# 5

# An Object Substitution Theory of Visual Masking

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

## Abstract

The brief presentation of a visual display, although clearly visible when shown by itself, can be rendered invisible by the subsequent presentation of a second visual display in the same location. Recent studies of *backward masking* of this kind have revealed several new effects that are not predicted by standard theories of masking. Among these are masking by four small dots that surround but that do not touch the target object, and masking by a surrounding object that remains on display after the target has been turned off. A critical ingredient in both these masking effects is the focus of spatial attention: almost no masking occurs if attention can be rapidly focused on the target, but powerful masking ensues if attention to the target is delayed. A new theoretical framework, inspired by recent developments in neuroscience, is described to explain both these and more traditional forms of masking. In addition to accounting for backward masking, this framework sheds light on several other active areas of vision research, including work on the attentional blink, inattentional blindness, and change blindness.

## INTRODUCTION

Among the most widely used and powerful tools available to researchers of visual perception is that of masking. At the most general level, masking refers to a reduction in the visibility of a visual object (the target) that is caused by the presentation of a second object (the mask) close to the target in space or time. In the purely spatial domain, the visibility of a target is greatly reduced if nontarget items are also presented nearby, an effect aptly referred to as *crowding* (Bouma, 1970). Introducing a temporal interval between the target and mask objects leads to wonderfully complex interactions between spatial and temporal variables. For example, a target that is highly visible when presented briefly in isolation, can be made completely invisible by the later presentation of a nontarget object in the same spatial location, or even in nearby but nonoverlapping locations. *Backward masking* of this kind has its strongest influence, not when target and mask objects are presented simultaneously, as intuition might suggest, but rather when a brief interval of 50-100 milliseconds intervenes between the presentation of target and mask.

Spatial-temporal interactions such as these provide a unique glimpse into the workings of the visual system. They inform us of processes involved in the emergence of a percept that would otherwise remain hidden to the behavioral researcher. These include the amount of time required to form a percept that is immune from the influences of later presented objects (Averbach & Coriell, 1961), the spatial range of influence between multiple objects in the visual field (Brietmeyer, 1984), and the extent to which visual information is processed outside of conscious awareness (Debner & Jacoby, 1994).

In this chapter we summarize insights that have recently been gained through studies of visual masking. However, before setting out, it is important to point out that there are two rather distinct ways in which masking is used in modern studies of vision. On the one hand, masking is a tool of convenience in a large number of vision experiments. It is no secret among researchers that masking is a handy way to adjust the difficulty of a task so that accuracy falls into a measureable range (somewhere between chance levels of responding and perfection). An informal survey of a recent volume of *Perception & Psychophysics* (1999, Vol. 61) indicated that 14 of 93 (15%) articles on vision used some form of backward masking. A typical stated rationale for backward masking is that it limits visual access to the target for a controlled period of time; the mask is said to erase the target from the mind. A second and much smaller group of researchers use masking to investigate the fine-grained spatial and temporal aspects of perception. The same journal volume contained only 5 (5%) articles in which masking itself was under investigation. We therefore hasten to add that although the present chapter may at times seem written primarily for this smaller second audience, the new insights into visual masking have implications that are equally important for those using masking as a tool of convenience, as they are for those studying the perceptual mechanisms of masking.

## THE STANDARD VIEW OF VISUAL MASKING

Visual masking that involves both spatial factors (patterns or forms) and temporal factors (pattern durations and interstimulus intervals) is typically divided into two types, based on the spatial relations that exist between the contours in the target and mask patterns. Masking that involves spatial superposition of contours (although these contours may be separated in time) is commonly refered to as *pattern masking*; when it involves closely adjacent but nonoverlapping contours it is called *metacontrast*. These two forms of masking are differentiated not only by these well-defined spatial relations between target and mask contours, but they are generally believed to depend on different underlying mechanisms.

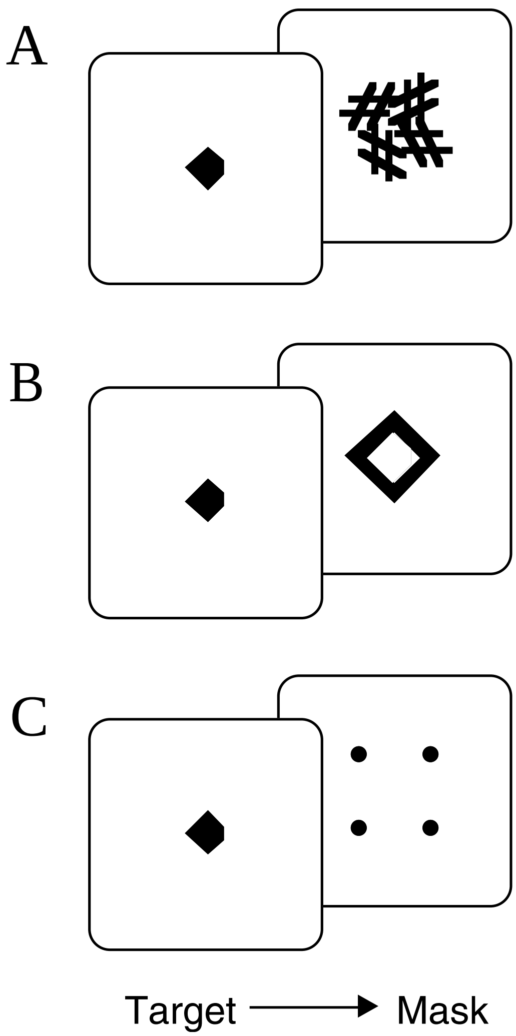


Figure 1. Three different potential masking stimuli in experiments in which observers attempt to identify a briefly presented target followed by the mask. The observer s task is to indicate which corner of the diamond has been removed. (A) Pattern mask: Contours of the mask are spatially superimposed on the contours of the target. (B) Metacontrast mask: Mask contours fit snugly around but do not overlap the target contours. (C) Four-dot mask: Four small dots surround the target. There are no standard theories which predict that masking will occur with the four-dot mask.

Typical examples of stimuli used in each of these kinds of masking experiments are shown in Figures 1a and 1b. In contrast to these masks, Figure 1c is a new type of masking display that we will have more to say about after reviewing the standard masking effects.

