

# Rapid serial visual distraction: Task-irrelevant items can produce an attentional blink

TROY A. W. VISSER

*University of British Columbia, Vancouver, British Columbia, Canada*

WALTER F. BISCHOF

*University of Alberta, Edmonton, Alberta, Canada*

and

VINCENT Di LOLLO

*University of British Columbia, Vancouver, British Columbia, Canada*

When two sequential targets (T1 and T2) are presented within about 600 msec, perception of the second target is impaired. This *attentional blink* (AB) has been studied by means of two paradigms: rapid serial visual presentation (RSVP), in which targets are embedded in a stream of central distractors, and the *two-target* paradigm, in which targets are presented eccentrically without distractors. We examined the role of distractors in the AB, using a modified two-target paradigm with a central stream of task-irrelevant distractors. In six experiments, the RSVP stream of distractors substantially impaired identification of both T1 and T2, but only when the distractors shared common characteristics with the targets. Without such commonalities, the distractors had no effect on performance. This points to the subjects' attentional control setting as an important factor in the AB deficit and suggests a conceptual link between the AB and a form of nonspatial contingent capture attributable to distractor processing.

In everyday experience, we are exposed to a vast array of images that stream forth at a rate that far exceeds the processing capability of the visual system. Because of the high input rate, not all stimuli can be processed completely; some are processed only in part, and others are missed completely. This cost is seen in a phenomenon known as the *attentional blink* (AB), which reveals a striking imbalance in the identification accuracy of two targets presented in rapid succession. Identification is almost perfect for the first target but is substantially reduced for the second, especially when it follows the first by 100–500 msec. Two paradigms, detailed below, have been used most frequently to study the AB. In one, the two targets are embedded in a stream of distractors; in the other, the targets are presented without distractors. Although both paradigms yield robust AB deficits, the role of distractors in determining the time course and magnitude of the AB is still unclear. The latter issue was the focus of the present work.

Perhaps not surprisingly, we found that distractors can interfere with target identification. More important, we

discovered that the degree of interference depends very much on the subject's attentional control settings (Folk, Remington, & Johnston, 1992; i.e., how the visual system is configured in order to process task-relevant stimuli efficiently). For example, when targets were letters, subjects established an attentional control setting to process letter-like stimuli efficiently while ignoring nonletter stimuli. As a result, distractors that shared letter-like characteristics (e.g., digits) interfered with target identification, whereas dissimilar distractors (e.g., random-dot patches) did not. The present work builds on Folk et al.'s (1992) findings by establishing a direct link between the subject's attentional control setting and the role of distractor processing in the AB.

## Two Main Paradigms of the AB

Investigations of the AB have been conducted most commonly with two paradigms. In one, known as *rapid serial visual presentation* (RSVP), two targets (e.g., letters) are inserted in a stream of distractors (e.g., digits). All the items are displayed sequentially in the same location at a rate of one every 100 msec or so, such that each target is masked by the next item in the stream (Chun & Potter, 1995; Raymond, Shapiro, & Arnell, 1992). In contrast, no distractors are employed in a second method, known as the *two-target* paradigm. Instead, two targets, each masked by a single temporally trailing mask, are displayed at different screen locations, separated by a variable temporal lag (Duncan, Ward, & Shapiro, 1994).

Similar AB deficits have been obtained with the two paradigms. In both cases, identification of the second tar-

---

This work was supported by an Izaak Walton Killam Memorial Pre-doctoral Fellowship to the first author and by research grants from the Natural Sciences and Engineering Research Council of Canada to the second and third authors. We thank Kimron Shapiro, Marisa Carrasco, Charles Folk, Bernhard Hommel, and one anonymous reviewer for helpful comments on earlier versions of this report. We also thank Jeanette Lum for assistance in data collection. Correspondence concerning this article should be addressed to T. A. W. Visser, Department of Psychology, University of Melbourne, 12th Floor, Redmond Barry Building, Parkville 3010, Australia (e-mail: tvisser@unimelb.edu.au).

get is markedly impaired at short intertarget lags, with accuracy improving as lag is increased. Consistency of outcomes, however, does not necessarily imply that the two paradigms tap the same underlying mechanisms. As Ward, Duncan, and Shapiro (1997) pointed out, it is possible that the AB obtained with the RSVP method may stem from difficulties in selecting the targets from the stream of distractors. The AB obtained with the two-target paradigm, on the other hand, may reflect difficulties in repositioning focal attention from the location of the first target to that of the second target.

To check on these options, Ward et al. (1997) employed a hybrid paradigm, called *skeletal RSVP*, in which two targets, each followed by a mask, were displayed in the same central location over a range of temporal lags. The skeletal RSVP paradigm involved neither distractor interference nor spatial switching, yet it yielded a conventional AB. This warranted the conclusion that the RSVP and the two-target paradigms tap the same underlying mechanism—namely, a common attentional limitation that impairs identification of the second of two targets presented in rapid succession.

### Focus of the Present Work

Building on the study of Ward et al. (1997), the present experiments address some persisting questions regarding the role of distractors in the AB. The evidence in Ward et al. indicates that an AB can be obtained in the absence of distractors even when all the stimuli are presented at the same location. But this need not mean that the stream of distractors used in the RSVP paradigm does not influence the magnitude or time course of the AB. Indeed, suggestive evidence has come from studies in which a conventional RSVP paradigm has been used. For example, Chun and Potter (1995) found a smaller AB for letter targets when distractors were mathematical symbols than when they were letters. Similarly, Maki, Bussard, Lopez, and Digby (2003) found that the magnitude of the AB for word and letter targets was significantly reduced when distractors were presented in a “false font.” These findings raise significant questions about the extent to which the presence of distractors may modulate the AB deficit and the mechanisms that underlie this effect.

Distractors could influence performance in at least two ways. One is the way noted by Ward et al. (1997)—namely, by interfering with target selection from the RSVP stream. In this case, interference might occur because the targets are masked by both preceding and subsequent items in the stream, allowing for both forward and backward masking. This differs from the two-target paradigm, in which the targets are never forward masked. Thus, the presence of distractors in the RSVP paradigm might introduce a source of interference—forward masking—that does not exist in the two-target paradigm. This option is also consistent with the results of Maki et al. (2003), who found that increasing the pixel density of distractors increased the magnitude of the AB. Given that forward masking is thought to be mediated by a summation of the contours of

the target and the mask (e.g., Bachmann & Allik, 1976; Breitmeyer, 1984; Scheerer, 1973), it seems likely that a more dense mask would lead to increased masking of T2 and, thus, a larger AB.

Another way in which distractors could interfere with target identification is through contingent capture, a process by which an extraneous stimulus that shares defining characteristics with a target diverts attention, thereby slowing target processing and reducing identification accuracy (Bacon & Egeth, 1994; Folk & Remington, 1998; Folk et al., 1992; Gibson & Kelsey, 1998). Contingent capture has conventionally been explained in terms of unnecessary shifts of spatial attention. For example, Folk et al. (1992) found that responses to red targets were slower when they were preceded by a nonpredictive red cue presented at a different spatial location. They explained this result by arguing that because the defining characteristic of the target was its red color, all red objects were capable of capturing attention. Thus, when the red cue was presented, attention was involuntarily shifted to its location, even though it did not predict the target location. This attentional shift was beneficial when the target and the cue occurred in the same location, but it was detrimental when the stimuli occurred in different locations, because an additional spatial shift was required to redirect attention to the location of the target.

More relevant to the present work are the results of Folk, Leber, and Egeth (2002), who found evidence for contingent capture in a task similar to a conventional AB paradigm. In their experiments, observers monitored a central RSVP stream for a target letter of a specific color. Prior to target onset, four distractors were presented at spatial locations adjacent to the RSVP stream. When the distractors were a different color than the target, performance was unaffected. However, when one of the distractors was the same color as the target (i.e., shared its defining feature), identification accuracy dropped sharply. Folk et al. (2002) attributed this decline to the capture of spatial attention by the distractor and went on to suggest that distractors in their paradigm might have played a role similar to that of T1 in the conventional AB paradigm.

Although there is considerable evidence for contingent capture mediated by spatial shifts of attention, recent work has shown that much the same delay in target processing can also be obtained under conditions in which spatial factors are ruled out. In this case, the delay arises from the time taken to process a leading distractor that shares the target's defining attribute. Ghorashi, Zuvic, Visser, and Di Lollo (2003) presented a stream of distractors that were either letters or random dots. The stream terminated in a target that was always a single letter. As would be expected on the basis of contingent capture, they found that responses to the target were slower when the distractors were letters than when they were random dots. More important, the contingent delay occurred whether the target was presented in the same location as the distractors or in a different location. This demonstrated that contingent capture could occur even when spatial shifts in attention were ruled out.

