Available online at www.sciencedirect.com
SCIENCE DIRECT®
Vision Research 44 (2004) 1321–1331
Vision Research
www.elsevier.com/locate/visres

Object substitution and its relation to other forms of visual masking

James T. Enns *

Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4

Received 15 July 2003; received in revised form 30 October 2003

Abstract

Three experiments compared letter identification accuracy over a wide range of target-mask intervals and mask types, including metacontrast, random dot noise, four surrounding dots, digits and letters. These comparisons were motivated by *object substitution theory* which makes three general predictions about visual masking: (1) very different looking backward masks will be equivalent in their effects when spatial attention is distributed, such that target identification is delayed, (2) masks will differ most in their effects on target identification when they are temporally integrated with the target, and (3) backward masking will be minimized when attention can be pre-focused on the spatial location of the target and the mask does not interfere with target identification. Results strongly supported the predictions and pointed to a new understanding of masking based on the separate processes of *object formation* and *object substitution*.

1. Introduction

Masking by object substitution is a term that has been coined to describe several features of visual masking that are difficult to explain by standard theories (Di Lollo, Enns, & Rensink, 2000; Enns & Di Lollo, 1997, 2000). The standard theories are based on one or more of the following ideas: (1) the visual integration and therefore perceptual confusion of events occurring in close spatiotemporal proximity (Di Lollo, 1980; Kahneman, 1968; Turvey, 1973), (2) the masking of stimuli by interruption of processing (Kolers, 1968; Michaels & Turvey, 1979; Spencer & Shuntich, 1970; Turvey, 1973), and (3) the masking of visual patterns by competitive neural interactions (Breitmeyer & Ganz, 1976; Keysers & Perrett, 2002; Weisstein, Ozog, & Szoc, 1975).

Four features of masking have been singled out as especially difficult to explain within these standard theories. First, local contour interactions between target and mask are not required for profound masking to occur. For example, when a briefly presented target shape is followed by four dots that surround the target, but that do not touch it, masking occurs that is comparable to that obtained with a snugly fitting frame, the standard metacontrast masking shape (Enns & Di Lollo, 1997). Furthermore, unlike the masking obtained with a metacontrast frame, masking by four dots is surprisingly immune to the spatial proximity between the contours of the target and those of the four dots (Di Lollo et al., 2000). The four dots are simply too inconsequential as contours, both in their intensity and proximity, to play the role that has been proposed for the mask in previous theories of metacontrast masking by channel interaction.

Second, masking by four dots is strongly modulated by spatial attention. When attention can be focused on the target location before the mask arrives little, if any, masking occurs. Yet, if the same stimulus sequence occurs unpredictably in one of three locations, masking occurs that is comparable in strength to that obtained with a snugly fitting metacontrast frame (Enns & Di Lollo, 1997). When the effects of stimulus discriminability and spatial pre-cuing were compared directly (Di Lollo et al., 2000), both factors were shown to reduce the degree of masking. In fact, the benefit of increased target visibility gained from a spatial location cue preceding the target by 100 milliseconds was similar to the benefit gained from reducing the visual similarity between the target and non-target shapes to that of a highly distinctive pop-out feature. Thus, although spatial attention plays no role in standard theories of visual masking, it is clearly critical to this new form of masking.

* Tel.: +1-604-822-6634; fax: +1-604-822-6923.
E-mail address: jenns@psych.ubc.ca (J.T. Enns).

0042-6989/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2003.10.024

Third, the temporal conditions under which masking by four dots can occur are strikingly different from those proposed as critical for standard theories of masking. In the strongest violation of these predicted temporal relations, the masking stimulus (either four dots or the metacontrast frame) was presented on the screen for the same brief time (15–45 ms) as the target shapes to be identified (Bischof & Di Lollo, 1995; Di Lollo et al., 2000). When both of these patterns were terminated together target identification accuracy was very high. However, when the mask continued to be on view after the target display, deficits in target accuracy began to emerge. This masking became asymptotically stronger as a function of the duration of the mask following target termination. This temporal relation runs counter to standard theories of metacontrast masking, where the importance of a stimulus onset asynchrony between target and mask has been called a 'law' (Breitmeyer & Ganz, 1976; Kahneman, 1968), which states that maximum masking occurs when there is a positive stimulus onset asynchrony between target and mask. This temporal relation is also counter to standard theories of pattern masking, which are designed specifically to predict maximal masking when there is a short temporal offset between target and mask (Kolers, 1968; Turvey, 1973).

The importance of continued mask duration has been recently emphasized in a study in which metacontrast mask duration and intensity were systematically varied (von Muehlenen, Enns, & Di Lollo, 2003). Despite the fact that the luminance contrast of the mask was made progressively weaker as mask duration was increased, masking continued to increase with mask duration, regardless of mask contrast. This result undermines all theories premised on critical temporal relations between target and mask transients, be they onsets of offsets, because low contrast masks have diminishingly small transients at their onset or offset.

Finally, the perceptual status of the target and the mask shapes as individual 'objects' has been shown to be critical in visual masking. For example, the strict spatial superposition of the four masking dots and the target is not necessary for masking to occur, provided that apparent motion is used to create the perception that the target and mask belong to the same object (Lleras & Moore, 2003). Another study used motion and color to segregate the target from the four masking dots, thereby sharply reducing the degree of masking that was observed (Moore & Lleras, in press). Other studies have shown that factors governing the perceptual organization of targets and mask play an important role in metacontrast masking (Enns, 2002; Ramachandran & Cobb, 1995).