Pattern masking presents the visual system with two different kinds of spatio-temporal conflict. One conflict occurs when the target and mask are perceived as part of the same unitary pattern, a consequence of the imprecise temporal resolution of the visual system. In this case, masking is akin to the addition of spatial noise (the mask) to the signal (the target) at early levels of visual representation and so is referred to as masking by integration (Breitmeyer, 1984; Kahneman, 1968; Scheerer, 1973; Turvey, 1973). The temporal signature of this form of masking is approximate symmetry around a maximum level of masking at a stimulus onset asynchrony (SOA) of 0 ms, with a complete release from masking beyond an SOA of about 100 ms in each direction.

A second conflict arises when processing of a first pattern (the target) is interrupted by a second pattern (the mask) that appears in the same spatial location before the target pattern has been fully processed. This conflict does not involve the early stages of processing where contours are defined, but instead involves a competition for the higher-level mechanisms involved in object recognition. It is referred to as masking by *interruption* and is aptly described by an analogy in which targets and masks are customers coming before the clerk of the visual system (Kolers, 1968). The amount of time the clerk spends with the first customer (the target) is sharply curtailed if a second customer (the mask) enters the store. The temporal characteristics are also very different from masking by integration. Interruption masking can occur only when the mask follows the target in time. Therefore, the masking function is referred to as U- or J-shaped, because target accuracy is often lowest at SOAs that are greater than zero, before it begins to improve at even longer SOAs (Bachmann & Allik, 1976; Michaels & Turvey, 1979). An illustration of how perception varies with mask SOA is shown in Figure 2.

In addition to being distinguishable by their temporal signatures, pattern masking processes are dissociable by manipulations of stimulus properties (which influence integration masking) and informational attributes (which influence interruption masking). For example, increasing the luminance contrast of the target diminishes the integration masking effect, whereas increasing the contrast of the mask increases it. At the same time, manipulations of contrast have little if any effect on interruption masking (Breitmeyer, 1984; Spencer & Shuntich, 1973). Conversely, presenting a varying number of potential target items from trial to trial (i.e., a manipulation of set size) has very little effect on masking by integration, although it works powerfully to increase masking by interruption (Breitmeyer, 1984; Spencer & Shuntich, 1973; Turvey, 1973).

Metacontrast masking occurs when a target shape is displayed that fits closely around

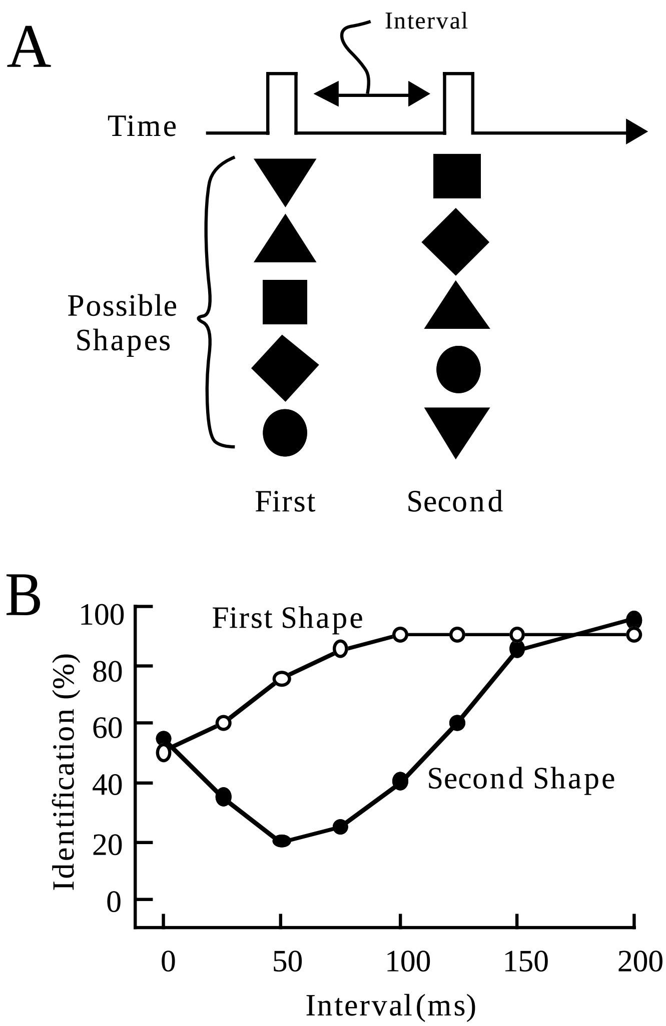


Figure 2. Backward pattern masking is often U- or J-shaped, referring to the finding that the largest influence of the mask occurs when the mask follows the target with a stimulus onset asychrony (SOA) of 100-150 milliseconds. (A). Two shapes are presented briefly in the same visual location, separated by a variable interval. (B). Identification accuracy for the first (open symbols) and second (closed symbols) as a function of interstimulus interval. Redrawn based on Bachmann & Allik, 1976.

the contours of one or more masking shapes (Alpern, 1953; Breitmeyer, 1984; Werner, 1953). Importantly, it only occurs within a narrow band of temporal intervals. When the interval between the brief presentations of target and mask is either very brief or very long, the target is clearly visible. At intermediate SOAs, however, perception of the target is impaired, leading to a U-shaped function of accuracy over SOA. The main mechanisms thought to be at work here involve inhibitory interactions between neurons representing the contours of the target and the mask (Breitmeyer & Ganz, 1976; Macknik & Livingstone, 1998; Weisstein, Ozog & Szoc, 1975 ). The main idea in these two-channel theories is that the onset of each shape initiates neural activity in two channels; one that is fast acting but short lived, the other slower acting but longer lasting. The faster channel signals stimulus onset; while the slower channel contains information about shape and color. Metacontrast masking occurs when the faster acting signals from the mask onset inhibit the activity of slower signals carrying information about the earlier target. In addition to the temporal aspects of metacontrast, a key piece of evidence consistent with this theory concerns the exquisite relationship between masking strength and contour proximity (Breitmeyer, 1984; Growney, Weisstein & Cox, 1977). Target visibility is sharply increased as the proximity between target and mask contours is increased, with little masking occuring beyond separations of 0.3 degrees near the fovea. The same contour proximity relations exist outside the fovea, although they appear scaled to the decreased spatial acuity that occurs with eccentricity.