Given the evidence implicating nonspatial mechanisms, Ghorashi et al. (2003) attributed the delay in target processing to the time spent processing leading distractors that shared the target's defining attribute. The reasoning was as follows. When a target-like distractor was presented directly before the target, it engaged the same mechanisms as those required for processing the target itself. As a consequence, target processing was delayed while the system was busy processing the distractor. During the delay, the target was vulnerable to the ensuing mask, and identification accuracy suffered accordingly. On this reasoning, when the distractors shared the target's defining characteristics, distractor processing caused a delay in target processing, thereby reducing identification accuracy. In contrast, when the distractors were random dots that did not share defining characteristics with the letter target, they were never processed. As a result, targets were processed immediately, and identification accuracy was high.

There is an obvious parallel between the study of Ghorashi et al. (2003) and a large number of the AB studies in which the RSVP paradigm has been employed. Namely, it is conceivable that the leading distractors in AB studies might trigger the kind of nonspatial attentional capture illustrated by Ghorashi et al., thus compounding the effects of the first target in producing an AB deficit. On this view, distractors in the RSVP stream are sometimes processed as though they were targets, thus delaying allocation of attention to the real target. During this delay, the target is vulnerable to masking (Chun & Potter, 1995; Giesbrecht & Di Lollo, 1998), and identification accuracy suffers correspondingly. Needless to say, this can occur with the RSVP paradigm, but not with the two-target paradigm, because distractors are never used in the latter.

In principle, this conjecture could be tested simply by comparing AB studies in which the RSVP paradigm has been employed with those in which the two-target paradigm has been employed. However, such a comparison is not entirely straightforward, because the two paradigms differ in ways that go beyond the presence or absence of an RSVP stream. For example, targets are displayed centrally in the RSVP paradigm, but peripherally in the two-target paradigm. Also, as was noted above, distractors may impair target processing through forward masking, and this source of impairment would be confounded with any effect due to contingent capture, whether spatial or nonspatial.

For these reasons, we chose to investigate the role of contingent capture in the AB deficit, using a method that,

in some sense, is the converse of the skeletal RSVP paradigm of Ward et al. (1997). To wit, Ward et al. took the RSVP paradigm and stripped it of distractors. Conversely, we took the two-target paradigm and *added* a central stream of distractors. Thus, our displays consisted of an RSVP stream of task-irrelevant distractors presented at fixation and two targets presented at peripheral locations. Masking by the distractors is avoided in this method, because the targets and the distractors are displayed in different spatial locations. More important, this paradigm makes it possible to study whether the magnitude and temporal course of the AB are influenced by the presence of an ostensibly irrelevant stream of distractors, as might be expected on the basis of contingent capture.

A complete summary of the targets, masks, and distractors used in each experiment is presented in Table 1. To anticipate our results, in Experiments 1A and 1B, we found a much larger AB when two letter targets were embedded within an RSVP stream of digits than when the stream was omitted. This was true whether or not the targets and the distractors appeared at the same (Experiment 1B) or different (Experiment 1A) spatial locations. On the basis of these results, forward masking was rejected as a significant source of distractor interference. Instead, we pursued the option that processing of target-like distractors delays allocation of attention to the target, leaving it vulnerable to decay and visual masking.

In Experiments 2 and 3, we found that the relationship between the targets and the distractors mediates the magnitude of the AB, such that the magnitude of the AB is maximal when the targets and the distractors are in the same stimulus category (Experiments 2 and 3) and declines as target-distractor similarity decreases. Finally, in Experiments 4 and 5, we illustrated the generality of our results, using a different class of targets, and demonstrated that our results cannot be explained solely on the basis of interference from neighboring distractors, as in the flanker effect (Eriksen, 1995).

## EXPERIMENT 1A

Experiment 1A was an exploratory study designed to find out whether the AB obtained with the two-target paradigm is affected by the presence of a task-irrelevant stream of distractors displayed at fixation. We employed two conditions. One was a conventional two-target condition in which the subjects identified two letter targets, each followed by a mask, presented at unpredictable eccentric

**Table 1**  
**Types of Distractors, Targets, and Masks Used in Experiments 1–5**

	Experiment				
	1A, 1B	2	3	4	5
Distractors	digits	letters	none, dots, digits, pseudoletters	digits, letters, pseudoletters	dots, letters
Targets	letters	letters	letters	digits	letters
Masks	digits	digits	digits	letters	digits

locations. The second condition was the same as the first, except that the central fixation cross was replaced by an RSVP stream of digits that the subjects knew would never contain a target and were instructed to ignore.

## Method

**Subjects.** Twenty-four undergraduate psychology students at the University of British Columbia participated for course credit. All reported normal or corrected-to-normal vision.

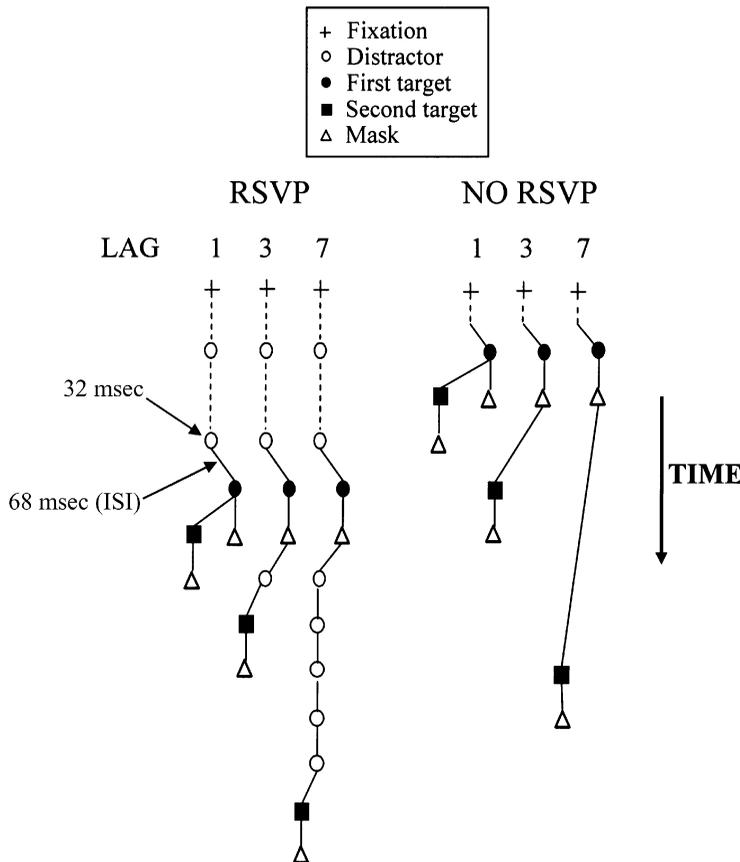
**Apparatus and Stimuli.** All the stimuli were displayed on a Tektronix 608 oscilloscopic point plotter equipped with fast P15 phosphor. The luminance was set at 25 cd/m<sup>2</sup>, measured by a Minolta LS-100 luminance meter. At the viewing distance of 57 cm, set by a headrest, the stimuli subtended approximately 1° of visual angle. The background and surrounding visual field were dark, except for dim illumination of the keyboard. The targets were uppercase letters from the English alphabet, and the distractors were digits.

**Procedure.** All the subjects served in two conditions, RSVP and no RSVP. Stimulus presentation in both conditions was based on the two-target paradigm used by Duncan et al. (1994). In both conditions, all the items were displayed for 32 msec, and were followed by a 68-msec blank interstimulus interval (ISI). On any given trial, two target letters were selected randomly without replacement from all the

letters of the English alphabet, except I, O, Q, and Z, which were omitted because of structural similarity to the digits 1, 0, 2, and 7.

In the no-RSVP condition, a trial began with the presentation of a small fixation cross in the center of the screen. The subjects pressed the space bar to initiate a trial. Following a 500- to 1000-msec blank screen, the first target letter was presented in one of four locations: 1° above, below, left, or right of the center of the screen. The location of the target was chosen randomly on each trial, with the restriction that each of the four locations was chosen an equal number of times within a session. A mask, consisting of a digit selected randomly from 0 to 9, was then displayed at the same location as the first target at a stimulus onset asynchrony (SOA) of 100 msec. The second target, also a letter, was presented at one of the remaining three locations, followed by a digit mask at an SOA of 100 msec. It followed the first target at one of three temporal lags corresponding to intertarget SOAs of 100, 300, and 700 msec. Thus, on trials in which the intertarget SOA was 100 msec, the second target and the first-target mask were presented simultaneously. The subjects were required to identify the two target letters, in either order, and to enter them on the keyboard at the end of the display sequence. Next, the fixation cross reappeared to indicate that the next trial was ready to begin.

