Journal of Experimental Psychology:
Human Perception and Performance
2003, Vol. 29, No. 1, 78–91

0096-1523/03/$12.00 DOI: 10.1037/0096-1523.29.1.78

Focal Distraction: Spatial Shifts of Attentional Focus Are Not Required for Contingent Capture

S. M. Shahab Ghorashi, Samantha M. Zuvic, Troy A. W. Visser, and Vincent Di Lollo
University of British Columbia

Contingent capture occurs when distractors that share the target's defining attribute capture attention and slow down target identification. This slowdown has been attributed to an involuntary attentional shift to the location of a pertinent distractor. The present study examined an additional source of delay: the time spent in processing pertinent distractors. In 7 experiments, distractors were presented at fixation, and targets were presented either at fixation or peripherally. Contingent capture invariably occurred when a salient distractor was presented within about 600 ms before the target, even when spatial shifts in attentional focus were ruled out. A 2-stage model is proposed in which stimuli must pass an input filter tuned to the target's defining attribute before gaining access to a high-level stage that is unavailable while a distractor is being processed.

When individuals first learn to drive, they quickly become acquainted with the basic operations of traffic lights. A green light means that they may proceed through the intersection, whereas a green arrow means that they may proceed only in the indicated direction. Nonetheless, despite extensive experience with traffic lights, a driver is often seen lurching straight ahead upon seeing a green left-turn arrow. The driver was obviously set to move ahead upon seeing a green light and did so inappropriately upon a green light of the wrong shape. This type of occurrence points to an attentional phenomenon known as *contingent capture,* often observed in studies of visual search, in which a target must be found among distractors. In such studies, a distractor captures attention when it shares the target's defining characteristic, resulting in slower identification of the target (Folk & Remington, 1998; Folk, Remington, & Johnston, 1992; Gibson & Kelsey, 1998).

Our objective was to investigate the mechanisms underlying contingent capture. In the experimental literature, the delay in target identification associated with contingent capture has been attributed exclusively to the time taken by an involuntary shift in the focus of attention to a distractor location. However, in our experiments we show that capture can also be obtained when distractors are presented at the focus of attention. This suggests that spatial shifts in attention are not the sole source of delay in contingent capture. Rather, the evidence points to the time taken to process the distractors as a second, independent, source of delay. Furthermore, we found that the mere presence of distractors, even if they share the target's defining attribute, does not automatically bring about contingent capture. Rather, capture occurs only during a brief period of less than about 600 ms while newly presented distractors are being processed. These findings are explained by a two-stage model in which incoming stimuli must pass an input filter in order to gain access to a capacity-limited, high-level processing stage.

Contingent Capture and Spatial Shifts of Attention

Contingent capture has been studied using a paradigm in which a nonpredictive cue is followed by a search display containing a target. For example, Folk et al. (1992) used a cue display in which sets of four dots surrounded each of four possible target locations. All sets of cue dots were white except one set that was red. The cue display was followed by a search display consisting of a red target among white distractors. When the red cue and red target appeared in different locations, response times were slower than when they appeared in the same location. Folk et al. ascribed this temporal deficit to an involuntary shift of attention to the location of the red cue, mediated by the observers' attentional control setting. Being set to attend to a red target, the observers' attention was captured by other red objects, such as the red cue. This slowed target identification because of the time taken to shift the attentional focus to the location of the cue when the target was somewhere else.

Slower response times, in this account, are attributed exclusively to the delay arising from an involuntary attentional shift to a nontarget location. It is plausible, however, that the delay may also be due, at least in part, to the time spent in processing the distractor itself. In this case, the delay would have two components: one arising from the time taken to perform the unintended attentional shift to the distractor's location, and the other arising from processing the item at that location. The latter source of delay is akin to what Theeuwes and Burger (1998) referred to as "identity intrusion effect."

S. M. Shahab Ghorashi, Samantha M. Zuvic, Troy A. W. Visser, and Vincent Di Lollo, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada.

This work was sponsored by a research grant from the Natural Sciences and Engineering Research Council of Canada to Vincent Di Lollo and by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada to Troy A. W. Visser. We acknowledge the expert assistance of Mark Rempel, who assisted in collecting and analyzing the data of Experiments 6 and 7.

Correspondence concerning this article should be addressed to Vincent Di Lollo, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada. E-mail: vince.dilollo@ubc.ca

Of relevance to the present work is the fact that the temporal delays arising from attentional shifting and involuntary processing were inextricably bound together in earlier experiments (e.g., Folk et al., 1992). This raises the question of whether contingent capture can be mediated solely by processing a distractor without an attendant spatial shift in attentional focus. In the present work, we examined this question using a nonsearch paradigm in which involuntary spatial shifts of attention to the location of a distractor could not occur because a stream of distractors was presented at fixation, followed by a target presented at an unpredictable eccentric location. Under these conditions, no spatial shifts to the distractors are to be expected because distractors are always presented at fixation. Therefore, any evidence of contingent capture must be attributed to the processing of the distractors themselves, and not to spatial shifts in the focus of attention.

Experiment 1

The purpose of Experiment 1 was to determine whether involuntary processing of a distractor can mediate contingent capture in the absence of a spatial shift of attentional focus. Each trial began with a stream of task-irrelevant distractors displayed at fixation, followed immediately by a target in an unpredictable eccentric location. To examine contingent capture, we used two conditions. In one condition, the distractors and the target were both letters. Therefore, the distractors shared the target's defining characteristic. In the other condition, the distractors were random-dot patterns that did not share the defining characteristic of the letter target.

When the distractors were letters (the *variable letter* condition), their identity was changed randomly from one frame to the next in the rapid serial visual presentation (RSVP) stream. When the distractors were dots (the *variable dot* condition), their configuration changed randomly from one frame to the next. Consistent with expectations based on contingent capture, response times were slower when the distractors were letters, suggesting that distractors that shared the target's defining characteristic captured attention.

Method

Observers. Thirty undergraduate students at the University of British Columbia participated for class credit. All reported normal or corrected-to-normal vision.

Apparatus, stimuli, and procedure. All stimuli were displayed on a Tektronix 608 oscilloscopic point plotter equipped with P15 phosphor. Observers viewed the displays from a distance of 57 cm, which was set by a headrest. The background and surrounding visual field were dark, except for dim illumination of the keyboard.

The experiment comprised two conditions: *variable letter* and *variable dot.* In the variable letter condition, the display consisted of an RSVP sequence of letter distractors presented in the center of the screen, and one letter target presented unpredictably in 1 of 12 locations, which were arranged as in a clock face with radius of 3°, centered on the center of the screen. The letters in the RSVP stream were selected randomly with replacement from all letters of the English alphabet, except *C* and *G*, with the constraint that the selected letter was not one of the two immediately preceding items. Each letter in the RSVP stream was displayed for 30 ms and was separated from the next letter by an interstimulus interval (ISI) of 70 ms, during which the screen was blank. This yielded a presentation rate of 10 items/s. The length of the RSVP stream was varied randomly between 5 and 10 items from trial to trial. The target was either a *C* or a *G*, displayed for 30 ms at an ISI of 70 ms after the last item in the RSVP stream. The luminance of the distractors was 15 cd/m^2, and that of the target was 10 cd/m^2, as measured by a Minolta LS-100 luminance meter. All letters subtended 0.78° vertically.

The variable dot condition was the same as the variable letter condition, except that each distractor was made up of 30 dots positioned randomly within an imaginary square of 0.78° side. The configuration of the dots was changed randomly from frame to frame in the RSVP stream. The luminance of the dots was 15 cd/m^2.

At the beginning of each trial a small fixation cross, subtending 0.25° of visual angle, was presented in the center of the screen at a luminance of 2 cd/m^2. Observers initiated each trial by pressing the spacebar, at which point the fixation cross disappeared and the RSVP sequence began. Observers were required to identify the target letter as quickly as possible by pressing either the left- or the right-arrow key, marked *C* or *G*, respectively. The fixation cross then reappeared to indicate that the next trial was ready to begin. The display sequence on a typical trial is illustrated in Figure 1.

Observers participated in the two conditions during one 30-min session. Each condition consisted of one block of 192 trials in which the target appeared in each of the 12 locations equally often. Each condition began with 10 practice trials, during which no data were recorded. The order of the conditions was counterbalanced across observers.

Results and Discussion

Trials on which errors were made were discarded from the analysis. This amounted to 5.2% of trials in the variable letter condition and 5.9% in the variable dot condition. No observer made more than 20% errors. Outliers were removed from the remaining trials using a procedure described by Van Selst and Jolicoeur (1994) that uses a floating criterion, based on sample size, to determine outliers. This resulted in the removal of a further 1.9% of trials from the variable letter condition and 2.2% from the variable dot condition.

