FOCAL DISTRACTION:
SPATIAL SHIFTS OF ATTENTION ARE NOT REQUIRED FOR CONTINGENT CAPTURE

by

SAMANTHA MARIJA ZUVIC

B.Sc., The University of British Columbia, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Psychology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 2001

© Samantha Marija Zuvic, 2001
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology
The University of British Columbia
Vancouver, Canada
Date May 24, 2001
ABSTRACT

Contingent capture occurs when distractors that share distinguishing characteristics with a target capture attention and slow down target identification. Conventionally, this slowdown has been attributed to the time wasted by an inappropriate attentional shift to the location of a distractor. To examine this account, we obviated the spatial shift by presenting distractors at fixation, and measured contingent capture both directly by measuring response times and indirectly by estimating the duration for which the target remains vulnerable to backward masking.

Contingent capture invariably occurred when a salient distractor was presented within about 600 ms before the target. Because spatial shifts were ruled out using our procedure, the conventional account is insufficient. We augment that account with a two-stage model in which stimuli must pass an input filter tuned to the target's distinguishing characteristic before gaining access to a high-level stage which is unavailable for targets while distractors are being processed.
# TABLE OF CONTENTS

Abstract ...................................................................................... ii

Table of Contents ........................................................................ iii

List of Figures .............................................................................. iv

General Introduction .................................................................... 1

Experiment 1 ................................................................................ 3
  Method ....................................................................................... 4
  Results and Discussion ............................................................... 5

Experiment 2 ................................................................................ 8
  Method ....................................................................................... 9
  Results and Discussion ............................................................... 9

Experiment 3 ............................................................................... 12
  Method ..................................................................................... 12
  Results and Discussion ............................................................ 13

Experiment 4 ............................................................................... 16
  Method ..................................................................................... 18
  Results and Discussion ............................................................ 19

Experiment 5 ............................................................................... 20
  Method ..................................................................................... 20
  Results and Discussion ............................................................ 20

General Discussion ..................................................................... 21

References .................................................................................. 27

Figure Captions .......................................................................... 30
LIST OF FIGURES

Figure 1. Schematic representation of the stimulus sequence in Experiment 1 . . . 31
Figure 2. Mean results of Experiment 1 ....................................................... 32
Figure 3. Schematic representation of the stimulus sequence in Experiment 2 . . . 33
Figure 4. Mean results of Experiment 2 ....................................................... 34
Figure 5. Schematic representation of the stimulus sequence in Experiment 3 . . . 35
Figure 6. Mean results of Experiment 3 ....................................................... 36
Figure 7. Mean results of Experiment 4 ....................................................... 37
Figure 8. Mean results of Experiment 5 ....................................................... 38
Focal distraction:
Spatial shifts of attention are not required for contingent capture

When we first learn to drive, we quickly become acquainted with the basic operations of traffic lights. For example, a green light means that we may proceed through the intersection, whereas a green arrow means that we may proceed only in the indicated direction. Nonetheless, despite extensive experience with traffic lights, one often observes a driver lurching straight ahead upon seeing a green left-turn arrow. Obviously, the driver was 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 amongst 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, & Johnston, 1992; Folk & Remington, 1998; 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 inappropriate 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 contingent capture can arise merely from processing a distractor item without the need for an accompanying spatial shift in the focus of attention. Further, we found that the mere presence of distractors, even if highly similar to the target, does not automatically bring about contingent capture. Rather, contingent capture occurs only during a brief period of less than about 600 ms, while newly-presented distractors are being processed. These findings can be explained by a two-stage model
of contingent capture, 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 non-predictive cue is followed by a search display containing a target. For example, Folk et al. (1992) employed 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 which was red. The cue display was followed by a search display consisting of a red target amongst 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. (1992) 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 inappropriate attentional shift to a non-target location. More recently, however, Theeuwes and Burger (1998) have argued that the slowing arises not only from the shift itself but also from processing of the distractor at the shift location. They referred to this additional source of delay as the identity intrusion effect. In the experiments of Theeuwes and Burger (1998), observers searched for one of two possible target letters among distractor letters. All letters were the same colour except one uniquely-coloured letter, called the singleton, which was always one of the two target letters. Observers were instructed to ignore the singleton. Nonetheless, response times to identify the target were longer when the singleton was the alternative target letter than when it was the same letter as the target. From this, Theeuwes and Burger drew two
conclusions. First, that attention had indeed been shifted to the location of the singleton and, second, that the singleton letter was involuntarily processed despite instructions to the contrary.

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 Theeuwes and Burger’s study. This is because processing of the singleton could not occur without an associated shift in attentional focus. This raises the question of whether contingent capture can be mediated solely by processing a distractor without a spatial shift in attentional focus. In the present work, we examined this question using a non-search paradigm in which inappropriate spatial shifts of attention could not occur. A stream of distractors was presented at fixation, followed by a target presented eccentrically. 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 varied the degree of similarity between the distractors and the target. There were two conditions: in one the distractors were letters and in the other they were random dots. The target was always a letter. When the distractors were letters (the Variable Letter condition), their identity was changed randomly from one frame to the next in the 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 with features in common with the target 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, 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 centre of the screen, and one letter target presented unpredictably in one of 12 locations arranged as in a clock face with radius of 3°, centered on the centre of the screen. The letters in the RSVP stream were selected randomly with replacement from all letters of the English alphabet excepting 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 inter-stimulus interval (ISI) of 70 ms, during which the screen was blank. This yielded a presentation rate of 10 items/sec. 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/m², and that of the target was 10 cd/m², 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/m².