In the RSVP condition, the sequence of events was identical to that in the no-RSVP condition, with the following exceptions. Each trial began with a fixation cross, which disappeared when the sub-



**Figure 1.** Schematic representation of the stimulus sequence in each experiment at Lags 1, 3, and 7. The segmented line at the beginning of each sequence represents the set of 5–10 leading distractors that were displayed at the center of the screen before the first target on each trial. The symbols shown as displaced from the center line represent items (targets or masks) that were displayed eccentrically.

ject pressed the space bar. After a 500-msec blank screen, a stream of digits was displayed in the center of the screen at an interitem SOA of 100 msec. The number of digits in the stream varied randomly on each trial from 5 to 10. Thus, on any trial, T1 was presented between 1,000 and 1,500 msec after the subject pressed the space bar to begin the trial. As in the no-RSVP condition, T2 followed T1 at one of three temporal lags corresponding to intertarget SOAs of 100, 300, and 700 msec. On any given trial, the digits were selected randomly with replacement, with the constraint that the selected digit was not one of the two immediately preceding items. The subjects were instructed to maintain fixation on the digit stream, but to ignore its contents, because the stream would never contain a target.

The sequence of events in the two conditions is illustrated schematically in Figure 1. All the subjects participated in both the RSVP and the no-RSVP conditions. The order of the two conditions was counterbalanced, such that half the subjects received the RSVP condition first. Each condition began with 15 practice trials, during which no data were recorded, followed by 144 experimental trials. The two targets were presented in each of the four locations an equal number of times at each of the three lags.

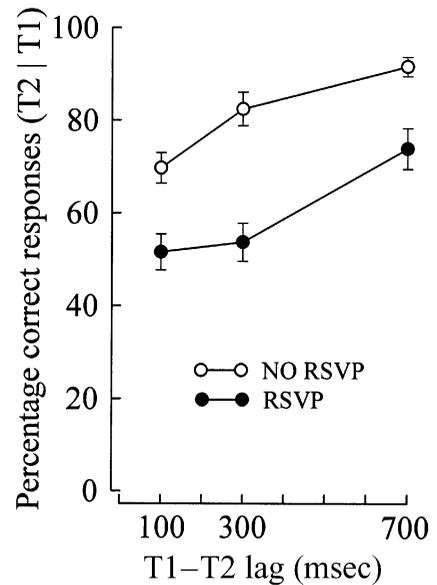
### Results and Discussion

In all the present experiments, estimates of second-target identification were based on only those trials in which the first target had been identified correctly. This procedure is commonly adopted in AB experiments on the grounds that, on incorrect trials, the source of the error is unknown, so its effect on second-target processing cannot be estimated. Responses were recorded as correct regardless of the order of report.

Mean percentages of correct identifications of the first target, collapsed across lags, were 88.6% and 95.5% for the RSVP and the no-RSVP conditions, respectively. A *t* test for related samples showed these values to differ significantly from each other [ $t(23) = 2.96, p < .01$ ]. Mean percentages of correct identifications of the second target as a function of lag, averaged over all the subjects, are presented in Figure 2. The results were analyzed in a 2 (condition)  $\times$  3 (lag) analysis of variance (ANOVA), which revealed significant effects of condition [ $F(1,23) = 47.08, MS_e = 353.44, p < .001$ ], and of lag [ $F(2,46) = 63.25, MS_e = 95.88, p < .001$ ]. The interaction also was significant [ $F(2,46) = 8.18, MS_e = 57.87, p = .001$ ].

Of major interest for the present purpose is the mean accuracy of second-target identification in the RSVP and no-RSVP conditions. It is immediately obvious from Figure 2 that the presence of distractors resulted in substantially lower performance in the RSVP condition. This is notable because the RSVP stream never contained a target, and the subjects had been instructed to ignore it. Far from being irrelevant, the RSVP stream caused accuracy to drop across all lags by an average of more than 20%.

In the foregoing, we have considered two sources of interference, forward masking and contingent capture, which might account for this type of result. Of these, forward masking must be ruled out on the grounds that the items in the RSVP stream were too far removed from the targets to act as effective masks (although this does not imply that forward masking cannot occur when targets and distractors are presented in the same location; see, e.g., Maki



**Figure 2. Experiment 1A:** mean percentage of correct identifications of the second target, separately for the RSVP and the no-RSVP conditions, given that the first target had been identified correctly. Error bars represent one standard error of the mean.

et al., 2003). Contingent capture, on the other hand, remains a viable option. It is possible that, despite instructions, the subjects may have been unable to completely ignore the distractor stream because it shared features with the targets. Another way of saying this is that the items in the RSVP stream may occasionally have captured attention. On those occasions, a distractor might be processed as though it were a target. If a target were then presented while the distractor was being processed, identification accuracy would suffer, because the distractor would have preempted attentional resources required for processing the target. This can be regarded as a form of AB deficit in which a distractor plays the role of the first target.

A corollary of this line of reasoning is that the presence of distractors should impair accuracy not only for the second target, but also for the first. To wit, if attention is captured by a distractor just before the first target is presented, fewer resources will be available for processing the target, and identification accuracy should suffer correspondingly. The results of the present experiment are in line with this expectation. Accuracy of first-target identification was significantly lower in the RSVP than in the no-RSVP condition. However, the results also show that the difference between the two conditions was smaller for the first target (6.7%) than for the second (21.5%). One possible reason is that in the no-RSVP condition, the level of first-target identification (95.5%) was very close to the 100% limit imposed by the response scale, thereby preventing a larger difference between the two conditions from becoming apparent. Also, it is possible that second-target accuracy may have been selectively more impaired

in the RSVP than in the no-RSVP condition. This could occur if the processing demands of the distractors combined with those of the first target to further reduce the resources available for the second target. At any rate, the fact that the distractors impaired the identification of both the first and the second targets is consistent with the hypothesis that the impairment may have been mediated by attentional capture.

Before reaching a definitive conclusion that the distractor interference seen in the present experiment was indeed nonspatial, a possible account based on a spatial spotlight/zoom lens metaphor (Sperling & Weichselgartner, 1995) needs to be considered. Suppose that at the start of each trial, subjects adopt a wide attentional spotlight of a diameter sufficient to encompass all possible target locations. Presentation of the central RSVP stream might cause the spotlight to become narrowly focused on center screen. Then, when the target arrives, the spotlight would need to be moved and refocused on the location of the target, leading to a cost of an inherently spatial nature. On this account, the delay in target processing is attributable not to the time wasted in processing a distractor in the RSVP stream, but to the time wasted in moving the spotlight. Needless to say, such a space-based account would not apply to the no-RSVP condition.

To check on this possibility, we replicated Experiment 1A but displayed both the targets and the distractors in the same spatial location. We reasoned that if spatial shifts of attention were influencing performance in Experiment 1A, displaying all the items in the same location, as was done in Experiment 1B, should reduce or eliminate the difference between the no-RSVP and the RSVP conditions. On the other hand, if distractor processing were the primary source of impairment in target identification, displaying all the items in the same location should yield deficits similar to those in Experiment 1A.

## EXPERIMENT 1B

### Method

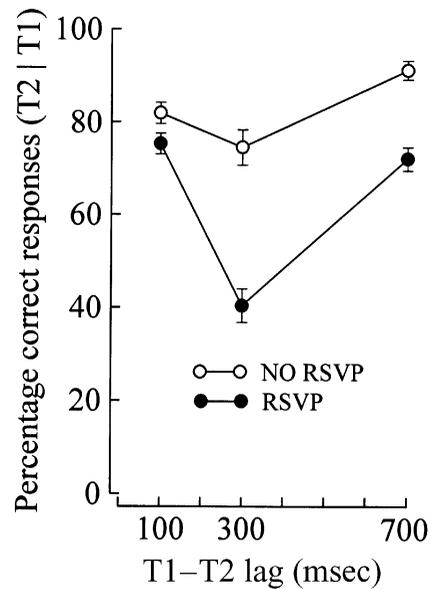
**Subjects.** Twenty-four undergraduate psychology students at the University of British Columbia participated for course credit. All reported normal or corrected-to-normal vision. None had participated in the previous experiment.

**Apparatus and Stimuli.** The apparatus and stimuli were the same as those in Experiment 1A.

**Procedure.** The procedure was the same as that in Experiment 1A, with the exception that in both the RSVP and no-RSVP conditions, all the items were displayed in the same central location on the screen. Thus, displays in the RSVP condition were identical to those used in most conventional AB studies (e.g., Raymond et al., 1992), whereas displays in the no-RSVP condition were identical to those in the skeletal RSVP employed by Ward et al. (1997).