Mean response times are illustrated by the segmented line in Figure 2 (the continuous line illustrates the results of Experiment 2). A paired-samples *t* test confirmed the graphical evidence that response times in the variable letter condition were slower than in the variable dot condition, $t(29) = 5.34$, $p < .001$.

Slower responding in the variable letter condition strongly suggests that observers were unable to ignore the distractors when they shared critical attributes with the target. This temporal deficit must be ascribed to the time taken to process the distractors

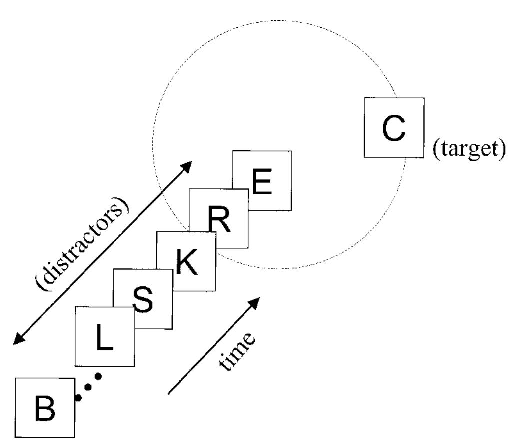

Figure 1. Schematic representation of the stimulus sequence in Experiment 1.

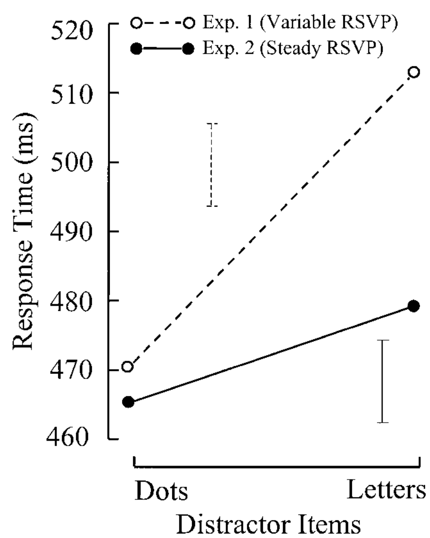

Figure 2. Segmented line: mean results of Experiment 1, averaged over observers, separately for the dot-distractor and letter-distractor conditions. The vertical segmented bar represents one standard error of the mean, averaged over both conditions. Continuous line: mean results of Experiment 2, averaged over observers, separately for the dot-distractor and letter-distractor conditions. The vertical continuous bar represents one standard error of the mean, averaged over both conditions. Exp. = experiment; RSVP = rapid serial visual presentation.

themselves, not to the time taken by misdirected spatial shifts in attentional focus. This is because by presenting all distractors at fixation and only the target in another location, an attentional shift away from fixation could only be toward the target, not toward a distractor. This is an important consideration because in conventional contingent-capture experiments, distractors were always presented away from fixation. This generated two potential sources of delay: the time taken to shift attention to the distractor and the time taken to process the distractor. Conventionally, the longer response times obtained when the distractors shared critical attributes with the target were ascribed exclusively to the misdirected shift in attentional focus. While not impugning the importance of spatial shifts in attention, the present results highlight the importance of involuntary processing of the distractors themselves in causing the temporal deficit seen in contingent capture. Our results thus substantiate Theeuwes and Burger's (1998) claim that the deficit obtained in contingent capture is due not only to the time taken by the involuntary attentional shift to the distractor, but also to the time taken to process the distractor. Furthermore, our results indicate that a spatial shift in attentional focus is not necessary for contingent capture to occur.

A simple account of the contingent capture obtained in this experiment is provided by a class of two-stage models in which processing is said to occur in two broadly sequential stages: an early stage, in which features relevant to the target are analyzed, followed by a capacity-limited serial stage, in which the stimuli are fully identified and encoded in a form suitable for subsequent report (Broadbent & Broadbent, 1987; Chun & Potter, 1995; Di Lollo, 1980; Duncan, 1980; Shapiro & Raymond, 1994; Visser, Bischof, & Di Lollo, 1999). Here, our principal objective is to make use of some basic tenets common to all these models to provide an account of the present results.

In this class of models, initial processing is said to be performed by input-filtering mechanisms whose functional characteristics are programmable under the control of higher brain regions. Programming the input filter is 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. Incoming stimuli that match the setting of the input filter gain direct access to further processing stages and, therefore, are processed rapidly and efficiently. Other stimuli do not gain direct access to the higher stage and, therefore, are processed less efficiently or are excluded altogether.

When a task involves searching for a target hidden among distractors, the input filter is optimally tuned to the defining characteristics of the target. This means that distractors that have similar characteristics may also pass the filter. For example, if an observer is set to look for a red target, the input filter will be configured to pass red objects. This will permit distractors to gain access to further processing if they are red in color. An important aspect of this model is that processing at the stage beyond input filtering is held to be strictly serial: only one item can be processed at a time. Thus, if a stimulus arrives while the high-level 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 class of two-stage filtering models is straightforward. Given that the target was a letter, it can be assumed that the input filter was optimally configured to pass letterlike stimuli. This means that in the variable letter condition, distractors matched the filter setting and gained access to high-level processing. It is likely that, on some trials, the distractor immediately prior to the target passed the input filter, thus preempting the high-level stage. As a result, high-level processing of the target was delayed until processing of the distractor was completed. This delay in target processing gave rise to the slower response times in the variable letter condition. In contrast, no delay occurred in the variable dot condition because the distractors did not match the setting of the input filter and, therefore, could not gain access to the high-level stage. This means that the target always gained immediate access to high-level processing, resulting in faster response times than in the variable letter condition.

A question that arises directly from the results of the present experiment is why were observers unable to ignore the central RSVP stream of distractors, given that they knew that the target would never be presented at that location. A plausible answer to that question can be based on the spatial distribution of attention adopted by the observers. On any given trial, the target could appear anywhere on the perimeter of an imaginary circle centered at fixation. Perception of the target, therefore, would be optimized if the spatial distribution of attention were to encompass the area defined by the imaginary circle. Because that area included the center of the screen, the central RSVP stream of distractors fell within the attended area and was processed, at least to some extent, along with the target. At first, this account may seem to run afoul of recent evidence that observers may be able to attend selectively to rings around fixation (Linnell & Humphreys, 2000). In fact, what the present evidence suggests is that, although observers may be able to attend selectively to the periphery of a circular area, they may not be able to ignore completely stimuli presented at the center of that area.

In the remainder of the present work, we report six additional experiments designed to test specific predictions from the two-stage model outlined earlier and to provide converging evidence obtained with different experimental paradigms. We began by testing a simple corollary implicit in this account. If it is true that response times in the variable letter condition are slow when the target arrives while the high-level stage is busy, then response times should be faster if the target were to arrive while that stage is free. This is because the target would gain immediate access to high-level processing, thereby obviating any delay at the input-filtering stage. This prediction was tested in Experiment 2.

Experiment 2

Our objective in Experiment 2 was to replicate Experiment 1 under conditions in which the high-level stage was likely to be free when the target arrived. To this end, we altered the contents of the central RSVP stream. Whereas in Experiment 1 a different item was presented in each successive frame, in Experiment 2 the item remained the same from frame to frame. For example, in the *steady letter* condition, a single letter was chosen randomly at the beginning of each trial and was flashed repeatedly throughout the RSVP stream. We reasoned that presenting the same letter repeatedly would allow its processing to be completed early in the stream, well before the arrival of the target. Thus, by the end of the RSVP stream, the high-level stage should be free, and the target should gain immediate access to high-level processing. In practice, this means that it should no longer matter whether the distractors shared the target's defining feature. Response times to the letter target should be approximately the same, whether the distractor stream contains a letter or a group of random dots.

Method

Observers. Thirty undergraduate students at the University of British Columbia participated for class credit. All reported normal or corrected-to-normal vision.

Apparatus, stimuli, and procedure. Apparatus, stimuli, and procedures in Experiment 2 were the same as in Experiment 1 with the following exceptions. When the distractors were letters, the RSVP stream consisted of a single letter, selected randomly at the beginning of each trial, as in Experiment 1, and flashed repeatedly throughout the RSVP stream (*steady letter* condition). When the distractors were random dots, the stream consisted of one random configuration of dots, selected at the beginning of each trial and flashed repeatedly throughout the stream (*steady dot* condition). The display sequence on a typical trial in the steady letter condition was the same as illustrated in Figure 1, except that the same letter was presented repeatedly throughout the distractor stream.

