms of visual masking than do the long-held feed-forward views based on inhibitory contour interactions.

From the time a stimulus first enters the eye to the time a percept emerges into consciousness, the initial stimulus has been coded at several levels in the visual system. One of the main goals in studying visual information processing is to specify the representations at each level and the temporal sequence between them (Lachman, Lachman, & Butterfield, 1979). Historically, that sequence was regarded as being mainly feed forward, with processing advancing from simple to increasingly complex attributes, along brain pathways that converge to a common area in which conscious perception occurs. Here, we present an alternative approach, one more in keeping with recent advances in visual neurophysiology. Because we now know that reentrant signals are a major form of communication between brain regions, it is time to begin making explicit use of reentrant processing in theories of perception. To show that this can be done effectively in one domain, we report on a series of psychophysical experiments involving visual masking. In our view, a similar approach could be adopted in other areas of visual perception, especially those in which temporal aspects of processing are at issue.

## PART I: THE CASE FOR A NEW THEORETICAL ORIENTATION

Most theories of perception hold to some form of interaction between seeing and understanding, or stimulus-driven versus conceptually driven processing, in predicting what will be perceived (Broadbent, 1958; Hebb, 1949; Helmholtz, 1866/1962; Neisser, 1967). A well-known and easily understood behavioral example of a top-down influence on perception is Bruner and Goodman's (1947) finding that poor children overestimate the size of coins in comparison with wealthy children.

In dealing with the theoretical problem of how the lower and higher levels interact, two options must be considered. On one option, the sequence of events is deemed to be unidirectional, so that various perceived attributes of a stimulus accrue as processing ascends from lower to higher levels. Processing takes place discretely at each level, meaning that once information has been transferred from a given level to another, processing at that level is complete. Top-down influences, such as the differential perception of coin size in Bruner and Goodman's (1947) experiment, are exerted entirely within a relatively high level in the system, after the results of lower level processes have been transferred to higher levels.

On the second option, processing is accomplished through iterative exchanges of neural signals among levels. An initial wave of stimulation ascends rapidly through the system, followed by descending signals between levels. Together, the ascending and descending pathways form part of an iterative-loop system, aimed at noise reduction and hypothesis verification, thereby establishing the most plausible perceptual interpretation of the incoming stimulus. In this view, the perception of the coins in Bruner and Goodman's (1947) experiment was based on concurrent interactions among all levels. The ascending activity initially provided some basis for the perception of coin size, but this activity was

---

Vincent Di Lollo and James T. Enns, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada; Ronald A. Rensink, Nissan Research & Development Inc., Cambridge, Massachusetts.

This research was supported by grants from the National Sciences and Engineering Research Council of Canada. We are indebted to Marvin Chun, Julian Hochberg, Marcel Just, Victor Lamme, and Steve Yantis for comments on earlier versions of this article. We are grateful to Rob Gordon, Margolee Horn, and Jamie Richards for assistance in data collection.

Correspondence concerning this article should be addressed to either Vincent Di Lollo or James T. Enns, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia Canada V6T 1Z4. Electronic mail may be sent to vince.dilollo@ubc.ca or jenns@psych.ubc.ca.

quickly modified by descending activity, reflecting the conceptual importance of coins to the observers.

Among theories of visual information processing, some form of the unidirectional option has been adopted far more frequently than the reentrant option. This may have been prompted by the widely publicized progression in the receptive field properties of cortical visual neurons, aptly characterized as simple, complex, and hypercomplex (Hubel & Wiesel, 1977). However, a corresponding body of neuroanatomical and physiological evidence has recently emerged which provides the appropriate physiological correlates for models of reentrant processing. In the present work, we explore ways in which this neural wiring may be incorporated into models of visual processing. Our intent is not to provide an exclusive alternative to a unidirectional approach, but to explore the extent to which current theoretical approaches can be augmented with notions of iterative reentrant processing, to provide more comprehensive accounts of visual information processing.

### Reentrant Processing: A Design Feature Neglected by Theory

Communication between two brain areas that are connected to each other is seldom one-way. If a source area sends signals to a target area, then the target area sends signals back to the source area through reentrant pathways (Felleman & Van Essen, 1991; Zeki, 1993). Two aspects of this connectivity are especially salient for our purpose. First, all major visual centers have reentrant connections with area V1, which is where stimulation first enters the cortex. Thus, neurons in the primary visual area can be activated by reentrant fibers from extrastriate cortex (Bullier, McCourt, & Henry, 1988; Mignard & Malpeli, 1991). Second, it is a general rule that the size of the receptive fields is smallest in V1 and increases progressively at higher centers. Because its receptive field is small, any given unit in an orientation column in V1 has no way of "knowing" whether the external stimulus is an isolated line segment or part of a more complex configuration. By the same token, a high-level unit might know the total configuration, but not its exact spatial location. Free exchange of information between the two levels would resolve the issue. Commenting on this pattern of reentrant organization, Damasio (1994) has remarked that the interconnectivity between visual centers is suggestive less of a "forward-moving stream" than of a "perpetual recurrence" of processing.

Major advances have been made in recent years in what is known about the topography of the pathways connecting lower with higher areas in the brain. As a general rule, ascending signals originate in Layers II and III of the lower area and terminate at the level of the granular cells in Layer IV of the higher area. Descending signals, on the other hand, originate in the pyramidal cells in Layer V of the higher area and terminate in Layers I and VI of the lower area (Felleman & Van Essen, 1991; Perkel, Bullier, & Kennedy, 1986). These general rules of interconnectivity apply throughout the visual pathways, from sensory input to the highest cognitive areas. Thus, ascending and descending signals in reentrant loops follow distinct pathways before coming into contact with each other. In this way, more abstract reentrant codes may be convolved with simpler low-level codes, perhaps to test alternative perceptual hypotheses regarding the incoming stimuli (Grossberg, 1995a; Mumford, 1992).

A splendid example of cortical interconnections and their likely functions has been provided by Shipp and Zeki (1989), who traced the units that project from V1 to V5 and the corresponding reentrant connections that return to V1 from V5. A notable aspect of this two-way link is that the reentrant fibers do not merely complete a feedback loop, from lower level neuron to higher level neuron and back, but are distributed widely on their return, including many neurons in the spaces between the source neurons. This arrangement constitutes a powerful means by which a target area not only can influence the outputs of currently active cells in the source area but also can selectively sensitize or inhibit other neighboring cells to modulate future incoming signals.

Reentrant connections from area V5 of the macaque monkey to lower visual areas have also been studied by Hupe et al. (1998), who found that the reentrant signals served to amplify and focus the activity of units in the lower areas and to facilitate figure-ground differentiation. A similar pattern of reentrant connectivity has been studied between Layer VI of visual cortex and lateral geniculate nucleus in cat (Sillito, Jones, Gerstein, & West, 1994). These researchers have suggested that the function of cortical feedback is to test for the presence of specific patterns in the incoming activity. The feedback circuit actively searches for correlations between a descending code representing a perceptual hypothesis and the initial pattern of low-level activity. When such a correlation is found, the ensemble is "locked" onto the stimulus.

Behavioral implications of neuroanatomical and physiological evidence for reentry were realized early by Hebb (1949), who saw in them a way to make concrete the then-subjective ideas of attention and imagery. A more recent theory that incorporates a form of reentrant processing was proposed by Treisman and Gelade (1980). They hypothesized that the features of a stimulus were first registered, in parallel, in a series of retinotopic maps similar to those found in extrastriate visual regions. To determine whether a given location in the visual field contained a conjunction of features (e.g., red and vertical), some information about these features was communicated to a common representation in which spatial location was explicitly coded. This was called the *master map of locations*, and it was described as having properties akin to area V1. Thus, this theory implied a form of reentry to conjoin the visual attributes that had been registered separately.

Many more studies have provided existence proof of top-down effects. The general strategy in such studies has been to demonstrate that some attribute of a stimulus is detected more accurately or rapidly when it is presented in a context that is familiar and meaningful (Reicher, 1969; Weisstein & Harris, 1974). Some of these effects have even been simulated using parallel distributed networks in which reciprocal activation of units at different levels has been a central feature (McClelland & Rumelhart, 1981). The main emphasis in these studies, however, has been on the existence and importance of top-down influences. Much less emphasis has been placed on the temporal course of the hypothesized reentrant loops and on the question of whether activity in lower levels is actually modified by descending activity.

In what follows, we describe a psychophysical method, recently developed in studies of motion perception and metacontrast masking (Di Lollo, Bischof, & Dixon, 1993), which we believe can be used more widely to address these questions. Before describing the experiments in detail, we first describe the general aspects of the methodology, including how we believe it can be used to inves-

tigate the temporal course of reentrant processes in visual perception. Such timing estimates from psychophysical data are critically important in establishing links with neurophysiological mechanisms.

### The Common-Onset Paradigm

Our approach is illustrated by a class of visual displays in which brief stimuli are perceived in ways that are not predictable using strictly feed-forward processes. In an experiment by Di Lollo et al. (1993), the target consisted of a square outline with a small gap at either the top, bottom, left, or right. This target square was displayed within a slightly larger square with a gap on each side, which acted as a mask. The observer's task was to identify the location of the gap in the inner square. The sequence began with a simultaneous display of both squares for a brief interval (10 ms). Then the inner target square was turned off, and the outer (masking) square remained on the screen for durations of up to several seconds. This stimulus sequence is illustrated in Figure 1A.

The stimulus sequence seen by observers differed markedly from the actual display. If perception were veridical, one would have expected observers to see a combined two-square display, part of which (the inner square) disappeared a short time later while the outer square remained visible for a further period of time. Instead, the phenomenal appearance of the display was quite different. As shown in Figure 1B, the combined two-square configuration was seen clearly, and the location of the gap in the inner square was identified accurately, only if the target and the mask both began and terminated together. However, if the initial two-

square configuration continued with a display of the masking square alone, perception of the target was severely impaired and the location of the gap could no longer be determined accurately, as shown in Figure 1C. Throughout this article, we refer to this form of backward masking as *common-onset masking*.

### Inhibitory Processes and Common-Onset Masking

Impaired perception of the target, as illustrated in Figure 1, must be regarded as a case of backward masking, because the perception of a display that is clearly visible when presented in isolation is severely impaired by events that occur subsequently. However, it must also be distinguished from other forms of backward masking, including *masking by light*, *by pattern*, and *by metacontrast* (Breitmeyer, 1984).

Masking by light is most easily differentiated from common-onset masking because although the former depends on a uniform change in light level across the visual field, the latter occurs even though the luminance of the display remains constant. Because there are no such changes in light level in common-onset masking, masking by light cannot play a role.

Backward masking by pattern bears slightly more resemblance to common-onset masking. This form of masking occurs when perception of a target is impaired by a temporally trailing, spatially overlapping, stimulus consisting of a field of noise or structured contours. Two major processes are hypothesized to be involved: *integration* of contours and *interruption* of processing (Kahneman, 1968; Scheerer, 1973; Spencer & Shuntich, 1970). Integration occurs if the target and the mask are combined into a unitary percept in which the image of the target is degraded by the overlapping contents of the mask. It is therefore strongest when the target and mask are presented simultaneously (stimulus-onset asynchrony [SOA] = 0), its strength diminishing as the SOA is increased. However, integration is easily dismissed as a factor in common-onset masking because there is never any spatial overlap between the target and the mask.

The second process involved in pattern masking, interruption, occurs when processing of the target is disrupted by the arrival of the mask, which takes over those processing mechanisms that are required in common by both stimuli. In this case, the target is not degraded; rather, its processing is cut short by the onset of the mask. This process is aptly described through an analogy in which targets and masks are customers coming before the clerk of the visual system (Kolers, 1968). Kolers surmised that the amount of time the clerk spent with the first customer (the target) would be sharply curtailed if a second customer (the mask) entered the store.

At a first glance, it might seem reasonable to regard common-onset masking as an instance of interruption masking, because the temporally trailing mask seems to interrupt processing of the target. But, again, there are serious problems. First, all theories of interruption are predicated on the mask following the target in time; it is the onset of the mask that interrupts the processing of the target (e.g., Michaels & Turvey, 1979; Turvey, 1973). In common-onset masking, there is no mask onset that is distinct from the onset of the target. In Kolers's (1968) analogy, two customers enter the store as a family unit. Second, and as important, the contours of the mask never overlap with those of the target, an essential ingredient in pattern masking. Finally, target and mask begin together in

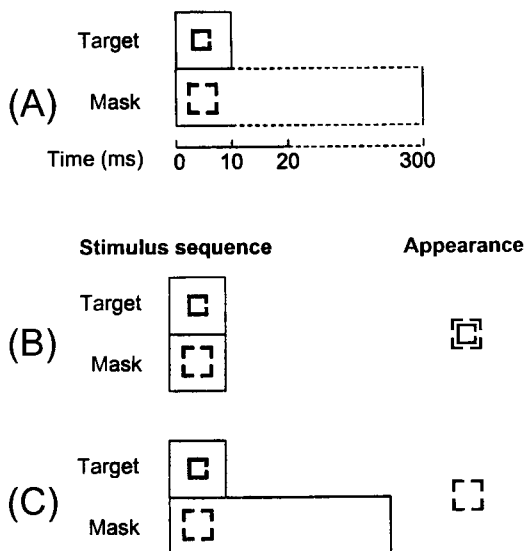


Figure 1. Schematic representation of the stimulus sequence in the common-onset masking paradigm. A: The sequence begins with a combined display of the target and the mask. After a brief period, the target is turned off, and the mask remains on view alone for various durations, which can include a duration of 0. B: When the duration of the trailing mask is equal to 0, the target and the mask both start and end together. In this case, both stimuli are seen clearly and distinctly. C: When the duration of the trailing mask exceeds about 100 to 150 ms, the target is not seen, and the area within the mask appears empty.

common-onset masking. This is the defining characteristic of masking by integration, not by interruption.

The form of masking that most closely resembles common-onset masking is *metaccontrast masking*, in which a target (such as the inner square in Figure 1A) and a mask (the outer square) are displayed sequentially in close spatial proximity (Alpern, 1953; Breitmeyer, 1984; Werner, 1935). The basic finding is that when the interval from the onset of the target to the onset of the mask is either very brief or very long, the target is perceived clearly and accurately. At intermediate SOAs, however, perception of the target is impaired, leading to a U-shaped function of accuracy over SOA. Strict temporal succession of target and mask is deemed essential for metaccontrast masking. Namely, the onset of the target must precede the onset of the mask by some temporal interval. Estimates of the optimal interval tend to be in the range of 50 to 150 ms. Because of its stability, this temporal relationship has come to be known as the *onset-onset law* (Kahneman, 1967) or the *SOA law* (Breitmeyer, 1984).

Except for the timing of the stimuli, the procedures for obtaining masking with common onset are the same as those for metaccontrast masking. From a theoretical perspective, however, the difference in timing is critical because masking with common onset is not predicted by current models. Whether based on inhibitory interactions (Breitmeyer & Ganz, 1976; Weisstein, Ozog, & Szoc, 1975) or on principles such as interactive trace decay (Ganz, 1975), current accounts consider metaccontrast to be an effect based on transient neural events triggered by the onsets of the stimuli and to depend critically on the SOA law. This is especially true for inhibitory models because they are predicated on excitatory and inhibitory processes that interact appropriately only if the stimuli are presented sequentially.

Inhibitory models cannot account for common-onset masking because they were designed to exclude the possibility of masking when the target and mask are presented simultaneously (Alpern, 1953; Breitmeyer & Ganz, 1976; Matin, 1975; Weisstein et al., 1975). This was done to account for the universal finding that metaccontrast masking occurs not at an SOA of zero, but at much longer SOAs. We should note, incidentally, that masking with common onset had never been seen in earlier work because in the conventional control condition the target and mask not only started together but also ended together. This prevented the factor truly essential for masking from being identified, namely, that the mask be seen alone at some appropriate time after the initial display. Be that as it may, current theories are explicit in asserting that metaccontrast masking cannot occur when the target and the mask have common onsets. The mechanisms that ensure that metaccontrast masking does not occur when SOA is equal to zero have been illustrated graphically by Breitmeyer and Ganz (1976, Figure 11b) and by Weisstein et al. (1975, Figure 5).

This said, one might consider revising the tenets of these models to accommodate masking with common onset. However, it is impossible to develop a theoretical account of common-onset masking that is linked to the key feature of all current models: transient neural signals. Neither *on* nor *off* transients can be linked to the cause of masking when the target and the mask are presented simultaneously. Consider each possible transient in turn.

First, because the target and the mask start together, there is no reason why the *on* transient of the mask, and not that of the target, should prevail as a source of masking. Second, *off* transients also

are inadequate. Were masking to arise from the transient produced by target offset, it would impair the visibility of the mask, not that of the target. But what of the *off* transient produced by the mask? Perhaps common-onset masking is an instance of offset-offset masking, in which the *off* transient produced by the termination of the mask interferes with perception of the target (e.g., Breitmeyer & Kersey, 1981). In fact, the *off* transient of the mask is demonstrably ineffectual in common-onset masking. In a study by Di Lollo et al. (1993), the mask remained on display until after the observer had made a response, thus eliminating the *off* transient as a contributor to perception of the target. There was no reduction in the strength of common-onset masking, confirming that common-onset masking is not an instance of offset-offset masking. Third, it is also not the case that the target is missed because it is displayed too briefly. The target is clearly visible when the two stimuli start and end together in displays that are just as brief. Finally, it cannot be said that masking occurs because the brief target is overwhelmed by the longer mask (e.g., the longer stimulus might be weighted more heavily or be given greater prominence in perceptual processing). This option is denied by the fact that no matter how long the mask or how brief the target, masking never occurs if the display begins with the mask alone and ends with a simultaneous display of target and mask.