## FINDINGS THAT DO NOT FIT THE STANDARD VIEW

Although these standard views are able to account for a vast portion of the empirical data on visual masking, there are several persistent findings that complicate the picture. Consider first the perceptual fate of masked targets. In the standard view, backward masking by pattern is said to terminate the processing of the target at a precategorical level (Kolers, 1968; Turvey, 1973). However, a phenomenon known as *masked priming* indicates that, despite the masking, processing of the target continues to lexical and even semantic levels. In typical priming experiments, a target is identified more easily if it is preceded by a semantically-related prime word (Cheesman & Merikle, 1986; Meyer, Schvaneveldt, & Ruddy, 1975). In masked priming, the prime word is followed by a spatially superposed pattern so that observers cannot report it. Yet, the same facilitation is found in masked priming as in typical priming (Debner & Jacoby, 1994; Marcel, 1983a; 1983b; Cheesman & Merikle, 1986). Masked priming has also been obtained when primes and targets are inserted in a rapid visual stream of distractor items that act as masks (Shapiro, Driver, Ward & Sorensen, 1997). Clearly then, backward masking does not interrupt target processing at an early level, as claimed by extant theories. What appears to be disrupted is not the visual analysis of a masked target, but the access to this analysis by conscious processes. Despite being unavailable to consciousness, masked targets nonetheless influence events at advanced stages of processing.

A second finding that is difficult to reconcile with the standard view is that there is little evidence of suppression of the target signal by a backward mask when direct neurophysiological measures are taken. An example involving pattern masking comes from a study of event-related potentials (ERPs) in which observers tried to detect words embedded in a rapid visual stream of random letter strings (Luck, Vogel & Shapiro, 1996; Vogel, Luck & Shapiro, 1998). In this case, each subsequent letter string acts as a backward mask for the previous item, serving to sharply reduce accuracy for the target word in some conditions. Yet many of the ERP signals associated with the target word (e.g., P1, N1, N400) were indistinguishable for the conditions in which behavioral masking did and did not occur. The one ERP signal that did show evidence of target suppression was the P3 component, a signal widely believed to reflect the conscious contents of working memory.

A similar story applies to metacontrast masking. Single-unit microelectrode recordings from cat and monkey visual cortex show that masking is associated with a reduction in peak responses occurring only beyond 80 ms and as late as 400 ms after stimulus onset (Bridgeman, 1975; 1980; 1988; von der Heydt, Friedman, Zhou, Komatsu, Hanazawa & Murakami, 1997). Contrary to expectations from two-channel theories, early visual components are affected minimally, if at all. Homologous results have been obtained for ERP recordings in humans (Jeffreys & Musselwhite, 1986; Vaughn, 1969). At the same time, optical imaging techniques show that cells in area V1 in monkey that are positioned to signal a target shape are suppressed when closely surrounded by masking shapes that lead to perceptual masking (Macknik & Haglund, 1999). This leaves us with the paradox that while some measures indicate no suppression of the the early visual signal from a masked target, other suggest that the earliest stages of visual processing are indeed influenced by a masking stimulus. As we will see, this paradox is resolved by the idea of cortical multiplexing (Bridgeman, 1980, Mumford, 1992), with the same neurons participating in different computations at various stages in the processing of a visual display.

Finally, current thinking is not easily accommodated to the notion that spatial attention plays an important role in visual masking. We mentioned earlier that one of the distinguishing characteristics of integration versus interruption mechanisms was their differential sensitivity to attentional manipulations such as set size (Spencer & Shuntich, 1970). Yet, there have been no theories of interruption masking in which increases in set size are predicted to result in larger masking effects. The same is true for metacontrast masking. It was reported as early as 1961 (Averbach & Coriel) that a ring used as a backward mask had no influence on the report of a single letter, while the same ring effectively masked the letter when it was accompanied by 3 other letters in the target display. However, this finding was not followed up by subsequent theoretical developments (Breitmeyer, 1984). Recently, it has also been demonstrated that metacontrast masking is modulated by the way in which the observer subjectively organizes an ambiguous display (Ramachandran & Cobb, 1995). Clearly, all forms of masking await an account in which spatial attention plays an integral role in the masking that occurs.

## FORMS OF MASKING THAT DEFY THE STANDARD VIEW

In our lab we have been exploring two new forms of masking over the past several years that are particularly difficult to assimilate into the standard views. Here we give only a brief introduction to these forms of masking in order to give the reader a flavor of the phenomena that need to be explained. Interested readers can experience these masking effects first hand through demonstrations on the world wide web (http://www.interchange.ubc.ca/enzo/osdescr.htm) and can read about them in greater detail in other papers (Di Lollo, Bischof & Dixon, 1993; Di Lollo, Enns & Rensink, in press; Enns & Di Lollo, 1997).

The first form of masking occurs when a briefly presented target is followed by four dots that surround, but that do not touch, a target shape (Enns & Di Lollo, 1997 ). Standard theories predict no masking in this case, since the four dots are neither a pattern that is superimposed on the target, nor do they contain contours that closely surround the target contours. Nonetheless, strong masking does occur. Furthermore, four dot masking is critically dependent on the target shape appearing in unpredictable locations, or on the presence of other nontargets in the display. When attention can be focused on the target location prior to target-mask sequence, no masking occurs. On the other hand, when attention cannot be focused prior to the arrival of the target, backward masking ensues. An illustration of stimuli from a four dot masking experiment is shown in Figure 1C.