Results and Discussion

Trials in which errors were made were discarded from the analysis. This amounted to 6.0% of trials in the steady letter condition and 5.9% in the steady dot condition. No observer made more than 22% errors. The remaining trials were screened for outliers using the same procedure as in Experiment 1. This resulted in a further 2.2% of trials being removed from the steady letter condition and 2.3% from the steady dot condition.

Mean response times are illustrated by the continuous line in Figure 2. A paired-samples *t* test between the two conditions failed to reach statistical significance, $t(29) = 1.97$, $p > .05$. To examine the effects of steady versus variable presentation, an analysis of variance (ANOVA) was performed on the combined results of Experiments 1 and 2 with one between-subjects variable (mode of presentation: varied vs. steady) and one within-subject variable (distractor type: letters vs. dots). The analysis revealed a significant effect of distractor type, $F(1, 58) = 27.53$, $p < .001$, $MSE = 889.74$, with slower response times when distractors were letters. Importantly, the interaction effect between mode of presentation and distractor type was also significant, $F(1, 58) = 6.62$, $p < .05$, $MSE = 889.74$, indicating that presenting the same distractor repeatedly reduced response times when the distractors were letters but not when they were random dots. The effect of mode of presentation was not significant, $F(1, 58) = 1.99$, $p > .05$, $MSE = 5{,}479.16$.

The pattern of results in Figure 2 is entirely in keeping with predictions from a filtering model. Especially notable is the interaction effect between mode of presentation and distractor type. With random-dot distractors, mode of presentation (variable vs. steady) had no effect on response times. In contrast, with letter distractors, response times were significantly slower in the variable letter condition, in which each frame in the RSVP stream contained a new letter, than in the steady letter condition, in which the same letter was flashed repeatedly throughout the stream. From the standpoint of a filtering model, these results can be explained as follows. When the distractors were random dots, the high-level stage was always free because random dots could not pass the input filter that was set to pass only letterlike stimuli. Therefore, regardless of whether the random-dot pattern was steady or changing, the target always gained immediate access to high-level processing, and response times were correspondingly short. This explains why mode of presentation had no effect on response times when distractors were random dots.

The high-level stage was also likely to be free in the steady letter condition, in which the distractor was a single letter flashed repeatedly throughout the RSVP stream. Being a letter, the distractor passed the input filter and gained access to further processing. However, the high-level stage was engaged only during the early frames in the display sequence. By the end of the RSVP stream, the letter had been processed and the high-level stage was again free and available for processing the upcoming target. Consequently, response times in the steady letter condition were relatively short and comparable to those in the random-dot conditions. The opposite was true in the variable letter condition, in which a new letter was presented in each successive frame. In that condition, the probability was high that a letter near the end of the stream had passed the input filter and was being processed when the target arrived. As a result, the target was delayed at the input stage until the high-level stage was again free, and response times increased correspondingly. This explains why target identification was substantially faster in the steady letter than in the variable letter condition.

This line of reasoning leads to a testable prediction. Response times in the variable letter condition were said to be relatively slow because the target arrived while the high-level stage was busy processing a distractor that had been presented near the end of the RSVP stream. This is because the closer a new letter is to the end of the stream, the more likely it is to delay the processing of the target by preempting the high-level stage. Conversely, the greater the number of frames in which the same letter is presented repeatedly near the end of the stream, the higher the probability that the high-level stage will be free and directly accessible by the target, thus decreasing response times. This supposition was tested in Experiment 3 by varying systematically the proximity of a new letter to the end of the RSVP stream.

Experiment 3

The proximity of a new letter to the end of the RSVP stream was manipulated in Experiment 3 by varying the number of frames for which the same letter was repeated near the end of the stream. This was done by modifying the display sequence in the steady letter condition of Experiment 2. Instead of flashing a single letter throughout the stream, we used two different letters. The stream began with one letter flashed repeatedly throughout the early frames. The display was then switched to the second letter for the remaining frames. We reasoned that as the number of postswitch frames was increased, the probability would also increase that processing of the second letter would be completed before the target letter arrived. For example, processing of the second letter would be more likely to be completed when it was presented for the last six frames than for only the last frame in the RSVP stream. On this reasoning, the probability of the high-level stage being free would increase with the number of postswitch frames. In turn, this would increase the probability of the target gaining immediate access to high-level processing and decrease response times correspondingly.

Method

Observers. Twenty undergraduate students at the University of British Columbia participated for class credit. All reported normal or corrected-to-normal vision.

Apparatus, stimuli, and procedure. Apparatus, stimuli, and procedures in Experiment 3 were the same as in Experiment 2 with the following exceptions. On each trial, the length of the RSVP stream was varied randomly between 10 and 15 distractors. As in the previous two experiments, there were two types of distractors: letters and random dots. When the distractors were letters, the stream contained only two letters, selected at random on each trial. The display sequence began with the presentation of one of these letters for several successive frames and was then switched to the other letter for the remaining frames. This sequence is illustrated schematically in Figure 3. The number of preswitch frames was varied randomly on each trial between 4 and 15. The number of postswitch frames was six, three, one, or zero (a condition identical to the steady letter condition in Experiment 2). That is, a switch from the first to the second letter occurred in the sixth-to-last frame of the RSVP stream, in the third-to-last frame, in the last frame, or not at all. In the condition in which the distractors were random dots, the RSVP stream began with a given configuration that was switched to a different configuration at the same points in the stream as when the distractors were letters.

Observers participated in two blocks of 384 trials each: one containing only letter distractors and the other containing only dot distractors. The order of the two blocks was counterbalanced across observers. The four lengths of the postswitch stream were presented randomly across trials within a block, with 96 trials for each of the four lengths.

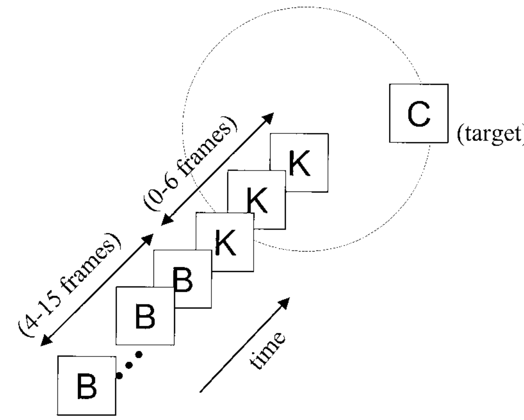

Figure 3. Schematic representation of the stimulus sequence in Experiment 3.

Results and Discussion

Trials on which errors were made were discarded from the analysis. This amounted to 5.3% of trials in the letter conditions and 5.7% in the dot conditions. The remaining trials were screened for outliers using the same procedure as in the previous two experiments. This resulted in the removal of a further 2.3% of trials in the letter conditions and 2.0% from the dot conditions. No observer made more than 17% errors in the experiment.

Mean response times are shown in Figure 4, separately for the letter and dot conditions and the number of postswitch frames. The two conditions in which there were no postswitch frames (i.e., the conditions identical to the steady letter and steady dot conditions in Experiment 2) were used as baselines against which to compare the remaining levels in the letter and in the dot conditions, respectively. The baseline levels are illustrated as continuous or segmented thin lines in Figure 4. The data in Figure 4 were analyzed in a 2 (distractor type: letter vs. dot) $\times$ 4 (number of postswitch frames: zero vs. one vs. three vs. six) within-subject ANOVA. The analysis revealed significant effects of distractor type, $F(1, 19) = 23.42$, $p < .001$, $MSE = 1{,}068.56$, and number of postswitch frames, $F(3, 57) = 19.56$, $p < .001$, $MSE = 93.50$. Notably, the interaction effect was also significant, $F(3, 57) = 12.39$, $p < .001$, $MSE = 98.77$. This interaction effect is consistent with the graphical evidence in Figure 4 that increasing the number of postswitch frames reduced response times when the distractors were letters but not when they were random dots. Separate analyses were performed on the letter and dot conditions to examine response times as a function of the number of postswitch frames. A significant effect was revealed with letter distractors, $F(3, 57) = 24.10$, $p < .001$, $MSE = 124.97$, but not with random-dot distractors, $F(3, 57) = 0.60$, $p > .05$, $MSE = 67.30$. A paired-samples t test revealed that the two baselines differed significantly from one another, $t(19) = 3.04$, $p < .05$.