Taken together, these findings call for a new theoretical framework for visual masking. Object substitution theory begins with the general premise that all of perception is the consequence of ongoing recurrent communication between neurons at lower- and higher levels of processing (Di Lollo et al., 2000). Initial sensory input from a new scene activates the spatially local and geometrically simple receptive fields of lower-level units, which, in the so-called *feedforward sweep*, activate units at higher levels that are sensitive over larger regions of the visual field and are tuned to more complex properties. In order to resolve ambiguity between alternative pattern activations at the higher level, and in order to bind patterns at the higher level to specific spatiotemporal locations, one or more *feedback sweeps* are required. Pattern hypotheses generated at the higher level are compared with the ongoing activity at the lower level. If the visual image remains stable over the iterations required to match the contents of these two levels to some criterion, conscious perception of the stimulus will ensue. However, if the input activity is altered before these iterations are complete, a mismatch will be detected and the iterative processes will begin again, this time based only on the sensory input that is currently activating the lower-level neurons.

Visual backward masking occurs, in this view, because of the reentrant checking that is required between higher and lower levels. When the target shape and the mask are presented briefly and terminated simultaneously, there is no inconsistent input from the lower levels. Target identification can be conducted to the extent that ongoing neural activity (i.e., fading visible persistence) allows reentrant checking to be completed. On the other hand, when the mask lingers beyond the target, the sensory information at the lower levels is no longer consistent with any of the target hypotheses initially suggested at the higher levels. If the target has not been identified by the time only the mask remains, processing will focus on the mask, which occupies the spatiotemporal position formerly occupied by the target. Note that in this view, the critical role of attention in masking concerns the speed with which the target can be identified. If attention is already focused on the target location prior to the onset of the display, it will speed target identification. If, on the other hand, spatial attention is misdirected or diffusely distributed prior to display onset, there will be a delay in the onset of target identification, leaving it more vulnerable to masking.

Object substitution theory does a reasonably good job of explaining the four empirical features of masking that are difficult to explain with the standard theories of integration, interruption, and competitive channel interactions (Enns & Di Lollo, 1997, 2000). The present study addresses the converse issue, namely, whether this framework can also account for masking effects obtained with conventional paradigms. According to the reentrant hypothesis, 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 emerging representation of a temporally leading target will be replaced in consciousness by that of the mask if it follows the target closely in time and appears before target identification is complete. However, it is also possible that there will also be differences in each form of masking, with for example, metacontrast masking producing specific types of contour interactions that are not shared by pattern masking or masking by four dots.

The approach taken to this question in the present study involved a systematic comparison of the effects of the four dot mask with the more 'classic' masks used in metacontrast, noise, and pattern masking. To accomplish this, a simple letter identification task and identical target-mask sequences were adopted. Three critical predictions were tested. First, under conditions of spatially distributed attention, the reentrant hypothesis predicts that all forms of backward masking will have, at a first approximation, an equal effect on target accuracy. This prediction derives directly from the idea that the contents of the mask will replace those of the target if it has not been identified prior to its replacement on the screen by the mask.

A second prediction is that the differences among masks will be most apparent when the target and mask are in close spatiotemporal proximity and therefore become temporally integrated with one another (Di Lollo, 1980). Under these conditions, the problems of target identification will not concern those of object substitution. Rather, they will involve problems of camouflage, in the case of noise and pattern masking, and local contour interactions, in the case of metacontrast masking. The problems of camouflage can be seen, in a sense, as an artifact of the task given to observers under these conditions. The formal task requirements are to identify as the 'target' only one aspect or component of the target-mask 'object' that has been perceptually fused, because of the limited temporal resolution of the visual system. The problems of local contour interactions can also be understood within the reentrant framework (Di Lollo et al., 2000; Enns, 2002; von Muehlenen et al., 2003), but critically, these 'fast loop' iterative processes are proposed to occur over a much shorter time scale than the 'slow loops' relevant to spatial attention and object substitution.

The third prediction is that backward masking will be minimized when attention can be pre-focused on the spatial location of the target. This prediction follows directly from the idea that if target identification can be completed before only the mask remains on view, object substitution will not occur. The problems of camouflage and of local contour interactions, on the other hand, that are thought to occur within the temporal integration window, are predicted not to benefit in the same way from prior attentional focus.

These three predictions were tested in this study. Experiment 1 tested the first two predictions by comparing target accuracy for six different masking conditions. In each case the mask itself acted as the target 'probe,' indicating to the observer which letter was to be identified. The distribution of attention was manipulated by varying the potential number of targets randomly between 1, 4, and 7 items. The temporal relation between target and mask was varied from −150 ms (mask before target) to +600 ms (target before mask). This meant that for all the positive temporal intervals, the letter to be identified was only indicated after the mask had been presented, preventing spatial attention from focusing on the target prior to the arrival of the mask. The next two experiments tested the third prediction, concerning focused attention, by placing a spatial cue in the target location either simultaneously with the target display (Experiment 2) or 100 ms prior to its arrival (Experiment 3).

2. Experiment 1: partial report

Observers identified the letter indicated by the mask on each trial and were asked to guess when they were uncertain. An example display from the 7-letter condition is illustrated in Fig. 1a. Fig. 1b shows the six

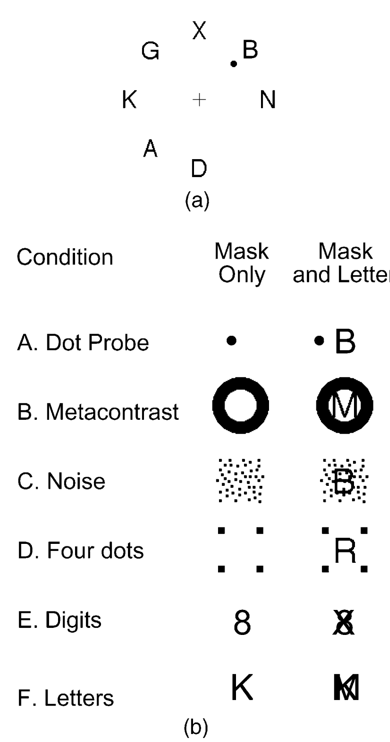

Fig. 1. (a) Example target display. The cross at the center indicates the fixation point; the black dot indicates the target letter to be reported, in this case the letter B. The target letter was preceded by, presented concurrently with, or followed by, one of the masks shown in (b).

different masking conditions that were compared. A dot probe condition (no mask) was included in this experiment to provide a baseline measure of iconic memory with as little inference from any form of masking as possible.