The second form of masking occurs when a brief display of a metacontrast-like target shape plus mask is continued with the mask shape alone (Di Lollo, Bischof & Dixon, 1993). The temporal relations among stimuli in this form of masking are illustrated in Figure 3. We refer to this as masking by *common onset* because there is no longer a temporal interval between the onsets of the target and mask. Target and mask come into view simultaneously,

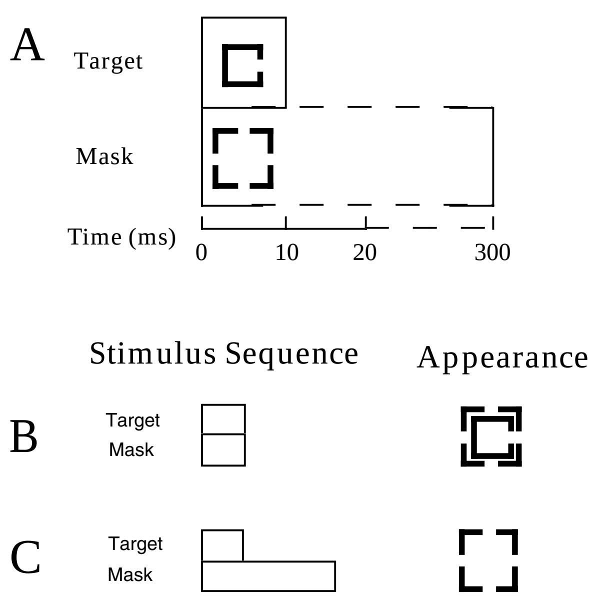


Figure 3. Schematic representation of the stimulus sequence in common-onset masking. (A) The sequence begins with a combined display of the target and the mask. After a brief period, the target is turned off, and the mask remains on view alone for various durations which can include a duration of zero. (B) When the duration of the trailing mask is equal to zero, 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 100-150 ms, the target is not seen, and the area within the mask appears empty.

followed by the termination of only those parts of the display that belong to the target. Notably, no masking occurs if the brief presentation of target and mask terminate simultaneously, indicating that the duration and contrast of the display items are sufficient to support perception. Yet, a small postponement in the termination of the mask begins to yield masking, with masking increasing in strength along with mask duration, until an asymptote is reached at a mask duration of 150 ms or so. This is clearly a form of masking that is not predicted by the previously described mechanisms of integration (which would incorrectly predict maximum masking at a mask duration of 0 ms), interruption (which depends on a second pattern to interrupt processing of the first), or metacontrast (which depends on a fast acting signal associated with the mask co-occuring with the slower signal of the target). Instead, the key ingredient for common onset masking appears to be that the mask remains on view following the termination of the target.

These two new forms of masking were recently combined in a series of experiments in which an initial brief display, consisting of several potential targets plus four small dots surrounding one of the items, was followed without interruption by a second display containing only the same four small dots. The observer s task was to report the item singled out by the four dots (Di Lollo, Enns & Rensink, in press). As shown in Figure 4, little or no

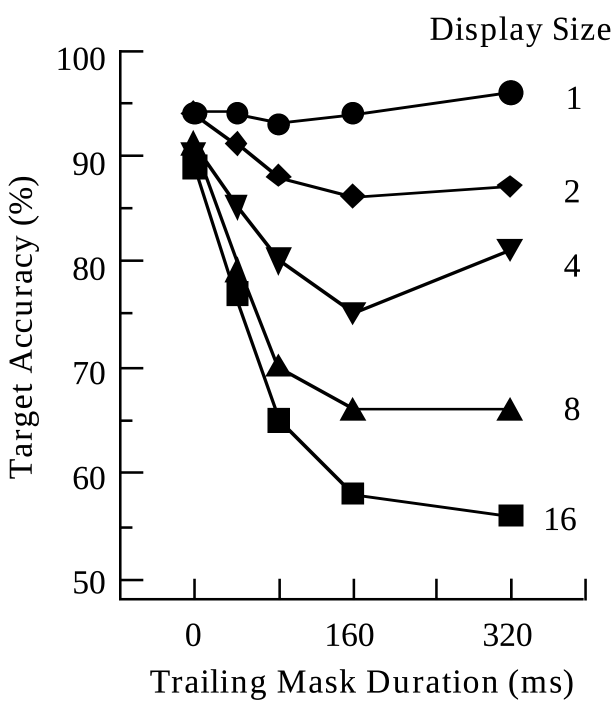


Figure 4. Target identification accuracy in a four-dot masking study (Di Lollo, Enns & Rensink, in press, Experiment 3). No masking occurs when attention can be rapidly deployed to the target location, as occurs for set size equals one. Accuracy is also affected little by increases in set size, provided that the four-dot mask terminates with the target display. However, pronounced masking occurs as both set size and mask duration are increased. According to the object substitution hypothesis, this occurs because the representation of the unattended target has been replaced by the mask representation before target identification could be completed.

masking occurred when the number of potential targets was one or when the mask terminated along with the target display. In contrast, pronounced masking occurred when several potential targets were displayed and when the mask remained on view following target display termination. Other experiments in the same study indicated that little masking occurred, even with a large display size, if the target differed from all other nontarget items by a very distinctive feature, or when the four dots preceded the target display by 90 ms. These findings indicate that common onset masking by four dots is critically dependent on the focus of spatial attention.

## AN ACCOUNT OF MASKING BASED ON CORTICAL REENTRY

Our views on masking emerged from a desire to incorporate recent advances in neuroscience into thinking on the psychophysics of masking. Our starting point was the principle that communication between two brain areas is never 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 (Damasio, 1994; Felleman & Van Essen, 1991; Zeki, 1993). Two aspects of this connectivity are especially salient for masking. First, most visual centres have reentrant connections with area V1, which is where stimulation first enters the cortex. Thus, neurons in V1 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 in higher centres. Because a receptive field in V1 is small, any given unit has no way of "knowing" whether the external stimulus is an isolated edge 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. However, an ongoing exchange of information between levels can be used to resolve this issue.

Numerous examples illustrating these principles have now been reported using single-unit recordings in visual cortex of cat and monkey (Hupe, James, Payne, Lomber, Girard & Bullier, 1998; Sillito, Jones, Gerstein & West, 1994; Lamme, Zipser & Spekreijse, 1997). They suggest that cortical reentrant signals are used to test for the presence of specific patterns in the incoming activity. It is as though the circuits actively search for a match between a descending code representing a perceptual hypothesis and the initial pattern of low-level activity. When such a match is found, the neural ensemble is "locked" onto the stimulus. For example, in one study, responses from V1 neurons were recorded in awake monkeys fitted with chronic microelectrode implants, viewing a textured figure on a background (Lamme et al., 1997). In the first part of the study, three stages of processing were revealed in the neural response. In the first 80 ms following stimulus onset, neurons responded only to local features presented within their receptive fields. Between 80-120 ms the same neurons began to respond to figure boundaries outside of their traditionally defined receptive fields. Finally, after about 120 ms the neurons responded to the surface of the figure. The authors concluded that the neurons in area V1 participate first in local feature detection, then in high-order boundary detection, and then in figure-ground assignment. In the second part of the study direct evidence of reentrant processing was obtained after the animals underwent extensive lesioning to extrastriate cortex ipsilateral to the recording site. After surgery, the V1 activity corresponding to the first two stages was still very much in evidence, but the V1 activity corresponding to figure-ground processing was missing. Behaviourally, the animals were no longer capable of distinguishing figure from ground. This indicates that reentrant processing is critical for establishing the sensitivity of V1 neurons to global attributes of a display.