The main obstacle in accounting for common-onset masking by feed-forward inhibitory theories is therefore the absence of any relevant transient neural signals in the stimulus sequence. The only remaining hypothesis is that masking with common onset is not mediated by transients at all, but by that portion of the display that remains visible after the target has been turned off. This strongly suggests that the emerging neural activity from the briefly presented target has been rapidly reinterpreted on the basis of the current sensory evidence favoring a solitary mask.

### A Hypothesis Involving Iterative Reentry

We believe that an account of common-onset masking can be developed in the broader context of how the visual system copes with rapid input. Here, we present a brief outline of the conceptual framework to provide a rationale for the experiments to be reported in Part II and the computational model in Part III. In Part IV we return to the question of whether this view can encompass other forms of masking.

In keeping with current thinking, we suppose that conscious perception of a stimulus emerges from preconscious neural events in one or more visual areas. This activation is caused not only by feed-forward signals from the retina but also by iterative exchanges among multiple visual areas. On the first cycle, the input from the initial display is encoded at a low level within the system (i.e., it is fine-grained in resolution, and the attributes coded are simple) and then proceeds to higher levels in which tentative cognitive representations are produced. These representations must be regarded as being in need of confirmation for at least two reasons. First, the initial representation may be incomplete and ill defined, or equivalently, the ascending signals might activate more than one representation at the higher level. This creates an ambiguity for perception that can be resolved by comparing the high-level codes with the initial pattern of activity at the lower level. Reentrant pathways would enable such a comparison. Second, because the receptive fields at the higher levels are large by

necessity, to permit more complex attributes to be coded, sensitivity to location is reduced. However, it can be regained by sending the high-level codes back to the lower areas in which small receptive fields maintain high-precision spatial maps.

We assume that on the second and later iterations, the high-level representation reenters the lower level and is compared with what is currently there, perhaps through a process of correlation (e.g., Eggermont, 1990). If the image on the screen has not changed, a match is found, and processing continues. If the screen has been blanked after the initial display, there is nothing on the screen to compete with the decaying initial stimulus representation. Thus, for a short time, a match can still be found with this decaying trace. If, however, the reentrant information does not match the current information at the lower level (i.e., if the input has changed), then a new tentative representation emerges, and processing continues as for a new stimulus. We propose that masking occurs when an initial display (target + mask) is replaced on the screen by a different configuration (mask alone) before the required processing iterations have taken place for the target to be identified. Thus, the emerging percept of the compound image (target + mask) is replaced in consciousness with the percept of the mask alone. It is as though the visual system treats the trailing configuration as a transformation or replacement of the earlier one (Kolers, 1972).

Implicit in this description is a general account of backward masking in which a process of perceptual substitution plays a critical role. It is our view that this type of masking occurs when the emerging representation of a target is replaced by the emerging representation of a trailing mask as the object occupying a given spatial location. We have referred to this process as masking by *object substitution*, to distinguish it from masking processes that involve low-level factors such as contour proximity and luminance (Enns & Di Lollo, 1997). This has been done in two ways. First, masking by object substitution is greatest when visual attention is misdirected or distributed over multiple items in the visual field. Second, masking by object substitution is largely insensitive to contour proximity and stimulus luminance. In the present work, we therefore continue to use the term *object substitution* to denote these two aspects of common-onset masking.

### Rationale of the Present Experiments

From the perspective of iterative reentrant processing, it is predicted that any variable that increases the number of iterations required to identify a target will also increase the strength and the temporal course of backward masking. Variation in set size is one such factor that is ubiquitous in the literature on selective attention. In visual search tasks, *set size* refers to the total number of candidate items in which the target item is embedded (Duncan & Humphreys, 1989; Treisman & Gelade, 1980; Wolfe, Cave, & Franzel, 1989). In nonsearch flanker tasks, *set size* refers to the number of extraneous or nontarget items that surround the known target location (Eriksen, 1995). In partial report tasks, *set size* refers to the number of potential items that may be subsequently probed for identification (Sperling, 1960).

In common-onset masking, increments in set size are expected to have an adverse effect on the accuracy of target reports. With displays containing a single item, albeit randomly located in the display, it is assumed that attention can be rapidly deployed to the target location, because there is no interference in this process

from surrounding items. Conversely, with displays containing multiple items, attention will be deployed to the target location more slowly, and therefore interference from involuntary processing of the surrounding items is expected. On the reentrant hypothesis, a larger number of iterations will be required to identify a target among increasing numbers of distractors. As identification time increases, the probability also increases that the initial display of the target-mask combination will have been changed to mask alone. Any additional processing of the display after the target has been turned off would thus only serve to confirm the perceptual hypothesis that the masking stimulus alone had been shown in that location.

To make the rationale of our approach entirely clear, two separable aspects of the set-size manipulation must be distinguished. One is the well-established finding that target identification becomes progressively more impaired (longer reaction times, more errors) as the number of items in the display is increased. This result is obtained when all elements in the display start and end together. In the present experiments, this corresponds to the condition in which the mask ends at the same time as all other elements in the display. In this case, a conventional set-size effect is expected as the number of distractors is increased (Bergen & Julesz, 1983).

A second aspect of the set-size manipulation is more central to our views on reentrant processing. Its distinguishing characteristic is progressively larger effects of set size as the exposure duration of the trailing mask is increased. This follows from the joint premises (a) that the number of iterations required to identify the target item increases with set size and (b) that the continued presence of the mask after target offset leads to the perceptual substitution of the combined target-mask display by the mask alone. These masking contingencies predict an increase in the set-size effect that is significantly larger than that found when all display items begin and end together.

## PART II: PROBING REENTRANT PATHWAYS WITH PSYCHOPHYSICS

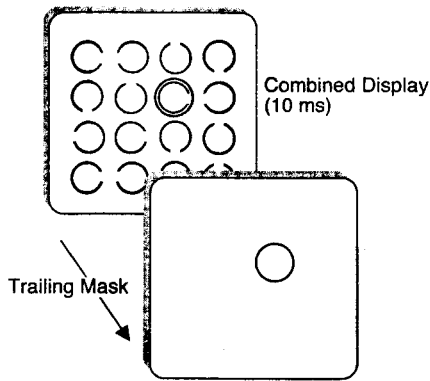
### Experiment 1: Common-Onset Masking Is Influenced by Set Size

The effect of set size on the strength of masking was examined in Experiment 1. We used brief displays containing up to 16 outline rings, each containing a gap either at the top, bottom, left, or right. One of the rings was singled out as the target by a surrounding ring, which also served as a mask (see Figure 2). Observers named the orientation of the gap in the target. All displays began with a brief simultaneous presentation of all elements and continued with the mask alone for varying durations. We found that masking was strongly affected by set size, in keeping with the hypothesis of reentrant processing.

#### Method

##### Observers and Apparatus

One male and one female undergraduate student with corrected-to-normal vision served as paid observers. Both were highly practiced psychophysical observers before the experiment began. They sat in a light-proof observation room and viewed the displays with natural pupils from



**Figure 2.** Schematic diagram of the stimulus sequence in Experiments 1, 2, and 3. The target was a broken ring singled out by a complete larger concentric ring, which also served as a mask. In Experiment 3, the larger ring was replaced with four dots located at the corners of an imaginary square concentric with the target ring (see Figure 8). The remaining broken rings served as distractors. In any given display, the number of broken rings varied from 1 (target with no distractors) to 16 (target with 15 distractors, illustrated). The sequence began with a combined display of the target, mask, and distractors for 10 ms and continued with a display of the mask alone for durations of 0, 40, 80, 160, or 320 ms.

a distance of 57 cm, set by a headrest. Stimuli were displayed within a  $4^\circ \times 4^\circ$  viewing area in the center of the screen of a Tektronix 608 oscilloscope equipped with P15 phosphor. The screen was front illuminated with a Kodak Carousel projector fitted with a 500-W General Electric Quartz-line projection lamp, attenuated to  $10 \text{ cd/m}^2$  by neutral-density filters. All stimuli were constructed with closely spaced dots. With front illumination turned on, the dots were seen as green-blue on a uniform white background. In dark-adapted viewing, as in Experiment 2, the dots were seen as light gray on a uniformly black screen ( $0 \text{ cd/m}^2$ ). The X, Y, and Z (intensity) coordinates of each dot were stored in a fast-plotting buffer that displayed them to the screen at a rate of one dot/ $\mu\text{s}$  (Finley, 1985). Photometric measurements were made with a Minolta LS-100 luminance meter.

### Stimuli

The stimuli are illustrated in Figure 2. On any given trial, the display consisted of one or more rings, each with a gap either at the top, bottom, left, or right. In units of visual angle, the thickness of each ring was approximately 1 min arc, the radius was  $0.4^\circ$ , and the size of the gap corresponded to a chord of  $0.4^\circ$ . One of the rings was singled out as the target by a slightly larger concentric ring, which also served as a mask. The radius of the masking ring was  $0.5^\circ$ , thus allowing for an interstice of approximately  $0.1^\circ$  between target and masking rings. The observer's task was to identify the orientation of the target's gap. The other rings in the display functioned as distractors. They were identical to the target, except that the gap was oriented randomly either to the top, bottom, left, or right, separately for each distractor. The viewing area was partitioned into a notional matrix of 16 square cells, each with  $1^\circ$  side. Every ring, whether target or distractor, was displayed in the center of one of the notional cells.

### Design and Procedures

The design consisted of the factorial combination of two independent variables: set size (number of rings in the display) and mask duration (the period for which the mask remained on the screen alone, after all other elements had been turned off). Set size was varied over five levels: 1, 2, 4,

8, or 16 rings, including the target. When set size was equal to 1, the display contained only the target. When set size was greater than 1, the display contained the target plus the appropriate number of distractors, distributed randomly amongst the remaining 15 locations, with the restriction that within a session, the target had to appear an equal number of times in each of the 16 locations.

The display sequence, illustrated in Figure 2, consisted of two frames presented sequentially without interruption. The first frame contained the target, the mask, and the distractors, if any. The exposure duration of the first frame was set at 10 ms. The second frame contained only the mask. Its duration was either 0, 40, 80, 160, or 320 ms. Thus, on any given trial, the sequence began with a display of all stimulus elements. After 10 ms, all elements except the mask were turned off, and the display continued with the mask alone for a variable duration, including zero. Trials were separated by a minimum interval of 5 s.

At the beginning of each trial, a small fixation cross was presented in the center of the screen. A trial was started when the observer pressed a button in a handheld box. The observer then indicated the location of the gap in the target by pressing the appropriate button (top, bottom, left, or right) in the handheld box, guessing if not sure. One session consisted of 160 trials. Within one session, the target was displayed 10 times in each of the 16 matrix locations, in a sequence that varied randomly across sessions and observers. In any given session, the number of distractors was fixed, as was the duration of the trailing mask. The 25 sessions resulting from the factorial combination of 5 set sizes and 5 mask durations, were ordered randomly for each observer. The experiment was then replicated, thus yielding a total of 20 estimates per cell per condition. Each observer thus contributed a total of 8,000 trials, being the product of 25 conditions, 16 matrix locations, and 20 trials per location.

### Luminance Calibration

Stimulus luminance denotes the photometric measure of the amount of visible light coming to the eye from a source that is extended in space and continuous in time, such as diffuse sunlight. It has been noted that conventional luminous calibration in units of  $\text{cd/m}^2$  is not suitable for oscilloscopic displays because the source is discontinuous in both space and time (Sperling, 1971). Instead, the appropriate luminous metric for oscilloscopic displays is the *luminous directional energy per point* (LDE), expressed in units of millicandle microsecond ( $\text{mcd}\cdot\mu\text{s}$ ). The luminance of all displays in the present work is expressed in units of  $\text{mcd}\cdot\mu\text{s}$ , obtained following Sperling (1971). In that procedure, a photometer is used to measure the luminance of a test patch plotted on the display surface. That luminance reading, with parameters such as the refresh rate of the display, are then used to calculate the LDE. In the present calibration, we used a square test patch whose side was  $1^\circ$  of visual angle when viewed at a distance of 57 cm. The square patch contained 961 dots, which were evenly distributed and refreshed at a rate of 1 kHz.

For experiments using an oscilloscope, we report both the luminance values of the test patch (in  $\text{cd/m}^2$ ) and the LDEs of the stimuli that were displayed. It must be stressed that the luminance values reported for the test patch are substantial overestimates of the luminance of the actual stimuli. This is because the test patch contained a relatively large number of dots; therefore, the light flux was much greater than that of the corresponding experimental stimuli, which had many fewer dots spread over a much larger area.

In addition, the visibility of the stimuli was influenced by the time-intensity relationship known as *Bloch's law*, which acted to reduce the brightness (i.e., the perceived intensity) of the stimuli at short exposure durations. This is because, up to a critical duration, the perceived brightness of a stimulus is determined jointly by its luminance and its duration. Thus, if all stimuli were displayed at a fixed luminance, time-intensity reciprocity would cause the mask to appear progressively brighter as its duration was increased. This would confound the duration of the mask with

its brightness. In earlier investigations, this confound has been resolved by displaying briefer stimuli at higher luminance values, so that all stimuli appear to be of equal brightness regardless of duration (Bowen, Pola, & Matin, 1974; Servière, Miceli, & Galifret, 1977). This procedure yields equal-energy stimuli up to the critical duration. Accordingly, we varied the luminance of the mask so that it appeared to be of the same brightness at all exposure durations. The process of brightness equalization was carried out in a separate psychophysical procedure, as described by Servière et al. (1977) and by Di Lollo and Finley (1986).

With the exceptions noted below, the luminance of all stimuli was that which yielded a test patch reading of  $239 \text{ cd/m}^2$ . This corresponded to an LDE of  $60.07 \text{ mcd-}\mu\text{s}$ , which made the stimuli comfortably visible at the exposure duration of 10 ms. The equalization procedure carried out for the different mask durations in Experiment 1 yielded test patch luminances of 239, 51, 36, 29, and  $28 \text{ cd/m}^2$ , for additional mask durations of 0, 40, 80, 160, and 320 ms, respectively. These corresponded to LDEs of 65.07, 13.90, 10.08, 7.90, and  $7.63 \text{ mcd-}\mu\text{s}$ , respectively. This range of LDE values overlaps with that used by Bischof and Di Lollo (1995) under conditions of light-adapted viewing.

### Results and Discussion

The results of Experiment 1, averaged over all 16 target locations, are shown in Figure 3, separately for each observer. These results agree with, and supplement, earlier masking results with common onset of target and mask (Bischof & Di Lollo, 1995; Cohene & Bechtold, 1974, 1975; Di Lollo et al., 1993; Markoff & Sturr, 1971). A brief comment is required on the shape of the curves in Figure 3. In classical metacontrast, target accuracy first declines as the interstimulus interval (ISI) between target and mask is increased from zero, and then it returns to normal at longer ISIs. The recovery occurs because as the ISI is increased, there is a longer mask-free period in which to process the target. The descending portions of the curves in Figure 3 match the descending

portions in conventional metacontrast-masking functions. However, no recovery is seen in Figure 3 because no mask-free period exists in the common-onset paradigm. Indeed, increasing the duration of the mask after target offset increases the probability that masking will occur, thus keeping accuracy at a low level.

To obtain estimates of masking strength at different retinal eccentricities, we divided the 16 cells of the notional  $4 \times 4$  display matrix into three parts. The inner part comprised the 4 cells surrounding the fixation cross; the eccentricity at the center of any inner cell was approximately  $0.7^\circ$ . The outer part comprised all peripheral cells except the corner cells (8 cells in all); the eccentricity was approximately  $1.6^\circ$ . The corner part consisted of the four corner cells whose eccentricity was approximately  $2.8^\circ$ . When analyzed with eccentricity as a factor, the results show that the average amount of masking increased with eccentricity, a result found also in classical metacontrast (Alpern, 1953; Kolers & Rosner, 1960). Within each eccentricity, however, the strength of masking increased directly with the same two factors: set size and duration of the trailing mask. In other words, the pattern of results illustrated in Figure 3 is representative of the pattern obtained at each of the three eccentricities. Thus, eccentricity did play a part, but factors in addition to eccentricity intervened to produce the pattern of results obtained in the present experiment.