### Results and Discussion

Mean percentages of correct identifications of the first target, collapsed across lags, were 90.4% and 97.5% for the RSVP and the no-RSVP conditions, respectively. A *t* test for related samples showed these values to differ significantly from each other [ $t(23) = 5.24, p < .001$ ]. Mean percentages of correct identifications of the second target



**Figure 3. Experiment 1B:** mean percentage of correct identifications of the second target, separately for the RSVP and the no-RSVP conditions, given that the first target had been identified correctly. Error bars represent one standard error of the mean.

as a function of lag, averaged over all the subjects, are presented in Figure 3.

The results were analyzed in a 2 (condition)  $\times$  3 (lag) ANOVA, which revealed significant effects of condition [ $F(1,23) = 73.72, MS_e = 194.65, p < .001$ ], and of lag [ $F(2,46) = 62.80, MS_e = 132.43, p < .001$ ]. The interaction also was significant [ $F(2,46) = 44.52, MS_e = 51.43, p < .001$ ].

The results of the present experiment replicate the principal findings of Experiment 1A: Namely, the presence of distractors in the RSVP condition substantially reduced identification accuracy for both T1 and T2. Given that all the items were presented in the same spatial location, the reduction in target identification accuracy cannot be explained by shifts of spatial attention. Instead, the results are consistent with the notion that processing of distractor items prevented allocation of attention to the target, thereby impairing identification accuracy.

Two notable differences between the results of Experiments 1A and 1B deserve comment. First, in Experiment 1B, T2 performance at Lag 3 was about 20% worse than that at the comparable lag in Experiment 1A. This difference points to additional deficits in target identification that occur when targets and distractors are in the same spatial location. This deficit may be mediated by forward masking. Alternatively, because of the spatial superimposition of the targets and the distractors in Experiment 1B, T1 selection may have been more difficult, thus increasing dwell time for T1 and, hence, the magnitude of the AB.

Also notable is that in Experiment 1A, T2 performance was worst at Lag 1 and improved steadily as lag increased.

In contrast, in the RSVP condition in Experiment 1B, T2 performance was check mark shaped, with relatively accurate identification at Lag 1, poor performance at Lag 3, and substantial improvement again at Lag 7—a pattern of results commonly referred to as *Lag 1 sparing* (Potter, Chun, Banks, & Muckenhoupt, 1998; Visser, Bischof, & Di Lollo, 1999). Why did Lag 1 sparing occur in Experiment 1B, but not in Experiment 1A? An analysis of this phenomenon by Visser et al. (1999) points to target location as the critical factor. They found that Lag 1 sparing commonly occurs when the targets are presented in the same spatial location, as in Experiment 1A. In contrast, Lag 1 sparing never occurs when targets are presented in different spatial locations, as in Experiment 1B.

Collectively, Experiments 1A and 1B indicate that a task-irrelevant RSVP stream of digit distractors impairs the identification of letter targets. This result is consistent with an explanation in terms of nonspatial contingent capture by distractors that share defining characteristics with targets. However, because we did not manipulate target-distractor similarity systematically, it is also possible that the mere presence of distractors, regardless of their similarity to the target, would have impaired performance. To examine this possibility, in Experiments 2 and 3, we varied the similarity between targets and distractors systematically and examined the influence of this manipulation on target identification.

## EXPERIMENT 2

Involuntary processing of the RSVP stream may have mediated contingent capture in the previous two experiments because the letters that the observers were set to identify shared common structural features with digit distractors. That is, both the letters and the digits were composed of broadly similar line segments and visual features. It is possible that, at least at an early processing stage, the subjects' attentional control setting may have operated at the level of structural features, as distinct from whole-letter configurations. In this case, some structural features of individual digits in the RSVP stream may have matched the control settings for the letters, causing that distractor to be processed as though it were a letter. In Experiment 2, we increased the similarity beyond the level of structural features by using letters, instead of digits, as distractors in the RSVP stream. To the extent that the strength of attentional capture increases with the similarity between targets and distractors, we expected identification of both of the targets to be more impaired in Experiment 2 than in Experiments 1A or 1B. The outcome confirmed this expectation.

### Method

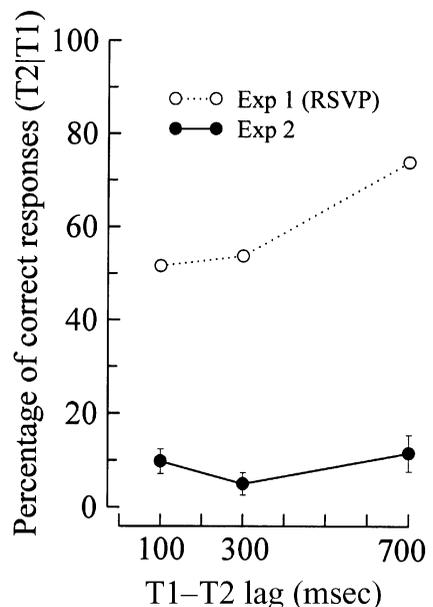
A new group of 24 subjects served in Experiment 2. The apparatus, stimuli, and procedure were the same as those in Experiment 1A, except for the following. The experiment contained only one condition, identical to the RSVP condition in Experiment 1A, save that the distractors were uppercase letters instead of digits. The masks were randomly chosen digits, as in Experiment 1A (see Table 1). This was done to ensure that the masking conditions remained identical be-

tween the two experiments, with variation only in target-distractor similarity.<sup>1</sup>

## Results and Discussion

Mean percentage of correct identifications of the first target, collapsed across lags, was 29.9%, as compared with a chance level of less than 5%. A *t* test for independent samples showed this value to be significantly lower than the corresponding value (88.6%) in the RSVP condition of Experiment 1A [ $t(46) = 12.36, p < .001$ ]. Mean percentages of correct identifications of the second target as a function of lag, averaged over all the subjects, are presented in Figure 4, which also includes the results of the RSVP condition in Experiment 1A for ease of comparison. The second-target data of Experiment 2 were analyzed in a one-way ANOVA, which revealed a significant effect of lag [ $F(2,46) = 3.24, MS_e = 83.83, p < .05$ ]. The results of Experiments 2 and 1A (Figure 4) were compared in a 2 (experiment)  $\times$  3 (lags) mixed ANOVA. All the effects were significant [experiment,  $F(1,46) = 120.36, MS_e = 772.88, p < .001$ ; lag,  $F(2,92) = 29.13, MS_e = 90.62, p < .001$ ; experiment  $\times$  lag,  $F(2,92) = 14.56, MS_e = 90.62, p < .001$ ].

Accuracy of second-target identification plunged dramatically in Experiment 2, as compared with Experiments 1A and 1B. This is precisely what would be expected on the basis of contingent capture. Given that the subjects' attentional control setting was to identify letter targets, the probability of attentional capture was greater when the distractors belonged to the same lexical class as the targets (Experiment 2) than when they merely shared structural features with the targets (Experiments 1A, 1B).



**Figure 4.** Experiment 2: mean percentage of correct identifications of the second target, given that the first target had been identified correctly. The results of Experiment 1 (RSVP condition) have been included for ease of comparison. Error bars represent one standard error of the mean.

Also in line with contingent capture is the finding that accuracy was severely impaired not only for the second target, but also for the first. The reasoning is straightforward: Attentional capture by temporally leading distractors reduced the processing resources available for the first target, thus producing an AB-like deficit in accuracy of first-target identification. A similar account can be offered for the finding that, in contrast to conventional AB studies, second-target accuracy in Experiment 2 showed little sign of improvement even at the longest lag (Figure 4). That is, attentional capture by the distractors presented during the intertarget lag reduced the processing resources available for the second target, and an AB deficit ensued. On this account, the AB deficit seen in Figure 4 should continue unabated at even longer lags, provided that distractors continued to be displayed throughout the lag.

Considered together, the outcomes of Experiments 1 and 2 strongly suggest that an ostensibly irrelevant stream of distractors can increase the magnitude and time course of the AB. Furthermore, this evidence shows that the degree of interference with target identification increases with the similarity between targets and distractors, as would be expected on the basis of contingent capture. This line of investigation was pursued in Experiment 3, where target–distractor similarity was varied at multiple levels.

### EXPERIMENT 3

Experiment 3 comprised four conditions that differed in the degree of target–distractor similarity. All four conditions were administered to each subject. The first was a conventional two-target condition (i.e., without a distractor stream), which provided a baseline for the remaining conditions. All other conditions contained an RSVP stream. In the second condition, each item in the RSVP stream consisted of a group of random dots whose configuration changed randomly and abruptly from one frame to the next. This condition defined the lowest level of similarity with the letter targets and, therefore, was expected to produce a correspondingly low level of contingent capture. In the third condition, each item in the RSVP stream was drawn from the set of pseudoletters illustrated in Figure 5. The pseudoletters were meaningless, but shared features with the letter targets and were intended to reveal the effect of structural similarity in the absence of meaningfulness. In the fourth condition, the items in the RSVP stream were digits. This condition was included in order to maintain comparability with Experiment 1A.