2.1. Method

Observers. Between 10 and 12 different observers were tested in each condition of the experiment. Observers were recruited from the UBC Human Subject pool and were given extra course credit in exchange for their participation. All reported normal or corrected-to-normal visual acuity.

Stimulus displays. Displays were presented on an Applevision monitor controlled by a Macintosh computer running VScope software (Enns & Rensink, 1992). The background screen was white (all pixels lit) and the letters and masks were black (no pixels lit). The target display consisted of 1, 4, or 7 letters (drawn from the set of 26 uppercase letters in Helvetica Font, excluding I, O, P, Q, and S) that could appear in any of eight positions. The letters were positioned at one of eight equally spaced positions on the circumference of an imaginary circle, 3° of visual angle in radius from the fixation point. Each letter subtended approximately $1 \times 0.67°$. The six different conditions tested in this experiment differed only in the nature of the probe (condition A) or mask (conditions B–F) that was presented in the same spatial location as the target letter to be identified.

For condition A, the dot probe was a black disc, 0.25° in diameter that was presented on the radius connecting the fixation point to the target letter, 2° toward the center. For condition B, the mask was an annulus, 0.5° wide with a 1° opening through which the letters were completely visible. For condition C, the mask consisted of 50 dots, each 0.10° in diameter that were sprinkled randomly on a 2° square. Condition D consisted of four dots, 0.15° in diameter, positioned at the corners of virtual square of 2°. The mask in condition E consisted of one of the digits between 2 and 8, inclusive, drawn in the same font as the letters. The mask in condition F consisted of a letter in the set B, C, E, F, G, H, K.

Procedure. Participants were instructed to identify the letter denoted by the probe dot (condition A) or the mask (conditions B–F) and to guess when uncertain. Each trial began with the fixation point presented for 500 ms. The offset of the fixation point coincided with the onset of the target-mask sequence. Both the target and the mask displays were presented for 30 ms and were separated by intervals of −150, −50, 0, +50, +150, 300 and 600 ms. The only exception was condition F, where intervals of +30, +90, +150, 300 and 600 ms were tested instead, because the nature of the task ("report the first letter in the sequence") made it impossible to test negative or zero letter-mask intervals. Observers viewed the displays with their heads in a chin rest, 57 cm from the screen. Each observer was tested on a total of 600 trials in each condition, separated into 10 blocks of 60 trials. Observers were encouraged to take short breaks of 1–2 min between blocks of trials.

2.2. Results

Mean proportion target accuracy is shown in Fig. 2, separately for the dot probe condition (Fig. 2A) and each of the five masking conditions (Fig. 2B–F). The data in each condition were examined with a repeated measures analysis of variance (ANOVA) in which the factors were display size (1, 4, 7 letters) and target-mask interval (−150, −50, 0, 50, 150, 300 and 600 ms). The only exception was in the letter masking condition, where only positive target-mask the intervals were tested. Follow up tests regarding specific hypotheses were conducted using standard simple effects procedures (Keppel, 1991). All effects described and discussed in this report were significant at $p < 0.05$.

Dot probe baseline. As shown in Fig. 2A, accuracy was influenced significantly by the interaction of display size and interval, $p < 0.001$. For intervals of −150 to 50 ms, accuracy varied only slightly with display size (mean difference in accuracy under 10%, $p < 0.05$) and did not vary with interval ($p > 0.05$), whereas for the remaining intervals accuracy declined much more rapidly for larger Displays Sizes than smaller ones ($p < 0.001$). The accuracy level in the largest display, at the longest interval, was referenced with a dashed line in Fig. 2A

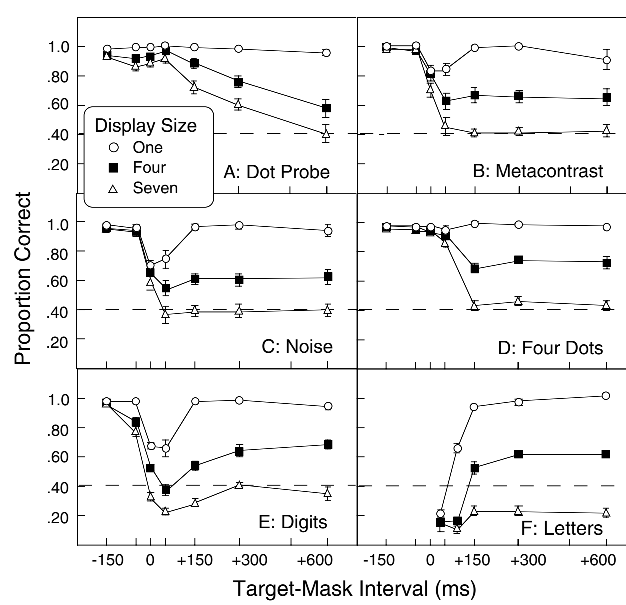

Fig. 2. Mean letter identification accuracy in Experiment 1: partial report. (A) Dot probe baseline, (B) metacontrast, (C) noise, (D) four dots, (E) digits, and (F) letters (error bars = 1 SE).

and the remaining figures, in order to facilitate comparison between masking conditions. This dashed line is not the chance-guessing rate, which is near 5% in this task (1/21). Instead, it represents the identification accuracy that is possible from the fading trace of the icon after 600 ms have elapsed, based on a 7-letter display.