### Synergy of Mask Duration and Set Size

It is clear from Figure 3 that masking developed as a joint function of set size and mask duration. Considered in isolation, neither factor can account for the entire pattern of results. The simple effect of mask duration is seen in the modest decrement in accuracy when the target was the only item in the display (Figure 3, set size = 1). But as set size increased, the effect of the mask

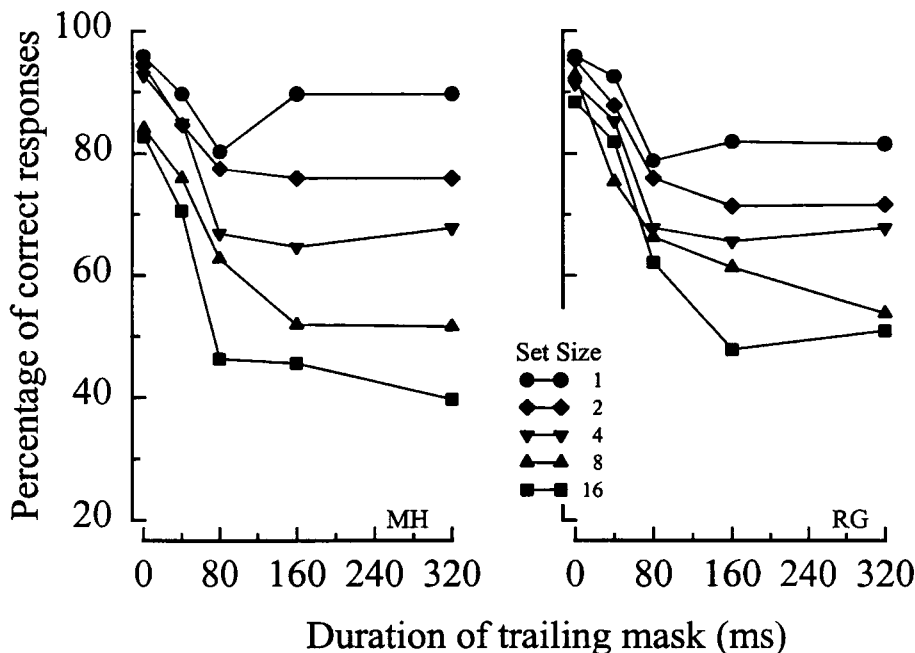


Figure 3. Results of Experiment 1: Light-adapted viewing with a ring mask. Mean percentage of correct identifications of the location of the gap in the target, separately for the 2 observers (MH and RG).



increased dramatically at durations beyond 40 to 80 ms, revealing the interdependence of the two factors. Similarly, the simple effect of set size is represented in Figure 3 by the vertical spread of the five points at a mask duration of zero, showing a modest decline in target identification as set size increased. But the effect of set size became much stronger at mask durations of 40 to 80 ms, again pointing to the synergy of the two factors.

To understand this interaction, we must consider how set-size effects may be mediated. Before proceeding with theoretical accounts, however, it is prudent to consider whether the interaction may have arisen spuriously through a ceiling effect. It may be suggested that the effect of set size was actually the same at all mask durations but that it was not revealed fully at a mask duration of zero because accuracy was compressed against the 100% limit imposed by the response scale.

This ceiling-effect account is not supported by the results in Figure 3. For both observers, set-size effects remained approximately constant at the shorter mask durations even though accuracy was below ceiling. This is especially evident in the results of Observer RG: At mask durations of 40 to 80 ms, accuracy was clearly unconstrained by a ceiling, yet the effect of set size remained modest. Thus, a ceiling effect could not have been fully responsible for keeping the effect of set size artificially small at the shorter mask durations. In the same vein, the nonmonotonicity seen for set size of 1 in Figure 3 is inconsistent with a ceiling effect.

Among the theoretically interesting options to account for the interaction of set size and mask duration is the time that is needed for spatial attention to focus on the target location. Studies of visual search show that this time increases with increases in set size (Duncan & Humphreys, 1989; Treisman & Gelade, 1980). Thus, if there is only 1 item in the display, attention can be deployed to it very rapidly; with larger set sizes, it will take longer for attention to focus on the target item. In this account, the decrement in accuracy shown in Figure 3 occurs because the memorial representation of the display begins to fade before spatial attention can be focused on the target location before the disappearance of the main display.

This account has plausible aspects to it, and these form an essential part of the reentrant-processing account that we favor. But memorial decay alone cannot account for the full pattern of results in Figure 3. Indeed, the only portion of Figure 3 that can be explained by the adverse effects of memorial decay is the vertical spread of the five data points at a mask duration of zero. This shows that target accuracy was very high for the small display sizes but that targets were occasionally missed as display size was increased. The relatively small magnitude of this effect indicates that the memorial representation that remained available after the termination of the display was quite adequate for target identification even at the largest set size. Thus, memorial decay from display onset to target identification was not in itself the most important factor.

Another way in which increments in set size can reduce accuracy is through increased crowding of the display items. As the number of distractors is increased, the target becomes progressively more crowded, and accuracy suffers accordingly (Bouma, 1970; He, Cavanagh, & Intriligator, 1996; Toet & Levi, 1992). A major reason for these crowding effects is that the resolution of attention is not sufficiently narrow to permit processing of the

target to the exclusion of neighboring items (Eriksen, 1995; Eriksen & St. James, 1986; LaBerge & Brown, 1989; Yeh & Eriksen, 1984). Thus, initial processing is not confined solely to the target but also includes features from nearby items. The extraneous features must be excluded from further processing if the target is to be identified correctly. This can be done rapidly if the target stands out distinctly, but the process takes longer if the target's features are easily confused with those of nearby items.

It is likely that crowding contributed to the modest decrement in accuracy seen with increasing set size in Figure 3 when the mask duration was zero. But the massive reductions in accuracy that developed as mask duration was increased, on an otherwise blank screen, cannot be attributed to crowding. The important point is the same as that for the passage of time hypothesis. For any given curve in Figure 3, set size—and hence crowding—was exactly the same at each duration of the trailing mask. Therefore, to produce the results in Figure 3, crowding must have acted in conjunction with mask duration. We believe this cooperative relationship can be understood on the basis of iterative reentrant processing.

### *An Account Based on Iterative Reentrant Processing*

For clarity of exposition, we outline the course of our proposed iterative reentrant processes at two durations of the mask: 0 and 320 ms.

*Mask duration equal to 0.* In this condition, the onset of the display triggered a period of activity at a low level in the visual system, leading to the formation of a tentative representation at a higher level. Such a representation required verification for several reasons. For example, the ascending signals may activate more than one representation at the higher level, or the initial high-level representation may be ambiguous or ill defined. This ambiguity can be resolved by comparing the high-level codes with the ongoing low-level activity generated by the initial stimulus. Also, the larger receptive fields at the higher levels may not permit highly specific spatial information to be preserved. This might produce errors such as illusory conjunctions in which visual features are assigned incorrectly to objects or locations. This also can be resolved through reentrant comparisons.

Broadly conceived, the iterative reentrant activity can be regarded as a binding process in which specific visual features are bound with the appropriate objects in space. When set size is manipulated, the precise number of iterations required to identify the target will vary. If the target is the only element in the display, the process of verification can be accomplished quickly in few iterations. If the target is crowded among distractors, however, more iterations are required to ensure that the correct features are bound with the corresponding objects. However, as noted above, set size and crowding did not constrain accuracy to any great extent when all elements of the display ended together. This is shown by the modest spread of the five data points in Figure 3 at a mask duration of zero.

*Mask duration equal to 320 ms.* The enduring mask in this condition introduces a powerful new source of interference in the target identification process. The interference arises from the continued presence of the mask in the space previously occupied by the target-mask combination. When the target is turned off before processing is complete, an ambiguity is created as to the identity of the object in that location. The ambiguity is brought about by the



ongoing low-level activity, which now consists both of an image of the target, which is decaying rapidly in time, and an image of the mask, whose strength is undiminished because it is supported by the current sensory input.

Given this pattern of activity, successful performance will depend on the number of iterations required for target identification. If set size is small and few iterations are required, the task can be accomplished on the basis of the relatively undecayed memorial image of the target. This is shown in Figure 3 by the relatively high level of accuracy with a set size of 1 at a mask duration of 320 ms. If set size is large, however, more iterations are required for correct identification. As the activity initially triggered by the target continues to decay over successive iterations, the correspondence between the initial reentrant code (target + mask) and the ongoing low-level activity diminishes. In contrast, the correspondence between the reentrant code for mask alone and the strong low-level activity maintained by the continued presence of the mask grows. During this sequence of events, the probability of mask alone replacing target + mask as the successful high-level code increases with the exposure duration of the trailing mask. Thus, at long mask durations, only the mask is eventually perceived, and the area within it appears empty, a perceptual outcome concordant with the phenomenological appearance of the displays. This outcome is represented in Figure 3 by the low accuracy with a set size of 16 at a mask duration of 320 ms. These sequential contingencies have been formalized in a computational model presented in Part III.

### Experiment 2:

#### Two Components of Common-Onset Masking

We have argued that the masking effects obtained in Experiment 1 can be attributed largely to a high-level process of object substitution. This is not to deny, however, that low-level mechanisms also contributed to the decline in target accuracy. Indeed, the measurable amount of masking obtained with single-element displays (Figure 3, set size = 1) suggests that contour interactions are leading to some masking even when attention can be focused on the target rapidly and without crowding. It is therefore worth entertaining the hypothesis that the masking effects illustrated in Figure 3 may consist of two components: a low-level component that operates in much the same way for displays of all set sizes and involves such early visual processes as contour formation and a high-level component that is differentially sensitive to the number of potential target items in the display and involves more complex aspects of perception, including object identification.

One way in which the two components may be decoupled is suggested by the recent finding that metacontrast masking, which is fully evident in light-adapted viewing, is all but absent in dark-adapted viewing (Bischof & Di Lollo, 1995). Especially important for the present purpose is the additional finding that no masking was obtained in dark-adapted viewing even if the target and the mask had a common onset, as in the present work. A theoretical account of those findings was premised on the absence of low-level inhibitory contour interactions in scotopic vision (Barlow, Fitzhugh, & Kuffler, 1957; von Békésy, 1968).

If it can be assumed that the effects of adapting luminance are confined largely to processes in low-level vision, as suggested by the findings of Bischof and Di Lollo (1995), then it should be

possible to decouple the low-level from the high-level masking components by conducting the experiment under conditions of dark adaptation. If dark adaptation abolishes masking for a single item while leaving masking in larger display sizes unaffected, results obtained in dark-adapted viewing should reflect only the effects of object substitution.

### Method

Observers and procedures were the same as in Experiment 1, except that the experiment was carried out in total darkness. The screen's front illumination was turned off, and the observers sat in the dark room for 30 min before beginning the experiment. The luminance of all stimuli was set using a test patch reading of 3.0 cd/m<sup>2</sup>. This corresponded to an LDE of 0.82 mcd- $\mu$ s, which made the stimuli barely visible under fully dark-adapted conditions at the exposure duration of 10 ms. If the intensity of the displays was reduced to that corresponding to a test patch of 2.0 cd/m<sup>2</sup>, the stimuli became indistinguishable, and the task was no longer feasible. The equalization procedure carried out for the different mask durations in Experiment 2 yielded test patch luminances of 3.0, 0.6, 0.4, 0.3, and 0.2 cd/m<sup>2</sup>, for additional mask durations of 0, 40, 80, 160, and 320 ms, respectively. These corresponded to LDEs of 0.82, 0.16, 0.11, 0.08, and 0.05 mcd- $\mu$ s, respectively. This range of LDE values overlaps with that used by Bischof and Di Lollo (1995) under conditions of dark-adapted viewing. Under these low-luminance conditions, any light adaptation that might have occurred during a trial was not photochemical but neural (e.g., Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). In this respect, note that recovery from neural light adaptation is very fast: Baker (1963) estimated a recovery of two log units within 250 ms.

### Results and Discussion

The results of Experiment 2, averaged over all 16 target locations, are shown in Figure 4, separately for each observer. Comparison of Figures 3 and 4 reveals three notable aspects of the results. First, adapting luminance had little, if any, effect on the incidence of masking with the larger display sizes. The strength of object substitution, as indexed by the interaction of set size and mask duration, remained very much in evidence in both light-adapted and dark-adapted conditions. Because adapting luminance is known to have a powerful influence on low-level visual processes (Barlow et al., 1957; von Békésy, 1968; Ikeda, 1965; Sperling & Sondhi, 1968) but not on higher level processes (Adelson & Jonides, 1980; Coltheart, 1980; Scharf & Lefton, 1970), this supports the hypothesis that object substitution is not a phenomenon of low-level vision.

Separability of low-level from high-level masking components is highlighted in Figure 5, where the light-adapted and the dark-adapted curves, averaged over the two observers, are superimposed on each other. At mask durations beyond about 40 ms, each dark-adapted curve is consistently higher than the corresponding light-adapted curve ( $p < .001$ , by a simple sign test for related samples). This difference is, if anything, an underestimate of the true differences because the luminance of the displays had to be set above absolute threshold to make the task feasible. This might have introduced a small inhibitory component, which, if present, acted to reduce the differences between the two sets of curves. At any rate, the results support the hypothesis that low-level and high-level masking combined in their effect on light-adapted accuracy but that only (or mostly) high-level masking was active in dark-adapted viewing.

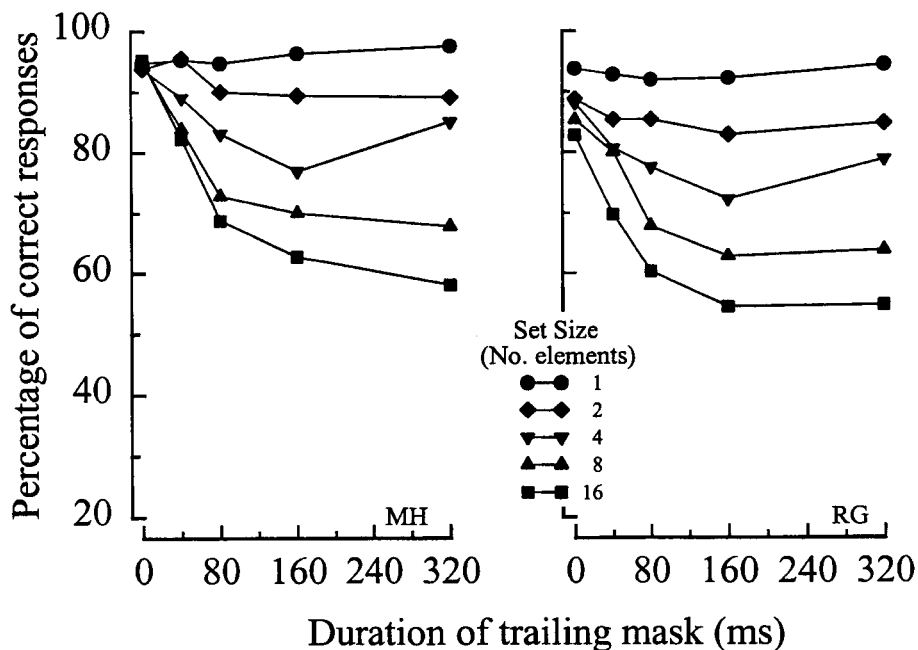


Figure 4. Results of Experiment 2: Dark-adapted viewing with a ring mask. Mean percentage of correct identifications of the location of the gap in the target, separately for the 2 observers (MH and RG).

The light- and dark-adapted curves also differ in revealing ways at the shorter mask durations (0 and 40 ms for MH; 0, 40, and 80 ms for RG). The light-adapted curves in Figure 3 exhibit an early drop, with set size having only a modest and relatively constant effect on accuracy. In contrast, the dark-adapted curves in Figure 4 do not exhibit the same early drop but show a more gradual

development of masking, with the effect of set size growing progressively over mask durations. This pattern of results suggests that light-adapted masking interferes with processes that occur early in vision and are relatively unaffected by set size. In contrast, the dark-adapted results are suggestive of masking that occurs later in the chain of events and involve higher level processes, such as object identification.

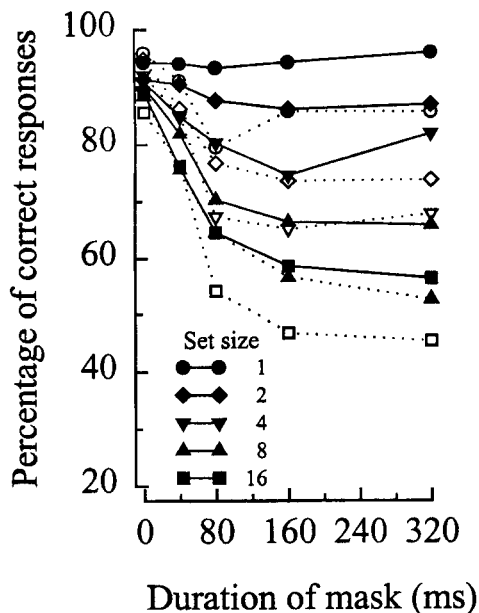


Figure 5. Comparison of means results of Experiments 1 (light adapted; open symbols) and 2 (dark adapted; filled symbols), averaged over the 2 observers.

Collectively, Experiments 1 and 2 provide evidence for at least two masking processes: low-level contour interactions and higher level object substitution. Of these, object substitution is of greatest interest in the present work and is examined more closely in the next four experiments. Before relinquishing the topic of low-level masking, however, a comment is in order regarding the mechanisms that may subserve the low-level masking effects revealed in the present work. These masking effects were obtained with the common-onset paradigm. We have argued that this rules out any account based on feed-forward inhibitory mechanisms that are time locked to the onsets of the stimuli. This is not to deny, however, that masking could be based on other inhibitory mechanisms activated by reentrant pathways in early visions, which are disabled in dark-adapted viewing (Barlow et al., 1957; von Békésy, 1968). Indeed, inhibitory interactions in reentrant cortico-geniculate pathways such as described by Sillito (1992) are capable of mediating the low-level masking effects revealed in the present work.

### Experiment 3: Masking by Object Substitution

In Experiments 1 and 2, low-level masking was separated from object substitution on the rationale that low-level inhibitory mechanisms are disabled in dark-adapted viewing. This leaves open the possibility that object substitution may be an effect obtained only

in dark-adapted viewing. A second issue is whether interference with early processes of contour formation can confidently be ruled out as determinants of the masking obtained in Experiment 2. Namely, it is still possible that some form of interference, involving not inhibition, but contour interactions, may play a role in object-substitution masking, as it does in classical metacontrast masking (Breitmeyer, Rudd, & Dunn, 1981; Burr, 1984; Growney, Weisstein, & Cox, 1977).