#### Method

A new group of 24 subjects served in Experiment 3. The apparatus, stimuli, and procedure were the same as those in Experiment 1A, with the following exceptions. The experiment contained four conditions. The blank condition and the digit condition were the same as the no-RSVP and the RSVP conditions, respectively, in Experiment 1A. The random-dot condition was the same as the digit condition, except that each item in the RSVP stream consisted of a patch of 100 dots distributed randomly within an imaginary square sub-



Figure 5. Set of pseudoletters used as distractors in the central RSVP stream in the pseudoletter condition in Experiment 3.

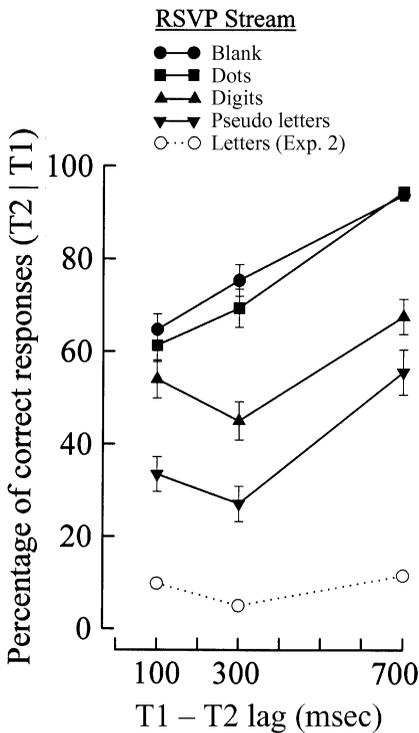
tending 1° of visual angle, centered at fixation. Each frame in the RSVP sequence contained a different random configuration of the dots. Finally, the pseudoletter condition was the same as the digit condition, except that each item in the RSVP stream consisted of a pseudoletter drawn randomly from the set illustrated in Figure 5. The sequence of events on any given trial is illustrated schematically in Figure 1.

#### Results and Discussion

Mean correct identifications of the first target, collapsed across lags, were 97.2%, 97.0%, 89.0%, and 73.0% for the blank, random-dot, digit, and pseudoletter conditions, respectively. A within-subjects ANOVA revealed a significant difference among these means [ $F(3,69) = 44.60$ ,  $MS_e = 209.35$ ,  $p < .001$ ]. Mean percentages of correct identifications of the second target, averaged over all the subjects, are presented in Figure 6, which also includes the results of Experiment 2 for ease of comparison. The data in Experiment 3 were analyzed in a 4 (condition)  $\times$  3 (lag) ANOVA which revealed significant effects of condition [ $F(3,69) = 101.70$ ,  $MS_e = 237.37$ ,  $p < .001$ ], and of lag [ $F(2,46) = 70.98$ ,  $MS_e = 260.24$ ,  $p < .001$ ]. The interaction also was significant [ $F(6,138) = 6.09$ ,  $MS_e = 125.31$ ,  $p < .001$ ]. The mean results for all five conditions in Experiments 2 and 3, averaged over lags, are shown in Figure 7, separately for the first and the second targets.

The evidence in Figure 7 is unambiguous. As target–distractor similarity is increased, identification accuracy for both targets decreases correspondingly. This is precisely what would be expected on the basis of the contingent-capture hypothesis: the greater the target–distractor similarity, the greater the probability that a distractor will be processed as though it were a target. If a real target is then presented within a brief interval, its identification is impaired, because the distractor has preempted the resources required for processing the target. In other words, an AB deficit occurs that is indistinguishable from a conventional AB mediated by the first target.

It is interesting that the blank and the random-dot conditions yielded very similar results (Figure 6). The data for the two conditions were examined in a 2 (condition)  $\times$  3



**Figure 6.** Experiment 3: mean percentage of correct identifications of the second target, separately for each condition, given that the first target had been identified correctly. The results of Experiment 2 (letter condition) have been included for ease of comparison. Error bars represent one standard error of the mean.

(lag) ANOVA, which revealed a significant effect of lag [ $F(2,46) = 59.46$ ,  $MS_e = 205.43$ ,  $p < .001$ ]. But neither the effect of condition [ $F(1,23) = 2.41$ ,  $MS_e = 129.49$ ,  $p > .1$ ], nor the interaction [ $F(2,46) = 1.51$ ,  $MS_e = 93.22$ ,  $p > .1$ ], was significant. The similarity between the blank and the random-dot conditions strongly suggests that the mere presence of an RSVP stream in the display is not sufficient for producing attentional capture. That is, in the present paradigm, capture was not produced merely by early visual events, such as onset transients triggered by the items in the RSVP stream. Instead, capture depended on the degree of structural and/or conceptual similarity between the targets and the distractors.

In the four experiments reported thus far, the targets always consisted of alphabetical characters. This raises the issue of the generality of the present findings. That is, the possibility must be considered that the present evidence favoring contingent capture may be unique to the domain of letter targets. This possibility was examined in Experiment 4, in which the targets consisted of digits instead of letters.

#### EXPERIMENT 4

Experiment 4 was designed to examine the generality of the findings reported thus far. Specifically, we asked

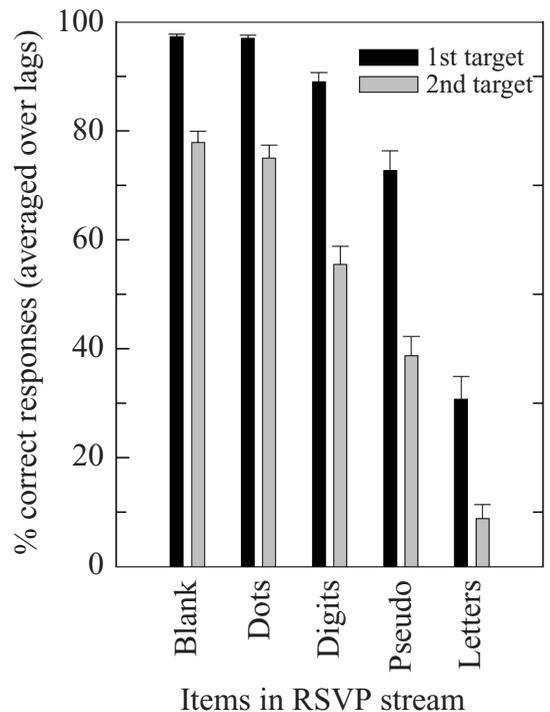
whether results obtained with digit targets would be homologous with those obtained with letter targets. On the contingent-capture hypothesis, just as the identification of letter targets is maximally impaired when the distractors are letters (Figure 7), identification of digit targets should be maximally impaired when the distractors are digits. This expectation was verified in Experiment 4, using three conditions in all. The targets were digits in every case, but the distractors were digits, letters, or pseudoletters.

#### Method

The same 24 subjects as those in Experiment 2 also served in Experiment 4. The order of participation in the experiments was counterbalanced. The apparatus, stimuli, and procedure were the same as those in Experiment 3, except for the following. The digit and the pseudoletter conditions were the same as those in Experiment 3, except that the targets were digits instead of letters, and the trailing masks were letters instead of digits. Similarly, the letter condition was the same as that in Experiment 2, except that the targets were digits and the masks were letters (see Table 1).

#### Results and Discussion

Mean correct identifications of the first target, collapsed across lags, were 92.5%, 93.2%, and 46.8% for the letter, pseudoletter, and digit conditions, respectively. A within-subjects ANOVA revealed a significant difference among these means [ $F(2,46) = 88.84$ ,  $MS_e = 573.99$ ,  $p < .001$ ]. Mean percentages of correct identifications of the



**Figure 7.** mean percentages of correct identifications of the first and second targets, averaged over all lags, when the distractors in the central RSVP were blank (i.e., omitted), dots, digits, pseudoletters, or letters. Error bars represent one standard error of the mean.

second target, averaged over the subjects, are presented in Figure 8. The data in Figure 8 were analyzed in a 3 (condition)  $\times$  3 (lag) ANOVA. All the effects were significant [condition,  $F(2,46) = 146.23$ ,  $MS_e = 466.24$ ,  $p < .001$ ; lag,  $F(2,46) = 21.18$ ,  $MS_e = 167.18$ ,  $p < .001$ , and condition  $\times$  lag,  $F(4,92) = 4.14$ ,  $MS_e = 104.91$ ,  $p < .01$ ].

Considered together with the results of the previous three experiments, the results of Experiment 4 attest to the generality of the principle that the level of impairment in target identification is directly related to the degree of target-distractor similarity. In Experiment 4, identification of the digit targets was severely impaired by digit distractors, whereas letter distractors had only a modest effect (Figure 8). The converse was true in Experiments 2 and 3, where letter distractors had a devastating effect on the identification of letter targets and digit distractors had a relatively modest effect. Patently, the critical factor across all the experiments was not the nature of the targets (letters vs. digits), but the similarity between the targets and the distractors.