Metacontrast mask. Fig. 2B shows that accuracy was influenced jointly by display size and interval, $p < 0.001$. In contrast to the dot probe condition, accuracy was near the baseline level (dashed line) for display size = 7 when the mask followed the display by 50 ms or more. A second contrast with the dot probe condition was evident in display size = 1, where the metacontrast mask produced significant masking (15%, $p < 0.01$) at intervals of 0 and 50 ms before regaining the baseline high level of accuracy ($p < 0.01$).

Noise mask. Fig. 2C shows that accuracy was influenced jointly by display size and interval, $p < 0.001$. Similar to metacontrast masking (Fig. 2B), accuracy for display size = 7 was at baseline levels when the mask followed the display by 50 ms or more. Accuracy in Display Size = 1 was also reduced for intervals of 0 and 50 ms (by 20%, $p < 0.01$) before regaining high levels of accuracy at longer intervals ($p < 0.01$).

Four dot mask. Fig. 2D shows that accuracy was very high and almost unaffected by display size ($p > 0.05$) until the target-mask interval was 150 ms. At that point, accuracy for display size = 1 remained unaffected by the mask ($p > 0.05$) while accuracy for display size = 7 was reduced to baseline 'iconic' levels ($p < 0.001$).

Digit mask. Fig. 2E shows that digits serving as masks reduced accuracy significantly for display size = 1 at target-mask intervals of 0 and 50 ms (by 25%, $p < 0.01$). For the larger Display Sizes the digit mask reduced accuracy even more at these intervals, driving accuracy to the 'iconic' baseline level for display size 4 at 50 ms and even below the baseline level for display size 8 at intervals between 0 and 150 ms ($p < 0.01$). Accuracy levels at the longest target-mask intervals were very similar to those of all the other masks.

Letter mask. Fig. 2F shows that letters acting as masks yielded the lowest levels of accuracy in the experiment. Accuracy in display size 8 hovered around the 15–20% range and never reached the baseline 'iconic' level for any interval ($p < 0.01$). In contrast, accuracy for display size 1 reached high baseline levels of accuracy by 150 ms and accuracy for display size 4 did the same by 300 ms.

2.3. Discussion

This experiment tested the prediction, derived from object substitution theory, that all forms of backward masking would be similar in their effects at positive target-mask intervals, provided that attention was distributed. This prediction derives from the general idea that perception of the mask will replace perception of the target if only the mask is found to be on view prior to complete identification of the target. An examination of accuracy in the no-mask baseline condition (Fig. 2A) and the five different masks that were tested (Fig. 2B–F) shows that this prediction was confirmed. When spatial attention was distributed most widely prior to target identification (display size 7), and there was a relatively long interval between target and mask (intervals of +150 to +600 ms), the influence of most of the masks was very similar. Target accuracy was reduced to the same low levels as those obtained when a simple dot was used to probe the visual representations of the display (Fig. 2A). The effect of the masks was to reduce accuracy to this low level at much shorter intervals than without a mask.

However, there were two notable exceptions to this general trend of equal backward masking for all mask types. First, unlike all the other masks, which had their maximum influence at an interval of 50 ms and beyond, the four dot mask had its full effect only at intervals of 150 ms and longer. This suggests that although backward masks are equal in their effects at intervals of 100 ms or more, they contain important differences in their effects at shorter intervals. This suggests that all masks other than the four dots have at least two components to their influence on target identification: an early or fast-acting component associated with object formation and a later or slower-acting component associated with object substitution.

A second deviation from this general pattern was seen in two of the mask types, digits and letters, which reduced target accuracy much more severely than the simple decay of information from iconic memory. Although this appears superficially to be a violation of the prediction derived from reentrant processing, a closer look reveals that the specific errors made in these two cases are actually consistent with it. Recall that the theory states that if only the mask remains on view prior to the complete identification of the target, processing will become focused on the mask. This means that if the mask itself activates target-relevant features or properties, these features may come to control the observer's response. This is exactly what happened. An examination of the responses made when targets were incorrectly identified indicated that many of them were related to the target-relevant features in the mask rather than being random. For digit masks this meant that, for example, that the digit 4 led to target responses that were visually similar to the digit (e.g., A); for letter masks, the identity of the mask was often incorrectly reported as the target letter. In both cases the mask seemed to be replacing the target as the object of conscious report by the observer.

The second main prediction tested in this experiment was that differences among masks would be most evident when the target and mask were in close temporal proximity and when attention was focused on a single target location. This is because it is in the range of 50–100 ms that issues of temporal integration and local contour interaction come into play that are little influenced by the distribution of spatial attention (Di Lollo et al., 2000). This prediction was also clearly confirmed. All of the masks, with the exception of the four dots, yielded significant reductions in accuracy at target-mask intervals of 0–50 ms. Some of these reductions in accuracy were larger than others, with digit and letter masks resulting in the largest impairments. This is consistent with idea that target identification in these intervals requires the 'breaking of camouflage' that has occurred through temporal integration (noise, digit, letter masks) and local contour competition (metacontrast). On this account, the strength of masking should be related directly to the degree of similarity between targets and masks and this is consistent with what was found. On a continuum of shape similarity, letter targets should be most confusable with other letters, somewhat less confusable with digits, and least confusable with random noise dots.

3. Experiment 2: simultaneous cue

This experiment was designed to test the third prediction of object substitution theory, namely, that with focused attention on the target location backward masking of all kinds would minimized. This was accomplished by repeating the conditions of Experiment 1 with the addition of one simple detail. Along with the onset of the target display, a spatial cue was presented to indicate the target location. This spatial cue eliminated the need for observers to remember the display until the mask appeared, as in Experiment 1. Instead, it allowed observers to begin target identification immediately with the presentation of the letter display.

3.1. Method

The method was identical to Experiment 1, except that a spatial cue was presented along with target letter on each trial. This cue was the same dot used to indicate the target letter in the Dot Probe baseline condition in Experiment 1 (Fig. 1B) and it appeared 1° from the target on the interior radius of the circular display.