Both these issues were addressed in Experiment 3, which was performed in light-adapted viewing under conditions that all but ruled out contour interactions. This was done by reducing the prominence of the masking contours in two ways. First, the amount of contour in the mask was reduced drastically. The annulus used in Experiments 1 and 2 was reduced to just four dots, an amount of contour that was demonstrably insufficient for classical metacontrast masking. Second, the distance between the four dots and the contour of the target was varied systematically from a small group inside the target to four widely spread points at the corners of an imaginary square surrounding the target. Spatial proximity of target and mask contours is a variable to which classical metacontrast masking is exquisitely sensitive.

### Method

Observers and procedures in Experiment 3 were the same as in Experiment 1, except for the following two changes. First, the masking ring was replaced by four dots located at the four corners of an imaginary square concentric with the target ring. Each dot was a small square patch with sides equal to 2 min arc. Second, the spatial separation between the four dots and the nearest part of the target ring was varied systematically at five levels. The separation could be either 40, 30, 20, or 10 min arc or the four dots could be clustered in a group at the center of the target ring.

The design of Experiment 3 was a 5 (set sizes)  $\times$  5 (mask durations)  $\times$  5 (spatial separations)  $\times$  16 (matrix locations) factorial, with 20 replications in each cell. This yielded a total of 40,000 trials, 160 per session. Each of the resulting 250 sessions took approximately 15 min to complete. Each observer participated in 8 sessions per day, grouped in two 1-hr periods, separated by a rest interval of at least 1 hr. In any given session, the number of distractors was fixed, as was the mask duration and the spatial separation between target and mask. Within each session, the target was displayed 10 times in each of the 16 possible locations, in a sequence that varied randomly across sessions and observers.

### Results and Discussion

Figure 6 shows the results averaged over all target locations and target-mask separations, separately for each observer. It is clear from Figure 6 that the interaction of set size and mask duration was as much in evidence in Experiment 3 as it was in the previous two experiments. Bearing in mind that Experiment 3 had been designed to minimize the effects of masking contours, we can be confident that the cooperative relationship between set size and mask duration obtained in all three experiments was due to factors that went beyond contour interference.

Degree of separation between the target and the four-dot mask had a negligible effect on the strength of masking. This was most evident at the two longer mask durations, in which accuracy was well below ceiling. Figure 7 shows the mean results at each target-mask separation for each set size, averaged over mask durations of 160 and 320 ms, all observers, and target locations. If contour interactions were the basis for masking, increasing spatial separation would have systematically improved target accuracy, as has often been shown for metacontrast masking (Breitmeyer et al., 1981; Growney et al., 1977; Levine, Didner, & Tobenkin, 1967).

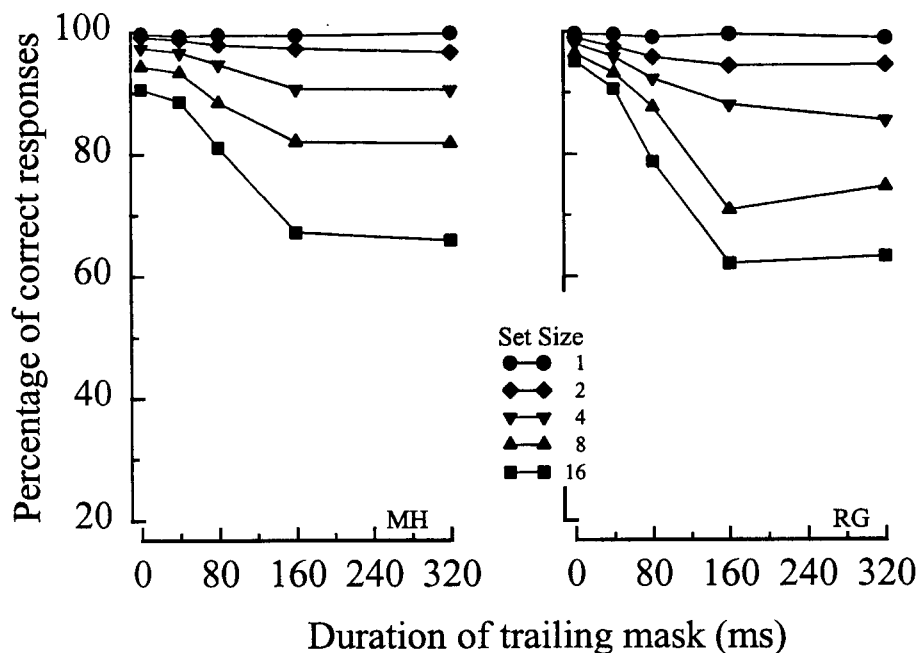


Figure 6. Results of Experiment 3: Light-adapted viewing with a four-dot mask. Mean percentage of correct identifications of the location of the gap in the target, averaged over all target-mask separations, separately for the 2 observers (MH and RG).

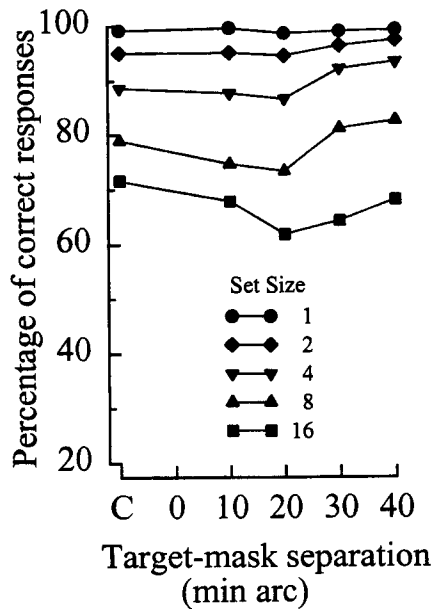


Figure 7. Mean percentage of correct responses at each target-mask separation, averaged over mask durations of 160 and 320 ms, observers, and target locations. The scores for mask durations shorter than 160 ms were too close to 100% to be useful in this figure. On the abscissa, C indicates the condition in which the four dots were clustered in a group at the center of the target ring.

Instead, the weak trends revealed in Figure 7 for some set sizes indicated first a small decrease in accuracy with increasing separation and then a small increase in accuracy. We conclude that contour interactions did not play a prominent part in the masking by four dots.

The similarity of the results in Figures 4 and 6 confirms that masking by object substitution can occur under both light- and dark-adapted viewing conditions. As was the case in Experiments 1 and 2, strength of masking in Experiment 3 developed as an interaction between set size and duration of the trailing mask. This is consistent with the hypothesis that masking occurred in the course of iterative reentrant processing, through a process of object substitution. That is, although the four-dot mask was insufficient as a source of contour interference, it was entirely adequate for defining a trailing configuration (a square surface) that replaced the target as the object of perception.

It is notable that the phenomenal appearance of the displays was consistent with this hypothesis. When all elements terminated simultaneously (duration of trailing mask = 0), the target was clearly visible with the four dots, and the orientation of the gap could be identified with ease. At the longer durations of the trailing mask, however, the four dots appeared to be clearly visible, but the target location appeared empty. The items surrounding the target did not seem to be so affected. Observers often indicated that they could have reported the gap orientation in any one of the surrounding items when only four dots could be seen in the target location.

The masking illustrated in Figure 6 supplements earlier results (Enns & Di Lollo, 1997) in revealing a remarkable novel finding. Four dots displayed simultaneously with the target can act as a powerful mask, provided that two conditions are met: first, that the

dots continue to be displayed for some time after the target has been turned off and, second, that attention be divided among several items in the display. Considered collectively, the results thus far clearly suggest the action of two masking components: a low-level component, active only in light-adapted viewing, and dependent on inhibitory contour interactions. The other, object substitution, occurring in both light- and dark-adapted viewing, independent of contour interactions, and influenced by attentional manipulations, such as variations in set size.

We noted earlier that common-onset masking cannot be explained by inhibitory feed-forward theories because they are transient dependent and bound to the SOA law. These theories are further questioned by the outcome of Experiment 3. In addition, the present finding that powerful masking is obtained with a mask consisting of only four dots cannot be explained even by those feed-forward theories that are not transient dependent, but rely on lateral inhibition as a basis for masking (e.g., Bridgeman, 1978; see also Bischof & Di Lollo, 1995). Such a mask does not have enough contours, and its location is too remote, to support the lateral inhibitory interactions that are required for such theories.

In all the experiments reported thus far, object substitution was found to depend critically on set size, a widely used attentional variable. In the remaining experiments, we asked whether this form of masking is also affected by other attentional manipulations, such as the degree of target-distractor similarity and the duration of a spatial precue. In addition, we asked whether this form of masking is as readily obtained with naive as with trained observers.

#### Experiment 4: Generality of Object Substitution

The next three experiments were designed to explore the generality of masking by object substitution. As a first step, we asked whether the strength of masking was affected by the degree of difficulty of the target-identification task. For this purpose, two changes were made. The first was aimed at exploring the type of visual feature that can be masked in object substitution. In Experiments 1-3, the distinguishing visual feature was constructed by deleting a segment from the target's contour. This is known to make the search task relatively difficult (Treisman & Gormican, 1988). In the present experiment, the critical feature was constructed by adding an orthogonal line to the target's contour. This is known to make the search task easier (Treisman & Souther, 1985). We also simplified the observer's response requirements by using a detection task rather than a discrimination task. In so doing, the response alternatives were reduced from four (gap orientations) to two (presence or absence of the target feature).

Along with these substantive changes, we made two procedural changes. First, instead of practiced psychophysical observers, we recruited naive observers from the undergraduate population. Second, the stimuli were presented on a standard raster-scan computer display. The aim was to check on the generality of object-substitution masking across participant populations and display conditions.

#### Method

Nine undergraduate students from the University of British Columbia participated for course credit. All had normal or corrected-to-normal vi-

sion, and none took part in any of the other experiments reported here. The experiment was carried out in a testing room with standard overhead fluorescent lighting.

All stimuli were black-on-white images displayed on an Applevision 1710 monitor driven by a PowerMacintosh and viewed from 57 cm. On any given trial, either 1, 8, or 16 complete rings, whose dimensions were similar to those in Experiment 1, were displayed randomly in the cells of a notional  $4^\circ \times 4^\circ$  matrix in the center of the screen. In displays containing more than one ring, half the rings had a vertical bar across the bottom, as shown in Figure 8. The target was always singled out by four dots, which also served as a mask, as in Experiment 3. The four dots were located approximately 20 min arc from the target's contour. On half the trials, the target had a vertical bar, and on the remaining trials, it did not. The observer's task was to say whether the target had the vertical bar.

At the beginning of each trial, a fixation marker was displayed in the center of the screen. The display sequence, similar to that in Experiment 3, consisted of two frames presented sequentially without interruption. The first frame contained the target, the four-dot mask, and the distractors, if any. The exposure duration was 45 ms. The second frame contained only the four-dot mask. Its duration was either 0, 45, 90, 135, or 180 ms. Observers were given 4 s to indicate target presence or absence by pressing one of two keys. Their responses were acknowledged by a central feedback symbol (+ for correct, - for incorrect), which was displayed for 1 s and also served as the fixation marker for the next trial. Each observer contributed 24 trials in each of the 15 conditions resulting from the factorial combination of three set sizes (1, 8, or 16 display elements) and five durations of the trailing mask. The 360 trials were completed within a single 1-hr session.

### Results and Discussion

Illustrated in Figure 9 is the mean target accuracy for those trials on which the target contained the vertical segment. Note that the lower limit of accuracy in Figure 9 is more properly regarded as being zero rather than the 50% chance level. This is because the observers indicated whether they had seen the vertical segment in the target. Thus, a score below 50% would indicate that the vertical segment, although present, was not seen because it had been masked. On trials in which the target did not contain the vertical

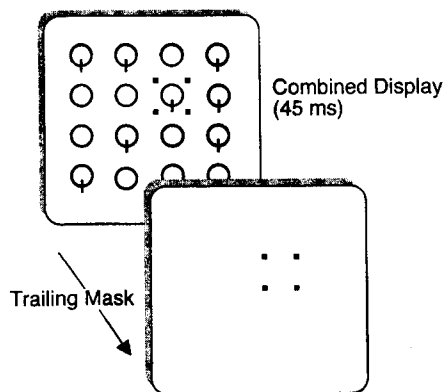


Figure 8. Schematic diagram of the stimulus sequence in Experiment 4. The display contained up to 16 rings, half of which had a vertical bar across the bottom. The target was singled out by four dots, as shown, which also served as the mask. Observers indicated whether the target contained the vertical bar. The sequence began with a combined display of the target, mask, and distractors for 45 ms and continued with a display of the mask alone for durations of 0, 45, 90, 135, or 180 ms.

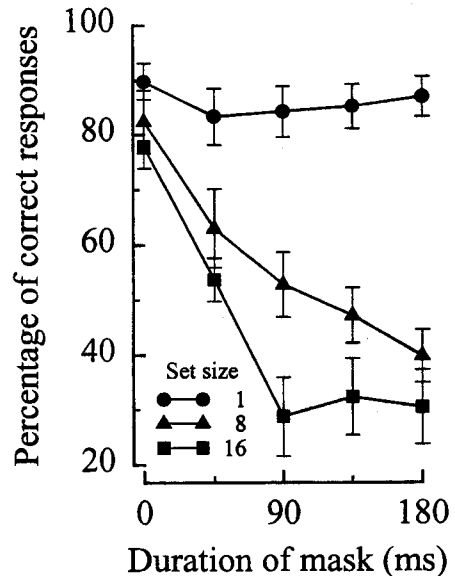


Figure 9. Results of Experiment 4: Mean percentage of correct target identifications when the target contained a vertical segment, averaged over all observers. Error bars indicate standard error of the mean.

segment (not shown in Figure 9), accuracy was at ceiling except at a mask duration of zero, where the results were comparable to those obtained when the vertical segment was present (Figure 9, mask duration = 0). Ceiling effects for accuracy on target-absent trials are commonly found in visual search experiments because observers are reluctant to guess that a feature they did not see was actually present (Enns & Rensink, 1991; Treisman & Gormican, 1988).

The results in Figure 9 were analyzed in a 3 (set sizes)  $\times$  5 (mask durations) within-subjects analysis of variance (ANOVA). All effects were significant: Set size,  $F(2, 16) = 41.30, p < .0001, MSE = 0.049$ ; mask duration,  $F(4, 32) = 21.76, p < .0001, MSE = 0.021$ ; Set Size  $\times$  Mask Duration,  $F(8, 64) = 7.18, p < .0001, MSE = 0.014$ .

There is a striking similarity in overall pattern between the results in Figures 6 and 9. This is not surprising because both sets of data were obtained with a four-dot mask in light-adapted viewing. In both figures, the result of principal interest was the interaction between set size and mask duration. In the present experiment, as in Experiment 3, no masking occurred when the target was the only item in the display, no matter how long the four-dot mask remained on the screen after the target had been turned off. In contrast, when additional items were displayed along with the target (Figure 9, set sizes of 8 and 16), masking increased progressively with mask duration. The phenomenological appearance of the display was similar to that described in Experiment 3. At the longer mask durations, the space inside the four dots was seen as empty.

Object substitution was very much in evidence in the present experiment even though the critical target feature was far easier to see than was the missing contour segment in the previous three experiments. Furthermore, the results obtained with a detection task were essentially the same as those obtained with the more complex four-alternative discrimination task in the other experi-

ments. These commonalities attest to the generality of masking by object substitution across visual features, task characteristics, and the observer's sophistication. In the present experiment, as in the previous three experiments, masking could not be attributed to local inhibitory interactions for several reasons. Perhaps the most telling was that masking depended critically on the presence of additional items with minimal contour, whose spatial separation from the target was beyond what could reasonably be expected to mediate inhibitory contour interactions.

The results reported thus far strongly implicate attentional processes. The two experiments reported next were designed to corroborate this link. This was done by exploring how the strength of masking is influenced by two well-known manipulations of spatial attention: *target-distractor similarity* (pop out) and *spatial precuing*.

### Experiment 5: Object Substitution Is Reduced by Target Pop Out

It is known that a target that stands out from a field of distractors can be identified quickly and accurately regardless of the number of distractors. This effect, known as *pop out*, is related to target-distractor discriminability. Search times become progressively faster as the target is made more discriminable from the distractors (Duncan & Humphreys, 1989).

From the perspective of iterative reentrant processing, less masking would be expected when the target pops out. This prediction is based on the assumption that fewer iterations are required to identify a target that is easily discriminable from the distractors (i.e., when pop out occurs). Features that belong to the distractors should be easier to reject, thus facilitating the binding of the appropriate features with the target object. In the common-onset paradigm, this means that fewer iterations need to be performed after the target has been turned off while only the mask is on display. Because the strength of masking is held to be directly related to the number of iterations performed while the mask alone is present on the screen, reducing the number of those iterations will result in a corresponding reduction in masking.

In the present experiment, pop-out conditions were achieved by increasing the structural difference between the target and distractors. The displays were as in Experiment 4, except that the critical target feature—the vertical line segment—appeared only in the target itself, never in the distractors. This increased the structural differences between the target and the distractors to such a large degree that pop out was obtained, as in standard visual search experiment (Treisman & Gormican, 1988).