Although these findings are consistent with contingent capture as the key factor underlying the deficit in target identification, an alternative account must be considered. It is possible that the present results may be an instance of the class of events denoted by the term *flanker effect*. The substantial literature that has accrued on this topic has recently been reviewed by Eriksen (1995). In a nutshell, the flanker effect is typically obtained in nonsearch experiments in which identification of a target is impaired by the presence of neighboring (flanking) distractor items. Importantly, the degree of impairment has been found to be directly related to the similarity between the target and the

distractors (Eriksen & St. James, 1986; Yeh & Eriksen, 1984).

On the face of it, there is a close parallel between the present findings and those for the flanker effect, especially as regards the dependence of the impairment on the similarity between targets and distractors. The theoretical interpretations, however, differ substantially from one another. Whereas contingent capture is said to represent events occurring mainly during the initial processing of the stimuli, the flanker effect is said to arise from interference with response planning and execution (Eriksen, 1995). From this perspective, it is important to distinguish between these alternative accounts of the present results. This was done in Experiment 5.

## EXPERIMENT 5

The main objective of Experiment 5 was to determine whether the present results were due to contingent capture or flanker interference. This was done by manipulating the spatial separation between the targets and the distractors. It is known that the flanker effect is critically dependent on the spatial proximity of targets and distractors and that the effect vanishes at separations beyond about  $1^\circ$  of visual angle (Pan & Eriksen, 1993). For example, in their experiment, Pan and Eriksen used two letters, slightly larger than  $0.2^\circ$ , that were mapped onto different keyboard responses. When the two letters were separated by  $0.5^\circ$ , presenting both letters simultaneously slowed responses to the target letter; however, separating the simultaneously presented letters by  $2.0^\circ$  eliminated the interference completely. In contrast, contingent capture is still fully in evidence at much larger separations (e.g.,  $4.7^\circ$  in Folk et al., 1992).

In the preceding four experiments, the targets and the distractors were always presented at adjacent screen locations, with center-to-center separations of approximately  $1^\circ$  of visual angle. Thus, it is possible that deficits in target identification might have been caused by flanker interference, not contingent capture. This possibility was investigated in Experiment 5 by separating the targets and the distractors by a distance sufficient to eliminate flanker interference and comparing performance with that in a condition in which spatial separation was within the range at which flanker interference is known to occur.

A  $2 \times 2$  factorial design was implemented, in which two types of distractors (letters or random dots) were crossed with two levels of spatial separation (near or far). As in the previous experiments, in the near condition, the targets (approximate size:  $1^\circ$  of visual angle) were separated by approximately  $1^\circ$  of visual angle. This was well within the range of visual separations at which robust flanker interference is obtained. In the far condition, the targets were separated by approximately  $3^\circ$  of visual angle, which is outside the range of flanker interference, but well within the spatial separation at which contingent capture effects are obtained.

Letter targets were used in all the conditions. On the basis of the preceding experiments, we expected target iden-

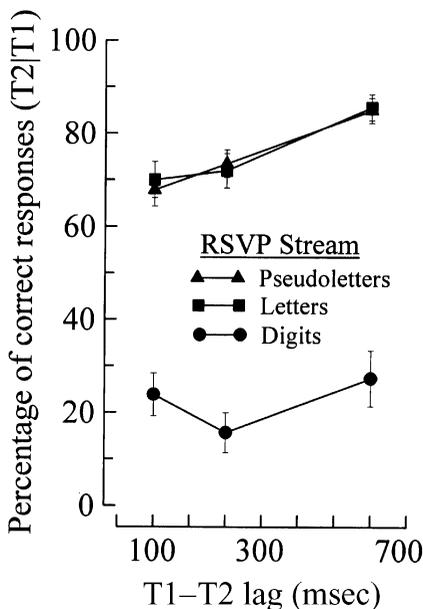


Figure 8. Experiment 4: mean percentage of correct identifications of the second target, separately for each condition, given that the first target had been identified correctly. Error bars represent one standard error of the mean.

tification to be more impaired with letter distractors than with random-dot distractors in the near condition. To the extent that the differential impairment was still in evidence in the far condition, contingent capture, rather than flanker interference, should be regarded as the chief underlying factor. The results favored contingent capture.

## Method

A new group of 24 subjects served in Experiment 5. The apparatus, stimuli, and procedure were the same those as in Experiment 1A, except for the following. The experiment comprised four conditions, resulting from the factorial combination of two types of distractors (letters or random dots, as in Experiments 2 and 3, respectively) and two levels of target–distractor separation (either 1° of visual angle from center to center, as in all the preceding experiments, or 3° of visual angle center to center). The targets were always letters, regardless of condition.

## Results and Discussion

Mean correct identifications of the first target, collapsed across lags, were 94.1%, 90.7%, 40.4%, and 55.5% for the dot-near, dot-far, letter-near, and letter-far conditions, respectively. A 2 (distractor)  $\times$  2 (separation) within-subjects ANOVA revealed the following significant effects: distractor,  $F(1,23) = 158.08$ ,  $MS_e = 899.52$ ,  $p < .001$ ; separation,  $F(1,23) = 16.68$ ,  $MS_e = 150.31$ ,  $p < .001$ ; and distractor  $\times$  separation,  $F(1,23) = 50.29$ ,  $MS_e = 121.16$ ,  $p < .001$ . Mean percentages of correct identifications of the second target, averaged over all the subjects, are presented in Figure 9. The data in Figure 9 were analyzed in a 2 (distractor)  $\times$  2 (separation)  $\times$  3 (lag) within-subjects ANOVA, which revealed the following significant effects:

distractor,  $F(1,23) = 482.87$ ,  $MS_e = 291.60$ ,  $p < .001$ ; separation,  $F(1,23) = 7.03$ ,  $MS_e = 332.07$ ,  $p < .05$ ; lag,  $F(2,46) = 82.19$ ,  $MS_e = 213.80$ ,  $p < .001$ ; distractor  $\times$  separation,  $F(1,23) = 45.45$ ,  $MS_e = 147.62$ ,  $p < .001$ ; distractor  $\times$  lag,  $F(2,46) = 16.20$ ,  $MS_e = 221.66$ ,  $p < .001$ ; and separation  $\times$  lag,  $F(2,46) = 10.77$ ,  $MS_e = 104.33$ ,  $p < .001$ . No other effects were significant.

Of principal interest is the finding that identification of the second target was poorer with letter distractors than with dot distractors not only when targets and distractors were close together (Figure 9A) but also when they were far apart (Figure 9B). This was also true for the first target. A separate ANOVA performed on the scores in Figure 9B, averaged across lags, confirmed that accuracy was lower with letter than with dot distractors [ $F(1,23) = 661.28$ ,  $MS_e = 158.01$ ,  $p < .001$ ]. On the premise that the flanker effect does not occur at spatial separations beyond about 1° of visual angle (Pan & Eriksen, 1993), these results support an interpretation in terms of contingent capture.

By the same token, the possibility is not ruled out that flanker interference could have been a factor at the near separation. This is implied by the finding that the difference between the letter and the random-dot conditions was greater when the stimuli were close together (Figure 9A) than when they were farther apart (Figure 9B), a pattern confirmed by the significant distractor  $\times$  separation interaction. Further analyses showed that, when the distractors were random dots, performance averaged across lags was about the same in the near as in the far conditions [ $F(1,23) = 2.30$ ,  $MS_e = 245.50$ ,  $p > .1$ ]. In contrast, when the distractors were letters, performance was significantly

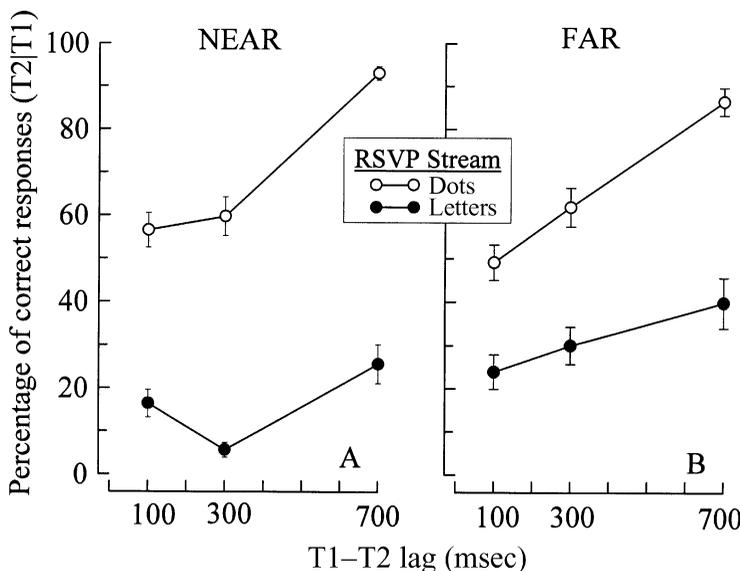


Figure 9. Experiment 5: mean percentages of correct responses when the distractors in the rapid serial visual presentation (RSVP) were dots or letters. (A): Near condition, in which targets were presented 1° from the central RSVP stream. (B): Far condition, in which targets were presented 3° from the central RSVP stream. Error bars represent one standard error of the mean.

better in the far than in the near condition [ $F(1,23) = 36.21$ ,  $MS_e = 234.20$ ,  $p < .001$ ]. Thus, increasing the spatial separation between the stimuli had no effect when target–distractor similarity was low but led to better performance when the similarity was high. Either flanker interference or contingent capture could account for the results at the near distance, but the results at the far distance are beyond what can be explained on the basis of flanker interference alone.