At the beginning of each trial a small fixation cross, subtending .25° of visual angle, was presented in the centre of the screen, at a luminance of 2 cd/m². Observers initiated each trial by pressing the space bar, 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-minute 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) which employs 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 in Figure 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 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 towards the target, not towards 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 inappropriate attentional shift to the distractor, but also to the time taken to process the distractor. Further, 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 the filtering model of Visser, Bischof, and Di Lollo (1999). This model belongs to 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).

In the model of Visser et al. (1999), 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 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 amongst 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 colour. 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 filtering model is straightforward. Given that the target was a letter, it can be assumed that the input filter was optimally configured to pass letter-like stimuli. This meant that in the Variable Letter condition, distractors matched
the filter setting and gained access to high-level processing. On some trials, it is likely that the
distractor immediately prior to the target preempted the high-level stage. As a result, processing
of the target was delayed at the input stage until processing of the distractor at the higher stage
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 meant that the target always gained immediate access to high-level
processing, resulting in faster response times than in the Variable Letter condition.

In the remainder of the present work, we report four experiments designed to test specific
predictions from this model and to provide converging evidence obtained with different
experimental paradigms. We began by testing a simple corollary implicit in this account. If it is
ture that response times in the Variable Letter condition are slow when the target arrives while
the high-level stage was busy, then response times should be faster if the target were to arrive
while that stage was 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, long
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 target-distractor similarity should no longer be an important factor. That is, response times 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 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 is illustrated in Figure 3.

Results and Discussion

Trials on 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 in Figure 4, along with the results of Experiment 1 for ease of comparison. 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 vs. variable presentation, an analysis of variance was performed on the combined results of Experiments 1 and 2 with one between-subject factor (Mode of Presentation: varied vs. steady), and one within-subject factor (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, $F(1,58) = 1.99$, $p > .05$, $MSE = 5479.16$.

The pattern of results in Figure 4 is entirely in keeping with predictions from the 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 the 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 which was set to pass only letter-like 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 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 responses time 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. Another way of saying this is that 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 through the early frames. The display was then switched to the second letter for the remaining frames. We reasoned that as the number of post-switch 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 post-switch 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 then was switched to
the other letter for the remaining frames. This sequence is illustrated schematically in Figure 5. The number of pre-switch frames was varied randomly on each trial between 4 and 15. The number of post-switch frames was either 6, 3, 1, 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 either 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 which 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 only dot distractors. The order of the two blocks was counterbalanced across observers. The four lengths of the post-switch stream were presented randomly across trials within a block, with 96 trials for each of the four lengths.

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 6, separately for the Letter and Dot conditions and number of post-switch frames. The two conditions in which there were no post-switch 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 segmented lines in Figure 6. The data in Figure 6 were analyzed in a 2 (Distractor Type: letters vs dots) x 4 (Number of Post-Switch Frames: 0, 1, 3, 6) within-subject analysis of variance. The analysis revealed significant effects of Distractor Type, \( F(1,19) = 23.42, p < .001, \text{MSE} = 1068.56 \), and Number of Post-Switch Frames: \( F(3,57) = 19.56, p < .001, \text{MSE} = 93.50 \). Notably, the interaction effect was also significant: \( F(3,57) = 12.39, p < .001, \text{MSE} = 98.77 \). This interaction effect is consistent with the graphical evidence in Figure 6 that increasing the number of post-switch 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 post-switch frames. A significant effect was revealed with letter distractors, \( F(3,57) = 24.10, p < .001, \text{MSE} = 124.97 \), but not with random-dot distractors, \( F(3,57) = 0.60, p > .05, \text{MSE} = 67.30 \). 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 is that, with letter distractors, response times became progressively shorter as the number of post-switch frames was increased (Figure 6). 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). In contrast, 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 4). 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 post-switch frames.

It is interesting to note that response times at baseline were faster when the distractors were dots than when they were letters (Figure 6). 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 6 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 6, 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 (1 frame) before the target, were still marginally above baseline when the new letter was presented 300 ms (3 frames) before the target, but were at or just below baseline when a new letter was presented 600 ms (6 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.

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 given stimulus
that remains on view does not keep on being re-processed as though it were a new stimulus. A possible underlying mechanism for distinguishing between “new” and “old” stimuli has been described by von Holst (1954) as follows:

The efference (to the effector) leaves an “image” of itself somewhere in the CNS to which the re-afference ... compares as the negative of a photograph compares to its print; so that ... when the efference copy and the reaference exactly compensate one another, nothing further happens. When, however, the efference is too small or lacking [or] too great ... the difference ... can ascend to a higher centre and produce a perception. (von Holst, 1954, p. 91).