#### Method

Stimuli and procedures in Experiment 5 were the same as in Experiment 4, except that only the target could have a vertical bar. The distractors were all complete circles, as shown in Figure 10. Ten experimentally naive observers participated in Experiment 5. They indicated by a keypress whether the circle inside the four dots contained a vertical bar.

#### Results and Discussion

Mean target accuracy for those trials on which the target contained the vertical segment is illustrated in Figure 11. As was the case in Experiment 4, accuracy on segment-absent trials (not

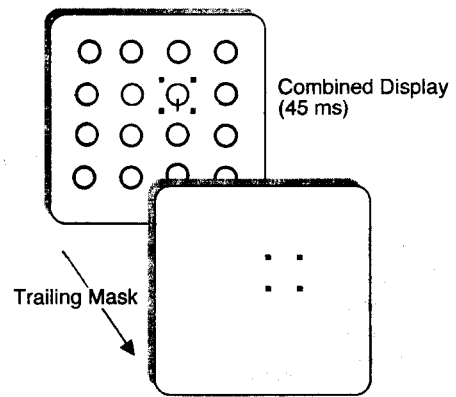


Figure 10. Schematic diagram of the stimulus sequence in Experiment 5. The display contained up to 16 rings. On half the trials, one ring contained a vertical bar across the bottom. On any given trial, the observer indicated whether one of the rings in the display contained a vertical bar. On every trial, the target ring was singled out redundantly by four dots, as shown, regardless of whether the target contained a vertical bar. The sequence began with a combined display of the target, mask, and distractors for 45 ms and continued with a display of the mask alone for durations of 0, 45, 90, 135, or 180 ms.

shown in Figure 11) was near ceiling throughout the domain. The results in Figure 11 were analysed in a 3 (set sizes)  $\times$  5 (mask durations) within-subjects ANOVA. All effects were significant: Set size,  $F(2, 18) = 13.91, p < .001, MSE = 0.024$ ; mask duration,  $F(4, 36) = 23.55, p < .0001, MSE = 0.007$ ; Set Size  $\times$  Mask Duration,  $F(8, 72) = 2.69, p < .01, MSE = 0.010$ .

Direct comparison of Figures 9 and 11 indicates that masking was stronger in Experiment 4, where target-distractor similarity was high, than in Experiment 5, where similarity was low. An ANOVA was carried out on the combined results of Experiments 4 and 5, with one between-subjects factor (experiment) and two within-subjects factors (set size and mask duration). The interaction effect between set size and experiment was significant,  $F(2, 2) = 5.64, p < .01, MSE = 0.050$ , confirming that, for corresponding set sizes, masking was weaker under the pop-out conditions of Experiment 5 than in Experiment 4.

Weaker masking with pop-out targets is in accordance with our predictions based on iterative reentrant processing. It was therefore of interest that we could find no experiments in the masking literature in which strength of masking was studied in relation to the discriminability of a target from its background. This is understandable because current theories do not provide a compelling rationale for such studies. In contrast, the present conceptual framework offers a cogent rationale and provides a consistent account of the experimental outcome.

In addition to strengthening the link with the attentional literature, Experiment 5 successfully decoupled the cuing from the masking functions served by the four dots. In Experiments 1–4, the four dots had served both as a cue and as a mask. In Experiment 5, the target was specified uniquely by the vertical line segment, rendering the four dots redundant as a cue. Thus, in the present experiment, the four dots functioned principally as a mask. The substantial amount of masking illustrated in Figure 11 shows that masking by object substitution can be obtained when the cuing

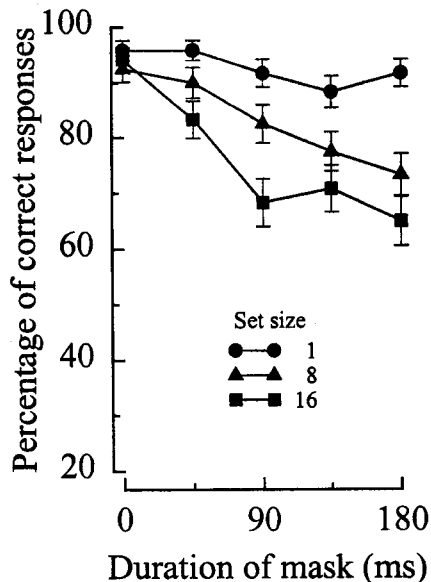


Figure 11. Results of Experiment 5: Mean percentage of correct target identifications when the target contained a vertical segment, averaged over all observers. Error bars indicate standard error of the mean.

function of the mask is redundant. More generally, it can be said that masking occurs whether or not the mask also serves as a cue.

#### Experiment 6: Object Substitution Is Reduced by Spatial Cuing

Experiment 6 was designed to strengthen the link between masking and attention by showing that masking responds in predictable ways to changes in yet another widely used attentional manipulation: duration of a spatial precue. It is known that a target among distractors can be identified more rapidly and more accurately if its location within the display is indicated ahead of time by a spatial precue (Eriksen & Hoffman, 1973; Van der Heijden & Eerland, 1973).

From the viewpoint of reentrant processing, the effect of spatial precuing is homologous to that of pop out. That is, the initial iterations required to locate the target within the display can be carried out before the target is presented, to an extent that varies with the period of precuing. This is tantamount to saying that attention can be focused at the target's location ahead of time. The upshot is that, as soon as the target is displayed, all processing activity can be directed immediately to identifying the target rather than to finding its location. This means that fewer iterations need to be done while the mask alone is on the screen, thus reducing the likelihood of the mask replacing the target as the object of perception.

#### Method

Stimuli and procedures in Experiment 6 were the same as in Experiment 4, except for the following. The display sequence, illustrated in Figure 12, consisted of three frames presented sequentially without interruption. The first frame contained the four-dot mask, which indicated the screen location where the target was about to be presented. The exposure

duration of the first frame was either 0, 45, 90, 135, or 180 ms. The second frame was a 45-ms display of the target, the four-dot mask, and the distractors, if any. The third frame was a 90-ms display of the four-dot mask. The duration of the third frame was set at 90 ms because that was the shortest duration of the trailing mask at which the full effect of set size was obtained in Experiment 4. Ten experimentally naive observers served in Experiment 6.

#### Results and Discussion

Mean target accuracy for the trials on which the target contained the vertical segment are illustrated in Figure 13. As was the case in the previous two experiments, accuracy on segment-absent trials (not shown in Figure 13) was near ceiling throughout the domain. The results in Figure 13 were analyzed in a 3 (set sizes)  $\times$  5 (durations of the precue) within-subjects ANOVA. All effects were significant. Set size,  $F(2, 18) = 125.37, p < .0001, MSE = 0.010$ ; duration of the precue,  $F(4, 36) = 17.58, p < .0001, MSE = 0.020$ ; Set Size  $\times$  Precue Duration,  $F(8, 72) = 3.53, p < .01, MSE = 0.009$ .

It is known that spatial precuing facilitates target identification (Eriksen & Hoffman, 1973; LaBerge, Brown, Carter, Bash, & Hartley, 1991). The results in Figure 13 reveal a similar facilitation in respect to masking by object substitution. Namely, the strength of masking declined progressively as the duration of the precue increased. The dual role played by the four dots in the experiments reported thus far is again worth stressing. When displayed before the target, the four dots acted as a cue, which facilitated target identification. When displayed after the target, however, the four dots acted as a mask, which reduced target accuracy. These outcomes bear out expectations based on iterative reentrant processing: As the duration of the precue was increased, fewer iterations had to be carried out while the mask alone was on the screen, and

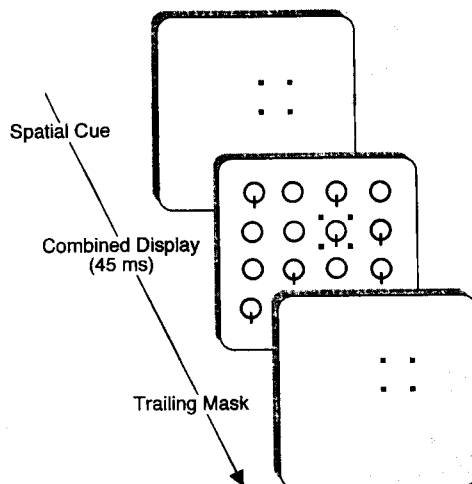


Figure 12. Schematic representation of the stimulus sequence in Experiment 6. The main display contained up to 16 rings, half of which had a vertical bar across the bottom. The target was singled out by four dots, as shown, which also served as a precue and as a mask. Observers indicated whether the target contained the vertical bar. The sequence began with a display of the four dots for durations of 0, 45, 90, 135, or 180 ms, continued with a display of the four dots, target, and distractors, and ended with a display of the four dots alone for 90 ms.



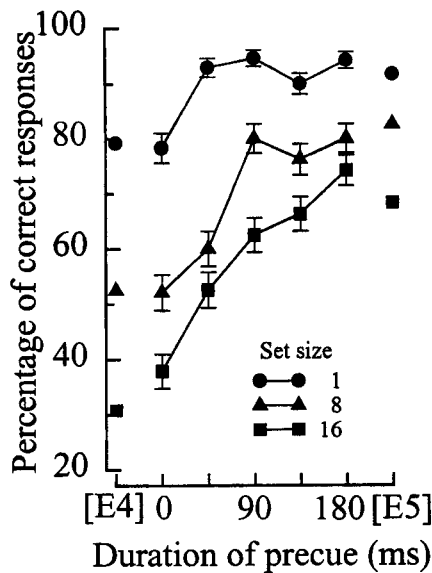


Figure 13. Results of Experiment 6: Mean percentage of correct target identifications when the target contained a vertical segment, averaged over all observers. Error bars indicate standard error of the mean. E4 = Experiment 4; E5 = Experiment 5.

masking decreased correspondingly. This is in accordance with the critical role of attention in masking by object substitution. The likelihood of the four dots replacing the target as the object to be perceived decreased as the duration of the precue was increased.

Also shown in Figure 13 are selected portions of the results of Experiments 4 and 5. The symbols on the left represent the results obtained in Experiment 4 with a mask duration of 90 ms. This condition was the same as the zero-precue condition in Experiment 6. The good correspondence between the two sets of points attests to the replicability of object-substitution masking across experiments and observers. The symbols on the right represent the results obtained in Experiment 5 with a mask duration of 90 ms. They match quite closely the results obtained in Experiment 6 with a precue of 180 ms. This correspondence underscores the equivalence of pop out and spatial precuing in their effects on object-substitution masking. This equivalence is expected on the basis of iterative reentrant processing. We have noted above how, in different ways, both manipulations can be said to reduce the number of iterations that need to be carried out while the mask alone is present on the screen. In turn, this reduces the likelihood that the mask will replace the target as the object of perception.

### PART III: A COMPUTATIONAL MODEL OF OBJECT SUBSTITUTION (CMOS)

Three important results emerged from the present study. First, in each of the six experiments, strong visual masking was obtained when an initial brief display of a target and a mask was continued with the mask alone. This is a form of masking that entails perceptual suppression of a pattern (target + mask) by those parts of the pattern (mask) that remain on view after other parts (the target) have been turned off.

A second important aspect of the experimental work was the dissociation of two masking components in Experiments 1 and 2. One component can be described as occurring relatively early in processing, active only under light-adapted viewing conditions, very sensitive to local contour interactions, and relatively insensitive to variations in set size. It is therefore similar in its behavior to classical metacontrast masking. A second component occurs later in processing, is equally effective under light- and dark-adapted viewing conditions, is insensitive to local contour interactions, and is readily influenced by the number of potential target items.

Finally, a critical link between spatial attention and object substitution was firmly established by studying masking under conditions of dark adaptation (Experiment 2) and by using masks that consisted of only four small dots (Experiments 3–6). Although the contours of these masks were insufficient to produce inhibitory contour interactions with the target, they were nonetheless very effective in suppressing the perception of the target item. However, these masks were able to do so only when spatial attention could not be easily focused on the target. Drawing spatial attention to the target by making it the only item in the display (all experiments), by providing it with distinctive visual features (Experiment 5), or by providing a spatial precue to the target location (Experiment 6) sharply reduced the effectiveness of the four-dot masks.

We regard this high-level suppression as an instance of masking by object substitution, to distinguish it from contour- and intensity-based masking processes in low-level vision (Enns & Di Lollo, 1997). Central to the object-substitution hypothesis is the view that visual representations of attended items are fundamentally different from those of unattended items. Especially important to the perception of rapid temporal sequences and visual masking is the increased spatiotemporal resolution and durability of attended items (Enns, Brehaut, & Shore, 1999; Moran & Desimone, 1985; Posner, 1980; Rensink, O'Regan, & Clark, 1997; Treisman & Gelade, 1980; Tsal, Meiran, & Lamy, 1995). The lower resolution and increased volatility of unattended items therefore leaves them more vulnerable to substitution by masking items. This holds true whether attention is reduced by distributing the stimuli over space, as in the present work and in the study of Enns and Di Lollo (1997), or over time, as in investigations of "attentional dwell time" (Duncan, Ward, & Shapiro, 1995).

Throughout the present work, we have put forward concepts of iterative reentrant processing to explain the empirical findings of object substitution. These concepts, however, have been expressed only in a qualitative way. In this section, we demonstrate that it is possible to embody these ideas within an explicit *computational model of object substitution* (CMOS). We describe a formal computational model that is similar in its architecture to many so-called closed-loop controllers that are common in several areas of industry and robotics (e.g., Carpenter & Grossberg, 1987; Edelman, 1978; Grossberg, 1995a; 1995b; Harth, Unnikrishnan, & Pandya, 1987; Mumford, 1992; Tononi, Sporns, & Edelman, 1992; Unnikrishnan & Venugopal, 1994).

In their most general form, these models describe a process in which some form of input is collected and coded before being sent on to an output device. However, the output device, in addition to sending an output signal, sends a copy of that signal back toward the input device. This involves another round of coding to facilitate

the comparison of this feedback signal with the current signals entering the input device. In this way, the output signal is influenced in an ongoing way by both current input and information already processed.

### Structural Aspects of CMOS

A basic assumption underlying CMOS is that perception emerges from the activity of a large number of modules similar to that illustrated in Figure 14, at least one for each spatial location. This module can be conceptualized in neurophysiological terms as akin to the circuit involving a cortical hypercolumn in striate cortex (area V1) and its connections to a corresponding region in one of the extrastriate visual areas. Because the present work is largely concerned with the visual attributes of shape and spatial location, we restrict our discussion to the modules of this kind that are devoted to pattern perception. We hasten to add that a complete model of visual functioning would require the description of similar modules for other visual attributes such as color, motion, and depth. We propose that these other modules would communicate with one another through common access to location-specific hypercolumns in area V1 (Treisman & Gelade, 1980; Zeki, 1993). In normal functioning, all modules would work simultaneously, each looking to achieve an optimal match between the activity at the higher and lower levels (see Desimone & Duncan, 1995).

To represent all the pattern information available in a retinal image, a large mosaic of pattern modules is required, each similar to the one shown in Figure 14. Each module is capable of generating as output a representation of the spatial pattern within its receptive field. We leave unspecified, for the present description, the extent to which these patterns are determined by fixed templates that have been learned (e.g., Carpenter & Grossberg, 1987), dynamic templates that can be modified within allowable margins (e.g., Mumford, 1992), or structural algorithms that implement more general principles of perceptual organization such as proximity, similarity, and common fate (e.g., Kubovy & Wagemans, 1995; Palmer & Rock, 1994). Regardless of which pattern-generating scheme is envisioned, the patterns are established through iterative exchanges between activity at the lower level, which has inherently finer spatial resolution but less pattern information, and that at the higher level, which has more pattern information but coarser spatial resolution.

The module shown in Figure 14 consists of three representational layers. Stimuli from the visual environment first arrive at

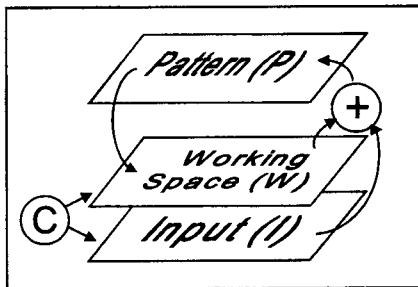


Figure 14. Schematic diagram of a processing module in the computational model of object substitution.

input layer *I*, which has many of the attributes found in striate cortex of the primate visual system. That is, the receptive fields of units within this layer are small, permitting fine spatial resolution, and the features coded are simple because the spatial region of influence is very small. At any given moment, the pattern of activation in this layer can be modeled as a pixel-like representation of the most recent stimulus. The activation triggered by a brief stimulus decays rapidly, within 100 to 200 ms, unless it is maintained by continued input from the environment. The output from this layer is summed together with the current contents of the intermediate layer, called the working space *W*, and sent to the pattern layer *P*.

The working space is regarded as still being part of striate cortex. Its spatial resolution is high, and its input consists of descending signals from the pattern layer. Although the input layer and working space do not output directly to each other, they are linked by virtue of their topographical registration, so that a point-by-point comparison can be performed between their contents. In neurophysiological terms, we imagine the input layer to perform input functions similar to those of Layer IV of neurons in Area V1 and to perform output functions similar to those of layers II and III of area V1. Similarly, the input and output characteristics of the working space would correspond to Layers I–III, respectively, of Area V1.

The uppermost pattern layer is regarded as part of extrastriate cortex that is sensitive to pattern, perhaps V3, V4, or both. It receives summed input from the input layer and working space, and it outputs back to working space. The receptive fields of units in the pattern layer are relatively large, so that spatial resolution is much lower than in the input layer. Rather than being pixel-like, the representation consists of pattern attributes.