3.2. Results

Mean proportion target accuracy is shown in Fig. 3, separately for each of the five masking conditions. The data were analyzed in the same way as Experiment 1.

Metacontrast mask. Fig. 3B shows that accuracy was influenced jointly by display size and interval, $p < 0.001$.

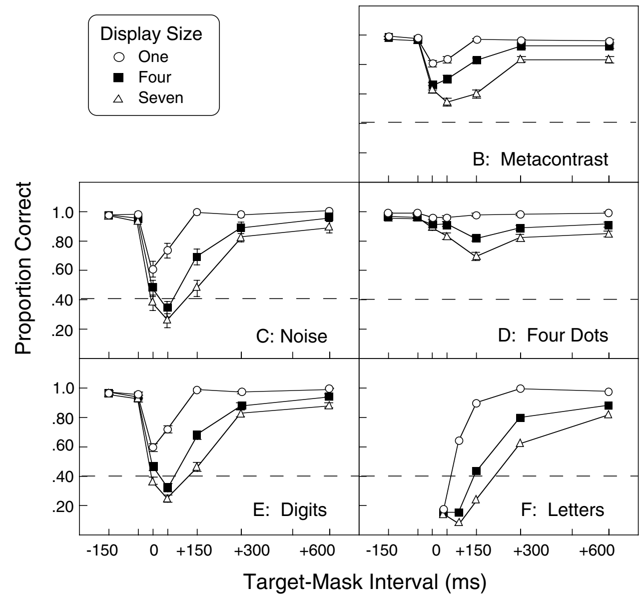

Fig. 3. Mean letter identification accuracy in Experiment 2: simultaneous cue (error bars = 1 SE).

In the two intervals in which the mask appeared before the target (−150 and 50 ms), accuracy was high and unaffected by display size or interval ($p > 0.05$). In the intervals 0, 50 and 150 ms, target accuracy was reduced ($p < 0.01$), with masking increasing with larger display sizes ($p < 0.01$). However, by 300 ms, accuracy was again substantially improved for each display size ($p < 0.01$), such that the remaining intervals yielded only relatively small display size effects (less than 10% overall, $p < 0.05$) but no effects of interval.

Noise mask. Fig. 3C shows a pattern of performance that was very similar to that for metacontrast masking. The only difference was a somewhat stronger masking effect overall at intervals of 0, 50 and 150 ms. Otherwise, with the exception of small display size effects (less than 10% overall, $p < 0.05$), there was no interval-dependent masking between 300 and 600 ms.

Four dot mask. Fig. 3D shows that under conditions of spatial pre-cuing there was a reduction in accuracy for the four dot mask only at 150 ms ($p > 0.05$). Aside from small but consistent effects of display size at all intervals ($p < 0.05$), this was the only masking seen in this condition.

Digit mask. Fig. 3E shows that the digit mask influenced accuracy through joint effects of display size and interval, $p < 0.001$. Beginning with the earliest interval, when the mask preceded the target by 150 ms, accuracy was high and display size effects were negligible ($p > 0.05$). However, as the mask came in closer temporal proximity to the target, accuracy decreased and display size effects grew ($p < 0.01$), until at the interval of +50 ms, there was a 30% difference in accuracy between display size 1 (70%) and display size 7 (40%). Yet, by the 300 ms interval, accuracy was again very high for display size 1 and only moderate display size effects remained (~10%, $p < 0.05$).

Letter mask. Fig. 3F shows that letters acting as masks still produce a large interaction between display size and interval ($p < 0.01$). Accuracy increased sharply along with target-mask interval ($p < 0.01$) and there were large display size effects in accuracy at intervals of 90, 150 and 300 ms ($p < 0.05$). It was not until the longest interval of 600 ms that accuracy reached the asymptotically high levels seen with the other masks.

3.3. Discussion

Observers did not have to retain more than one letter in memory in any of the conditions of this experiment. This meant that any masking effects remaining from Experiment 1 could not be attributed to failures of memory (i.e., failure to report letters that had already identified). Instead, the remaining masking effects could be attributed to interference in the object formation stage of letter identification. The results for all but one of the mask types indicated that this masking was effective at 0–150 ms intervals. In the case of letter masks it was effective even at the 300 ms interval.

However, it is possible that the design of this task still leads to an overestimation of the duration of the object formation stage. This is because it assumes that attention can be focused on the location of the target letter instantaneously with its presentation. If spatial orienting to the target location takes some time, then some of the remaining masking effects reflect a delay in onset of the object formation stage, rather than interference during the object formation stage itself. This was addressed in the next experiment.

4. Experiment 3: spatial pre-cue

This experiment was identical to Experiment 2, with the exception that the spatial pre-cue was presented for 30, 100 ms prior to the onset of the target display.

4.1. Results

Mean proportion target accuracy is shown in Fig. 4.

Metacontrast mask. Fig. 4B shows that accuracy was influenced jointly by display size and interval, $p < 0.001$. However, this interaction only applied to intervals of 0 and 50 ms. By 150 ms accuracy was only influenced by relatively small display size effects (less than 10% overall, $p < 0.05$) but not by any effects of interval.

Noise mask. Fig. 4C shows a very similar pattern. With the exception of small display size effects (less than 10% overall, $p < 0.05$), there was no interval-dependent masking between 150 and 600 ms.