## GENERAL DISCUSSION

The main objective in the present study was to determine whether the presence of distractors in the display affects the magnitude and time course of the AB deficit. We did this by comparing the AB deficits obtained with two separate paradigms. One was a conventional two-target paradigm in which two sequential targets are presented at eccentric locations without any distractors (Duncan et al., 1994). The other was a modified two-target paradigm consisting of a conventional two-target paradigm augmented by an RSVP stream of task-irrelevant distractors presented at fixation. Experiment 1A showed that the presence of task-irrelevant distractors impairs identification of both the first and the second targets. Experiment 1B showed that this impairment was not dependent on the distractors and the targets appearing in different spatial locations, strongly suggesting that the impairment is mediated by a nonspatial mechanism. The next two experiments revealed that the level of interference is directly related to the degree of target–distractor similarity. Experiment 4 extended the generality of these findings by showing that the impairment is not confined to the domain of letter targets. Last, in Experiment 5 we juxtaposed two possible accounts of the interference effect—contingent capture and the flanker effect—and obtained evidence favoring the former.

Two aspects of the results are especially relevant to the objectives of the present work. First, target identification was substantially impaired by an RSVP stream of task-irrelevant distractors, but only when the targets and the distractors were similar. When the distractors did not share any defining attribute with the target, identification accuracy was virtually the same whether or not the display contained a distractor stream (Figure 6, blank vs. dots conditions). This strongly suggests that the interference effect depended not on the onset transients triggered by the items in the RSVP stream, but on the degree to which the targets and the distractors shared defining characteristics. Second, given such common characteristics, the distractors appeared to be processed in an obligatory and automatic fashion, despite instructions to ignore them (for similar results obtained with other experimental paradigms, see Eriksen, 1995; Gatti & Egeth, 1978; Milliken, Lupiáñez, Debner, & Abello, 1999). Considered together, these two aspects of the results point to contingent capture as a plausible underlying mechanism. In what follows, we will consider ways in which contingent capture may bring about a deficit in target identification.

## Mechanisms of Contingent Capture by Distractors

A scheme capable of accounting for the present results can be devised on the basis of the input-filtering model proposed by Visser et al. (1999) and further developed by Ghorashi et al. (2003). In that model, processing is said to occur in two broadly sequential stages. The first is a high-capacity, parallel-processing stage whose main functions are to detect potential targets and to encode them in readiness for further processing. This initial stage is followed by a capacity-limited serial stage in which stimuli are fully identified and encoded in a form suitable for subsequent report.

The basic tenets of this model bear distinct similarities to other two-stage models, such as those proposed by Broadbent and Broadbent (1987), Chun and Potter (1995), Di Lollo (1980), Duncan (1980), Hoffman (1979), and Wolfe (1994). What distinguishes the model of Visser et al. (1999) from these earlier two-stage models is the explicit filtering function assigned to the first processing stage. Initial processing is said to be performed by input-filtering mechanisms whose functional characteristics are programmable under the control of higher cortical regions.

Programming the input filter is said to be part of a goal-directed process aimed at tuning the visual system to those attributes and characteristics of incoming stimuli that are likely to prove useful for performing the task at hand. Stimuli that match the setting of the input filter are tagged as potential targets and become candidates for admission to the second stage. Other stimuli are denied access to the second stage and remain at the input stage, where they are subject to decay and to masking by trailing stimuli. The tuning of the input filter is thought to be relatively broad. When a task involves searching for a target in a complex display, the filter is said to be optimally tuned to the defining characteristic of the target. This means that distractors that share the target's distinguishing characteristic may also pass the filter. For example, if an observer is set to look for a red target, other red objects, such as a red distractor, may pass the filter and gain access to the second stage. An important aspect of this model is that processing at the second stage is strictly serial: Only one item can be processed at a time. Thus, if a stimulus arrives while the second stage is busy, it is delayed at the input stage even if it matches the filter's characteristics.

Interpretation of the present results in terms of this two-stage model is straight forward. When the target was a letter, as in Experiments 1, 2, and 3, it can be assumed that the input filter was optimally configured to pass letter-like stimuli. This meant that, to the extent that the distractors possessed letter-like features, they could pass the filter and gain access to Stage 2. Similarly, distractors that possessed digit-like features could gain access to Stage 2 in Experiment 4, in which the target was a digit.

Now consider the case in which a distractor that shared the target's defining characteristic was presented directly before the target. Because it matched the setting of the input filter, the distractor gained access to Stage 2, thereby

making it unavailable for the ensuing target. As a consequence, the target was delayed at the input stage until processing of the distractor was completed. While so delayed, the target was vulnerable to masking by the trailing stimulus, with consequent loss in accuracy of identification (this is consistent with the findings of Giesbrecht & Di Lollo, 1998, who showed that the AB vanished if the mask after T2 was omitted). Importantly, this pattern of events occurred only when the distractors shared the target's defining characteristic and, therefore, matched the setting of the input filter. As target–distractor similarity was decreased, the probability of a distractor's matching the input filter was reduced correspondingly. This meant that distractors were less likely to gain access to the second stage and preempt its use. In that case, the target could gain immediate access to Stage 2, thus avoiding the potentially harmful delay in Stage 1. The results in Figure 7 are entirely in line with this interpretation.

Other aspects of the results can also be explained naturally within this conceptual framework. As can be seen in Figure 7, when target–distractor similarity was increased, accuracy of identification became progressively more impaired not only for the second target, but also for the first. This is to be expected, because a distractor presented directly before a target may pass the input filter and preempt the use of Stage 2. The target, whether first or second, then remains vulnerable to masking throughout the ensuing delay in Stage 1. In essence, the same account applies to the deficit seen in either target. To be sure, the deficit seems to be reliably greater for the second target than for the first. The reasons for this asymmetry remain to be investigated. One possibility is that it may take less time for Stage 2 to process and reject a distractor than to fully process an actual target. This would result in shorter Stage 1 delays for the first target, which is always preceded by a distractor, than for the second target, which is always preceded by a real target as well as by distractors, especially at the shorter lags. The compounded effects would translate into a longer period of vulnerability to masking for the second target.

Some support for this suggestion is yielded by comparing first-target accuracy in Figure 7 with second-target accuracy at Lag 1 in Figure 6. Regardless of the type of distractor, it is always the case that first-target accuracy is substantially higher than second-target accuracy. This is consistent with the notion that processing of an immediately preceding first target delayed second-target processing for a substantial period, whereas processing of an immediately preceding distractor delayed first-target processing for a much shorter period, limiting its vulnerability to masking.

The two-stage model also offers a natural account of the exceptionally long AB deficits obtained in Experiment 2 when both the target and the distractors were letters (Figure 4) and in Experiment 4 when both the target and the distractors were digits (Figure 8). In both cases, a pronounced AB deficit was in evidence, with only minimal improvement by a lag of 700 msec. This contrasts sharply

with most conventional studies of the AB, in which the deficit has been found to vanish at lags beyond about 500 or 600 msec. Such long-lived effects can be understood by noting that distractors continued to be presented throughout the period of the intertarget lag. Thus, a distractor presented directly before the second target could preempt the use of Stage 2 regardless of lag. The second target would then be delayed in Stage 1 and would remain vulnerable to masking throughout the period of delay. Accuracy of second-target identification would therefore be impaired regardless of intertarget lag.

### Implications for Theories of the AB

The main theoretical development arising from the present work is a link between the role of distractors in the AB and contingent capture. This link, however, is only indirectly relevant to general theories of the AB, because it does not drive a distinction between them. Nevertheless, it is of interest to examine how current theories of the AB can encompass the present findings within their conceptual frameworks.