### Functional Aspects of CMOS

At the outset of a new series of iterative cycles, corresponding to the onset of a new perceptual event, the contents of *W* are reset to zero. On each subsequent cycle, the contents of *I* are summed with the contents of *W*, and the sum is sent on to *P*. The relative importance of each of these contribution to the sum is specified by a constant,  $\lambda$ , which is used to weight the contents of *I*. Thus, because on the initial ascending loop, *W* is reset to zero, the sum of *I* and *W*, which is sent to *P* on the first iteration, consists solely of the weighted contents of *I*. Next, the contents of *P* are outputted to *W* by means of a simple overwriting operation, whereby the input from *P* replaces the current contents of *W*. In the course of the transfer, the information is translated from the pattern codes of *P* to the pixel codes of *W*. After the transfer, *W* contains the same pattern information as *P* but in a code that permits point-to-point comparison with the current contents of *I*.

To complete the first iteration, the contents of *W* are compared with the contents of *I*. This comparison serves at least two objectives. First, the initial high-level representation may be in need of confirmation because the previous ascending signal may have activated more than one initial representation or, equivalently, because the initial representation may be unclear or ambiguous. In either case, the ambiguity can be resolved by comparing the high-level codes with the initial pattern of activity at the lower level. Second, because the receptive fields in *P* are large, spatial resolution has been lost in that representation. Yet if the pattern

code is to be successfully bound to its corresponding display location, it is necessary that the reentrant pattern signals be placed in spatial registration with the active pixel signals.

These objectives are achieved by using the outcome of the comparison between  $W$  and  $I$  to guide a hill-climbing algorithm aimed at establishing the best fit between  $P$  and  $I$ . This is accomplished over successive iterations by repositioning within  $W$  the signal from  $P$ . That is, the outcome from the hill-climbing algorithm is used for repositioning the contents of  $W$  so as to optimize the correspondence with the contents of  $I$ . The overall objective is to determine whether the pattern encoded in  $P$  is a reasonably good match to the pixels currently active in  $I$ . It is worth pointing out that  $W$  does not necessarily contain a complete copy of the pattern in  $I$ . As noted above, the exact form of these representations is left unspecified within CMOS. For example, it is possible that the various features are represented in different modules similar to that in Figure 14 (e.g., Carpenter & Grossberg, 1987). Alternatively, and not incompatibly, the features may be represented by templates that can be dynamically modified in the course of reentrant processing (e.g., Mumford, 1992). In either case, these representations would allow the comparison to be based on the best matching features between  $W$  and  $I$ .

In the course of successive iterations, the contents of  $I$  change dynamically with new input from the environment. The contents of  $P$ , on the other hand, change more slowly because the input to  $P$  is a weighted sum of what is currently in  $I$  and what was in  $W$  on the previous iteration. Therefore, there is a degree of inertia in the system's response to changes in the external stimulus. The degree of inertia varies with the relative weight assigned to  $I$ , and therefore to  $W$ , by the constant in Equation 1, below. A relatively greater weight of  $W$  reflects a conservative tendency, in which additional new input in  $I$  is reflected only marginally in the contents of  $P$ , and therefore in  $W$ , over the next few iterations. This inertia permits random transient events to be discounted and, in so doing, it adds stability to the hill-climbing process. With less inertia,  $P$  would tend to change dynamically with  $I$ . The inherent noise would then interfere with the process of determining whether the representation in  $W$  was homing in on the correct patterns and locations.

### Quantifying CMOS

The tenets of CMOS are implemented in the following equations:

$$P_j(k) = W_j(k-1) + \lambda I_j(k-1) \quad (1)$$

$$W_j(k) = \frac{P_j(k-1)}{[\sum_j P_j^2(k-1)]^{1/2}}, \quad (2)$$

where  $P_j(k)$  is a vector representing signal  $j$  at iteration  $k$  at the pattern level (see Figure 14),  $W_j(k)$  represents signal  $j$  at the working space level, and  $I_j(k)$  represents signal  $j$  at the input level and is a weighting constant that specifies the relative contributions of  $I$  and  $W$  to the sum in  $P$ . Three distinct signals are used in implementing of CMOS: the target signal ( $j = 1$ ), the mask signal ( $j = 2$ ), and the noise inherent in the system ( $j = 3$ ). The term  $[\sum_j P_j^2(k-1)]^{1/2}$  represents the total energy of the signals at level  $P$ ; dividing through by this value in Equation 2 normalizes the vector in the working space ( $W$ ) to a constant value. This makes the

comparison largely independent from the strength of the patterns themselves, thus preventing mismatches arising from variations in the strength of back connections. This also permits processing to begin at the onset of a cycle rather than waiting for the signal to build up. Because of normalization, the outcome of the comparison between  $W$  and  $I$  reflects spatial registration alone.

By virtue of its generality, this formulation has the advantage of not requiring a precise specification of how the signal  $j$  is represented by the vector  $P_j$ . Normalization, however, does require that the relative strengths of the signals be specified. For simplicity, we have chosen to set the strength of the visible target signal ( $I_1 = I_{target}$ ) equal to that of the mask ( $I_2 = I_{mask}$ ).

Noise is introduced in the circuit both in the ascending pathway, from the external input, and in the descending pathway. In general, noise is best regarded as a random zero-sum variable. For simplicity, however, we represent it as a fixed pattern that is always present and does not contain any useful information. In the present implementation of CMOS, the noise signal is represented by a constant of strength  $v$ , set to 0.4 in the simulation.

### Linking CMOS to Perception

A critical factor in our experiments on masking was the number of potential target items in the display (set size). Identification of the target item denoted by the ring in Experiments 1–2 and by the four dots in Experiments 3–6 was increasingly impaired as set size was increased. In keeping with the large literature on set-size effects in visual perception (Duncan & Humphreys, 1989; Eriksen, 1995; Sperling, 1960; Treisman & Gelade, 1980), we modeled the time required for attention to be deployed to the location of the target as a linear function of set size:

$$t_c = S * n, \quad (3)$$

where  $t_c$  is the total time that elapses before attention "makes contact" with the target location,  $S$  is a search rate that is influenced by such factors as the degree of similarity among the items (Duncan & Humphreys, 1989; Wolfe et al., 1989) and whether a spatial precue has been presented before the onset of the search array (Eriksen, 1995), and  $n$  is the number of items in the display. Note that modeling attention as a linear function of set size does not make any commitment to serial versus limited-capacity parallel processes, because both predict a linear change in performance with set size (Townsend & Ashby, 1983).

In modeling the data from Experiments 1 to 5 with CMOS,  $t_c$  was computed by allowing  $S$  to be a free parameter that was adjusted to provide a best fit to the data. In Experiment 6,  $t_c$  was computed in two stages, corresponding to the two phases of the experiment. In the first phase, four dots were displayed in the location of the upcoming target for durations that varied from 0 to 180 ms. In the second phase, which was the same as the 90-ms condition in Experiment 4, the stimulus array was displayed briefly and was followed by the four dots for 90 ms. During the second phase,  $t_c$  was modeled in the same way as in Experiment 4, namely,  $t_c$  was increased at the rate  $S_{post}$ , whose value depended on the degree of similarity between the target and other items in the display. However, during the earlier phase in which only the precue was on view, the precue was assumed to reduce  $t_c$  as a function of  $S_{pre}$ , a free parameter denoting processing rate, multiplied by  $n$ , the size of the upcoming search array. The same type

of function (i.e., a linear approximation to the effects of attention) was used in both phases of Experiment 6 because the attention literature indicates that the benefit of a precue is a joint function of the duration of the precue and the number of items in the upcoming display (Cheal & Lyon, 1994a, 1994b; Eriksen, 1995). Note that in the case where precue duration was zero in Experiment 6,  $t_c$  was simply based on the  $S_{\text{post}}$  value determined in the second phase.

The probability of perceiving the target was assumed to be proportional to the fraction of target energy present at the pattern level ( $P$ ) at time  $t_c$ , namely, at the time attention was deployed to the target

$$\text{Prob} = \frac{1}{K} * \frac{P_1^2(t_c)}{\sum_j P_j^2(t_c)} \quad (4)$$

or, because  $P_{\text{target}} =$  target strength,  $P_{\text{mask}} =$  mask strength, and  $P_{\text{noise}} =$  noise strength,

$$\text{Prob} = \frac{1}{K} * \frac{P_{\text{target}}^2(t_c)}{P_{\text{target}}^2(t_c) + P_{\text{mask}}^2(t_c) + P_{\text{noise}}^2(t_c)}, \quad (5)$$

where  $K$  is the asymptotic value at which  $\text{Prob} = 1$ . Equation 5 can be seen as a first approximation to a signal-to-noise measure. The fraction of target energy also is proportional to the square of the cross-correlation of the target signal to the total signal at level  $P$ . This latter measure is similar to that used by Bridgeman (1978) for a similar purpose. Note that  $K$  represents a term whose value is inversely related to the observer's ability to extract useful information from the target signal. Here,  $K$  is treated as a free parameter that is adjusted to provide the best fit of the model to the data.

#### *Response of CMOS When All Parts of the Display Terminate Together*

We now consider the model's response when the target and the mask terminate together, as when the duration of the trailing mask was equal to zero in the experiments reported in Part III. Upon termination of the stimulus display, the contents of  $P$  begin to deteriorate because the signal becomes progressively more noisy with each successive iteration. As shown in Figure 14 and in Equation 1, when there is no new input to  $I$ , only noise is added to the contents of both  $I$  and  $W$ . Thus, the strength of the target signal in  $P$  (i.e.,  $P_{\text{target}}$ ) will decrease, and that of the noise component will increase with each new iteration. If identification can occur before the target signal decays below a critical level  $K$ , as indicated in Equation 5, a correct response will ensue. Otherwise, the response will be stochastic, detection being simply proportional to the level of remaining energy in the target signal.

#### *Response of CMOS When Part of the Display Continues as a Trailing Pattern*

Next, we consider the model's response when the mask remains on display for some time after the target has been turned off. In the present experiments, this was the case when the duration of the trailing mask was greater than zero. On target offset (i.e., when  $I_{\text{target}}$  becomes zero), the target signal at the pattern level ( $P_{\text{target}}$ ) begins to decrease for the reasons given above. On the other hand, the mask signal at that level ( $P_{\text{mask}}$ ) remains strong because it is supported by continued input  $I_{\text{mask}}$ . If the mask remains on view

beyond the target, successive iterations cause  $P_{\text{target}}$  to become progressively weaker, with the mask and noise signals becoming an increasingly larger fraction of the total energy at the pattern level. If the target can be identified before its signal decays below a critical level  $K$ , a correct response will be made. Otherwise, the mask signal will begin to determine what is seen, with the likelihood of missing the target increasing as  $P_{\text{mask}}$  continues to grow. In essence, the mask becomes increasingly likely to replace the target in level  $P$  as the object of perception.

#### Instantiating CMOS

We instantiated CMOS in a Pascal language program that is available from the authors. The results of the simulation are presented in Figure 15A–E, corresponding to Experiments 2 to 5, respectively. The data of Experiment 1 were not used because, as discussed in Part II, target accuracy in Experiment 1 was influenced by factors such as inhibitory contour interactions, which are not part of CMOS. Individual panels in Figure 15 contain the data of the relevant experiment and the corresponding CMOS simulation. For simplicity, the data for the two observers in Experiment 2 were averaged in Figure 15A, and the same was done for Experiment 3 (Figure 15B). In all simulations, the value of the noise factor was held constant at 0.4, the weighting factor was 0.025, and the duration of each iterative loop was set at 13 ms. A systematic exploration of cycle times ranging from 0 ms to 20 ms indicated that 13 ms provided a better fit to the data than either longer or shorter times. This is close to the 5–10 ms values reported by Sillito et al. (1994) for a single cycle between lateral geniculate nucleus (LGN) and area 17 of cat. Note also that in a biological system, the activity within a loop is likely to consist of several cascading signals, each graded—rather than discrete—in time. Therefore, the fixed value of 13 ms used in the present model must be regarded as a first approximation to the average duration of a single iterative cycle. The values of  $K$  and  $S$ , which were estimated from the data, are shown in Table 1, separately for each experiment. Also shown in Table 1 are the  $r^2$  values indicating the proportion of the variance accounted for, separately for each experiment.

The simulations in Figure 15 provide remarkably good fits to the data. The values of  $r^2$  in Table 1 indicate that the proportion of the variance accounted for by the CMOS simulation was between .78 and .98. A second interesting aspect of Table 1 is the range of values assigned to  $S$ , which indexes the time to contact with the target. In Experiments 2 and 3, the average value of  $S$  was 11, whereas in Experiment 4 and in the postcue phase of Experiment 6, the average was 35. This difference in the value of  $S$  is consistent with the fact that the participants in Experiments 2 and 3 were highly practiced psychophysical observers, whereas the participants in the other experiments were unpracticed. Time to contact with the target was therefore influenced by the difference in observational skill. In Experiment 5, the low value of  $S$  ( $S = 5$ , Table 1) reflects the fast time to contact when the target pops out, even if the observers are inexperienced. Note that the value of  $S$  during the precue period in Experiment 6 ( $S = 7$ , Table 1) was very close to that obtained with target pop out in Experiment 5. This high level of consistency adds to the overall plausibility of the model.

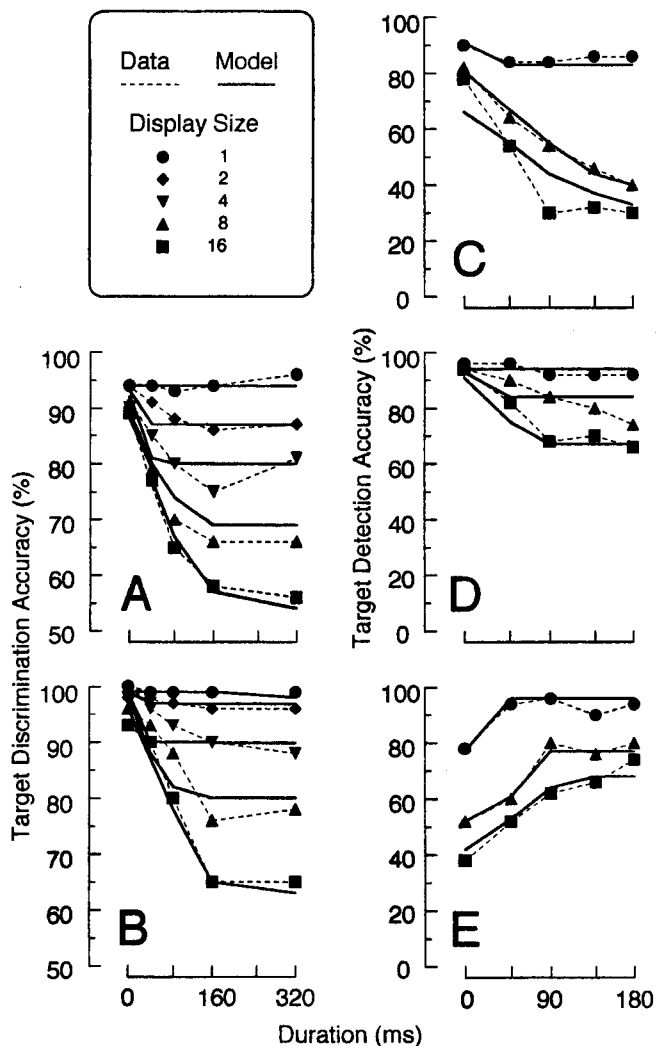


Figure 15. Panels A to E contain the empirical data (dashed lines) and the corresponding computational model of object substitution simulations (solid lines) of Experiments 2 to 6, respectively. In Panels A to D, the abscissa indicates the duration of the trailing mask. In Panel E, the abscissa indicates the duration of the precue. The results of the two observers in Experiment 2 were averaged to yield the data in Panel A. Similarly, the data in Panel B were obtained by averaging the results of the two observers in Experiment 3.

### Levels of Explanation

In its present version, CMOS is intended to model the overall functional architecture of the visual system, not the implementational details of its components. Two aspects of the model are essential for this purpose. One is the iterative recurrent signaling between brain areas that are linked by reentrant pathways. Reentry is required for comparing current input with past input. It may be suggested that the same objective could be achieved by a feed-forward system, such as two lines, one with a longer delay. But this is true only when the control system is known completely and to an arbitrarily high accuracy, which is clearly not the case in a visual system that guides action in the real world. Nor is this true

for the type of masking we describe, for which there are no extant feed-forward accounts. The second essential aspect of CMOS is the normalization carried out at the *P* level (Figure 14). This introduces a nonlinearity in the model and mediates the process of object substitution that is central to our account of common-onset masking. The normalization step at the *P* level has the consequence that the target and mask compete directly for signal strength: An increase in the strength of one means a decrease in the strength of the other.

Implementational details of other parts of the model are less specific and are left free to vary. For example, the precise mechanism by which the external input and reentrant signals are compared (C in Figure 14) has not been spelled out. The computer simulation is based on the assumption that the time taken for the comparison increases linearly with set size, an assumption buttressed by the large literature on set-size effects in visual perception (Duncan & Humphreys, 1989; Treisman & Gelade, 1980). It is likely that the comparison is done by a hill-climbing process based on correlation, as exemplified in the Alopex optimization algorithm (Unnikrishnan & Venugopal, 1994; see also Eggermont, 1990). The present, more general formulation has the advantage of not requiring a precise specification of how the comparison is actually carried out. This generality is a strength, not a weakness, because the model shows this architecture to be adequate to the assigned task, regardless of the specifics of the underlying mechanisms.