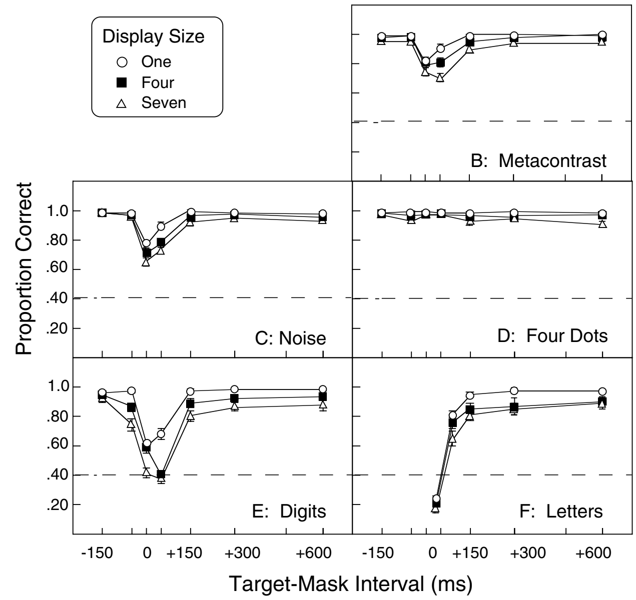

Fig. 4. Mean letter identification accuracy in Experiment 3: spatial pre-cue (error bars = 1 SE).

Four dot mask. Fig. 4D shows that there was no interval-dependent reduction in accuracy at all ($p > 0.05$). Only the small but consistent effects of display size were significant ($p < 0.05$).

Digit mask. Fig. 4E shows that the digit mask still influenced accuracy through joint effects of display size and interval, $p < 0.001$. At 150 ms, accuracy was high and display size effects were negligible ($p > 0.05$). However, as the mask came in closer temporal proximity to the target, accuracy decreased and display size effects grew ($p < 0.01$). Yet, by +150 ms and beyond, accuracy was no longer dependent on target-mask interval ($p > 0.05$).

Letter mask. Fig. 4F shows that letters acting as masks no longer yielded any interaction between display size and interval ($p > 0.05$). Accuracy increased sharply with increasing target-mask intervals ($p < 0.01$) and there were small display size effects in accuracy at all intervals ($p < 0.05$). By an interval of +150 ms accuracy had reached the asymptotically high levels seen with all the other masks.

4.2. Discussion

This experiment tested the prediction, derived from object substitution theory, that all forms of backward masking are minimal if spatial attention can be focused on the target location prior to the target-mask sequence. This prediction derives from the idea that object substitution will not occur if the target shape can be identified prior to the time that only the mask shape remains on view, which will then become the focus of the identification processes. Focused spatial attention serves to

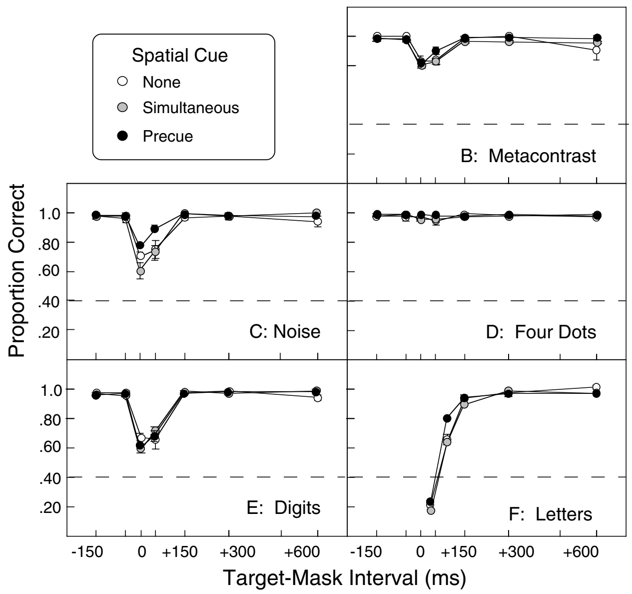

Fig. 5. Mean letter identification accuracy for display size = 1 across all experiments (error bars = 1 SE).

speed the onset and rate of target identification (Di Lollo et al., 2000).

The results show that this prediction was confirmed. When attention was focused on the target location 100 ms prior to the onset of the target, accuracy was unaffected by masks that followed after 150 ms or more. This suggests that letter identification was complete in this experiment by about 100 ms.

As in Experiments 1 and 2, substantial masking was observed in this experiment in the range of intervals between −50 and +50 ms, for all masks other than the four dots, even when there was only a single display item and when a spatial pre-cue indicated with 100% certainty where this item would appear. In this range of intervals, the focus of spatial attention had little influence on target identification, as can be seen in Fig. 5, where the masking effects for display size 1 can be compared across all three experiments. This attention-insensitive aspect of masking is consistent with the idea that temporal integration and local contour interaction are influenced relatively little by the distribution of spatial attention (Di Lollo et al., 2000).

5. General discussion

This comparison of the effects of spatial attention (display size and spatial cuing) on various types of visual masks (metacontrast, random noise, four dots, and patterns) provides strong support for the idea, derived from the object substitution theory of masking (Di Lollo et al., 2000), that there are at least two distinct visual masking processes.

The first process is active in the range of target-mask intervals of 0–100 ms. This form of masking interferes with *object formation*, and in the case of pattern masking, this seems to be through the mechanism of temporal integration. Although it is the formal task of the observer to identify the target object (letter), when targets and masks are presented in close temporal proximity, the first 'object' formed by the visual system is actually a composite of the target and mask patterns. For some masks, such as the four dots, such a composite object interferes little if at all with the task of identifying the target letter. However, for other masks, such as the random dots, digits and letters, the fusion of target and mask slows down target identification. For this early process, the effects of the mask are essentially those of 'camouflage,' which must be removed or segmented before the target can be identified.

This temporally early masking process is influenced very little by whether spatial attention is widely distributed, equally ready to select any one of eight different possible target locations, or whether it is already narrowly focused on only a single display location. This can be seen by comparing target accuracy for the smallest set size (one) across all three experiments, as shown in Fig. 5. The data points for each masking condition are almost all identical when this comparison is made, illustrating the point that this aspect of masking is uninfluenced by the spatial focus of attention and that it is over by the time 150 ms has elapsed between the presentation of target and mask.