We recently incorporated these neuroscience finding into a computational model of masking (Di Lollo, Enns & Rensink, in press). The central assumption is that perception is based on the activity of three-layer modules similar to that illustrated in Figure 5, arrayed over the visual field. Each module can be conceptualized as a circuit involving the connections between cortical area V1 and a topographically related region in an extrastriate visual area. The output of each module is a representation of the spatial pattern within its receptive field.

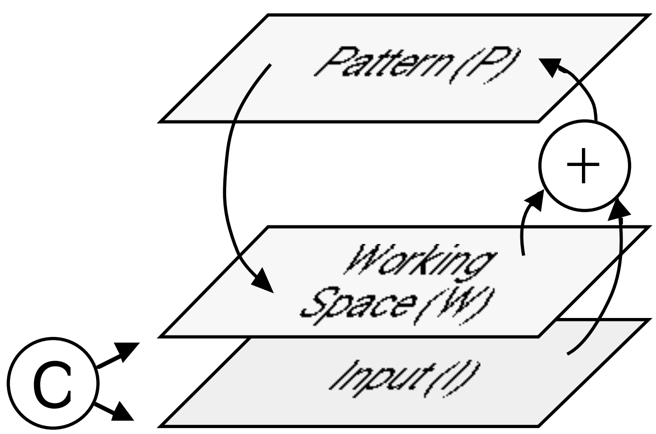


Figure 5. A computational model for object substitution (Di Lollo, Enns & Rensink, in press). A large number of three-layer modules, such as the one shown here, are arrayed over the visual field. Stimuli from the visual field arrive at the Input Layer, where the receptive fields are small, the features coded are simple, and activation decays rapidly unless maintained by continued visual input. The contents of the Input Layer are summed with the current contents of Working Space, and sent to the Pattern Layer, where the receptive fields are much larger and code for more complex patterns.

Although we do not describe the model in detail here, we summarize the behavior of the model in response to two different kinds of displays: a brief display in which the target and a four dot mask terminate together, and the identical display with the exception that the four dot mask remains on view after the target has been turned off. As shown in the data in

Figure 4 and in the demonstrations on the internet, only the latter display produces masking by four dots, which increases in strength as a function of mask duration.

The onset of a new visual event initiates activity in all layers on the first cycle. The activity in the Pattern Layer is then fed back to the Working Space by means of a simple overwriting operation. In this transfer, pattern information is translated back to the pixel-like codes of the Input Layer, permitting a direct comparison. This comparison is necessary because the initial wave sent to the Pattern Layer may have been unclear, it may activated more than one pattern, and because the spatial resolution of the Pattern Layer is inherently coarse. If the code in the Pattern Layer is to be successfully bound to its actual display location, it is necessary that the reentrant signals be placed in spatial registration with the active signals in the Input Layer.

Most important for masking is that the contents of the Input Layer change dynamically with new visual input. The contents of the Pattern Layer change more slowly because its input is a weighted sum of what is currently in the Input Layer and what was in the Working Space on the previous iteration. This produces a degree of inertia in response to changes in input that is an unavoidable consequence of reentrant processing. If the visual input changes during this critical period of inertia, masking will ensue. We refer to this process as *object substitution* because the emerging representation of a target in the Pattern Layer has been replaced by the emerging representation of the mask as the object occupying a given spatial location.

**A Brief Display in Which all Items Terminate Together**

The onset of the display triggers activity in the lower levels of the visual system, leading to the formation of a tentative shape hypothesis at higher levels. Such hypotheses require verification because the initial signal may have been unclear, it may even have activated more than one hypothesis, and because the larger receptive fields at the higher levels do not preserve precise location information. These ambiguities can be resolved by comparing the high-level codes with the ongoing low-level activity generated by the initial stimulus. Broadly conceived, these iterations of reentrant activity can be regarded as a binding process in which specific visual features are linked to the appropriate objects in space. When all items in the display terminate together, activity in the lower levels begins to subside, yielding the phenomal experience of rapidly fading visible persistence. As the common onset masking data show (see Figure 4), observers are able to identify a target very accurately when all briefly presented display items begin and terminate together.

**A Brief Display in Which the Mask Continues as a Trailing Pattern**

When the target items are turned off first, leaving only the four-dot mask in the target location, an ambiguity is created as to the identity of the target. The ongoing activity at low levels in the system now consist only of an image of the mask, despite the fact that earliest descending signals were generated by an hypothesis consistent with the target. Given this kind of a conflict, what is perceived will depend on the number of iterations required for target identification. If only a few iterations are required, the task may be completed before the signal from the target has faded completely and the four-dot percept has grown too strong. However, if more iterations are needed the probability that the "mask alone" percept will replace the "target plus mask" percept grows rapidly. At the longest mask durations, only the four dots of the mask are perceived, with the area between the dots appearing empty.

The object substitution model therefore accounts in a very natural way for the relation between masking and spatial attention. Because the main mechanism involves successive iterations of reentrant processing, 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. One of the best known ways to delay target identification is to distribute spatial attention widely over the visual field or even to explicitly misdirect attention to nontarget locations (Posner, 1980). This delay seems to reflect a fundamental limit on the number of items that can be consciously attended at any time (Pashler, 1994; Pylyshyn & Storm, 1988). Our model incorporated this aspect of perception by assuming that attention is deployed to the location of the target as a joint function of set size (Treisman & Gelade, 1980), the degree of similarity among items (Duncan & Humphreys, 1989; Wolfe, Cave & Franzel, 1989) and whether a spatial pre-cue had been presented before the onset of the search array (Eriksen, 1995).