Two classes of models have been proposed to account for the AB. In one class, aptly named *bottleneck models*, the AB is said to arise from a processing bottleneck that occurs at a relatively late stage corresponding to Stage 2 in the present scheme. If access to that stage is preempted by the first target, a bottleneck develops, and the second target remains delayed at an earlier stage, where it is subject to masking and decay. The attendant impairment in second-target accuracy is manifested as the AB deficit.

In an alternative model, the AB deficit is said to arise from interference in visual short-term memory (VSTM; Shapiro & Raymond, 1994; Shapiro, Raymond, & Arnell, 1994). To gain access to VSTM, incoming stimuli must pass through preset templates set to match the two targets. The response item is then selected from all the stimuli currently in VSTM. Critical to this account is the assumption that stimuli other than the two targets may also gain access to VSTM, depending on their goodness of fit to the preset templates. Nontargets would then compete with the actual targets in VSTM for selection as the response items. An AB deficit is said to occur when an item other than the second target is selected. Thus, as the number of nontarget items admitted to VSTM is increased, the probability of successful second-target retrieval diminishes because of greater competition from distractor items, and the magnitude of the AB deficit is increased.

Direct comparisons between these two models have revealed broad structural and conceptual similarities (Isaak, Shapiro, & Martin, 1999; Shapiro, Arnell, & Raymond, 1997). It is immediately obvious, for example, that the input templates postulated in interference theory perform much the same selectivity functions as the first stage in two-stage models. In both cases, an initial selection is performed whose aim is to pass targets and ignore nontarget items. Thus, the probability of an incoming stimulus's gaining access to further processing will depend on its goodness of fit to the input template (interference theory)

or on how well it matches the filter setting at the input stage (two-stage models). In either case, distractors that share defining characteristics with the target are likely to gain access to further processing, whether VSTM or Stage 2.

In a nutshell, the two schemes offer broadly equivalent accounts of early processing but differ in their accounts of the events that underlie the AB at later processing stages. In the present experiments, the key variable under investigation was target–distractor similarity, a factor that both theories regard as affecting only the early input stage. For this reason, the present results can be encompassed equally well by both theories. We have seen how the increased magnitude of the AB with increasing target–distractor similarity (Figures 6 and 8) can be handled within a two-stage bottleneck model with dynamic input filtering. The interference model can handle those findings in much the same fashion on the basis of preset templates. That is, distractors that share features with the targets are more likely to match the templates and gain access to VSTM, thereby producing greater interference for target retrieval (e.g., Lavie & Tsai, 1994). The greater number of items in VSTM would interfere with retrieval not only of the second target, but also of the first. Again, this is in accordance with the present findings.

Clearly, a distinction between interference and bottleneck models cannot be achieved on the basis of factors that influence only the initial input stage, because the two schemes offer homologous accounts of processing at that stage. Rather, a distinction hinges on the manipulation of variables to which the models are differentially sensitive—notably, variables that affect processing beyond the initial input stage. From this perspective, the major contribution of the present work is not a distinction between extant theories of the AB, but the establishment of a nexus between the role of distractors in the AB and the processing of irrelevant items in contingent capture.

## REFERENCES

- BACHMANN, T., & ALLIK, J. (1976). Integration and interruption in the masking of form by form. *Perception*, *5*, 79-97.
- BACON, W. F., & EGETH, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, *55*, 485-496.
- BREITMEYER, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- BROADBENT, D. E., & BROADBENT, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*, 105-113.
- CHUN, M. M., & POTTER, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 109-127.
- DI LOLLO, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*, 75-97.
- DUNCAN, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, *87*, 272-300.
- DUNCAN, J., WARD, R., & SHAPIRO, K. L. (1994). Direct measurement of attentional dwell time in human vision. *Nature*, *369*, 313-315.
- ERIKSEN, C. W. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition*, *2*, 101-118.
- ERIKSEN, C. W., & ST. JAMES, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*, 225-240.
- FOLK, C. L., LEBER, A. B., & EGETH, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics*, *64*, 741-753.
- FOLK, C. L., & REMINGTON, R. W. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 847-858.
- FOLK, C. L., REMINGTON, R. W., & JOHNSTON, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 1030-1044.
- GATTI, S. V., & EGETH, H. E. (1978). Failure of spatial selectivity in vision. *Bulletin of the Psychonomic Society*, *11*, 181-184.
- GHORASHI, S. M. S., ZUVIC, S. M., VISSER, T. A. W., & DI LOLLO, V. (2003). Focal distraction: Spatial shifts of attention are not required for contingent capture. *Journal of Experimental Psychology: Human Perception & Performance*, *29*, 78-91.
- GIBSON, B. S., & KELSEY, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 699-706.
- GIESBRECHT, B. L., & DI LOLLO, V. (1998). Beyond the attentional blink: Visual masking by object substitution. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 1454-1466.
- HOFFMAN, J. E. (1979). A two-stage model of visual search. *Perception & Psychophysics*, *25*, 319-327.
- ISAAK, M. I., SHAPIRO, K. L., & MARTIN, J. (1999). The attentional blink reflects retrieval competition among multiple rapid serial visual presentation items: Tests of an interference model. *Journal of Experimental Psychology: Human Perception & Performance*, *25*, 1774-1792.
- LAVIE, N., & TSAI, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*, 183-197.
- MAKI, W. S., BUSSARD, G., LOPEZ, K., & DIGBY, B. (2003). Sources of interference in the attentional blink: Target–distractor similarity revisited. *Perception & Psychophysics*, *65*, 188-201.
- MILLIKEN, B., LUPÍÁÑEZ, J., DEBNER, J., & ABELLO, B. (1999). Automatic and controlled processing in Stroop negative priming: The role of attentional set. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1384-1402.
- OYAMA, T., WATANABE, T., & FUNAKAWA, M. (1983). Effects of test-mask similarity on forward and backward masking of pattern by patterns. *Psychological Research*, *45*, 303-313.
- PAN, K., & ERIKSEN, C. W. (1993). Attentional distribution in the visual field during same–different judgments as assessed by response competition. *Perception & Psychophysics*, *53*, 134-144.
- POTTER, M. C., CHUN, M. M., BANKS, B. S., & MUCKENHOUP, M. (1998). Two attentional deficits in serial target search: The visual attentional blink on an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 979-992.
- RAYMOND, J. E., SHAPIRO, K. L., & ARNELL, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 849-860.
- SCHEERER, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychological Research/Psychologische Forschung*, *36*, 71-93.
- SHAPIRO, K. L., ARNELL, K. M., & RAYMOND, J. E. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*, 291-296.
- SHAPIRO, K. L., & RAYMOND, J. E. (1994). Temporal allocation of visual attention: Inhibition or interference? In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 151-188). San Diego: Academic Press.
- SHAPIRO, K. L., RAYMOND, J. E., & ARNELL, K. M. (1994). Attention to visual pattern information produces the attentional blink in RSVP. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 357-371.
- SPELTING, G., & WEICHELGARTNER, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review*, *102*, 503-532.

- VISSER, T. A. W., BISCHOF, W. F., & DI LOLLO, V. (1999). Attentional switching in spatial and nonspatial domains: Evidence from the attentional blink. *Psychological Bulletin*, **125**, 458-469.
- WARD, R., DUNCAN, J., & SHAPIRO, K. (1997). Effects of similarity, difficulty, and nontarget presentation on the time course of visual attention. *Perception & Psychophysics*, **59**, 593-600.
- WOLFE, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.
- YEH, Y. Y., & ERIKSEN, C. W. (1984). Name codes and features in the discrimination of letter forms. *Perception & Psychophysics*, **36**, 225-233.

#### NOTES

1. A sharp-eyed reviewer noted that this manipulation also has the effect of varying distractor-mask similarity. That is, in Experiment 1A, both the distractors and the masks were digits, whereas in Experiment 2, the distractors were letters and the masks were digits. It could be suggested that accuracy in Experiment 2 was lower not because the distrac-

tors and the *target* were similar, but because the distractors and the *mask* were different.

Target-mask similarity has been studied extensively in the context of the AB (Isaak, Shapiro, & Martin, 1999), as well as in other contexts (Oyama, Watanabe, & Funakawa, 1983). To our knowledge, however, there have been no studies of the effects of similarity between *distractors* and masks. Thus, although not entirely outside the realm of possibilities, an account based on distractor-mask similarity would be ad hoc. Besides, distractor-mask similarity alone cannot encompass all of the present results. For example, in Experiment 3, letter target identification was better when the distractors and the masks were different (dots vs. digits) than when they were the same (both digits), which is the opposite of what was found in Experiments 1A and 2. This suggests that, if at all viable, any role of distractor-mask similarity must be strongly modulated by target-mask similarity.

(Manuscript received June 24, 2003;  
revision accepted for publication March 15, 2004.)