These results are consistent with the way that the role of attention was modeled in the *computational model for object substitution* (CMOS) presented in Di Lollo et al. (2000). In this model, the early processes of object formation operate in parallel over the visual field. Object contours are grouped together, consistent with the hypotheses concerning which objects are likely to be presented on any given trial. The critical time-limited aspect of target identification is not in this early object formation stage, but lies instead in the time that elapses before the focus of attention can be aimed at the display location where the target object is being formed. This parameter of spatial attention, referred to as 'time to contact' in the CMOS model, was modeled as a linear function of display size (Di Lollo et al., 2000), in keeping with a large literature on the effects of display size on target identification (Duncan & Humphreys, 1989; Eriksen, 1995; Sperling, 1960; Treisman & Gelade, 1980).

The second masking process revealed in this study, the one that seems to apply most uniformly to all forms of mask shapes and patterns, is that of *object substitution*. This masking effect comes about when the original target item is replaced in the display by a masking pattern before the target item has been identified for perceptual report. The perceptual consequence of this physical replacement in the display is a ‘substitution’ of the target by the mask in the consciousness of the observer. As such, the specific perceptual relationship that exists between the target and the mask is of far less importance in this form of masking than when the mask interferes with object formation. In the most extreme demonstration of this point, four small dots that merely surround but that do not overlap spatially with the target contours act as a backward mask that is as effective as any of the other masks that were compared (metacontrast frame, noise dots, confusable shapes), provided that attention cannot be focused rapidly on the target.

What the present results make clear is that masking by object substitution is critically dependent on the existence of a temporal delay between target presentation (a physical event) and target identification (a mental achievement). This is evident in three critical features of the data. First, in Experiment 1, when attention was widely distributed at the onset of the target display, there was no backward masking in evidence for a single display item, beyond a target-mask interval of 100 ms. This held true whether the mask was a metacontrast frame, random noise, or even a competing shape such as another letter. This means that aside from the expected masking effects during object formation, there were no additional masking effects that arose from a delay in selecting the correct display item for perceptual report.

Second, the four dot mask yielded no evidence of any masking during the object formation stage, either in Experiment 1 where attention was distributed, or in Experiment 3 where attention could be focused earlier on the target location. In none of the experiments involving the four dots were any masking effects in evidence for intervals shorter than 150 ms; neither was there any masking when only one item was in the display. Evidently, four dots pose no significant ‘camouflage’ problems for target identification.

Finally, the absence of all backward masking effects for target-mask intervals of 150 ms and more in Experiment 3, provided the target letter location was known in advance, is consistent with object substitution masking depending critically on a delay between display presentation and the identification of a specific item in that display.

It is, of course, still the case that a mask presented immediately following a target may interfere with target identification. But it is important to realize that in order to do so it must satisfy each of two conditions: (a) the mask must be presented prior to the completion of target identification, and (b) a temporally integrated target-mask composite must interfere with target identification. Masks that fail to satisfy both of these conditions will be ineffective, as was the case in the present study for digit or letter masks when they are presented 150 ms or more following a pre-cued target letter (i.e., they satisfied condition b but not a) and for the four small dots when they appeared concurrently with a spatially pre-cued target letter (i.e., they satisfied condition a but not b).

5.1. Relations to previous theories of masking

The object substitution framework is not the first theoretical approach to highlight the distinction between masking by ‘integration’ or ‘fusion’ versus masking by ‘interruption’ or ‘erasure.’ For example, in an extensive study of pattern masking, Turvey (1973) distinguished between target-mask interference in the transmission of information from peripheral sensory organs (which he called peripheral masking) and interference that occurs among representations in the central decision center itself (he called this central masking). He proposed that integration was the primary mechanism involved in peripheral masking while interruption was the primary mechanism involved in central masking.

More recently, Shih and Sperling (2002) reported experiments on spatial cuing and masking in which they made a similar distinction between interference at a perceptual level versus interference in the consolidation of information in visual short-term memory. They concluded that the presence of non-target or distractor items simultaneously presented with the target resulted in perceptual interference, whereas the presence of a backward pattern mask produced interference at the memory consolidation stage.

At an empirical level, the present experiments extend the earlier work of Turvey (1973), and of Shih and Sperling (2002) in several important ways. First, the present study examined not merely one specific pattern mask, but a wide range of different types of mask. Second, the range of temporal intervals tested included not only positive target-mask intervals (backward masking) but also negative intervals (forward masking) and zero intervals (simultaneous masking). Third, the present experiments examined the influence of direct spatial cuing in combination with the other factors of target-mask interval and set size (Turvey’s experiments ignored the role of attention while those of Shih and Sperling examined an indirect auditory spatial cue in the context of only backward masking).

At a theoretical level, while the present approach is deeply indebted to these previous efforts to understand pattern masking in terms of separable processes, it must be pointed out that the object substitution framework is the only one in which it makes any sense to make a direct comparison of masks with such different physical characteristics. Prior to object substitution theory, there has been no reason to suspect that metacontrast masks, pattern masks and four dots are so closely related to one another in the way they interfere with target identification. Moreover, object substitution theory leads to the specific predictions explored in this paper concerning the conditions under which these various forms of masking will lead to different results.

Object substitution theory also helps to bring coherence to findings in the literature that heretofore have been treated as little more than curiosities. Take, for example, a 30-year old study of the influence of display size and pattern mask intensity (contrast) on letter identification (Spencer & Shuntich, 1970). Displays consisted of either 1 or 12 letters, arranged on an imaginary circle, with the target being indicated by a small bar marker. A random-line pattern mask was presented briefly at the target's location over a wide range of target-mask intervals. The results revealed one masking component, in the intervals of −50 to +100 ms, which was sensitive to variations in mask contrast (greater masking with larger contrast) but not to variations in set size. A second component could be seen in the intervals beyond 100 ms, where variations in mask contrast had little influence, but where stronger masking was associated with the larger set of potential letters. Although this study has had no discernible influence on standard theories of masking (e.g., Breitmeyer, 1984) the results have a ready interpretation within object substitution theory. The early contrast-sensitive component maps onto the temporal integration that is relevant to object formation, while the later display size-sensitive component maps onto the processes of object substitution.