The finding of principal interest in Experiment 3 was that, with letter distractors, response times became progressively shorter as the number of postswitch frames was increased (Figure 4). When the new letter was presented for only one frame, response times were relatively long and comparable to those in the variable letter condition in Experiment 1 (Figure 2, segmented line). In contrast,

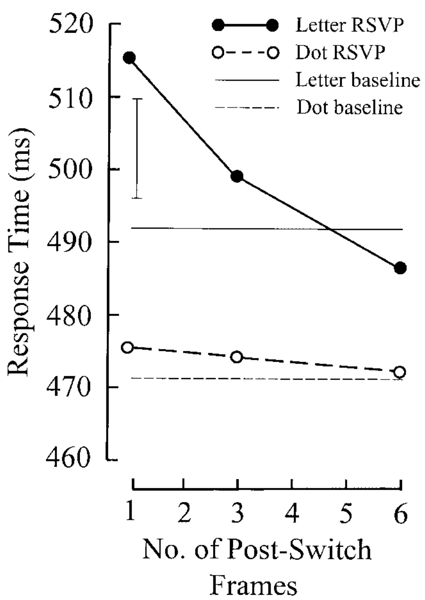

Figure 4. Mean results of Experiment 3, averaged over observers, separately for the letter-distractor and dot-distractor conditions. The letter baseline and dot baseline represent the mean response time when the number of postswitch frames was equal to zero in the letter-distractor and the dot-distractor conditions, respectively. The vertical bar represents one standard error of the mean, averaged over all conditions. RSVP = rapid serial visual presentation.

when the new letter was on view for six consecutive frames, response times were shorter and comparable to those in the steady letter condition in Experiment 2 (Figure 2, continuous line). It must be emphasized that this pattern of results was obtained only when the distractors were letters that shared the defining characteristic of the target. When the distractors were random dots that shared no salient features with the target, response times were unaffected by the number of postswitch frames.

It is interesting to note that response times at baseline were faster when the distractors were dots than when they were letters (Figure 4). A similar, though nonsignificant, relationship was obtained in Experiment 2. This suggests that some high-level processing capacity may have been preempted when the distractors were letters, even when the same letter was presented repeatedly throughout the RSVP stream. However, this in no way impugns the principal finding that, with letter distractors, response times were negatively related to the number of frames for which the same letter was presented near the end of the stream.

The results in Figure 4 are entirely in keeping with the hypothesis that processing of the target is delayed if it is presented while the high-level stage is busy processing a distractor. The actual period for which the high-level stage was busy with a new letter can be inferred directly from the data in Figure 4, bearing in mind that the SOA between successive frames in the RSVP sequence was 100 ms. With letter distractors, response times were significantly above baseline when the new letter was presented for only 100 ms (one frame) before the target, were still marginally above baseline when the new letter was presented 300 ms (three frames) before the target, but were at or just below baseline when a new letter was presented 600 ms (six frames) before the target. This suggests that the time taken to process a new item was in the range 400–600 ms, which is comparable with homologous estimates obtained in studies of the *attentional blink* (Raymond, Shapiro, & Arnell, 1992) and of the *dwell time of attention* (Duncan, Ward, & Shapiro, 1994). This comparison is pursued further in the General Discussion section.

Implicit in this account is the assumption that activity in the high-level stage is relatively short lived and time locked to the onset of a new stimulus. After an initial burst of activity triggered by a new stimulus, the high-level stage does not continue to be active if the stimulus remains on view beyond a certain duration. This means that, once processed, a stimulus that remains on view does not keep on being reprocessed as though it were a new stimulus. A possible mechanism underlying the distinction between "new" and "old" stimuli has been proposed by von Holst (1954), who suggested that any given stimulus leaves an image of itself in the visual system to which subsequent inputs are compared. If subsequent inputs match the internal image (i.e., if the stimulus is old), no further processing is required. But if a mismatch occurs, the latter input is processed as a new stimulus. A similar concept has been proposed by Robertson (1996), who suggested that incoming stimuli leave a multidimensional "footprint" in the visual system, which then influences the processing of subsequent inputs.

Regardless of underlying mechanisms, a distinction between new and old stimuli is essential for explaining why a distractor that has been on view for some time produces little, if any, contingent capture, even if it shares a defining characteristic with the target. To obtain contingent capture, the target must be presented while the distractor is new, namely, while it is in the course of being actively processed. For this reason, contingent capture was obtained in the variable letter condition in Experiment 1 and when the new letter was displayed for only one frame in Experiment 3, but not in the steady letter condition in Experiment 2 or when the new letter had been on view for an extended period in Experiment 3.

Experiment 4

In the preceding three experiments, the period for which the target was delayed at the input stage was estimated directly by measuring response times. In the next two experiments we took a different approach to the same goal. Instead of measuring response times, we estimated the period for which the target remains vulnerable to backward masking by a pattern presented shortly afterward in the same location. Our plan was to combine estimates of vulnerability to masking with direct estimates of processing delays to provide converging evidence for a two-stage account of contingent capture.

Backward masking refers to a reduction in the visibility of a brief stimulus (the target) caused by a second stimulus (the mask) whose onset trails the onset of the target by a brief interval of time known as *stimulus onset asynchrony* (SOA). It is known that the strength of masking is inversely related to the SOA. In general, masking becomes progressively weaker as the SOA is increased. In addition, recent evidence indicates that the strength of masking is negatively related to the degree of attention deployed to the target: Masking is strongest when the target is unattended, but its strength diminishes as attention is increasingly deployed to the target (Di Lollo et al., 2000; Giesbrecht & Di Lollo, 1998; Mack & Rock, 1998).

The negative relationship between masking and attention can be used as a basis for testing the present account of contingent capture if it is assumed that, while delayed at the input stage, the target is unattended. On this assumption, the period for which the target is vulnerable to masking will vary with the period of delay at the input stage. In the context of the present work, the period of vulnerability will be indexed by the response times obtained in the preceding three experiments. For example, the target will be vulnerable to masking over a longer SOA in the variable letter condition, in which response times were long, than in the steady letter condition, in which response times were shorter (Experiments 1 and 2; Figure 2).

Experiment 4 was a replication of Experiments 1 and 2, except that, instead of measuring response times, we measured the period for which the target was vulnerable to masking. The period of vulnerability was estimated by means of a dynamic threshold-tracking technique known as PEST (parameter estimation through sequential testing; Taylor & Creelman, 1967). Within a run, PEST varied the SOA between the target and a trailing mask dynamically so as to converge toward a predetermined level of identification accuracy. The SOA was automatically decreased when accuracy was too high, and it was increased when accuracy was too low. At the end of a run, PEST reported the critical SOA, namely, the SOA that yielded the predetermined level of accuracy. In agreement with the outcomes of Experiments 1 and 2, we found that the critical SOA was longer in the variable letter condition than in either the steady letter condition or in the conditions in which the distractors were random dots.

Method

Experiment 4 comprised the same four conditions as Experiments 1 and 2: variable letter, variable dot, steady letter, and steady dot. One group of 18 observers participated in the variable letter and variable dot conditions, and a separate group of 18 observers participated in the steady letter and steady dot conditions. The order of conditions was counterbalanced across observers within each group. The displays in each condition were identical to those in the corresponding condition in Experiments 1 and 2, except that a trailing mask, consisting of a digit drawn randomly from the set 2, 3, 4, 5, 7, and 8 was presented for 30 ms in the same location as the target, at a luminance of 10 cd/m^2. The digits 0, 1, 6, and 9 were never used as masks because of their similarity to some letters.

The SOA between the target and the mask was varied dynamically by a threshold-tracking procedure (PEST; Taylor & Creelman, 1967) set to converge on the SOA at which observers made approximately 80% correct identifications of the target. We refer to this as the *critical SOA.* One PEST run comprised 192 trials, with the initial SOA set at 300 ms. Throughout a run, PEST kept track of the observer's performance over trials and performed a statistical test (Wald, 1947) to determine whether accuracy was above or below 80% correct identifications. If it was above, the SOA was reduced by one step, initially set at 40 ms. If accuracy was still above 80%, the staircase continued to descend in steps of 40 ms. When accuracy fell below 80%, the direction of the staircase was reversed, and the SOA was increased by 20 ms. Every time the direction of the staircase was reversed, the step size was halved to a minimum of 2 ms. When the staircase continued in the same direction for more than two steps, the step size was doubled to a maximum of 40 ms. Thus, the resolution of the staircase became progressively finer as it converged on the critical SOA.