Other instances in which the details of the underlying mechanisms are left unspecified have been noted above. These include the exact form of the perceptual hypotheses tested in the course of an iterative cycle: Are they fixed templates (Carpenter & Grossberg, 1987; Unnikrishnan & Venugopal, 1994) or dynamic templates (Mumford, 1992)? Similarly, the representations at the *W* and *I* levels have been left free to take any of several forms without affecting the overall functioning of the model. In essence, the present version of CMOS was designed to demonstrate that a general scheme based on iterative reentrant processing could account successfully for a range of perceptual phenomena, even when the precise nature of the individual components was left unspecified. In casting our ideas in the form of a computational model, we aimed at avoiding the ambiguities that are all but inevitable when the model is expressed in the vernacular, as was done in Parts I and II, above. In essence, CMOS can be regarded as belonging to a class of similarly general schemes that have been developed to account for such perceptual and attentional phenomena as metacontrast masking, visible persistence, iconic memory,

Table 1  
Values of *S*, *K*, and *r*<sup>2</sup> in Each Simulation

Exp	Fig	<i>S</i>	<i>K</i>	<i>r</i> <sup>2</sup>
2	15A	12	0.525	.95
3	15B	10	0.475	.93
4	15C	32	0.520	.94
5	15D	5	0.510	.78
6	15E	7 <sup>a</sup> , 39 <sup>b</sup>	0.500	.98

Note. Exp = experiment; Fig = figure; *S* = search rate in milliseconds; *K* = asymptotic value at which *p* = 1.

<sup>a</sup> Precue. <sup>b</sup> Postcue.

and visual search (e.g., Bridgeman, 1978; Sperling, 1960; Wolfe et al., 1989).

#### PART IV: NEUROPHYSIOLOGICAL AND THEORETICAL CONSIDERATIONS

An issue of concern in developing CMOS as a model of iterative reentrant processing was to maintain neuroanatomical plausibility at the level of a general architecture. Here, we consider some likely neural correlates of the processes embodied in CMOS. We then consider the relation between CMOS and other proposals for masking and for reentry.

##### Neural Correlates of CMOS

The pattern of reentrant pathways that provide the neuroanatomical foundation for CMOS was outlined in Part I (Felleman & Van Essen, 1991; Shipp & Zeki, 1989; Sillito et al., 1994). We now examine specific neurophysiological evidence that is relevant to the timing of the reentrant process embodied in CMOS, and we relate that timing to the temporal course of masking by object substitution.

Within the network of reentrant pathways, there is evidence that information is processed in both local and global loops. Local loops span neighboring brain regions that communicate reciprocally through two-way links (Eckhorn et al., 1988; Gray & Singer, 1989). A good example of local cortical loops has been described by Hupe et al. (1998). Another example has been described between LGN and primary visual cortex in a cat (Sillito et al., 1994). Activity in this loop appears to be useful in optimizing low-level motion signals and suppressing motion smear. Of particular interest here is the temporal course of this activity. The authors estimated that one cycle from LGN to cortex and back to LGN takes about 10 ms. This is similar to the cycle time of 13 ms that yielded the best fit in CMOS (Figure 15). We hypothesize that many such local loops are triggered throughout the visual pathways in an initial burst of activity following stimulation. Once triggered, the local loops continue to operate concurrently, each attuned to a different attribute of the stimulus.

Superimposed on these fast local loops are slower global loops that cover larger brain areas. An example of a global loop encompassing most of the brain has been provided in a study of the event-related brain potentials (ERPs) associated with target detection and cue validity (Luck et al., 1994). ERPs were recorded from occipital, parietal, central, and frontal regions of the scalp. A sequence of three attentional effects were found, the first in the occipital region, the next in the frontal region, and the third back in the occipital region. The sequence began with a positive wave (P1) whose amplitude diminished progressively from occipital to frontal regions. A corresponding negative wave (N1), peaking in the frontal region some 120 ms after stimulus onset, peaked progressively later in the central, parietal, and occipital regions. The temporal pattern of the N1 wave indicated the presence of a relatively early anterior generator source and a later posterior source. This entire temporal pattern is consistent with the activation of successive components in a global loop extending from occipital to frontal and back to occipital regions. This pattern of activation was confirmed in a related experiment involving feed-

back loops between occipital and frontal brain regions (Luck & Hillyard, 1994).

On this hypothesis, activity proceeds from primary cortex to other visual areas where input signals are convolved with stored information. The coded signals are then returned to lower centers for further processing. The reentrant pathways, however, do more than merely complete a feedback loop. A more active function is implicated by the fact that the descending fibers terminate not only on the units that triggered the initial ascending signals but on neighboring and intermediate units as well. The significance of this anatomical distribution was noted by Shipp and Zeki (1989) and by Sillito et al. (1994), among others, who suggested that the reentrant signals may act to change the response properties of the target units. It is as though the system reconfigures itself with each iteration, so that the same cells can serve different functions at different stages of the processing cycle. Much the same conclusion was reached by Bridgeman (1980) on the basis of single-unit recordings from monkey's area V1 in a metacontrast experiment: "The same cell can participate in early, late, and preresponse phases of the cortical response, demonstrating a true multiplexing in single-cell coding" (p. 361). This could well offer a solution to the problem of the elusive "grandmother cell." Grandmother cells sometimes cannot be found because any given cell can participate, so to speak, in the entire range of behaviors from granddaughter to grandmother at different times in the processing cycle.

##### Reentrant Processing of Stimulus Attributes

Lamme, Zipser, and Spekreijse (1997) displayed textured stimuli containing a figure on a background and recorded responses from V1 neurons in awake monkeys fitted with chronic microelectrode implants. The experiment was conducted in two parts. The first provided evidence concerning the sequence in which different stimulus attributes are processed. The second provided evidence of reentrant processing. In the first part, three stages of processing were revealed in the V1 response. In Stage 1 (up to about 80 ms from stimulus onset), neurons responded only to local features presented within their receptive fields. In Stage 2 (80–120 ms), the same neurons began to respond to figure boundaries. In Stage 3 (beyond about 120 ms), the neurons responded to the surface of the figure. Lamme et al. concluded that "visual processing in V1 progresses from local feature detection to high-order boundary detection, culminating in a representation of figure-ground relationships of surfaces in the scene" (p. S969).

Evidence of reentrant processing was obtained in the second part of the experiment, in which the same monkeys were tested after having undergone extensive lesions to extrastriate cortex ipsilateral to the recording site. After surgery, the V1 activity corresponding to Stages 1 and 2 continued to be very much in evidence, but the V1 activity corresponding to Stage 3 was missing. Behaviorally, the monkeys were no longer capable of distinguishing figure from ground. Note that a very similar outcome was obtained under anesthesia. When the monkeys were anesthetized, neurons in V1 continued to respond to local features (Stage 1) and remained sharply tuned to stimulus orientation, but Stages 2 and 3 were missing. These findings confirm that, within a processing cycle, signals from higher centers reenter the primary visual cortex. More important, when the reentrant flow is disrupted, local

nonfigural attributes of the stimulus can still be processed, but global attributes relating to the overall configuration are lost.

The parallel with the results of Experiments 1–3 is compelling. Of the two processing stages revealed in our experiments, the first occurred within less than 80 ms from stimulus onset. At this stage, masking depended critically on local contour interactions, suggesting that processing was concerned with local contours. Moreover, the strength of masking was not affected by global figural properties of the display, notably set size. The corresponding Stage 1 of Lamme et al. (1997) also occurred within the first 80 ms, was concerned with local contours, and was not affected by the figural properties of the rest of the display. In contrast, our second stage, modeled by CMOS, occurred later in the processing sequence, was unaffected by local contours, and masking depended critically on global figural properties of the display. We suggest that our second stage corresponds broadly to Lamme et al.'s Stage 3, which was predicated on reentrant signals to V1 from extrastriate cortex.

The main function assigned to these reentrant signals in current theories (e.g., Grossberg, 1995a; Mumford, 1992; Sillito et al., 1994) and also implemented in CMOS is to test alternative hypotheses about the identity of the inducing stimuli and their locations in space. That is, the initial pattern of activity in primary cortex may be compatible with more than one representation at higher cortical levels or, equivalently, with a representation that is ill defined or ambiguous. Moreover, fine spatial resolution cannot be maintained by the large receptive fields at the higher levels. Both figural ambiguity and positional uncertainty can be resolved by convolving the reentrant signals with ongoing activity at the lower level. In the lesioned monkeys of Lamme et al. (1997), no such correlation was possible because there were no reentrant signals to be convolved with the V1 activity. Our theoretical conviction, set out in CMOS, is that object-substitution masking occurs when the reentrant signals arrive while the predominant activity in V1 is that produced by the mask rather than the target, thus maximizing the correspondence with the reentrant code for mask alone. In this respect, a further parallel between Lamme et al.'s Stage 3 and our results with the four-dot mask should be noted. Both are concerned with surfaces rather than contours. Under the appropriate conditions in our experiments, what was perceived instead of the target was the empty square surface bound by the four dots. In this process of object substitution, the object defined by the four dots replaced the object defined by the target.

### Comparing Feed-Forward With Reentrant Approaches

Feed-forward and reentrant approaches were contrasted at the beginning of the present article, not as mutually exclusive alternatives, but as complementary ways in which visual processing might be conceptualized. It is our view that although many forms of masking can be explained using feed-forward hierarchical stages, without recourse to reentrant processes, common-onset masking cannot.

We consider, first, conventional backward masking. Feed-forward accounts can deal with backward masking, provided that the mask trails the onset of the target by a suitable SOA. This includes the case in which the mask consists of only four dots. For example, it has been shown that identification of a target is severely impaired by a four-dot mask presented some 70 ms after target onset (Enns & Di Lollo, 1997). This can be explained

without recourse to reentry by hypothesizing that masking takes place through a process of object substitution at a relatively high processing level. For instance, suppose that the physical attributes of a stimulus are first encoded, rapidly and in parallel, at an early stage of processing. More meaningful attributes, such as overall form or configuration, would emerge at a higher level, where processing is held to be serial and vulnerable to interruption by temporally trailing stimuli. Backward masking of a target by a four-dot mask could be explained on two assumptions: First, at the higher level, the mask is represented not as an aggregate of four dots, but as a square surface delimited by the four dots; second, masking at the higher level occurs by interruption of processing. This means that if a trailing stimulus arrives while the processing of a leading stimulus is still incomplete, the leading stimulus is abandoned, and processing is redirected to the trailing stimulus. Thus, if the mask arrives while processing of the target is still incomplete, the mask object (a blank square surface) is substituted for the target object in conscious perception.

Attentional effects on backward masking could also be explained without recourse to reentry. All that is needed is to assume that stimuli presented at the focus of attention are processed more promptly and more rapidly than unattended stimuli. Thus, a target presented at a precued location will be processed more rapidly and will escape substitution by a trailing mask at shorter SOAs. For much the same reason, if attention can be deployed rapidly to a target that pops out, the period for which the target remains vulnerable to substitution by a trailing mask will be reduced correspondingly.

What is problematic for feed-forward accounts is masking by common onset of target and mask. We have seen in Part I that common-onset masking cannot be explained by onset-locked inhibitory processes in early vision, nor by accounts based on the temporal limits of spatial attention, nor by accounts based on lateral inhibitory processes that result in crowding. Instead, a complete account must include a sensory component, a memory component, and some process in which the two are compared. Consider first the consequences of trying to account for common-onset masking using only one of these components. If the memory component is abandoned, then the task of reporting the target in a display of 10-ms duration becomes impossible. Both electrophysiological and psychophysical measures indicate that the brain simply does not process visual information that quickly. The high degree of target accuracy in these displays therefore confirms the existence of a sufficiently rich visual memory to accomplish the task. Next consider the consequences of eliminating the sensory component. Without an obligation to process the four dots on the screen when they trail the display, no masking would occur, as is shown by the high accuracy when the trailing mask duration is zero. The very existence of common-onset masking therefore points to a visual process that is unable to inhibit current sensory input in favor of memory representation that would be adequate to the task on its own.

Yet, a question that still remains is whether the comparison process involving these two components needs to be a reentrant one. We think it does because the comparison is between a visual signal already in mind (the memory of the target display) and a visual signal currently in the eye (the sensory presence of the four dots). This is reentry by any definition. In most previous experiments, such interactions have most often been referred to simply as



*top-down* influences, because the time scale was such that it was not clear where the top-down effects were taking place. As mentioned in the introduction, in Bruner and Goodman's (1947) study on the apparent size of coins, the effect of attention may have occurred only at the highest levels of processing. The methodology of common-onset masking is unique in being able to trace the top-down influence all the way down to the emergence of a percept.

Finally, we note that the proposed reentry process need not necessarily occur between distinct brain regions. It could well occur entirely within a given topographical cortical map. This possibility does not create a problem in principle for the reentrant account. It is known that reentrant connections exist between brain centers located within the same general area (*intraareal connections*; Gilbert & Wiesel, 1989) as well as between more distant areas.

This said, a second question needs to be addressed. What are the inherent advantages in feed-forward accounts over reentrant accounts? Parsimony is often cited as the most obvious advantage, and intuitively, feed-forward accounts may seem more parsimonious than reentrant schemes. On closer inspection, however, reentrant processing reveals a degree of parsimony that goes beyond what can be achieved with feed-forward circuitry alone. For example, the experiments of Lamme et al. (1997), noted earlier, showed that reentrant signals can reconfigure the same neurons to perform several very different functions in successive phases of processing. Such multiplexing permits a leaner and more efficient system than one with enough neurons to do the same job in a feed-forward fashion.

At any rate, the assumption of parsimony in defense of the feed-forward view is itself questionable. As is true for most biological systems, brain mechanisms cannot be assumed to have evolved along what an engineer might regard as most parsimonious lines. Were theories to be developed with parsimony as the overriding constraint, we could well end up with models that were highly parsimonious but bore little resemblance to how the brain works. In contrast, the neuroanatomical evidence for reciprocal connections between brain centers is massive, and it points compellingly to reentrant processing as the brain's main *modus operandi*. This provides grounds for establishing a credible linking proposition (Teller, 1990) between the domains of psychophysics and neurobiology based on reentry.

### Reentrant Processing and Other Forms of Masking

We have argued that common-onset masking cannot be explained by current theories of backward masking. Indeed, theories of metacontrast, the form of masking most similar to common-onset masking, were designed explicitly to rule out masking when the target and mask had a common onset. Here we address the converse issue, namely, whether the principles of iterative reentrant processing espoused in CMOS can also account for masking effects obtained with conventional paradigms. In other words, is it possible that common-onset masking engages the same underlying mechanisms as classical metacontrast masking and backward masking by pattern?

According to the hypothesis of iterative reentrant processing, there is no difference in principle between masking with common onset and many aspects of classical metacontrast and pattern

masking. All forms of backward masking will be subject to the influences of object substitution, in that the representation of a temporally leading target will be replaced in consciousness by that of the mask if it follows closely in time and appears at the same location. Undoubtedly, there will also be minor differences in each form of masking, with, for example, metacontrast masking having specific types of contour interactions that are not shared by pattern masking or masking by four dots. However, the critical requisite for the object-substitution process that all these forms of masking share is that the mask be visible during the period in which the iterations between higher level pattern representations and lower level contour representations are likely to occur.

To help emphasize the similarities across the various forms of masking, it should be noted that masking with common onset is not itself dependent on the mask remaining visible throughout the display period. Masking is also obtained in display sequences that begin with a brief combined display of target and mask, continue with an appropriate ISI during which the screen is empty, and end with a brief display of the mask alone (Bischof & Di Lollo, 1995). Except for the presence of the mask in the initial step, this is the same sequence as in conventional metacontrast, and it yields the same U-shaped accuracy over SOA. To add further credence to the link between the two forms of masking, Bischof and Di Lollo have reported that individual differences were maintained across the two paradigms and masking was affected in similar ways by changes in viewing conditions.

The strongest evidence for reentrant processes in metacontrast masking has come from electrophysiological studies. Single-unit recordings from cat and monkey visual cortex show that metacontrast masking is associated with a reduction in peak V1 responses occurring beyond 80 ms and as late as 400 ms after stimulus onset (Bridgeman, 1975, 1980). Contrary to expectations from feed-forward inhibitory theories, earlier components are affected minimally, if at all—a result confirmed by von der Heydt et al. (1997). Bridgeman (1980) has interpreted the reduction in the late peak of the V1 response as representing the influence of reentrant signals from regions beyond the primary visual cortex.

Homologous results have been obtained with scale recordings of ERPs in humans (Bridgeman, 1988; Jeffreys & Musselwhite, 1986; Vaughn, 1969). Vaughn's evidence was especially clear in pointing to an extrastriate origin of the late component observed in the V1 response in metacontrast masking. In addition, both Vaughn and Bridgeman (1980) observed that in primary cortex, the spatial distribution of the late activity was far wider than that of the early activity. This is in accordance with a reentrant origin of the late activity. It is now known that in primary cortex of monkey, the spatial distribution of the reentrant fibers is far wider than that of the corresponding ascending units (e.g., Shipp & Zeki, 1989; Sillito et al., 1994).