In addition to providing a useful framework for masking by four dots and by common onset, the object substitution model can account for the standard masking effects. Indeed, seen from this perspective, there is no difference in principle between masking with common onset and the broad characteristics of metacontrast and pattern masking. All forms of backward masking will be subject to the consequences of having the representation of a temporally leading target being replaced by that of the mask if it follows closely in time and appears at the same location before target identification is complete. We fully expect that there will be minor differences unique to each form of masking, with for example, metacontrast masking having specific types of local contour interactions that are not shared by pattern masking or masking by four dots. However, the critical requisite for *object substitution* that all forms of backward masking share, is that the mask continue to be visible during the period in which the iterations between higher level pattern representations and lower level contour representations are likely to occur.

Finally, the object substitution model makes sense of a number of findings in the masking literature that until now have posed genuine puzzles for the standard view. For instance, it is now easier to understand why both pattern masking (Spencer & Shuntich, 1970) and metacontrast (Averbach & Coriel, 1961), which ostensibly have such different mechanisms, are similar when set size is varied: very little masking when the target is the only item on the display, along with pronounced masking of the same item when the target is one of several display items. As was the case for the masking by four dots (see Figure 4), by the time spatial attention has been deployed to the target location in the larger displays, only the mask item is available for conscious report.

A second result that is easily explained by the object substitution model is the somewhat curious finding that the effectiveness of a backward mask increases with mask duration (Dixon & Di Lollo, 1994). Neither the standard views of interruption masking (based on the termination of target processing) nor those of metacontrast (based on channel inhibition) predict that a mask s influence will increase with its duration. The object substitution model, however, makes precisely this prediction, since a mask of longer duration will be more likely to complete and reinforce the iterative pattern confirmation process. As the common onset masking experiments show, this prediction is borne out in the data (Di Lollo, Enns & Rensink, in press).

Third, the object substitution model predicts that a mask will do more than simply terminate target processing; the mask will itself become the new focus of object identification mechanisms. This predicted substitution has been observed quite directly in studies conducted on visual masking in rapid serial visual presentations (Chun, 1997; Martin, Isaak & Shapiro, 1995). When observers fail to correctly report the identity of a masked target, their false reports are usually for the item that follows, and therefore has masked, the target. Of course, this effect is much more difficult to observe in traditional studies of pattern and metacontrast masking because the observer is almost never asked to report directly on their perception of the mask.

Fourth, the object substitution model has a ready explanation for the large imbalance that it known to exist between forward and backward masking effects. As reviewed earlier, forward masking, when it is even observed, has a very narrow temporal window and is fully accounted for by the inherent temporal smearing of the visual system, which reduces it to the mechanisms of masking by spatial crowding and noise integration. Backward masking, on the other hand, has a much wider temporal window and is often much larger in magnitude. As we have also discussed, both of these characteristics can be easily exaggerated through manipulations of set size (Spencer & Shuntich, 1970) and target-mask similarity (Breitmeyer, 1984). This large bias favoring the effectiveness of backward masking is exactly what is predicted if the primary mechanism involves the replacement of an emerging object representation with another because the initial representation has been contradicted by subsequent input.

Fifth, it is important to note that object substitution is agnostic with respect to how conscious perceptual processes differ from non-conscious ones. Previous theories of masking made the prediction that non-conscious priming should not occur, because the mechanisms of masking are at a level prior to complete target identification. Object substitution is not burdened by such constraints, as it merely assumes that attention is space- and time-limited. If the sensory input has changed before attention can be devoted to the target item, then masking will occur, meaning that the identity of the target will not be consciously accessible. However, this in no sense rules out the possibility that mechanisms not available to consciousness may have processed the target to the extent that it influences other processes.

## OBJECT SUBSTITUTION AND VISUAL ATTENTION

One of the most important practical implications to arise from our understanding of object substitution is that backward visual masking does not merely terminate the processing of a target display. Rather the perceptual processes associated with conscious perception will now be actively engaged in perception of the mask. In addition to helping us understand why masks that are visually similar to the target are generally the most effective (Breitmeyer, 1984), this observation makes it clear that the characteristics of the mask chosen in a given experiment will themselves play an important role in the determination of task performance. As such, if a mask is being used to control task difficulty, one should always consider how the mask itself may be influencing performance.

An example of this cautionary principle can be seen in recent research on the *attentional blink,* in which perception of the second of two briefly presented targets is impaired if it is presented with a temporal lag of up to 500 ms after the first target (Shapiro, 1994). For example, observers may be asked to report the identity of two letters inserted into a visual stream of digits. Although accuracy of report for the first letter is nearly perfect, accuracy in identifying the second letter is reduced substantially. This deficit has been attributed to the second target being unattended while processing resources are devoted to the first target. However, it has long been recognized that the second target must be masked in order for the attentional blink to occur. Ostensibly, the purpose of masking is to increase the difficulty of processing the second target, thereby bringing accuracy within a measurable range. But if this were the principal function served by masking, then either simultaneous (integration) or backward (interruption) masking should produce the same effect, and the second target deficit should be found using either procedure.

In fact, the accuracy deficit occurs in this task only if backward masking is used (Brehaut, Enns & Di Lollo, 1999; Giesbrecht & Di Lollo, 1998). If a simultaneous mask is used, identification of the second target is impaired equally across all lags, but the lag-dependent deficit that is the hallmark of the attentional blink is missing. This points to object substitution as the mechanism of masking. That is, while unattended, the second target is vulnerable to replacement by the trailing mask. As a consequence of this replacement, the mask is substituted for the second target as the object for eventual conscious registration. On the basis of these results, it is clear that backward masking of the second target is more than a methodological convenience: it reveals the workings of mechanism that would have gone undetected had masking been used merely as a tool of convenience.

We believe that masking by object substitution also has direct relevance to the recently popularized phenomena of inattentional blindness and change blindness. Consider first *inattentional blindness*, which refers to objects that are presented to the visual system but not seen because the observer is attending to something else (Mack & Rock, 1998). A typical way of studying this involves a procedure in which observers are asked to report which of two intersecting lines is longer in a brief display. These lines can be presented either at the fovea or in the periphery, and the task can be made arbitrarily difficult by varying the difference in length between the horizontal and vertical lines. After 3 or 4 trials of this task have been completed, a critical stimulus is presented unexpectedly in one quadrant of the crossed lines, about 2 degrees from the center. The observer is then asked whether they have seen anything on the screen other than the cross figure, that is, anything that has not been present on previous trials. A large variety of stimuli presented in this way go undetected by observers. Not only do observers deny seeing anything new, but they are unable to describe what it looks like when asked to guess, nor are they able to select the correct item when it is shown to them again in a recognition test.