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 afterwards 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 the 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) that 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 between the target and the trailing mask: 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, Enns, & Rensink, in press; Giesbrecht & Di Lollo, 1998; Mack & Rock, 1998).

The negative relationship between masking and attention can be used as a basis for testing the two-stage 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 4).

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 towards 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; 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 7. The data were analyzed in an analysis of variance, with one between-subject factor (Presentation Mode: variable or steady) and one within-subject factor (Distractor Type: letters or dots). This analysis revealed significant effects of Presentation Mode, $F(1,34) = 5.80, p < .05, \text{MSE} = 895.27$, Distractor Type, $F(1,34) = 39.35, p < .001, \text{MSE} = 205.35$, and a significant interaction effect, $F(1,34) = 14.02, p < .001, \text{MSE} = 205.35$. 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 two-stage 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 7) 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 7 is very similar to that in Figure 4, suggesting that
response times and critical SOA provide equivalent 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, 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 which was shown repeatedly
for either 0, 1, 3, or 6 frames. The same sequence was used when the distractors were random
dots. The display sequence on a typical given trial is illustrated in Figure 5, 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.

Results and Discussion
Mean critical SOAs are shown in Figure 8, separately for each RSVP condition and number of post-switch frames. The data were analyzed in a within-subject analysis of variance comprising two factors: Distractor Type (letters or dots) and Number of Post-Switch Frames (0, 1, 3, 6). The analysis revealed significant effects of Distractor Type, $F(1,11) = 5.20, p < .05$, MSE = 1869.12, and Number of Post-Switch 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 7 was not significant, $t(11) = 0.83, p > .05$.

The results of Experiment 5 provide further converging evidence in support of the two-stage model of contingent capture. As expected, the number of post-switch 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 post-switch frame, and became progressively shorter as the number of post-switch 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.

GENERAL DISCUSSION

Our objective in the present work was to examine some of the mechanisms underlying contingent capture. Specifically, we asked whether contingent capture can be mediated solely by the involuntary processing of a distractor, without the need for a spatial shift in the focus of attention. Spatial shifts of attention were obviated in the present work by presenting all
distractors at fixation and the target at peripheral locations. A series of five experiments provided uniform support for the proposition that spatial shifts in attentional focus are not necessary for contingent capture. Further, we found that the temporal deficit associated with contingent capture occurs only if the target arrives within about 600 ms after the presentation of a salient distractor.

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 also gain access to high-level processing.

Theoretical constructs such as attentional control setting and input filtering can be used to explain why some distractors are processed as though they were targets while others are not. But 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 earlier 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 inappropriate attentional shift to the location of the distractor, which caused a corresponding delay in target processing. That account, however, is obviously insufficient in situations, such as the present study, in which the distractors are
presented at fixation. 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, such as outlined above, 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 inappropriate attentional shift to the distractor location. Indeed, this question is open to empirical investigation using a factorial design in which degree of target-distractor similarity 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 6 and 8) 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; Giesbrecht & Di Lollo, 1998). 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).

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 6 and 8 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 to explain 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 a distractor or the first target. 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, an 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 distinguishing characteristic 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 characteristic 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 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 share the same defining characteristic, instructions to
ignore the first target are ineffectual. Not only is an AB deficit reliably obtained, 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 coloured letter inserted in a central RSVP stream of grey
letters, and the distractor consisted of four # signs presented peripherally around one of the grey
letters. On some trials, all four # signs were grey, on other trials one of them was the same
colour 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 colour 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. Finally, the experiment of Egeth et al. (2000) demonstrates 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. (2000).

In summary, the results obtained with both the AB and the contingent-capture paradigms are in keeping with the two-stage model outlined in the foregoing, 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


Figure captions

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

Figure 2. Mean results of Experiment 1, averaged over observers, separately for the dot-distractor and letter-distractor conditions. The vertical bar represents one standard error of the mean, averaged over both conditions.

Figure 3. Schematic representation of the stimulus sequence in Experiment 2.

Figure 4. Mean results of Experiment 2, averaged over observers, separately for the dot-distractor and letter-distractor conditions (filled symbols). The vertical bar represents one standard error of the mean, averaged over both conditions. The results of Experiment 1 have been included for ease of comparison (segmented line).

Figure 5. Schematic representation of the stimulus sequence in Experiment 3.

Figure 6. 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 post-switch 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.

Figure 7. 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.

Figure 8. 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 post-switch 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.
Figure 2
Figure 4

Exp. 1 (Variable RSVP)

Exp. 2 (Steady RSVP)

Response Time (ms)

Dots  Letters

RSVP Type
Figure 6

Response Time (ms)

No. of Post-Switch Frames

- Letter RSVP
- Dot RSVP
- Letter baseline
- Dot baseline
Figure 7
Figure 8

Critical SOA (ms)

- Letter RSVP
- Dot RSVP
- Letter baseline
- Dot baseline

No. of Post-Switch Frames

1 2 3 4 5 6