The present results can therefore be seen as extensions of this more than 30 year old study showing that the distribution of spatial attention plays a critical role in whether backward pattern masking will occur in a letter identification task. The extensions include: (1) that this finding is not specific to any particular pattern mask, but is equally valid for masks as different as snugly fitting metacontrast frames and four small dots that surround the target location, (2) that spatial precuing is an effective means of eliminating all traces of object substitution masking, and (3) that a mask does more than merely 'interrupt' target processing. Indeed, it appears to replace the emerging representation of the target object in the consciousness of the observer with a representation based on the identity of the masking object.

5.2. Does target identification require reentrant processing?

Several recent reports have questioned the need for reentrant visual processing in order to accomplish what appear to be quite sophisticated tasks of visual categorization (Thorpe & Fabre-Thorpe, 2003; VanRullen & Koch, 2003; VanRullen & Thorpe, 2001). These experiments have used backward pattern masks and the main result has been that complex discriminations involving 'animal' versus 'non-animal' or 'cat ' versus 'dog' can be made when naturalistic scenes are presented for as little as 20 ms prior to the mask. Since manual responses indicating correct categorization of these same stimuli can be made in as little as 200 ms they have been called 'ultra-rapid categorization' and their speed has been used to support the claim that complex categorizations are possible using only feed forward mechanisms.

In evaluating the relevance of this claim to the current work, which is premised on reentrant processing being required for target identification, it is important to note that the methods used in ultra-rapid categorization ignore the time consuming and effortful mental processes involved in preparing the visual system to make a particular discrimination. The importance of appropriate mental preparation in the perception of brief visual scenes was documented in the seminal work of Potter (1976), who showed that scenes could be categorized as belonging to a pre-specified class in a much shorter time than they could be committed to memory for immediate recognition.

The methods of ultra-rapid categorization also ignore the fact that the mental preparation required to make a categorization occurs through reentrant or feedback processes. This is true almost by definition, since the preparation of the visual system for a briefly presented and masked scene, both in the new work and the earlier work of Potter (1976), comes about through a non-pictorial means. Observers have prepared their visual systems for these scenes by following written or spoken instructions.

In contrast to the rapid two-fold categorization of scenes, target identification in the present study involved selecting one of 21 equally likely targets and linking it to the correct location. In comparison to observers in a scene categorization task, observers in the present study were relatively unprepared for each target that was presented. The nature of the 'preparation' that was varied concerned the spatial location of the target letter, through variation in set size and spatial cuing, not any preparation as to its identity. What was also varied was the time at which the mask could begin interfering with target identification. Therefore, if target identification involves iterative reentry, as is assumed within the object substitution framework, then the minimal number of essential or required iterations were held constant in this study. The actual number of iterations, however, was likely manipulated by factors such as distractor letters, spatial pre-cues and masks.

5.3. Limitations and future directions

Object substitution theory is at present primarily a qualitative framework that helps to organize findings on visual masking and to make qualitative predictions about masking conditions that have not yet been tested.

A quantitative implementation of the theory has been presented for a particular set of masking experiments that involve masks of varying duration (Di Lollo et al., 2000). An online simulation will also soon be available, along with simulations of several other theories of backward masking, by Francis (in press), http://www.psych.purdue.edu/~gfrancis/Publications/BackwardMasking/. Although the quantitative model was originally designed to account for masking that occurs as a function of the mask duration in the common onset masking paradigm (Di Lollo et al., 2000) the online simulation promises to generate quantitative predictions with respect to masking as a function of the target-mask interval (Francis, in press). This means that it may soon be quite easy to compare both the present data and those from future masking experiments with a large range of quantitative models, including those of object substitution.

Acknowledgements

This research was supported by a research grant from the Natural Science and Engineering Research Council of Canada.

References

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.

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.

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General, 109*, 75–97.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481–507.

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

Enns, J. T. (2002). Visual binding in the standing wave illusion. *Psychonomic Bulletin and Review, 9*, 489–496.

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., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences, 4*, 345–352.

Enns, J. T., & Rensink, R. (1992). *VScope(tm): Vision testing software for the Macintosh*. Vancouver: Micropsych Software.

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.

Francis, G., (in press). Online simulations of models for backward masking. *Behavior, Research Methods, Instruments & Computers*.

Kahneman, D. (1968). Method, findings, and theory in studies of visual masking. *Psychological Bulletin, 70*, 404–425.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (third ed.). NY: Academic.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences, 6*, 120–125.

Kolers, P. A. (1968). Some psychological aspects of pattern recognition. In P. A. Kolers & M. Eden (Eds.), *Recognizing patterns*. Boston, MA: MIT Press.

Lleras, A., & Moore, C. M. (2003). When the target becomes a mask: Using apparent motion to isolate the object component of object-substitution masking. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 106–120.

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.

Moore, C.M., Lleras, A., (in press). On the role of object representations in substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 509–522.

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

Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109*, 260–305.

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).

Thorpe, S. J., & Fabre-Thorpe, M. F. (2003). Seeking categories in the brain. *Science, 291*, 260–263.

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.

VanRullen, R., & Koch, C. (2003). Visual selective behavior can be triggered by a feed-forward process. *Journal of Cognitive Neuroscience, 15*, 209–217.

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorization of natural and artifactual objects. *Perception, 30*, 655–668.

von Muehlenen, A., Enns, J. T., & Di Lollo, V. (2003). Metacontrast masking with target and mask energies equated across exposure durations. *Vision Sciences Society Abstracts*, 247.

Weisstein, N., Ozog, G., & Szoc, R. (1975). A comparison and elaboration of two models of metacontrast. *Psychological Review, 82*, 325–343.