One of the details of the inattentional blindness procedure that has not been given much consideration is the role played by a pattern mask, which is presented immediately following the display of the intersecting lines and remains on view for 1.5 s. The authors use the mask for the conventional purpose of preventing any additional processing of the display after it is removed from the screen. However, the object substitution hypothesis predicts that it is perception of the mask that interferes directly with access to the briefly presented and undetected targets. More specifically, it predicts that inattentional blindness will increase directly with the duration of the mask.

*Change blindness* refers to a phenomenon in which large changes to the visual world go undetected if attention is not already focused on the objects or area in which the change occurs. Some of the earliest reports involved observers who did not notice changes made during an eye movement while inspecting a photograph (e.g., a switch in the hats worn by two gentlemen), although these same changes were easily detected when they occurred during a fixation (Grimes, 1996; McConkie & Currie, 1996). Subsequent reports indicated that similar results could be obtained if the changes occurred during a brief visual interruption in the scene (Rensink, 2000; Rensink, O Regan & Clark, 1997), if the changes occurred during a cut in a movie (Levin & Simons, 1997; Simons, 1996), and even if they occurred during a real-world conversation between an unwitting participant and an actor. In this case it was the actor who exchanged places with another actor when a door being carried by other actors briefly interrupted the conversation (Simons & Levin, 1998).

In each of these cases of change blindness, it is clear that the focus of attention is an important predictor. Changes are detected more readily when they occur to objects that are of interest to the observer (Rensink, O Regan & Clark, 1997; Rensink, in press), and when they occur in locations that have had attention drawn to them by a salient cue such as local visual transient or a unique color (Rensink, O Regan & Clark, 1997; Scholl, 2000). However, there has been less systematic study so far on the role played by the visual image that replaces the retinal (or environmental) location of the original image. From the perspective of object substitution, the important role played by this new image is that it prevents access to the fading representation of the original image. Instead, ongoing iterative perceptual processes are now busy confirming that the newly presented image does indeed contain the patterns and colors hypothesized by recent input into the system. Only when attention is focused on an object prior to the changed image are conscious visual processes able to detect the difference.

Finally, a word of caution. Although in the foregoing examples of attentional blink, inattentional blindness, and change blindness, a backward visual mask is a critical factor in causing a failure to see objects that are clearly registered on the retina, it is important to be clear that the mechanism being proposed is not inherently tied to masking. Instead, backward masking is a methodological window that gives a glimpse into the mechanisms of iterative reentrant processing and the limited ability of conscious perception to monitor the perceptual products of these processes (Di Lollo, Enns & Rensink, in press). In support of this view, there have been recent reports in each of these areas of research, showing that failures of perception can be induced even in the absence of a mask. For instance, an attentional blink is observed without masking if the nature of the perceptual task is sufficiently different for the first and second targets (Enns, Visser, Kawahara & Di Lollo, in press), inattentional blindness has been reported when the unexpected and unattended stimulus is not followed by a mask (Mack & Rock, 1998); and change blindness can be induced by simply splashing mud unexpectedly onto parts of a picture other than the target (O Regan, Rensink & Clark, 1999). Importantly for the object substitution hypothesis, each of these manipulations involves a misdirection of attention, away from the visual target that goes undetected. While attention is focused on the task of identifying the first target in a visual stream, or has been captured by mudsplashes while viewing a scene, and before attention can be redeployed to the critical visual target, the iterative processes involving the critical target have lost all trace of second target or the change that occurred to the target object.

## BROADER IMPLICATIONS FOR VISUAL PERCEPTION

It is possible that what we have learned about masking from the study of object substitution may have wider currency in helping us to understand other aspects of visual perception. For example, a number of researchers studying the relations between visual imagery and perception have made the intriguing observation that sensory input seems to compete directly with the contents of imagination for the control of consciousness (Kosslyn &

Rabin, 1999; Farah, 1989). In one study observers were asked to recall from memory whether President Lincoln, who is depicted on the American penny, faces to the right or left (Kosslyn & Rabin, 1999). Although accuracy in this task was significantly above chance levels when this task was performed from memory alone, it did not differ from chance when observers were asked to choose from two outline drawings, one facing left, the other right. The authors interpreted this to mean that the image-based memory necessary to perform this task requires the same neural machinery as the perception of the two outlines in the recognition task. As in the case of masking by object substitution, the conscious brain has a tendency to resolve conflicts between top-down (reentrant) and bottom-up (sensory input) signals in favor of the immediate sensory input. Although such a bias is probably highly advantageous for most aspects of perception, since it permits one to see what is actually on view, it can lead to failures in being able to access visual memories; both very short-term ones in the case of backward masking, as well as longer term ones in the case of visual imagery.

We should also be humbled by the realization that what may appear to be newly uncovered principles of object substitution have been known to magicians and tricksters for thousands of years. At least they have been known in the sense that an audience can be reliably fooled if certain rules are followed. An especially good example based on object substitution involves an old card trick that has recently gained a new life on the internet. In this trick, the audience is shown six face cards and asked to choose one, but not to tell any other audience member, or the trickster, which card has been chosen. With each member of the audience having their card firmly and privately in mind, the trickster removes all six cards and says, with a flourish I have the power to read your mind. Your chosen card will now be removed from the set. Five cards are returned to view, and invariably the chosen card is not among them. The trick works very well, as evidenced by the large amount of speculation that it has generated on the internet. Much of this speculation concerns how a large audience (in this case world wide) can be coerced into chosing the same card, which appears to many to be the basis of the trick. But, as with all good tricks, the apparent basis is not the actual basis of the trick. The simple trick is that the cards in the second display are five entirely new face cards.