Studies of backward masking by pattern are also consistent with our view of reentrant processing. An experiment by Spencer and Shuntich (1970) is especially relevant because it revealed substantial set-size effects on both the strength and the temporal course of masking. The stimuli consisted of either 1 or 12 alphabetical characters, with the target being indicated by an adjoining bar marker. A pattern mask was presented briefly at the target's location over a range of SOAs. The results revealed an early masking component (up to about 100 ms), which responded to the variations in luminance but not to set size, and a late component

(up to 300 ms), which responded to set size but not to differences in luminance. This interaction between set size and SOA in the later component parallels that found in the present experiments between set size and duration of the trailing mask. In both cases, the late masking component depended on the presence of distractors in the display and is therefore amenable to explanation in terms of reentrant processes. We therefore regard the late masking component of Spencer and Shuntich as an instance of object-substitution masking in which the trailing mask replaced the target as the object to be perceived.

Special note should be made of a theory of visual attention and masking proposed by Bundesen (1990), in which the accuracy of target identification becomes progressively weaker as attention becomes more divided. Joint consideration of masking and attention makes this theory potentially useful in the present context. However, major theoretical development would be required if the theory is to account for the results reported in Part II.

As presently formulated, Bundesen's (1990) theory is based on only one form of masking, namely, backward masking by pattern. Not modeled is metacontrast masking, in which there is no spatial overlap between the contours of the target and those of the mask. An important consideration in modeling these two types of masking is the differential effect of SOA. As modeled by Bundesen, maximum masking by pattern is obtained at an SOA of zero, with strength of masking decreasing as the SOA is increased. Masking by metacontrast, on the other hand, is nonexistent at an SOA zero, with strength of masking increasing as the SOA is increased up to an optimal duration and diminishing thereafter (e.g., Breitmeyer, 1984). The treatment of SOA within Bundesen's theory is clearly inappropriate for any of the masking effects obtained in the present study (Part II), because none of these involved spatially overlapping contours. Therefore, Bundesen's theory would need to be expanded to include metacontrast-like masking paradigms if it is to be useful in accounting for the present results.

There is also a second aspect of Bundesen's (1990) theory, crucial to the relationship between masking and attention, that requires expansion and elaboration. Regardless of the temporal asynchrony between the target and mask, all masking is attributed to a single underlying factor: interruption of processing. This ignores masking by integration, which is based on different underlying mechanisms and is responsible for most pattern-masking effects at short SOAs (Michaels & Turvey, 1979; Scheerer, 1973; Turvey, 1973). The important issue is that masking by integration is not affected by the distribution of attention (Spencer & Shuntich, 1970) and, therefore, must be modeled differently from masking by interruption. The two sources of masking, which are currently confounded in Bundesen's theoretical fit to the data, should be modeled separately to specify accurately the role of attention in different forms of masking.

Considered collectively, the evidence from conventional metacontrast and pattern-masking paradigms is consistent with the pattern of evidence observed for common-onset masking. Because current theories of these conventional forms of masking cannot account for common-onset masking, whereas the reentrant processing theory can account for all forms of backward masking involving objects and patterns, our hypothesis is that the same mechanisms are being tapped in the various paradigms. As such, we hypothesize that all masking effects that are sensitive to temporal variables, including the time required to attend to the relevant

target stimulus and the time that elapses between successive stimuli, reflect the number and speed of the iterations that are required for stimuli to be perceived.

## CONCLUSION

Common-onset masking has been demonstrated in this study to be a powerful new tool for the study of visual perception. The simple act of watching a target display disappear from the screen can be used to investigate the intricate processes of perceptual reinterpretation that form our everyday perceptions. By these processes, an initial representation formed from a brief display (target + mask) can be observed to be discarded in favor of a representation based on the display currently in view (mask alone). If such reinterpretations did not occur, then the perceptual sequence seen by an observer would consist of the target-mask combination followed by the mask alone. Indeed, the target-mask combination is precisely what is seen unless the mask remains on view.

The present study has identified at least two distinct phases in the ongoing processes of perceptual reinterpretation: an early phase in the first 100 ms or so, probably involving intra-areal activity, dedicated to the processing of local stimulus attributes; and a later phase beyond 100 ms, perhaps involving interareal activity, dedicated to object perception. We surmise that conscious perception is associated most strongly with the second phase of activity. This assumption is consistent not only with the results of our experiments, where a percept that was reliably elicited by a brief stimulus (common onset and offset of target and mask shapes) was dramatically altered by the continued presence of only a part of the same pattern (mask alone) but also consistent with much other psychophysical research showing that conscious perception is associated with the relatively invariant attributes of the object rather than with the inherently unstable attributes of the image (e.g., size constancy, Rock, 1983; colour constancy, Mollon, 1995; shape constancy, Shepard, 1981; scene perception, Intraub, Gottesman, Willey, & Zuk, 1996).

This view leads to several important predictions with regard to visual masking that would otherwise be counterintuitive from the perspective of conventional models that are feed forward and contour based. These include the three main findings of this study: (a) The brief presentation of a pattern can be masked by the continued presence of only a part of that pattern, (b) a shape that does not overlap with the target and is only notionally defined can serve as a powerful mask, and (c) masking is very sensitive to the spatiotemporal deployment of attention.

## References

- Adelson, E. H., & Jonides, J. (1980). The psychology of iconic storage. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 486-493.
- Alpern, M. (1953). Metacontrast. *Journal of the Optical Society of America*, 43, 648-657.
- Baker, H. D. (1963). Initial stages of light and dark adaptation. *Journal of the Optical Society of America*, 53, 98-103.
- Barlow, H. B., Fitzhugh, R., & Kuffler, S. W. (1957). Change of organization of the receptive fields of the cat's retina during dark adaptation. *Journal of Physiology (London)*, 137, 338-354.

- Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, *303*, 69–698.
- Bischof, W. F., & Di Lollo, V. (1995). Motion and metacontrast with simultaneous onset of stimuli. *Journal of the Optical Society of America A*, *12*, 1623–1636.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Bowen, R. W., Pola, J., & Matin, L. (1974). Visual persistence: Effects of flash luminance, duration and energy. *Vision Research*, *14*, 295–303.
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychological Review*, *83*, 1–36.
- Breitmeyer, B. G., & Kersey, M. (1981). Backward masking by pattern stimulus offset. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 972–977.
- Breitmeyer, B. G., Rudd, M., & Dunn, K. (1981). Spatial and temporal parameters of metacontrast disinhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 770–779.
- Bridgeman, B. (1975). Correlates of metacontrast in single cells of the cat visual system. *Vision Research*, *15*, 91–99.
- Bridgeman, B. (1978). Distributed sensory coding applied to simulation of iconic storage and metacontrast. *Bulletin of Mathematical Biology*, *40*, 605–623.
- Bridgeman, B. (1980). Temporal response characteristics of cells in monkey striate cortex measured with metacontrast masking and brightness discrimination. *Brain Research*, *196*, 347–364.
- Bridgeman, B. (1988). Visual evoked potentials: Concomitants of metacontrast in late components. *Perception & Psychophysics*, *43*, 401–403.
- Broadbent, D. (1958). *Perception and communication*. Oxford, England: Pergamon.
- Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, *42*, 33–44.
- Bullier, J., McCourt, M. E., & Henry, G. H. (1988). Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Experimental Brain Research*, *70*, 90–98.
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.
- Burr, D. C. (1984). Summation of target and mask metacontrast stimuli. *Perception*, *13*, 183–192.
- Carpenter, G., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54–115.
- Cheal, M., & Lyon, D. R. (1994a). Allocation of attention in texture segregation, visual search, and location-pecking paradigms. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47A*, 49–70.
- Cheal, M., & Lyon, D. R. (1994b). A framework for understanding the allocation of attention in location-precued discrimination. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *47A*, 699–739.
- Cohene, L. S., & Bechtold, H. P. (1974). Visual recognition as a function of stimulus offset asynchrony and duration. *Perception & Psychophysics*, *15*, 221–226.
- Cohene, L. S., & Bechtold, H. P. (1975). Visual recognition of dot-pattern bigrams: An extension and replication. *American Journal of Psychology*, *88*, 187–199.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics*, *27*, 183–228.
- Damasio, A. R. (1994). *Descartes' error*. New York: Putnam.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75–97.
- Di Lollo, V., Bischof, W. F., & Dixon, P. (1993). Stimulus-onset asynchrony is not necessary for motion perception or metacontrast masking. *Psychological Science*, *4*, 260–263.
- Di Lollo, V., & Finley, G. (1986). Equating the brightness of brief visual stimuli of unequal durations. *Behavior Research Methods, Instruments, and Computers*, *18*, 582–586.
- Duncan, J., & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458.
- Duncan, J., Ward, R., & Shapiro, K. L. (1995). Direct measurement of attentional dwell time in human vision. *Nature*, *369*, 313–315.
- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analysis in the cat. *Biological Cybernetics*, *60*, 121–130.
- Edelman, G. M. (1978). Group selection and phasic re-entrant signaling: A theory of higher brain function. In G. M. Edelman & V. B. Mountcastle (Eds.), *The mindful brain* (pp. 51–100). Cambridge, MA: MIT Press.
- Eggermont, J. J. (1990). *The correlative brain*. Berlin: Springer.
- Enns, J. T., Brehaut, J., & Shore, D. I. (1999). The duration of a brief event in the mind's eye. *Journal of General Psychology*, *126*, 355–372.
- Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, *8*, 135–139.
- Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review*, *98*, 101–118.
- Eriksen, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*, 101–118.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, *14*, 155–160.
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, 225–240.
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cerebral Cortex*, *1*, 1–47.
- Finley, G. (1985). A high-speed point plotter for vision research. *Vision Research*, *25*, 1993–1997.
- Ganz, L. (1975). Temporal factors in visual perception. In E. C. Chartered & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 5, pp. 169–231). New York: Academic Press.
- Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, *9*, 2432–2442.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Sciences USA*, *86*, 1698–1702.
- Grossberg, S. (1995a). The attentive brain. *American Scientist*, *83*, 438–449.
- Grossberg, S. (1995b). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, *55*, 48–120.
- Growney, R., Weisstein, N., & Cox, S. I. (1977). Metacontrast as a function of spatial separation with narrow line targets and masks. *Vision Research*, *17*, 1205–1210.
- Harth, E., Unnikrishnan, K. P., & Pandya, A. S. (1987). The inversion of sensory processing by feedback pathways: A model of visual cognitive functions. *Science*, *237*, 184–187.
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Helmholtz, H. V. (1962). *Handbook of physiological optics* (3rd ed.,

- J. P. C. Southall, Trans.). New York: Dover. (Original work published 1866)
- Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque visual cortex. *Proceedings of the Royal Society, London (B)*, *198*, 1–59.
- Hupe, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and ground by V1, V2, and V3 neurons. *Nature*, *394*, 784–787.
- Ikeda, M. (1965). Temporal summation of positive and negative flashes in the visual system. *Journal of the Optical Society of America*, *55*, 1527–1534.
- Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed pictures: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language*, *35*, 118–134.
- Jeffreys, D. A., & Musselwhite, M. J. (1986). A visual evoked potential study of metacontrast masking. *Vision Research*, *26*, 631–642.
- Kahneman, D. (1967). An onset-onset law for one case of apparent motion and metacontrast. *Perception & Psychophysics*, *2*, 577–584.
- Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin*, *70*, 404–425.
- Kolers, P. A. (1968). Some psychological aspects of pattern recognition. In P. A. Kolers & M. Eden (Eds.), *Recognizing patterns* (pp. 4–61). Cambridge, MA: M.I.T. Press.
- Kolers, P. A. (1972). *Aspects of motion perception*. New York: Pergamon Press.
- Kolers, P. A., & Rosner, B. S. (1960). On visual masking (metacontrast): Dichoptic observations. *American Journal of Psychology*, *73*, 2–21.
- Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*, 225–234.
- LaBerge, D. L., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*, 101–124.
- LaBerge, D. L., Brown, V., Carter, M., Bash, D., & Hartley, A. (1991). Reducing the effects of adjacent distractors by narrowing attention. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 65–76.
- Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*. Hillsdale, NJ: Erlbaum.
- Lamme, V. A. F., Zipser, K., & Spekreijse, H. (1997). Figure-ground signals in V1 depend on extrastriate feedback [Abstract]. *Investigative Ophthalmology & Visual Science*, *38*, S969.
- Levine, R., Didner, R., & Tobenkin, N. (1967). Backward masking as a function of interstimulus distance. *Psychonomic Science*, *9*, 185–186.
- Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1000–1014.
- Luck, S. J., Hillyard, S. A., Mouloua, M., Woldorff, M. G., Clark, V. P., & Hawkins, H. L. (1994). Effects of spatial cuing on luminance detectability: Psychophysical and electrophysiological evidence for early selection. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 887–904.
- Markoff, J. I., & Sturr, J. F. (1971). Spatial and luminance determinants of the increment threshold under monoptic and dichoptic viewing. *Journal of the Optical Society of America*, *61*, 1530–1537.
- Matin, E. (1975). The two-transient (masking) paradigm. *Psychological Review*, *82*, 451–461.
- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- Michaels, C. F., & Turvey, M. T. (1979). Central sources of visual masking: Indexing structures supporting seeing at a single, brief glance. *Psychological Research*, *41*, 1–61.
- Mignard, M., & Malpeli, J. G. (1991). Paths of information flow through visual cortex. *Science*, *251*, 1249–1251.
- Mollon, J. D. (1995). Seeing colour. In T. Lamb & Janine Bourriau (Eds.), *Colour: Art and science* (pp. 127–150). Cambridge, England: Cambridge University Press.
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in extrastriate cortex. *Science*, *229*, 782–784.
- Mumford, D. (1992). On the computational architecture of the neocortex: II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*, 241–251.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29–55.
- Perkel, D. J., Bullier, J., & Kennedy, H. (1986). Topography of the afferent connectivity of area 17 in the macaque monkey. *Journal of Comparative Neurology*, *253*, 347–402.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 275–280.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- Scharf, B., & Lefton, L. A. (1970). Backward and forward masking as a function of stimulus and task parameters. *Journal of Experimental Psychology*, *84*, 331–338.
- Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung*, *36*, 71–93.
- Servièrè, J., Miceli, D., & Galifret, Y. (1977). A psychophysical study of the visual perception of “instantaneous” and “durable.” *Vision Research*, *17*, 57–63.
- Shepard, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 279–342). Hillsdale, NJ: Erlbaum.
- Shipp, S., & Zeki, S. (1989). The organization of connections between areas V5 and V1 in the macaque monkey visual cortex. *European Journal of Neuroscience*, *1*, 309–332.
- Sillito, A. M. (1992). GABA mediated inhibitory processes in the function of the geniculate-striate system. *Progress in Brain Research*, *90*, 349–384.
- Sillito, A. M., Jones, H. E., Gerstein, G. L., & West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, *369*, 479–482.
- Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology*, *85*, 198–203.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74* (11, Whole No. 498).
- Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instrumentation*, *3*, 148–151.
- Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America*, *58*, 1133–1145.
- Teller, D. (1990). The domain of visual science. In L. Spillmann & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 11–21). New York: Academic Press.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*, 1349–1357.
- Tononi, G., Sporns, O., & Edelman, G. M. (1992). Reentry and the

- problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex*, 2, 310–335.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, 95, 15–48.
- Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114, 285–310.
- Tsal, Y., Meiran, N., & Lamy, D. (1995). Towards a resolution theory of visual attention. *Visual Cognition*, 2, 313–330.
- Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 81, 1–52.
- Unnikrishnan, K. P., & Venugopal, K. P. (1994). A correlation-based learning algorithm for feedforward and recurrent neural networks. *Neural Computation*, 6, 469–490.
- Van der Heijden, A. H. C., & Eerland, E. (1973). The effects of cueing in a visual signal detection task. *Quarterly Journal of Experimental Psychology*, 25, 496–503.
- Vaughn, H. H., Jr. (1969). The relationship of brain activity to scalp recordings of event-related potentials. In E. Donchin & D. Lindsley (Eds.), *Average evoked potentials*, (pp. 45–94). Washington, DC: National Aeronautics and Space Administration.
- von Békésy, G. (1968). Mach- and Hering-type lateral inhibition in vision. *Vision Research*, 8, 1483–1499.
- von der Heydt, R., Friedman, H. S., Zhou, H., Komatsu, H., Hanazawa, A., & Murakami, I. (1997). Neuronal responses in monkey V1 and V2 unaffected by metacontrast. *Investigative Ophthalmology & Visual Science*, 38, S459.
- Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I., & Schnapf, J. L. (1990). The control of visual sensitivity: Receptor and postreceptor processes. In L. Spillmann & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 53–101). New York: Academic Press.
- Weisstein, N., & Harris, C. S. (1974). Visual detection of line segments: An object-superiority effect. *Science*, 186, 752–755.
- Weisstein, N., Ozog, G., & Szoc, R. (1975). A comparison and elaboration of two models of metacontrast. *Psychological Review*, 82, 325–343.
- Werner, H. (1935). Studies on contour: I. Qualitative analysis. *American Journal of Psychology*, 47, 40–64.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature-integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433.
- Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onset: New perceptual objects or visual masking? *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1505–1513.
- Yeh, Y., & Eriksen, C. W. (1984). Name codes and features in the discrimination of letter forms. *Perception & Psychophysics*, 8, 159–199.
- Zeki, S. (1993). *A vision of the brain*. Oxford, England: Blackwell.

Received October 15, 1998

Revision received April 20, 1999

Accepted December 20, 1999 ■