Results and Discussion

Within a PEST run of 192 trials, estimates of the critical SOA stabilized within about 50 trials. Therefore, the estimates obtained in the last 150 trials were averaged, separately for each observer in each condition, to yield the critical SOAs illustrated in Figure 5. The data were analyzed in an ANOVA, with one between-subject variable (presentation mode: variable vs. steady) and one within-subject variable (distractor type: letter vs. dot). This analysis revealed significant effects of presentation mode, $F(1, 34) = 5.80$, $p < .05$, $MSE = 895.27$, distractor type, $F(1, 34) = 39.35$, $p < .001$, $MSE = 205.35$, and a significant interaction effect, $F(1, 34) = 14.02$, $p < .001$, $MSE = 205.35$. Individual t tests revealed that the critical SOA was significantly longer in the variable letter than in the steady letter condition, $t(34) = 3.88$, $p < .001$. In contrast, the critical SOAs in the variable dot and steady dot conditions did not differ significantly from one another, $t(34) = 0.54$, $p > .05$. It should also be noted that the critical SOA in the steady letter condition did not differ significantly from either the steady dot condition, $t(17) = 1.68$, $p > .05$, or the variable dot condition, $t(34) = 0.69$, $p > .05$.

The results of Experiment 4 provide converging evidence consistent with the present account of contingent capture. Critical SOAs were longest in the variable letter condition, suggesting that vulnerability to masking extended over a longer period when the distractors shared a defining characteristic with the target. In contrast, critical SOAs in the remaining three conditions (Figure 5) were shorter and indistinguishable from one another, consistent with the claim that the period of target vulnerability was shorter when the same distractor was presented repeatedly (steady letter condition) or when the distractors shared no salient features with the target (random-dot conditions).

The pattern of results in Figure 5 is similar to that in Figure 2, suggesting that response times and critical SOA provide equivalent

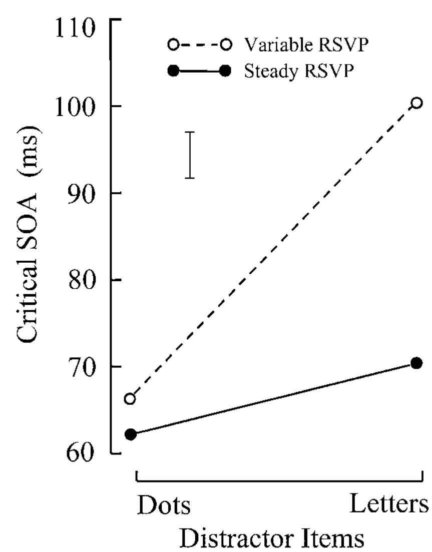

Figure 5. Mean results of Experiment 4, averaged over observers, separately for the variable and the steady RSVPs, in the dot-distractor and letter-distractor conditions, respectively. The vertical bar represents one standard error of the mean, averaged over all conditions. RSVP = rapid serial visual presentation; SOA = stimulus onset asynchrony.

estimates of the period for which the target is delayed at the input stage while the high-level stage is busy processing a distractor. In addition, the present results show that contingent capture can be obtained with response accuracy as the dependent measure, as well as with the more conventional response time measure. This parallels a similar finding by Egeth, Folk, Leber, and Nakama (2000) and underscores the robustness of the phenomenon. Further evidence for the equivalence of response time and response accuracy as indices of the target's delay at the input stage was obtained in Experiment 5, in which the design of Experiment 3 was replicated using critical SOA instead of response time as the dependent measure.

Experiment 5

Method

The design of Experiment 5 was the same as that of Experiment 3 except that the dependent measure was critical SOA instead of response time. Twelve observers served in two conditions, in counterbalanced order: in one the distractors were letters and in the other they were random dots. When the distractors were letters, the RSVP stream began with one letter flashed repeatedly in successive frames and then switched to another letter that was shown repeatedly for zero, one, three, or six frames. The same sequence was used when the distractors were random dots. The display sequence on a typical trial is illustrated in Figure 3, except that a mask was presented after the target, as in Experiment 4. The SOA between the target and the trailing mask was varied dynamically by means of the PEST procedure, as in Experiment 4.

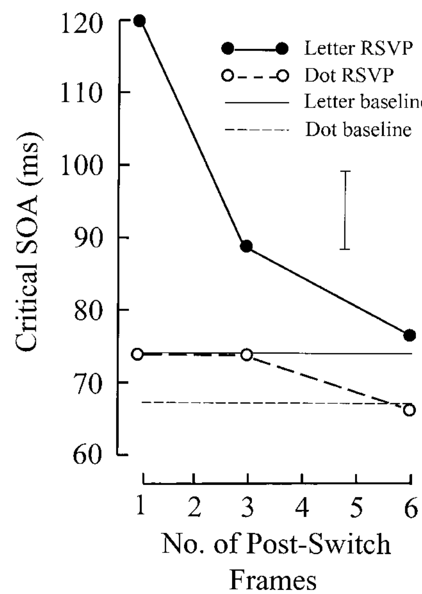

Figure 6. Mean results of Experiment 5, averaged over observers, separately for the letter-distractor and dot-distractor conditions. The letter baseline and dot baseline represent the mean response time when the number of postswitch frames was equal to zero in the letter-distractor and the dot-distractor conditions, respectively. The vertical bar represents one standard error of the mean, averaged over all conditions. RSVP = rapid serial visual presentation; SOA = stimulus onset asynchrony.

Results and Discussion

Mean critical SOAs are shown in Figure 6, separately for each RSVP condition and number of postswitch frames. The data were analyzed in a within-subject ANOVA comprising two variables: distractor type (letter vs. dot) and number of postswitch frames (zero vs. one vs. three vs. six). The analysis revealed significant effects of distractor type, $F(1, 11) = 5.20$, $p < .05$, $MSE = 1{,}869.12$, and number of postswitch frames, $F(3, 33) = 12.24$, $p < .001$, $MSE = 297.94$. Notably, the interaction effect was also significant, $F(3, 33) = 5.64$, $p < .05$, $MSE = 351.70$. A paired-samples t test indicated that the difference between the two baselines in Figure 5 was not significant, $t(11) = 0.83$, $p > .05$.

The results of Experiment 5 provide further converging evidence in support of a two-stage model of contingent capture. As expected, the number of postswitch frames in the RSVP stream had a substantial effect on critical SOA when the distractors were letters but not when they were random dots. With letter distractors, critical SOA was longest when the RSVP stream contained only one postswitch frame, and it became progressively shorter as the number of postswitch frames was increased. In contrast, with dot distractors, the location of the switch had no effect. As was the case in Experiment 4, these results are in keeping with the claim that response times and critical SOA are equivalent measures of the period for which the target is delayed at the input stage while the high-level stage is busy processing a salient distractor.

Experiment 6

We have noted in the foregoing how involuntary spatial shifts in attentional focus have been regarded as one source of time cost in contingent capture (Folk et al., 1992). The present work points to a second source: involuntary processing of a distractor that shares the target's defining feature. The evidence, however, is not entirely unambiguous. The fact that in our experiments the target was always a letter introduced a confound: In the conditions in which the distractors shared the target's defining feature, the distractors were necessarily always letters. In conditions in which the distractors did not share the target's defining feature, the distractors were always random-dot patterns. Therefore, the response time differences obtained in the present experiments (e.g., Figure 2) cannot be ascribed unambiguously to whether the distractors shared the target's defining feature. Rather, the differences could have arisen from factors inherent in the nature of the distractors themselves (letters vs. dots), regardless of whether they shared the target's defining feature.

In the following paragraphs, we consider two sources of ambiguity and show how each can be resolved by implementing a factorial design in which the type of distractors and the type of target are fully crossed.

First, we consider the possibility that the time costs seen in our experiments were mediated exclusively by spatial shifts in attentional focus. An argument along these lines can be constructed from a spotlight/zoom-lens metaphor of attention (Sperling & Weichselgartner, 1995; but see Driver & Baylis, 1989). Suppose that at the start of each trial observers adopt a wide zoom-lens type of attentional focus of some fixed diameter. Such a supposition is plausible because the target always appears somewhere on a constant-sized circle around fixation. Presentation of the RSVP stream would affect the dimensions of the spotlight in different ways, depending on whether the distractors are random dots or letters. Being unstructured and meaningless, random-dot patterns are unlikely to capture attention. Therefore, observers can maintain a wide zoom-lens mode until the target appears and then proceed to identify it. In contrast, letters may have a special status relative to random-dot patterns because they are well-known perceptual units with meaning attached to them, which makes them difficult to ignore. Thus, letter distractors may capture attention just because they are letters, not because they share the target's defining characteristic. This would cause the zoom lens to become narrowly focused on the RSVP stream. 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 RT cost associated with letter distractors is attributable not to the time wasted in processing a letter in the RSVP stream but to the time wasted in shifting the location of the spotlight. We refer to this as the *spotlight hypothesis.*

Second, we consider the possibility that letters and random-dot patterns may capture attention in equal measure, but random dots may be processed and rejected more quickly in the postselection stage due, for instance, to their lack of meaning. In this case, the RT cost would be attributable not to the time wasted in processing a distractor but to the fact that random dots can be processed more rapidly than letters. We refer to this as the *postselection hypothesis.*

Both of these sources of ambiguity can be resolved by implementing a factorial design in which the type of distractor is fully crossed with the type of target. In practice, it is difficult to use random dots as targets because they lack a readily identifiable characteristic that can be used in a two-alternative response. For that reason, we replaced the random-dot patterns with oriented line segments. In Experiment 6, the target was either a letter (*C* or *G*) or an oriented line segment (45° or 135°). The distractors were either letters (excepting *C* and *G*) or line segments whose orientation was varied randomly about the vertical or horizontal axis but well away from the diagonal axes.

Illustrated in Figure 7 are predictions derived from three hypotheses: the spotlight hypothesis, the postselection hypothesis, and the present hypothesis—to which we refer as the *contingent-processing hypothesis*—in which the time cost is said to arise from involuntary processing of a distractor that shares the target's defining characteristic.

Predictions from the spotlight hypothesis are illustrated in Figure 7A. According to this hypothesis, response times are determined by the nature of the distractors. In the condition in which the distractors are simple line segments, response times are expected to be relatively short because observers can maintain a wide zoom-lens type of attentional focus throughout a trial. In the condition in which the distractors are meaningful letters, however, response times are expected to be relatively long because the focus of the zoom lens becomes narrower during the RSVP stream and needs to be repositioned and refocused when the target arrives. In this account, therefore, response time to identify the target depends exclusively on the type of distractor. Furthermore, the relationship between distractors and target is not considered important. Also illustrated in Figure 7A are predictions from the postselection hypothesis: When the distractors are line segments, response times are relatively short because line-segment distractors can be processed and rejected faster than letter distractors that, therefore,

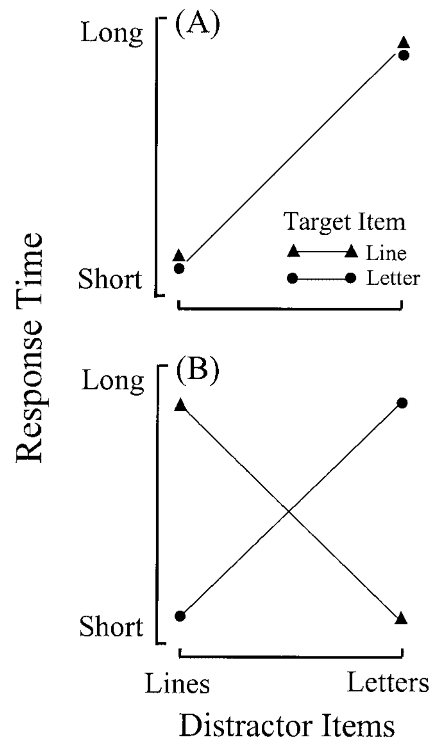

Figure 7. Schematic representation of theoretical predictions in Experiment 6. (A) Predictions based on the spotlight and postselection hypotheses. (B) Predictions from the contingent-processing hypothesis. See text for explanation.

yield longer response times. Again, the relationship between distractor and target is not considered important.

These two hypotheses make similar predictions because both are based on the nature of the items in the RSVP stream (lines or letters). In contrast, the contingent-processing hypothesis makes distinct predictions because it is based not on the nature of the distractors but on whether the distractors share the target's defining feature. For this reason, the contingent-processing hypothesis predicts an interaction effect, as illustrated in Figure 7B: Response times are expected to be relatively long when the distractors share the target's defining feature but relatively short when they do not, regardless of whether they are lines or letters.

Method

Experiment 6 was a virtual replication of Experiment 1, except for the following. There were four conditions: letter–letter (Lt–Lt), letter–line (Lt–Ln), line–letter (Ln–Lt), and line–line (Ln–Ln), in which the first item in each pair indicates the type of distractor in the RSVP stream, and the second item indicates the type of target. The letter displays, whether distractors or targets, were the same as in Experiment 1. The line displays were as follows. In the conditions in which line segments were used as targets (Ln–Ln and Lt–Ln), the target display consisted of a line 1° long and less than .1° wide, tilted either to the right (45°) or to the left (135°). Observers reported the tilt of the line by pressing either the left-arrow key or the right-arrow key on the keyboard that was marked \ or /, respectively. In the conditions in which line segments were used as distractors (Ln–Ln and Ln–Lt), the distractor display consisted of an RSVP stream of between 5 and 10 frames, each containing a line whose orientation was varied randomly between 0° and 165° in steps of 15° (except for orientations of 45° and 135°, which were the target orientations), with the restriction that a line could not have the same orientation as the line in either of the preceding two frames. The physical dimensions of the distractor lines were the same as those of the target lines. A new group of 36 observers served in Experiment 6. Each observer served in all four conditions that were sequenced randomly, separately for each observer. A session began with 100 practice trials in the condition that was scheduled to be done first. This was followed by 96 experimental trials in each of the four conditions. Each condition after the first was preceded by 25 practice trials.

Results and Discussion

The median response time was calculated for each observer, separately for each condition. Figure 8 illustrates the mean of the median response times, averaged over all observers, separately for each condition.

The data in Figure 8 were analyzed in a 2 (distractor type: line vs. letter) × 2 (target type: line vs. letter) within-subject ANOVA. The analysis revealed a significant interaction effect, $F(1, 35) = 19.02$, $p < .001$, $MSE = 2{,}250.66$. Neither the effect of distractor type, $F(1, 35) = 2.37$, $p > .13$, $MSE = 1{,}296.56$, nor the effect of target type ($F < 1$) reached significance.

Predictions from the contingent-processing hypothesis illustrated in Figure 7B were confirmed by the data in Figure 8: Response times were relatively long when the distractors shared the target's defining characteristic and relatively short when they did not, regardless of whether they were lines or letters. It is worth noting that the results obtained with letter targets in Experiment 6 (Figure 8, circular symbols) were similar to the results of Experiment 1 (Figure 2, segmented line). This strongly suggests that the line-segment distractors used in Experiment 6 were functionally equivalent to the random-dot patterns used in Experiment 1. It is to be expected, therefore, that the results of Experiments 2–5 would not be changed materially if the random-dot distractors were to be replaced with line-segment distractors.

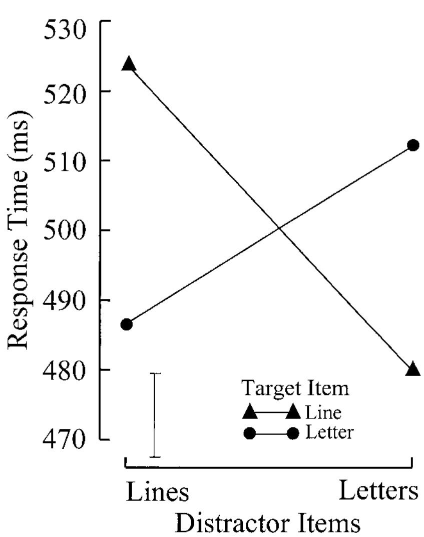

Figure 8. Mean results of Experiment 6, averaged over observers, separately for the four experimental conditions. The vertical bar represents one standard error of the mean, averaged over all conditions.

A smarter spotlight. On the face of it, the outcome of Experiment 6 is consistent with the hypothesis that the time cost seen in the present experiments is attributable, at least in part, to the time taken by involuntary processing of distractor items that share the target's defining characteristic. Before reaching a definitive conclusion, however, we need to consider an alternative account involving a "smart" attentional spotlight that generates time costs attributable exclusively to spatial shifts in the focus of attention. An argument along these lines can be formulated by postulating a spotlight endowed with preprocessing capabilities used for governing the beam's diameter and location. In the preprocessing stage, distractor items that do not share the target's defining attribute are rejected, and the beam need not be narrowed to process that item further. If, however, a distractor shares the target's defining attribute, the beam has to be narrowed to the location of the distractor for further processing. Eventually, this produces a time cost because the beam has to be repositioned when the target arrives. This option can be illustrated with reference to Experiment 6.

Suppose that, at the outset of any given trial, the beam diameter is set relatively wide. In the conditions in which the distractors do not share the target's defining attribute (Lt–Ln and Ln–Lt), the distractors are discarded in the preprocessing stage, and observers can maintain a wide-beamed spotlight until the target arrives. In the conditions in which the distractors share the target's defining characteristic (Lt–Lt and Ln–Ln), however, the distractors demand attention, and the beam has to be narrowed to the location of the distractors and then repositioned to the location of the target. This would give rise to a time cost of an inherently spatial nature.

On this hypothesis, the results of Experiment 6 can be explained entirely on the basis of the time taken by spatial shifts in the focus of attention, with no recourse to the time cost of processing the distractor items. Thus, letter distractors affected letter targets because the beam had to be narrowed, whereas line distractors allowed for a wide beam. Similarly, line distractors affected line targets because the beam had to be narrowed, whereas letter distractors allowed for a wide beam.

Still unresolved, then, is the critical issue of whether the results of Experiment 6 can be explained exclusively by spatial factors (repositioning of the spotlight) or whether nonspatial factors (processing of distractors) are also required. This issue was addressed in Experiment 7 in which spatial shifts in the focus of attention were obviated by presenting all items, whether distractors or targets, in a single location at the center of the screen. In every other respect, Experiment 7 was a replication of Experiment 6. We reasoned as follows: If the time cost seen in Experiment 6 (Figure 8) were attributable exclusively to spatial factors, eliminating the spatial shifts would also eliminate the time cost. Conversely, to the extent that the time cost is attributable to nonspatial factors such as distractor processing, the cost should remain even when the spatial shifts are eliminated.

Experiment 7

Method

Experiment 7 was a replication of Experiment 6, except that all stimuli were presented at the center of the screen. A new group of 24 observers participated in the experiment.

Results and Discussion

The median response time was calculated for each observer, separately for each condition. Figure 9 illustrates the mean of the median response times, averaged over all observers, separately for each condition.

The data in Figure 9 were analyzed in a 2 (distractor type: line vs. letter) × 2 (target type: line vs. letter) within-subject ANOVA. The analysis revealed a significant effect of distractor type, $F(1, 23) = 8.81$, $p < .01$, $MSE = 3{,}759.40$, and a significant interaction effect, $F(1, 23) = 86.93$, $p < .001$, $MSE = 1{,}975.72$. The effect of target type did not reach significance, $F(1, 23) = 1.11$, $p > .30$, $MSE = 3{,}913.49$.

The main question in the present experiment was whether a pattern of results similar to that seen in Experiment 6 can be obtained when all stimuli are presented in a single location, thus obviating spatial shifts in attentional focus. The results of Experiment 7 answered that question in the affirmative: the pattern of time costs seen in Experiment 7 (Figure 9) was similar to that in Experiment 6 (Figure 8). Identification of the target was delayed when it was presented directly after a stream of distractors that belonged to the same stimulus class. Thus, a letter target was identified more slowly when the leading stream contained other letters than when it contained oriented line segments. Similarly, identifying the orientation of a line segment took longer if it was preceded by a stream of other line segments than by a stream of letters. These are instances of attentional capture contingent on whether the distractors shared the target's defining characteristic. The important issue here is that in neither case can the time cost be attributed to the time taken in relocating an attentional spotlight because spatial factors were explicitly ruled out in the experimental design. Rather, the results are consistent with the hypothesis that the time cost arose from obligatory processing of distractors that belonged to the same category as the target. Processing of the target was delayed while the system was busy processing a relevant distractor, thus giving rise to the response time pattern seen in Figure 9.

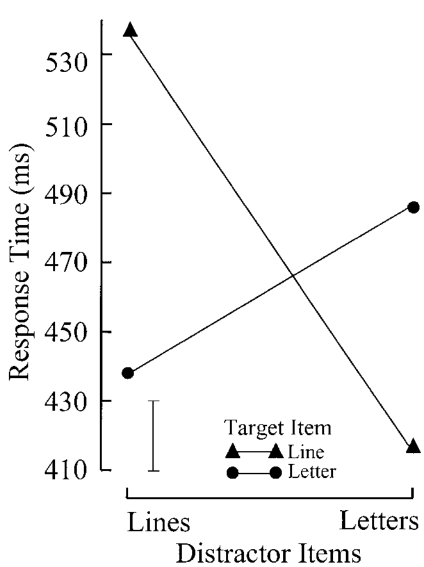

Figure 9. Mean results of Experiment 7, averaged over observers, separately for the four experimental conditions. The vertical bar represents one standard error of the mean, averaged over all conditions.

This is not to say that spatial factors play no role in studies in which targets and distractors are presented in different locations. What the present experiment strongly suggests, however, is that the temporal deficit in contingent capture is attributable, at least in part, to nonspatial factors such as the time taken to process distractor items. It is notable that the magnitude of the time cost obtained in Experiment 7 (Figure 9) was even more pronounced than that obtained in Experiment 6 (Figure 8). Please note the scale differences in the two figures. We compared the results of Experiments 6 and 7 directly by analyzing the combined results in a 2 × 2 × 2 ANOVA comprising two within-subject variables (distractor type: line vs. letter and target type: line vs. letter) and one between-subjects variable (experiment: Experiment 6 vs. Experiment 7). The outcome of the combined analysis was entirely consistent with the results of the separate analyses carried out on the results of Experiments 6 and 7. More important, the combined analysis revealed a highly significant second-order interaction effect among target type, distractor type, and experiment, $F(1, 58) = 16.88$, $p < .001$, $MSE = 2{,}241.64$.

This interaction effect is important because it highlights the powerful influence exerted on the magnitude of contingent capture by the degree to which distractors share common characteristics with the target. In Experiments 1–6, observers were set to process only certain letters in peripheral locations. This meant that spatial location was never a shared attribute between targets and distractors, and capture depended principally on whether a distractor belonged to the same stimulus category as the target. In contrast, in Experiment 7, all stimuli were presented in the same location. This had two main consequences. First, it deprived the observers of a spatial cue for separating peripheral targets from the central stream of distractors. This was an important cue because, had observers been able to filter out all items in the central RSVP stream in Experiments 1–6, no contingent capture could have occurred. Second, the number of shared characteristics was increased because targets and distractors shared spatial location as well as stimulus category. The increased commonality caused the magnitude of contingent capture to be correspondingly greater in Experiment 7 than in Experiment 6, as revealed by the second-order interaction effect noted earlier.

General Discussion

The principal objective of the present work was to examine the mechanisms that underlie contingent capture. Specifically, we asked whether the time cost associated with contingent capture can be mediated by the involuntary processing of a distractor, independently of a spatial shift in the focus of attention. Seven experiments provided uniform support for the proposition that the time cost that is the signature of contingent capture can be mediated by the involuntary processing of a distractor that shares the target's defining characteristic. Consistent with this interpretation, we found that the temporal deficit in target identification occurs only while a new distractor is being processed, namely, when the target arrives within about 600 ms after the presentation of a salient distractor.

Support for the present hypothesis does not rule out inappropriate spatial shifts in attentional focus as a source of delay in contingent capture (e.g., Folk et al., 1992). Rather, the present evidence points to an additional temporal cost—processing of a distractor item—that occurs when the distractors share the target's defining characteristic. Also left open in the present work is the question of how similar the distractors need to be to the target in order for capture to occur. For example, given that the targets are letters, it could be asked to what extent contingent capture would occur if the distractors were false fonts, or digits, or other stimuli that shared features with letters. A systematic investigation of this issue, however, is beyond the scope of the present work.

Collectively, the present study and earlier studies of contingent capture have addressed two fundamental questions. First, why is it that distractors that share a defining characteristic with the target are processed as though they were targets? Second, given that a salient distractor is present in the display, what causes the response to the target to be impaired? A uniform answer to the first question has been given, implicitly or explicitly, in all studies of contingent capture: A distractor is processed as though it were a target if it matches the observer's attentional control setting, which is aimed at optimizing the processing of the target (e.g., Folk & Remington, 1998). In the present work, the observer's attentional set was instantiated as an input filter tuned to the target's distinguishing characteristic. The input filter functions as the first stage in a two-stage processing sequence. Only stimuli that match the setting of the input filter can gain access to the second stage. This means that distractors that share the target's distinguishing characteristic may gain access to the second stage, thereby delaying processing of the target. A scheme such as this is entirely compatible with the distinction between attention and selection recently drawn by Remington and Folk (1999).

Theoretical constructs such as input filtering and attentional control setting can explain why some distractors are processed as though they were targets while others are not. However, these constructs alone cannot provide an answer to the second question; namely, why does the presence of a salient distractor slow down the response to the target. In experiments in which the distractors were presented away from fixation (e.g., Folk et al., 1992), the temporal deficit was attributed to the time wasted by an involuntary attentional shift to the location of the distractor. That scheme, however, is obviously insufficient to account for the outcomes of the present experiments, especially Experiment 7. Nor is the mere postulation of a further processing stage sufficient. For example, no temporal deficit should be expected if processing at the higher stage were done in parallel so that both the target and the distractors could be handled concurrently.

What is compellingly demanded by the present results is a capacity-limited second stage in which processing is strictly—or predominantly—serial. This is not to deny that in studies involving spatial shifts in the focus of attention, the temporal deficit in the response to the target may have arisen, at least in part, from the time wasted on an involuntary attentional shift to the distractor location. Indeed, this question is open to empirical investigation using a factorial design in which the distractors' sharing of the target's defining feature is crossed with whether the distractors are presented peripherally or at fixation. However, both an input-filtering stage and a capacity-limited serial second stage seem to be required for a complete account of the temporal deficit obtained in investigations of contingent capture, regardless of the location of the distractors.

Contingent Capture and the Attentional Blink

The results of Experiments 3 and 5 (Figures 4 and 6) invite comparison with a phenomenon known as the *attentional blink* (AB), which refers to an impairment in identifying the second of two targets presented in rapid succession (Raymond et al., 1992). The precise cause of this second-target deficit is still undetermined, but there is general agreement that it stems from the attentional drain involved in selecting the first target to the detriment of the second target (Chun & Potter, 1995; Shapiro, Arnell, & Raymond, 1997). This viewpoint is supported by the finding that if the requirement to process the first target is removed, the second-target deficit is much reduced or totally eliminated (Raymond et al., 1992; Seiffert & Di Lollo, 1997). To be sure, the present experiments contained only one target instead of two sequential targets commonly used in AB studies. Nevertheless, the option must be considered that the leading distractors in the present experiments may have caused an attentional drain akin to that caused by the first target in conventional AB experiments.

A notable point of contact between contingent capture and the AB is the temporal course of the two phenomena. The AB deficit is most pronounced when the temporal lag between the two targets is short, with performance returning gradually to baseline as the lag is increased to about 600 ms. Figures 4 and 6 reveal much the same temporal course for contingent capture. Just as in the AB, speed and accuracy of identification were substantially impaired when the target was presented shortly after a salient distractor and then gradually returned to baseline as the lag between the onsets of the distractor and the target was increased to about 600 ms.

One way of explaining the similar time courses of the two phenomena is by assuming that the role of the salient distractor in the contingent-capture paradigm corresponds to that of the first target in the AB paradigm. In both paradigms, attentional resources are preempted by the leading item, whether it is the first target or a distractor that shares the target's defining feature. Processing of an ensuing target is then impaired for a period that, as we have seen, is approximately the same in both paradigms.

On the face of it, this account seems to run afoul of one major procedural difference between the two paradigms. We noted earlier that the AB deficit is eliminated if the requirement to process the first target is waived. In contrast, with the contingent-capture paradigm, a deficit is reliably obtained even when the observer is explicitly instructed to ignore the leading distractor. If a valid parallel is to be drawn between these two paradigms, this apparent inconsistency must first be resolved: Why is it that instructions to ignore the first target eliminate the deficit in the AB paradigm, whereas similar instructions fail to eliminate a corresponding deficit in the contingent-capture paradigm?

On closer inspection, the inconsistency turns out to be more apparent than real. There is no question that the AB deficit vanishes when the requirement to process the first target is removed. However, the literature indicates this to be true only when the first target does not share the defining attribute of the second target. For example, in the study of Raymond et al. (1992), the first target was a white letter to be identified and the second a black *X* to be detected. Thus, the two targets differed distinctly from one another both in defining attribute and in the type of response required. The same was true in the study of Seiffert and Di Lollo (1997), in which the first target was a bright letter to be identified and the second a dim *X* to be detected. In all of these experiments, the two targets had different distinguishing characteristics, and the AB deficit was eliminated or much reduced when observers were instructed to ignore the first target.

In contrast, when the two targets have the same defining characteristic, instructions to ignore the first target are ineffectual. Not only is an AB deficit obtained reliably, but its magnitude is almost as large as when observers are required to process the first target. This was first reported by Chun (1997) in a study in which both targets were letters to be identified. Homologous results were obtained in pilot studies reported by Potter, Chun, Banks, and Muckenhoupt (1998). In both cases, pronounced AB deficits were obtained whether observers were required to report the first target or to ignore it. This strongly suggests that the first target is hard to ignore when it shares a distinguishing characteristic with the second target.

In light of these results, what appeared to be an inconsistency between the AB and the contingent-capture paradigms becomes a compelling similarity. In both paradigms, a to-be-ignored leading item, whether first target or distractor, will interfere with the perception of a trailing target only when the two have the same defining characteristic. Thus, in the present work, perception of a trailing target letter was impaired when the leading distractors were other letters but not when they were unrelated random dots. The parallel between contingent capture and the AB is further buttressed by the results of recent experiments by Egeth et al. (2000). In those experiments, the target was a colored letter inserted in a central RSVP stream of gray letters, and the distractor consisted of four # signs presented peripherally around one of the gray letters. On some trials, all four # signs were gray, on other trials one of them was the same color as the target letter. Accuracy of target identification was found to be impaired, but only on trials in which one of the # signs was the same color as the target. Just as important, the temporal course of the impairment resembled that obtained in conventional AB studies and in the present Experiments 3 and 5. Furthermore, the experiment of Egeth et al. demonstrated that a leading distractor that shares the target's defining characteristic cannot be ignored despite instructions to the contrary, whether it is displayed centrally, as in the present work, or peripherally, as in the experiments of Egeth et al.

In summary, the results obtained with both the AB and the contingent-capture paradigms are in keeping with a two-stage model in which stimuli must pass an input filter tuned to the target's defining characteristics before gaining access to a resource-limited serial second stage. In both paradigms, the processing of stimuli that match the setting of the input filter appears to be automatic and obligatory. As a consequence of this obligatory processing, the processing of temporally trailing stimuli is impaired, with the impairment following similar time courses in the two paradigms. The obvious advantage of bringing the AB and contingent capture within a single conceptual rubric is that the outcomes obtained with two ostensibly different paradigms can be explained on the basis of the same underlying mechanisms.

References

Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid visual presentation. *Perception & Psychophysics, 42,* 105–113.

Chun, M. M. (1997). Temporal binding errors are redistributed in the attentional blink. *Perception & Psychophysics, 59,* 1191–1199.

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 and Performance, 21,* 109–127.

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.

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 448–456.

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.

Egeth, H. E., Folk, C. L., Leber, A., & Nakama, T. (2000, June). *Contingent attentional capture and the attentional blink.* Paper presented at Attraction, Distraction, and Action: An Interdisciplinary Conference and Workshop on Attentional Capture, Villanova, PA.

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 and 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 and Performance, 18,* 1030–1044.

Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for display-wide visual features. *Journal of Experimental Psychology: Human Perception and 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 and Performance, 24,* 1454–1466.

Holst, E. von. (1954). Relation between the central nervous system and the peripheral organs. *Animal Behaviour, 2,* 89–94.

Linnell, K. J., & Humphreys, G. W. (2000). Eccentricity and object-based selection. *Perception, 29*(Suppl., S63).

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

Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and 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 and Performance, 18,* 849–860.

Remington, R. W., & Folk, C. L. (1999, November). *Evidence for a distinction between attention and selection.* Paper presented at the 40th annual meeting of the Psychonomic Society, Los Angeles.

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General, 125,* 227–249.

Seiffert, A. E., & Di Lollo, V. (1997). Low-level masking in the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1061–1073.

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 mechanisms in attention, memory and language* (pp. 151–187). San Diego, CA: Academic Press.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102,* 503–532.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *The Journal of the Acoustical Society of America, 41,* 782–787.

Theeuwes, J., & Burger, R. (1998). Attentional control during visual search: The effect of irrelevant singletons. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1342–1353.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology, 47A,* 631–650.

Visser, T. A. W., Bischof, W. F., & Di Lollo, V. (1999). Attentional switching in spatial and non-spatial domains: Evidence from the attentional blink. *Psychological Bulletin, 125,* 458–469.

Wald, A. (1947). *Sequential analysis.* New York: Wiley.

Received October 25, 2000
Revision received March 26, 2002
Accepted May 31, 2002 ■