Why does this trick fool people? Very simply, for two reasons that are at the heart of the object substitution hypothesis. First, observers do not detect changes to objects that they have not attended. All attention in the first display is focused on the card selected for memory by the observer. Secondly, observers assume that unless otherwise informed, the visual world consists of the objects that are currently on view. It does not occur to them to doubt that the second five cards are different from the first five that were not selected. While this may be a generally useful assumption, perhaps even an evolutionary adaptation in the service of visually guided action, it does leave us vulnerable to events that violate this belief in a stable world . Such events include the fun of magic, but they also include the more serious everyday realities of traffic accidents, human-computer interaction, and eye witness testimony.

## AKNOWLEDGEMENTS

The research described in this chapter was supported by grants from the Natural Sciences and Engineering Research Council of Canada to both authors. Correspondence may be addressed to either J. T. Enns (jenns@psych.ubc.ca) or V. Di Lollo (vince.dilollo@ubc.ca), Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B. C., Canada V6T 1Z4.

## REFERENCES

Alpern, M. (1953). Metacontrast. *Journal of the Optical Society of America, 43*, 648-657.

Averbach, E., & Coriel, A. S. (1961). Short-term memory in vision. *Bell Systems Technical Journal, 40,* 309-328.

Bachmann, T., & Allik, J. (1976). Integration and interruption in the masking of form by form. *Perception, 5*, 79-97.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177-178.

Brehaut, J.C., Enns, J.T., & Di Lollo, V. (1999). Visual masking plays two roles in the attentional blink. *Perception & Psychophysics, 61*, 1436-1448.

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.

Bridgeman, B. (1975). Correlates of metacontrast in single cells of the cat visual system. *Vision Research, 15,* 91-99.

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.

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.

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology, 40,* 343-367.

Chun, M. M. (1997). Temporal binding errors are redistributed in the attentional blink.

*Perception & Psychophysics, 59,* 1191-1199.

Damasio, A. R. (1994). *Descartes error.* New York: Putnam s.

Debner, J. A., & Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 304-317.

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., Enns, J. T., & Rensink, R. A. (in press). Competition for consciousness among visual events: The psychophsyics of reentrant pathways. *Journal of Experiental Psychology: General.*

Dixon, P., & Di Lollo, V. (1994). Beyond visible persistence: An alternative account of temporal integration and segregation in visual processing. *Cognitive Psychology, 26*, 33-63.

Duncan, J. & Humphreys, G. (1989). Visual search and stimulus similarity. *Psychological Review, 96*, 433-458.

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., Visser, T. A. B., Kawahara, J., & Di Lollo, V. (in press). Visual masking and task switching in the attentional blink. Chapter to appear in Shapiro, K. (Ed.), *The limits of attention: Temporal constraints on human information processing.* Oxford University Press.

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.

Farah, M.J. (1989). The neural basis of imagery. *Trends in Neurosciences, 12,* 395-399.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cerebral Cortex, 1*, 1-47.

Ganz, L. (1975). Temporal factors in visual perception. In E. C. Carterette & M. P. Friedman (Eds.) *Handbook of Perception, Vol. 5*, (pp. 169-231). New York: Academic Press.

Giesbrecht, B. L., & Di Lollo, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1454-1466.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.). *Perception* (pp. 89-110). New York: Oxford University Press.

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.

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.

Jeffreys, D. A., & Musselwhite, M. J. (1986). A visual evoked potential study of metacontrast masking. *Vision Research, 26,* 631-642.

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.* Boston, MIT Press.

Kosslyn, S. M., & Rabin, C. (1999). The representation of left-right orientation: A dissociation between imagery and perceptual recognition. *Visual Cognition, 6,* 497-508.

Lamme, V. A. F., Zipser, K., & Spekreijse, H. (1997). Figure-ground signals in V1 depend on extrastriate feedback. *Investigative Ophthalmology & Visual Science, 38,* S969 (abstract).

Levin, D.T., & Simons, D.J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review, 4,* 501-506.

Luck, SJ, Vogel, E.K., & Shapiro, KL (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383,* 616-383.

Mack, A., & Rock, I. (1998). *Inattentional Blindness.* The MIT Press, Cambridge, MA.

Macknik, S.L., & Haglund, M.M. (1999). Optical images of visible and invisible percepts in the primary visual cortex of primates. *Proc. National Acad. Sciences, 96,* 15208-15210.

Macknik, SL & Livingstone, MS (1998) Neuronal correlates of visibility and invisibility in the primate visual system. *Nature Neuroscience, 1,*144-149.

Marcel, A. J. (1983a). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197-237.

Marcel, A. J. (1983b). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology, 15,* 238-300.

Martin, J., Isaak, M. I., & Shapiro, K. L. (1995). *Probe identification errors support an interference model of the attentional blink in rapid serial visual presentation.* Poster presented at the Annual Meeting of the American Psychological Society, New York, NY.

McConkie, G.W., & Currie, C.B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 563-581.

Meyer, D. E., Schveneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98-118). New York: Academic Press.

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.

Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biological Cybernetics, 66,* 241-251.

O'Regan J.K., Rensink R.A., Clark J.J. (1999). Change-blindness as a result of 'mudsplashes'.*Nature, 398,* 34.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 16*, 220-224.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3-25.

Pylyshyn, Z., & Storm, R.W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179-197.

Ramachandran, V. I., & Cobb, S. (1995). Visual attention modulates metacontrast masking. *Nature, 373,* 66-68.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7,* 345-376.

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.

Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung, 36,* 71-93.

Scholl, B.J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition, 7,* 377-396.

Shapiro, K. L., Driver, J., Ward, R., & Sorensen, R. E. (1997). Priming from the attentional blink: A failure to extract visual tokens but not visual types. *Psychological Science, 8*, 95-100.

Shapiro, K.L. (1994). The attentional blink: The Brain's eyeblink. *Current Directions in Psychology, 3,* 86-89.

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.

Simons, D.J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301-305.

Simons, D.J., & Levin, D.T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review, 5,* 644-649.

Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology, 85,* 198-203.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

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.

Vaughn, H. H., Jr. (1969). The relationship of brain activity to scalp recordings of event-related potentials. In E. Donchin and D. Lindsley (Eds.), *Average evoked potentials,* (pp. 45-94), NASA: Washington, D. C.

Vogel, E.K., Luck, SJ, & Shapiro, KL (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1656-1674.

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.

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.

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell.