Categorically Defined Targets Trigger Spatiotemporal Visual Attention

Brad Wyble
Massachusetts Institute of Technology

Howard Bowman
University of Kent at Canterbury

Mary C. Potter
Massachusetts Institute of Technology

Transient attention to a visually salient cue enhances processing of a subsequent target in the same spatial location between 50 to 150 ms after cue onset (K. Nakayama & M. Mackeben, 1989). Do stimuli from a categorically defined target set, such as letters or digits, also generate transient attention? Participants reported digit targets among keyboard symbols in a changing array of 8 items. When 1 target preceded a second target in the same location at a stimulus onset asynchrony of 107 ms (but not 213 ms), the second target was reported more often than in a condition in which there was no leading target. When the 2 targets were at different locations, report of the second target was impaired. With both letters and digits as targets, the enhancement effect was shown not to be due to category priming. Critically, the attentional benefit was present whether or not participants reported the leading target. Transient attention, contingent attentional capture, popout, and Lag 1 sparing in the attentional blink may involve a common mechanism for orienting processing resources towards salient and task relevant stimuli.

Keywords: transient attention, attentional blink, RSVP, spatial cueing, attentional capture

How does a viewer rapidly identify a target in a visual field that is full of distracting stimuli? This question is commonly addressed by presenting subjects with an array of stimuli containing a target item and recording the speed and/or accuracy of identifying or detecting the target. Directing covert visual attention to the upcoming target's location in the visual field has been shown to help the visual system to detect or identify a target. The location of the target can be signaled either with an arrow presented at fixation that points in a particular direction (Colegate, Hoffman, & Eriksen, 1973; Posner, 1980; Posner, Snyder, & Davidson, 1980) or by placing a cue at the location of the target itself (Muller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner, 1980). Both types of cues affect the ability of subjects to report targets, facilitating performance when the cue indicates the correct location and producing an impairment when the cue indicates the incorrect location. These changes in target processing occur even when the eyes remain in the same position; attention can be directed to a location without making an eye movement (e.g., Posner, Nissen, & Ogden, 1978).

When the target stimulus is only briefly available, it is important to deploy attention at the correct moment in time as well as to the correct spatial location. The visual system seems well suited to using rapid bursts of attention to aid in the identification of a briefly presented target item. Specifically, when a particular location is cued, the cue has its greatest beneficial effect between 50 and 150 ms after cue onset, leading Nakayama and Mackeben (1989) to term the effect *transient attention*. This transient effect is specific to the cue's location (Muller & Rabbitt, 1989); thus, it is a spatiotemporal form of visual attention.

Even when a single stream of stimuli is presented at one location, as occurs with rapid serial visual presentation (RSVP), transient attention to one target may play a role in assisting the identification of a later target, and this may occur without the presence of obvious cues, such as color, luminance, or onset (Bowman & Wyble, 2007; Chun & Potter, 1995; Nieuwenhuis, Gilzenrat, Holmes, & Cohen, 2005; Weichselgartner & Sperling, 1987). Theories of RSVP target detection suggest that the visual system detects the match between a visual stimulus and the target set, which triggers a deployment of attentional resources rapidly enough to catch the target item before it is replaced by the following distractor. However, the process of detecting a match and deploying attention takes some time, which produces a temporal lag between the onset of the target and the onset of attention. Thus, when two targets are presented in rapid succession, the second target may receive a disproportionate benefit of the attention deployed in response to the first target. This delayed deployment of attention can enhance report of the second target relative to report of a lone target (Bowman & Wyble, 2007). It is this effect that inspired the present experiments.

Although it has been well established, as reviewed below, that attention can be captured by stimuli that share a target-defining feature with the target set, such as color, there have been many

fewer studies showing that attention is affected by categorically defined targets. The implications of the ability to manipulate attention with categorically defined task sets are especially important for understanding how we volitionally configure our attention to react to a sought after item in a list or a visual scene.

Here, we explore this question using digits and letters as targets. A first target is presented briefly to attract attention to a location, and a second, different target appears shortly afterward at the same or a different location. We ask whether the first target cues the second one and whether the cueing effect is dependent on the relative location, timing, and reportability of the first target. The results clearly show that the second target is reported more often if it occurs in the same location and within a short temporal window following the first target. Remarkably, this effect occurs regardless of whether or not the first target is identified.

Central Cues and Sustained Spatial Attention

Spatial attention operates under volitional control in many conditions. One can decide to attend to a given part of the visual field either by making an eye movement to that location (overt attention) or by directing attention to that location without moving the eyes (covert attention; Helmholtz, 1896; c.f. Nakayama & Mackeben, 1989; Posner, 1980). The metaphor of an attentional spotlight is used to describe this sustained attention. Such a spotlight can remain active for several seconds and is highly malleable. The attended area can have different spatial configurations, including multiple spotlights (Awh & Pashler, 2000; McMains & Somers, 2004) and striped patterns (Gobell, Tseng, & Sperling, 2004).

This form of attention is driven by endogenous (i.e., internally generated) commands in response to external instruction, such as "Attend to the location indicated by the arrow." A characteristic of sustained attention is its relatively slow speed of deployment. The delay between the onset of an endogenous visual cue (e.g., an arrow) and the onset of a behaviorally measurable effect at the indicated location is in the range of 300 ms (Muller & Rabbitt, 1989; Theeuwes, Godijn, & Pratt, 2004; but see Tipples, 2002, for an example of rapid directing of attention by arrows).

A classic demonstration of this effect is presented in Posner et al. (1978). Subjects saw an arrow pointing left or right at a central fixation location; and, 1 s later, a bright square appeared 7° either to the left or the right, to which subjects responded as rapidly as possible. The square appeared in the indicated location (valid trial) on 80% of the arrow trials and in the other location (invalid trial) on 20%. There were also neutral trials with a plus sign (+) at fixation instead of an arrow; in those cases, the target appeared equally often in each location. In the arrow condition, responses were faster on valid than on invalid trials.

Local Cues and Transient Attention

When the cue appears at the location of the upcoming target, the speed and/or accuracy of responding to a subsequent target at that location is enhanced. Posner and his colleagues (e.g., Posner & Cohen, 1984) used an exogenous spatial cue, such as a local increase in luminance, and found shorter reaction times to targets in the cued location. It is important to note that this effect can be observed even though the target appeared equally often in the cued and uncued location (i.e., the cue was uninformative; Liu, Pestilli, & Carrasco, 2005). The effect of spatial cueing can also be observed in the accuracy of unspeeded identification tasks. Subjects in studies by Nakayama and Mackeben (1989) identified a target in a briefly presented and strongly masked array. Cueing the location of the target enhanced the ability of participants to identify it. This enhancement was especially pronounced when the cue-target stimulus onset asynchrony (SOA) was approximately 50 to 150 ms, justifying the term "transient attention," as previously noted. The rapid onset of this enhancement effect is consistent with the onset of the cueing effect observed in the speeded response task of Posner and Cohen (1984).

Transient Attention as Contingent Capture

When transient attention is drawn to a nontarget location by a salient visual event, detection of the target suffers (Theeuwes, 1994a; Yantis & Jonides, 1984). This effect is called *attentional capture*. There is a debate about the extent to which the task set can influence which stimuli will capture attention. Theeuwes (1994b) has claimed that it is primarily the inherent salience of a stimulus that drives its ability to capture attention. Folk, Remington, and Johnston (1992), however, claim that the ability of any stimulus to capture attention is under top-down control, a hypothesis known as *contingent capture*. For instance, if a color specifies targets in a rapidly presented stream of letters, distractors (hash marks) of that same color can capture attention away from the target location, producing an impairment of target identification (Folk, Leber, & Egeth, 2002). Likewise, color-marked targets presented in a location that subjects are instructed to ignore can impair performance (Serences et al., 2005). Thus, attentional capture is not determined solely by intrinsic salience, but is modulated by task set. Such demonstrations of top-down control are critical for understanding how transient attention contributes to cognitive function, and it is that issue we explore here.

Cueing Attention With Categorically Defined Targets

The notion of cueing in most visual attention experiments involves giving the visual system an advance signal as to where a subsequent target item will occur with the use of a highly salient stimulus, such as a bar, frame, or colored marker that is itself not a target. Here we extend the definition of a cue to refer to any stimulus, even a potential target, that manipulates the deployment of attention either towards or away from a following critical stimulus. In these experiments, we will make use of categorically defined targets as attentional cues.

There is a comparatively small number of studies that attempt to cue attention to a particular spatial location with a categorically defined target. In the past, indirect evidence that categorically defined targets can manipulate the deployment of attention was found by presenting simultaneous search tasks and examining the slope of reaction time as a function of the number of distractors. Jonides and Gleitman (1972) asked subjects to search for the letters "A," "Z," and "O" among digit distractors or the digits 2, 4, and 0 among letter distractors and obtained flat search slopes, consistent with popout of the categorically unique targets. Critically, popout was seen for "O" and 0, respectively, despite the fact that in the experiment those two stimuli were physically identical. In control conditions in which the distractors were in the same

category as the targets, the 0 no longer popped out among digits, nor the "O" among letters: There was a search slope of over 20 ms/item. Another finding of categorically defined popout in visual search was found by Egeth, Jonides, and Wall (1972). They showed that the reaction time to detect a digit in an alphabetical display is basically constant with respect to the number of distractors (3.9 ms/item).

Schneider and Shiffrin (1977) explored categorically triggered processing with a paradigm similar to RSVP, in which subjects reported target letters from a stream of briefly presented frames, each containing one, two, or four letters. Among their findings were that subjects who were given the same set of letters as targets over hundreds of trials were eventually able to detect targets without regard to the number of distractor letters on the frame, whereas subjects who received a new set of targets on every trial continued to be affected by the number of distractors. These results again suggest that membership in a well-learned target category is a potent attentional cue.

Kyllingsbaek, Schneider, and Bundesen (2001) built on this methodology with a specially constructed set of alphabetic stimuli that were composed of a regimented set of line segments, so as to equate low-level featural information between target and distractor sets. In these experiments, subjects were trained to create an arbitrary categorical division of the alphabet (RDHZNTBGC versus PQFLXMSKV) over the course of thousands of trials, with one set or the other as defined targets. Afterwards, the set of letters that had formerly been targets were able to capture attention when presented just before, during, or after a critical target frame, as demonstrated by a reduction in target detection. Thus, attention was cued away from the location of the target (i.e., was captured) by a highly trained former target, despite the low-level featural similarity between targets, distractors, and former targets.

In a study by Shih (2000), subjects viewed two parallel RSVP streams to the left and right of fixation. They reported two digit targets presented among letter distractors in either the same or different streams at different temporal intervals. When the second target followed the first by 70–210 ms, there was an interaction between the relative spatial locations of the two targets (i.e., same versus different) resulting in enhanced ability to report the second target when the two were in the same location, relative to the different location condition. At longer SOAs the difference between the conditions disappeared. This work is important in demonstrating a spatial bias of attention that is effective within 100 ms of a presented target, although the experiment was not designed to demonstrate true enhancement of second target report; there was no baseline accuracy uncontaminated by T1 processing. A chief aim of the present manuscript is to show that this spatial bias is an enhancement when measured against a baseline.

Another important study in this domain is that of Barnard, Scott, Taylor, May, and Knightley (2004). Subjects searched a single RSVP stream of words for a semantically defined target category, a human occupation. Distractor words semantically related to the target category evoked an attentional blink (an increased difficulty in processing a target appearing within about 500 ms of an earlier target) and Lag 1 sparing (successful processing of an immediately following target). This result implies that, even at the level of word processing, stimuli can be rapidly evaluated according to their semantic proximity to a target set, producing a deployment of attentional resources with a time course comparable to detection of

letters or digits. These studies firmly establish that categorically defined targets can be used as attentional cues, but it remains to be shown that manipulating attention with a categorically defined target set can result in enhanced processing.

Categorical Detection of Targets

Debate persists regarding the question of whether the learned categorical boundaries of stimuli such as letters and digits underlie the search slope advantage of categorically defined targets (Egeth et al., 1972) or whether differences in perceptual features between natural categories allow the viewer to distinguish them. For example, digits tend to have more rounded edges than letters; and feature detectors tuned for curved elements could, on average, pick up digits more often than letters. Although Jonides and Gleitman (1972), in the study described above, report unequivocal evidence in favor of true categorical perception, the effect has not been easy to replicate (Duncan, 1983). The use of inverted characters is one promising way of distinguishing true categorical effects from featural confounds (Hamilton, Mirkin, & Polk, 2006), but the issue is difficult to resolve decisively because, apart from a few examples of ambiguous symbols (e.g., O and 0), the categorically distinct stimuli are also physically distinct. The present set of experiments does not address the debate of how categorically defined targets are detected; we focus instead on the attentional effects that occur in response to detecting a letter or digit target.

Briefly Presented Targets as Attentional Cues

Building on previous findings, the present experiments focus on the spatial and temporal aspects of attentional cueing with categorically defined target sets using a novel methodology. Attention is cued to a specific location by a stimulus that belongs to the target category (e.g., letters or digits), but this leading target is presented too briefly to be reliably perceived. We suggest that, with this manipulation, the leading target acts as an attentional cue. That is, the target item is present long enough to attract attention as a probable target but not long enough to be encoded as a reportable percept on a majority of trials.

In these experiments, the first target can appear either in the same location as the second target (acting as a valid cue that is expected to facilitate report of the second target), in a different location (acting as a misleading or invalid cue as in attentional capture), or not at all (allowing us to ascertain the baseline reportability of a target in the absence of an attentional manipulation).

Participants were instructed to report the identity of any targets they see and were told only that there would be one or two of them presented in each trial. Thus, we can compare trials on which the first target was reported with those on which it was not. This conditional analysis allowed us to test the hypothesis that the visual system can use categorically defined targets to deploy attention contingently. The initial detection of a match between a target and the target set precedes the deployment of selective attention and any subsequent encoding of that target. Therefore, this preattentive category detection need not result in a consciously reportable percept of the target item that triggered attention. Accordingly, this theory has a clear-cut prediction: If two targets are presented in close spatiotemporal proximity (i.e., in the same location and at an SOA of about 100 ms), identification of the

second target should be enhanced even if the first target is not itself reported. We explore this prediction throughout this article.

Experiment 1: Spatiotemporal Factors in Categorically Defined Transient Attention

Experiment 1 tested the hypothesis that a member of a categorically defined target set (i.e., a digit) can initiate transient attention at its location in a spatial array that is continually changing. We consider a critical test of transient attention to be the ability of a first, brief target (the *leading target*, LT) to (a) benefit a second target (the *critical target*, CT) in the same location, within about 100 ms of the onset of the first target but not at longer SOAs and to (b) interfere with the second target if the two are in different locations.

Method

Participants. Fifteen participants were volunteers from the Massachusetts Institute of Technology community of age 18–35 years, who were paid to participate in the experiment. All reported normal or corrected-to-normal vision. The results of 1 participant were excluded because of a false alarm rate that was more than two standard deviations above the mean for the group.

Apparatus and stimuli. The experiment was programmed using Matlab 5.2.1 and the Psychophysics Toolbox extension (Brainard, 1997), and it was run on a PowerMac G3. The Apple 17-in. monitor was set to a 1,024 × 768 resolution with a 75 Hz refresh rate. The stimuli consisted of targets and distractors (see Figure 1A for bitmap images of the stimuli) presented in eight locations around a fixation cross (see Figure 1B). Spacing between stimulus locations was adjusted to avoid crowding effects (Pelli, Palomares,

& Majaj 2004). At the viewing distance of 50 cm, the characters were approximately 1.3° by 2.1° in visual angle and spaced such that the entire display formed a rectangle approximately 10.5° in height and 8.8° in width. Stimuli were aligned horizontally by centering and vertically along the base of the character. Two of the distractors had descenders protruding slightly below the baseline.

Design and procedure. The LT was presented for 80 ms and the CT for 107 ms. There were five conditions, mixed randomly within a single block of 240 trials (48 in each condition). In the baseline condition, the LT was omitted and replaced by a distractor. In the other four conditions, there were two targets. In two conditions, the LT was presented in the same location as the upcoming CT at an SOA of 107 or 213 ms (valid trials); in the other two conditions, the LT was presented in a different, randomized location at an SOA of 107 or 213 ms (invalid trials).

A trial began with a fixation cross at the center of the screen. After 500 ms, randomly chosen distractors began to appear in the eight selected locations surrounding the fixation cross. One item, at a random location, changed every 27 ms, replacing the previous stimulus at the same location. This pattern continued through the entire trial, which lasted from 933 to 1,707 ms after the onset of the first item. The random change of one item every 27 ms ensured that target onsets had no intrinsic salience. Furthermore, the duration of each stimulus (other than the targets) was effectively randomized: The average duration of a distractor was 213 ms but could be as short as 27 ms or could last the entire trial. Thus, the visual system was unable to use onsets of new items or their durations as effective cues for the deployment of attention.[1]

Prior to each trial, a CT and LT were selected from the target set at random without replacement. The spatial location of the CT was selected randomly. The CT was presented at a random time-step between 400 and 1,200 ms after the onset of the array. The trial continued for at least 513 ms following the onset of the CT. The location and timing of the LT were determined relative to the CT. The temporal structure of a valid short-SOA trial is shown in Figure 2. Invalid trials were identical except that the LT was presented in a randomly chosen location different from that of the CT.

Participants were seated 50 cm from the monitor in a dimly illuminated room. At the end of the trial, the set of possible targets (2, 3, 4, 5, 6, 7, 8, 9) was presented across the top of the screen and a prompt instructed participants to "Enter any digits that you saw and press Enter." No feedback was given. There were three sets of eight practice trials, containing randomly mixed trials of each condition, in which the CT was presented for extended periods of time, progressively shortened from one practice block to the next (durations of 540, 270, and 189 ms).

Scoring. If participants reported more than two digits, only the first two digits reported were included in the analysis.

A.

Targets 2 3 4 5 6 7 8 9

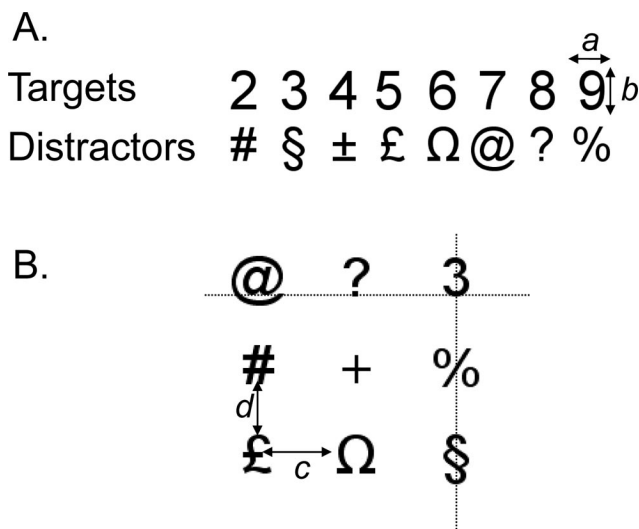Distractors # § ± £ Ω @ ? %

B.

@ ? 3

+ %

£ Ω $

Figure 1. Stimuli used in Experiments 1 and 2 (A) and one example arrangement (B) showing one target and seven distractors. The dashed lines, arrows, and letters are shown only to illustrate the alignment and spacing of characters; they were not present in the stimulus. Size and spacing in degrees of visual angle were as follows: a = 1.3, b = 2.1, c = 3.3, d = 1.3. Shown in this figure are screen captured images of the stimuli used, as created with the Arial font.

---

[1] Presentation time was recorded to allow us to exclude any trials for which frame update deviated from the 27 ms schedule by at least an entire monitor refresh cycle (13 ms) for the period ranging from 200 ms before the LT to 200 ms after the CT. This check excluded 0.5% of the trials. A similar check excluded 0.6% and 1.3% of trials in Experiments 2 and 3, respectively. No trials were excluded for Experiments 4 and 5.
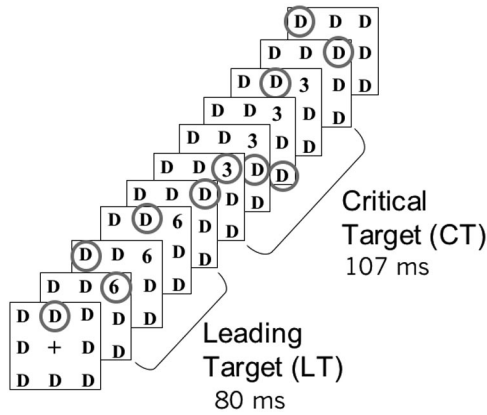
Figure 2. Excerpt of events surrounding the presentation of targets amidst an ongoing presentation of distractors (D). Each frame indicates a time step of 27 ms (two screen refresh cycles at 75 Hz). Onsets of new items, either a target, or a distractor that replaces a previous item, are graphically indicated by the gray circles, which are drawn for illustration. This is an example of a valid/107-ms SOA trial, in which the leading target (LT), 6, was presented in the same location as the critical target (CT), 3. The square outline frames are drawn for illustration. Each D represents a symbol distractor from the set shown in Figure 1.

Results and Discussion

The results were consistent with the idea that the LT captures attention: When the LT was in the same location as the CT at an SOA of 107 ms (but not at 213 ms), CT report was more accurate than when the LT was omitted; when the LT was at a different location at an SOA of 107 ms, the accuracy of the CT was reduced.

Responses. Participants made an average of 0.92 responses per trial out of a possible 2 responses: 0.69 of these were correct (i.e., the response matched the CT or LT) and 0.23 were false.

CT accuracy. Identification accuracy of the CT and the LT in the five conditions is shown in Table 1. Figure 3 graphically depicts CT accuracy in the five conditions. A 2 × 2 analysis of variance (ANOVA) of CT accuracy, using SOA and validity as factors while disregarding the uncued trials, found main effects of SOA, $F(1, 13) = 18.31$, $\eta_p^2 = .58$, $p < .0015$, and validity, $F(1, 13) = 71.5$, $\eta_p^2 = .85$, $p < .001$, and a significant interaction

Table 1
Proportion of Correct Reports in Experiment 1

| Report | SOA 107 ms | | SOA 213 ms | | CT only |
	Valid	Invalid	Valid	Invalid	
LT	.19 (.03)	.32 (.03)	.31 (.03)	.31 (.04)	
CT	.66 (.05)	.34 (.04)	.49 (.04)	.35 (.03)	.48 (.05)
CT|LT	.58 (.08)	.16 (.05)	.45 (.07)	.15 (.04)	
CT|!LT	.68 (.05)	.43 (.05)	.50 (.05)	.46 (.05)	

Note. Reports are broken down as leading targets (LT), critical targets (CT), CT conditional on LT report (CT\|LT), and CT conditional on no LT report (CT\|!LT). On valid trials, the LT was presented in the same location as the CT; on invalid trials, it was presented in a different location. Standard errors are in parentheses. There were 14 participants, each of whom performed 240 trials spread among the five conditions. SOA = stimulus onset asynchrony.
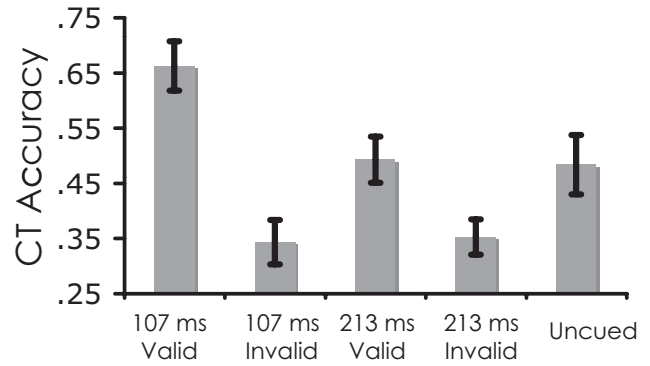

Figure 3. Accuracy of critical target (CT) report for the five conditions in Experiment 1. For the four conditions with two targets, the data are labeled with the SOA between the two targets and their relative locations (valid and invalid). Standard error bars are shown in this and all following figures.

between the two, $F(1, 13) = 15.69$, $\eta_p^2 = .57$, $p < .003$. These effects, and especially the interaction, show that manipulating the spatiotemporal relationship between the LT and the CT affected the ability to report the CT.

Subsequent focused analyses compared CT accuracy in each of the four cued conditions with CT accuracy in the uncued (i.e., baseline) condition. CT identification accuracy was enhanced when the LT was presented at a 107-ms SOA in the same location (valid trials) .66 versus .48, paired $t(13) = 5.44$, $p < .001$, Cohen's $d = .97$.[2] When the LT was in a different location (invalid trials) at the short SOA, CT accuracy was impaired, .34 versus .48, paired $t(13) = 4.15$, $p < .002$, Cohen's $d = .79$. These results suggest that attention is drawn to the location of the LT, improving accuracy in valid trials and impairing accuracy in invalid trials.

At the longer SOA, the accuracy advantage for valid trials disappeared, .49 versus .48, paired $t(13) = .18$, $p > .5$, Cohen's $d = .05$. This result suggests that transient attention is indeed short lasting. However, on invalid trials at the longer SOA, there was a drop in CT accuracy, .35 versus .48, paired $t(13) = 3.9$, $p < .002$, Cohen's $d = .79$. Evidently attention did not readily move to a new target location even after the positive benefit of transient attention had dissipated.

LT accuracy. On trials with two targets, the LT was reported on less than half the trials. Accuracy of LT report is shown in Table 1. A 2 × 2 ANOVA with SOA and validity as factors revealed main effects of SOA, $F(1, 13) = 9.1$, $\eta_p^2 = .41$, $p < .02$, and validity, $F(1, 13) = 5.4$, $\eta_p^2 = .30$, $p < .05$, modified by a strong interaction, $F(1, 13) = 21.7$, $\eta_p^2 = .62$, $p < .001$. Inspection of the data showed that the interaction was the result of a reduction in LT accuracy in the valid/107-ms SOA condition. A one-way ANOVA of LT accuracy over the other three conditions revealed no significant difference among them, $F(2, 13) = .126$, $\eta_p^2 = .01$, $p > .5$. The reduction in LT accuracy in the valid/107-ms SOA condition suggests a competitive relationship between the two targets.

---

[2] Cohen's d is computed as the ratio of the absolute value of the difference between the means and the square root of the average of the standard deviations of the two conditions.

Interaction between LT and CT. To explore the conditional relationship between LT and CT report, we examined CT performance conditionally on whether the LT had been reported (CT|LT) or had not been reported (CT|!LT). The beneficial effect produced by a valid LT was not contingent on the participants' ability to report it. CT report in the valid/107-ms SOA condition was not significantly different whether the LT was reported or not, paired $t(13) = 1.1$, $p > .25$, Cohen's $d = .39$, although there was a trend for CT report accuracy to be higher for those trials in which the LT was not reported. The fact that CT report can be strongly enhanced in the absence of LT report shows that the attentional effect is not critically dependent on successfully encoding the LT into working memory. In the invalid conditions, correct report of the LT markedly reduced report of the CT compared to the uncued condition for both short and long SOAs, both $ps < .002$, both $ds > .95$, whereas the CT at both SOAs was minimally affected when the LT was not reported, both $ps > .05$, both $ds < .2$. As predicted, transient attention to the LT location only benefited the CT at an SOA of 107 ms, and not 213 ms. Why, then, did an invalid LT interfere with the CT not only at an SOA of 107 ms but also at 213 ms? The key seems to be the differential effect of report of the LT in the two cases. The benefit of transient attention was minimally affected by whether the LT was reported, whereas the deficit in the invalid condition only occurred when the LT was reported, possibly because sustained attention to that location was associated with report.

Accuracy of the CT at 213-ms SOA was not impaired in the valid condition, even on trials in which the LT was reported. This is surprising, in that, at this SOA, one might expect evidence of an attentional blink induced by the LT. However, the blink can be attenuated by inducing a distributed state of attention (Olivers & Nieuwenhuis, 2005). The randomized presentation pattern used in the present experiments may have produced the same effect.

In summary, the results of Experiment 1 show that an LT presented too briefly to be reported on more than half the trials can nonetheless draw attention to its location, either benefiting a CT arriving 107 ms later at the same location or impairing report of the CT when it appears at a different location. Critically the benefit to the CT was not contingent on report of the LT, suggesting that target detection precedes encoding of a reportable percept.

Experiment 2: Transient Attention in Four Successive Arrays

Experiment 2 replicated Experiment 1 using a simplified procedure. In Experiment 1, the LT was always associated with a particular temporal pattern of transients: The LT appeared for 80 ms, followed by a distractor. Possibly the attentional system is able to pick up on a regular pattern of such stimulus transitions if it is predictive of a target (Turk-Browne, Junge, & Scholl, 2005) despite the random replacement of a distractor every 27 ms. In Experiment 2, we presented only four consecutive arrays, replicating three conditions of Experiment 1: valid and invalid trials with the LT and CT at a short SOA and the no-LT condition. Because all characters—distractors and targets—changed on each array, there was no specific temporal pattern associated with the LT.

Method

The method of Experiment 2 was similar to that of Experiment 1 with the following exceptions. Participants were shown only four successive arrays, each with eight items. The first array, which appeared for 67 ms, sometimes included an LT; the second array, for 27 ms, consisted of distractors; the third array, for 107 ms, included a CT; and the fourth array, for 27 ms, again consisted of distractors, All distractors were randomly shuffled on each frame, producing simultaneous transitions at all locations rather than a sequential pattern as in Experiment 1.

Participants. Twelve new participants were recruited from the same pool used in Experiment 1.

Design and procedure. There were three types of trials: valid, invalid, and uncued. Each trial had a CT. Valid and invalid trials contained an LT. Figure 4 depicts an example of a valid trial. CTs and LTs were chosen from the set of seven digits (2, 3, 4, 5, 6, 7, 8) such that the 42 different combinations of CT and LT (without repetition within a trial) were presented within each of the three conditions, for a total of 126 trial types (the LT designation was virtual in the uncued condition). These 126 trials were randomly intermixed within each of two blocks, for a total of 252 trials per participant. Eighteen practice trials with a longer target duration preceded the experimental trials.

Each trial began with a fixation cross presented for 500 ms, followed by four successive arrays of eight stimuli surrounding a cross in the same spatial configuration as Experiment 1. Each array contained one of each of the eight distractors in a randomized order except when a target replaced one of the distractors. The distractors in each location were shuffled on each successive array. The first array was present for 67 ms, the second for 27 ms, the third for 107 ms, and the final array for 27 ms. Thus, the SOA between the LT and the CT was always 93 ms. For valid trials, the

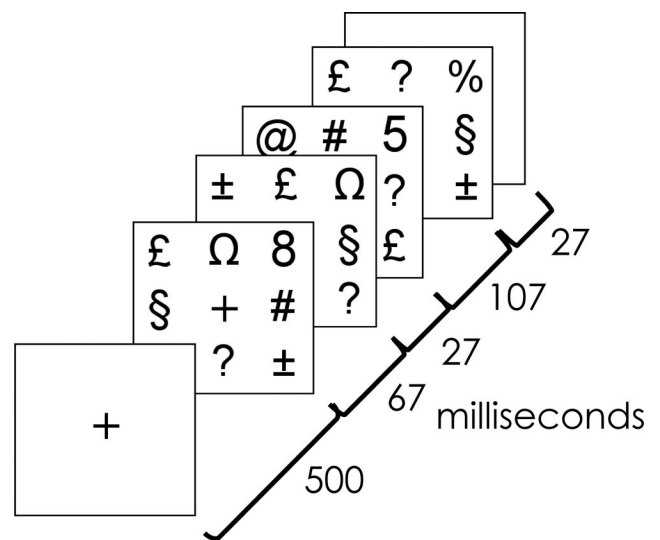*Figure 4.* Example of a valid trial in Experiment 2. The leading target (LT), 8, is presented amidst a shuffled array of symbol distractors. After a brief mask, the critical target (CT), 5, is presented in the same location. Invalid trials had the same temporal structure except that the LT was presented in a different, randomly chosen location. On uncued trials, the LT was replaced with a distractor.

LT was presented in the same (random) location as the CT. For invalid trials, the LT's location was chosen randomly from one of the other seven locations.

Results and Discussion

The results of Experiment 2 are shown in Table 2 and Figure 5. Participants made an average of 1.10 responses per trial, of which 0.80 were correct reports of one or the other target.

A one-way ANOVA of CT accuracy indicated a significant difference among the three conditions, $F(2, 22) = 33$, $\eta_p^2 = .75$, $p < .001$. The data replicated the pattern of results of Experiment 1 in the corresponding conditions. CT identification was better in the valid condition than in the uncued baseline, as shown by a paired t test, .64 versus .54, $t(11) = 4.8$, $p < .001$, Cohen's $d = .63$. CT identification in the invalid condition was worse than in the uncued baseline, .44 versus .54, $t(11) = 3.67$, $p < .005$, Cohen's $d = .77$. As in Experiment 1, LT identification was worse in the valid than in the invalid condition, .32 versus .45, paired $t(11) = 3.84$, $p < .003$, Cohen's $d = .57$, which suggests again that there is a competition between the CT and the LT. Critically, the CT enhancement effect on valid trials was unaffected by whether the LT was or was not reported. CT accuracy was .67 when LT was reported, .64 when it was not. Most importantly, on trials in which participants failed to report the LT, report of the CT was enhanced relative to trials in which no LT was present, .64 versus .54 paired $t(11) = 3.07$, $p < .02$, Cohen's $d = .38$,

As in Experiment 1, the cost incurred by invalid cueing was exclusively present on trials in which the LT was reported. For invalid CT|!LT trials, CT accuracy was in fact weakly enhanced relative to baseline, .60 versus .54, $t(11) = 3.07$, $p < .02$, Cohen's $d = .38$. This could reflect the fact that trials for which participants failed to report the LT would have tended to be those in which attention was directed away from the LT location. Thus, on invalid trials without report of the LT, the participants may have been more likely to be attending to the CT location.

Experiment 2 demonstrates that the effects on CT accuracy produced by valid and invalid presentation of an LT in Experiment 1 were not an artifact of the temporal pattern of presentation. In Experiment 2 all eight stimuli, including the LT, appeared simultaneously, and yet the LT was able to enhance the subsequent CT when in the same location and interfere with the CT when it was in a different location.

Table 2
Proportion of Correct Reports in Experiment 2

Report	Valid	Invalid	CT only
LT	.32 (.06)	.45 (.07)	
CT	.64 (.04)	.44 (.03)	.54 (.04)
CT|LT	.67 (.03)	.25 (.05)	
CT|!LT	.64 (.05)	.60 (.05)	

Note. Reports are broken down as leading targets (LT), critical targets (CT), CT conditional on LT report (CT|LT), and CT conditional on no LT report (CT|!LT). Standard errors are in parentheses. On valid trials, the LT was presented in the same location as the CT; on invalid trials, it was presented in a different location. There were 12 participants, each of whom performed 252 trials spread among the three conditions.
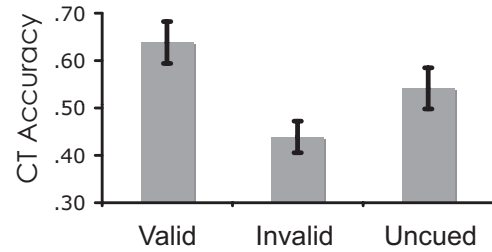

Figure 5. Accuracy of critical target (CT) report in the three conditions of Experiment 2. In valid trials the two targets appeared in the same location. For invalid trials, the targets appeared in different locations and reduced subsequent CT identification accuracy. In uncued trials, the leading target (LT) was replaced by a distractor.

Experiment 3: Transient Attention or Category Priming?

In Experiments 1 and 2, the effective cue was a member of the same category–digits –as the critical target. Could the cueing effect be a form of category priming in which digits prime other digits? An LT might temporarily activate items in the target category, leading to a facilitation of target report due to a short-lasting priming effect. We considered this hypothesis to be unlikely, however, as the cueing effects in Experiments 1 and 2 were obtained only when the cue was in the same location as the target. This strict spatial correlation is not characteristic of semantic priming effects. Moreover, the cueing effect was eliminated with a slightly longer SOA (213 ms), whereas semantic priming effects tend to be prolonged (Chiarello, Liu, Shears, Quan, & Kacinik, 2003).

Nevertheless, it is important to investigate the possible contribution of category priming to the observed enhancement of target report accuracy. In Experiment 3, participants were instructed to report any targets from two categories– digits and letters–allowing us to compare the benefit of cueing the CT with an LT of the same versus a different target category. If this improvement in performance results from category priming, the enhancement effect should be more prominent when the CT and LT are from the same category: letters or digits. On the other hand, if the benefit is due to spatial attention, the enhancement should be blind to the categorical similarity between the leading and the critical targets.

Method

The method was the same as that of Experiment 1 except as noted. In this experiment, participants were instructed to report targets from either of two distinct categories (letters and digits). CTs (letters or digits) were preceded by a leading item that was either a letter, a digit, or a distractor.

Participants. The 15 participants were volunteer students or staff at the University of Kent, Canterbury, United Kingdom, or the Massachusetts Institute of Technology, of ages 18–35 years, reporting normal or corrected-to-normal vision. Participants were paid for their participation. The results of 1 participant were excluded because of a false-alarm rate that was more than two standard deviations above the mean for the group.

Apparatus and stimuli. The targets were eight letters and eight digits from the following set: B C D E F G H R 2 3 4 5 6 7 8 9.

The character font for targets and distractors was Arial, as in Experiments 1 and 2.

Design and procedure. There were two blocks of 120 trials each. The design was a 3 × 2 factorial design and all trials were structured in the manner of the short-SOA/valid condition of Experiment 1, meaning that a CT was always preceded by a leading item in the same location at an SOA of 107 ms. The LT was presented for 80 ms and the CT for 107 ms. The leading item was a letter, a digit, or, for the no-cue control condition, a distractor symbol (the three levels of the first factor). The CT was either a digit or a letter (the two levels of the second factor).

The leading item and the CT were randomly selected without replacement from the set of eight possible values to ensure that the CT and leading item were always different. Furthermore, when the leading item was a distractor (in the uncued condition) that item was removed from the pool of available distractors for the rest of the stream.

Results and Discussion

The results of Experiment 3 can be seen in Figure 6 and Table 3. Participants made an average of 0.81 responses out of a possible two responses per trial, of which 0.64 were correct.

An ANOVA of the accuracy of report of the CT with the category of the CT and the LT as factors showed that there was a main effect of CT category (letter or digit), with digits better perceived than letters, .66 versus .45, $F(1, 13) = 35.74$, $\eta_p^2 = .73$, $p < .001$. There was also a highly significant effect of the category of the leading item (letter, digit, distractor) on CT report, .58, .62, and .48, respectively, $F(2, 26) = 17.8$, $\eta_p^2 = .58$, $p < .001$, indicating a strong cueing effect for leading items drawn from the target set as opposed to the distractor set. However, there was no interaction between the category of the CT and the LT, $F(2, 26) = .335$, $p > .72$, $\eta_p^2 = .025$.

The purpose of this experiment was to determine whether LTs of a particular category (letter or digit) are more effective cues of CTs of the same category than of a different category. The failure

Table 3
Proportion of Correct Reports in Experiment 3

Report	Letter CT			Digit CT		
	Leading item			Leading item		
	Letter	Digit	Symbol	Letter	Digit	Symbol
CT	.46 (.04)	.51 (.05)	.38 (.03)	.69 (.03)	.72 (.03)	.57 (.04)
CT|LT	.48 (.10)	.39 (.09)		.56 (.11)	.72 (.10)	
CT|!LT	.47 (.04)	.54 (.04)		.69 (.03)	.71 (.03)	

Note. Results are separated according to the category of the critical target (CT) and the leading target (LT), CT conditional on LT report (CT|LT), and CT conditional on no LT report (CT|!LT). Standard errors are in parentheses. There were 14 participants, each of whom performed 240 trials spread among the six conditions. In the symbol condition, the LT was replaced by a distractor.

to find an interaction in the preceding analysis suggests there was no such relationship between the categories of the CT and the LT. To explore the possibility further, a more focused set of analyses was performed. We first determined whether a cueing effect was present for all possible combinations of CT and LT. In a planned series of comparisons, CT accuracy in each of the four conditions containing an LT (either digit or letter) was found to be significantly higher than accuracy for CTs that were not preceded by a target: Letter–Letter, $p < .02$, $d = .69$; Digit–Letter, $p < .002$, $d = .92$; Letter–Digit, $p < .002$, $d = .88$; Digit–Digit, $p < .001$, $d = 1.19$. Critically, despite the fact that there was sufficient power to detect cueing effects on all four target-cued conditions, a 2 × 2 ANOVA on CT accuracy with LT and CT category as factors (excluding trials without an LT) found no interaction, $F(1, 13) = .026$, $\eta_p^2 = .002$, $p > .5$, no significant effect of cue set, $F(1, 13) = 3.876$, $p > .06$, $\eta_p^2 = .23$, and a strong effect of target set, $F(1, 13) = 36.477$, $p < .001$, $\eta_p^2 = .74$. Thus, there is no evidence that an LT elicits a form of attention or priming that favors a CT of the same category.
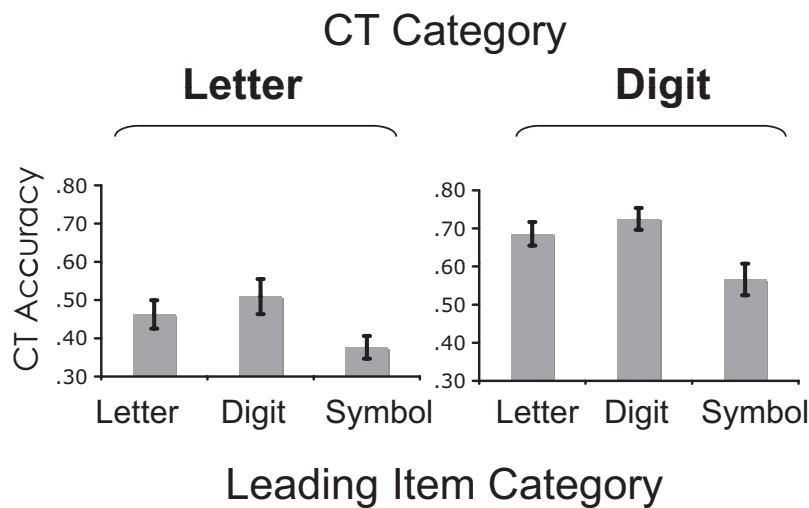

Figure 6. Critical target (CT) accuracy as a function of the category of the leading target (LT) and CT in Experiment 3. Symbol refers to distractor symbols, corresponding to uncued trials.

As with the previous experiments, the enhancement effect on the CT was not contingent on having reported the LT. Overall, on LT-containing trials, the proportion of correct CT|LT was .56 and CT|!LT was .60, paired $t(13) = .53$, $p > .6$, Cohen's $d = .18$. A series of t tests was performed to show that the enhancement effect of LT presence occurred for trials in which the LT was not perceived, as in the previous experiments. Four tests on the values comparing CT|!LT to uncued CT scores in Table 3 produced the following values: Letter LT/Letter CT, $t(13) = 2.7$, $p < .02$, $d = .66$; Digit LT/ Letter, CT $t(13) = 4.4$, $p < .001$, $d = 1.3$; Letter LT/Digit CT, $t(13) = 10.1$, $p < .001$, $d = .96$; Digit LT/Digit, CT $t(13) = 9.0$, $p < .001$, $d = 1.12$. Thus, as in the earlier experiments, a cueing benefit for the CT was not contingent on correct report of the LT.

Experiment 4: Use of a Spatially Uninformative LT

In the preceding experiments the LT, when it appeared, was positively correlated with the location of the following CT. In Experiments 1 and 2, the LT, when present, appeared in the same location as the CT on 50% of the trials and in another of the seven locations on the remaining 50% of trials. In Experiment 3, the LT was always in the same location as the CT.

It is possible that the visual system learned this statistical relationship; similar partially valid associations have been shown to be learned in visual tasks (Fiser & Aslin 2002; Turk-Browne et al., 2005). To control for this possibility, in Experiment 4 we removed the correlation between the LT and the CT locations. We predicted that the effects of LT validity on CT identification scores would still be found after this change.

Method

Participants. The 10 participants were volunteer students or staff at the Massachusetts Institute of Technology, of ages 18-35 years, reporting normal or corrected-to-normal vision. Participants were paid for their participation.

Apparatus and stimuli. Targets were the digits 2–9, and distractors were the symbols used in Experiment 1 (see Figure 1A). There were four stimuli locations arranged around the fixation in a square formation. Stimuli were approximately 1.4° by 2° in visual angle with their centers 3.4° above/below and right/left of the center of a fixation cross, which measured 0.6° wide and 0.3° in height.

Design and procedure. Participants saw a series of brief spatial arrays containing one or two targets. Each trial consisted of 14 arrays that changed at 53-ms intervals in all of the four locations, except that the CT remained for 107 ms. Distractors were chosen randomly from the same set of symbols used in Experiment 1 with the restriction that no location contained the same distractor on two consecutive temporal arrays. The digit targets, both the LT and the CT, were chosen at random without replacement. On 20% of the trials, the CT was presented alone to serve as a baseline. On the remaining 80% of trials, both an LT and a CT were presented, with the four possible locations fully counterbalanced such that the two targets were in the same location on 25% of the trials. That is, the location of the LT was uninformative as to the location of the following CT. There were 16 practice and 320 experimental trials per participant.

The LT could appear at positions 6, 7, 8, or 9 in the temporal sequence of distractor locations and was presented for 53 ms. The CT was presented 107 ms after the onset of the LT, so that there was always a 53-ms distractor presentation between the LT and the CT. The CT lasted for 107 ms. On the 20% of trials without a leading target, the LT was replaced by a randomly chosen distractor, subject to the prohibition of distractor repetition in the same location. Participants were instructed, as before, that they would see one or two digits and to report them in order. No feedback was provided and participants were not shown the target set.

Results and Discussion

The results of this experiment are presented in Table 4 and replicate the general finding seen in previous experiments. Participants made an average of 1.10 responses per trial, of which 0.87 were correct. A one-way ANOVA of CT accuracy indicated a significant difference among the three conditions, $F(2, 18) = 63.6$, $\eta_p^2 = .87$, $p < .001$.

CT identification was better in the valid condition than in the uncued baseline, as shown by a paired t test, .79 versus .64, $t(9) = 6.8$, $p < .001$, Cohen's $d = 1.3$. CT identification in the invalid condition was worse than in the uncued baseline, .64 versus .57, $t(9) = 3.9$, $p < .004$, Cohen's $d = .61$. As in Experiment 1, LT identification was worse in the valid than in the invalid condition, .25 versus .33, paired $t(9) = 3.6$, $p < .01$, Cohen's $d = .58$, which suggests again that there is a competition between the CT and the LT. Finally, as in Experiment 1, the CT enhancement effect on valid trials was unaffected by whether the LT was or was not reported. CT accuracy was .80 when LT was reported, .78 when it was not. As in previous experiments, even for trials in which the LT was not perceived, the CT was perceived more often than in the baseline condition, .78 versus .64, $t(9) = 7.2$, $p < .001$, Cohen's $d = 1.3$.

Experiment 4 demonstrates that the effects on CT accuracy produced by valid and invalid presentation of an LT in the previous experiments were not tied to the spatially informative nature of the LT. Making the LT spatially uninformative, with 75% of trials invalid, produced results that were similar to those in the previous experiments, in which the LT was positively correlated with the location of the CT.

Table 4
Proportion of Correct Reports in Experiment 4

Report	Valid	Invalid	CT only
LT	.25 (.04)	.33 (.05)	
CT	.78 (.03)	.56 (.03)	.63 (.04)
CT|LT	.80 (.05)	.30 (.05)	
CT|!LT	.78 (.03)	.69 (.04)	

Note. Reports are broken down as leading targets (LT), critical targets (CT), CT conditional on LT report (CT|LT), and CT conditional on no LT report (CT|!LT). Standard errors are in parentheses. In this experiment, the LT location was spatially uninformative, so there were three times as many invalid as valid trials. The number of CT-only trials equaled the number of valid trials. There were 10 participants who performed 320 trials spread among the three conditions.

Experiment 5: Redefining the Distractor Set

The distractor set in the previous four experiments had been carefully chosen to be of similar density and featural composition as the target sets so that targets did not stand out from distractors at early levels of visual analysis. However, the symbols used as distractors were less familiar than either letters or digits. The effects we observe could be due to enhanced intrinsic salience of highly familiar stimuli against a background of less familiar symbol distractors or the result of an unplanned but subtle featural distinction between the target and distractor character sets.

To ensure that the properties we have observed are not particular to the symbols used above, Experiment 5 replicated Experiment 4 using a distractor set composed of letters.

Method

The design of this experiment was identical to that of Experiment 4 with the exception that distractors were chosen from the following set of letters: A B C D E F G H. Thirteen participants were chosen from the same pool as that used in Experiment 4. The data of 1 participant with accuracy near chance levels (.15 correct in the uncued condition) were discarded, as were data of 3 other participants with accuracy that was near to ceiling (.97, 1.0 and .97 in the uncued condition).

Results and Discussion

The results of this experiment are presented in Table 5. They replicate the general findings seen in previous experiments. Participants made an average of 1.3 responses per trial, of which 0.98 were correct. A one-way ANOVA of CT accuracy indicated a significant difference among the three conditions, $F(2, 18) = 49.5$, $\eta_p^2 = .84$, $p < .001$.

CT identification was better in the valid condition than in the uncued baseline, as shown by a paired t test, .79 versus .73, $t(9) = 2.8$, $p < .025$, Cohen's $d = .49$. CT identification in the invalid condition was worse than in the uncued baseline, .73 versus .56, $t(9) = 8.1$, $p < .001$, Cohen's $d = 1.3$. LT identification was worse in the valid than in the invalid condition, .37 versus .44, paired $t(9) = 4.5$, $p < .002$, Cohen's $d = .99$. Finally, and most importantly, the enhancement effect of LT presentation on CT identification was especially pronounced on trials in which the LT was not reported, .81 versus .73, $t(9) = 4.5$, $p < .002$, Cohen's $d = .66$.

Table 5
Proportion of Correct Reports in Experiment 5

Report	Valid	Invalid	CT only
LT	.37 (.04)	.44 (.03)	
CT	.79 (.04)	.55 (.04)	.73 (.04)
CT|LT	.73 (.06)	.30 (.05)	
CT|!LT	.81 (.4)	.73 (.04)	

Note. Reports are broken down as leading targets (LT), critical targets (CT), CT conditional on LT report (CT|LT), and CT conditional on no LT report (CT|!LT). Standard errors are in parentheses. In this experiment, the LT location was spatially uninformative, so there were three times as many invalid as valid trials. The number of CT-only trials equaled the number of valid trials. There were 10 participants who performed 320 trials spread among the three conditions.

These results demonstrate that the same effects are observed even when targets and distractors are both highly familiar to participants. Thus, the cueing effects observed in the previous experiments cannot be ascribed to differing levels of familiarity between the target and distractor sets nor to gross featural differences between the symbols and the letter or digit target sets.

General Discussion

The results of the present experiments support the hypothesis that stimuli can trigger a spatially restricted, transient deployment of attention by virtue of belonging to a categorically defined target set. For participants seeking a categorically defined set of stimuli, presentation of one target cued attentional resources to its own location; enhancing report of a subsequent target at the same location 107 ms later while impairing report at a different location. At an SOA of 213 ms, this benefit was not observed. Critically, in all of these experiments, the enhancement effect that was observed for valid cueing at a short SOA was pronounced and in some cases exaggerated on trials in which participants failed to report the leading target. This finding supports the hypothesis that a preattentive mechanism detects a member of the target category and deploys attention that enhances identification of a second target arriving within about 100 ms.

By focusing on trials on which the briefly presented LT is missed, the present results reveal rapid attentional shifts of spatial attention produced by an unreported target. This is in contrast to earlier studies, (e.g., Duncan, Ward, & Shapiro, 1994; Reeves & Sperling, 1986; Weichselgartner & Sperling, 1987) in which the subject encodes a first target before attention is diverted to another location, resulting in a prolonged attentional dwell time of up to several hundred milliseconds at the first target's location.

Convergent Evidence Against Priming

We conclude that the enhancement of CT identification is not due to category priming, as the data presented here differ from priming effects in several major respects. For one, the effect of the LT on the CT is very short, gone by 213 ms, whereas semantic priming effects are observed at SOAs as long as 800 ms (Chiarello et al., 2003). Further, the effect is restricted to the location of the LT, as we observed in Experiments 1, 2, 4, and 5. If this enhancement were due to category priming, the effect should not be so strongly constrained to one location. Finally, and most importantly, in Experiment 3 the effect was shown to be unaffected by a categorical difference (letters vs. digits) between the LT and CT. If the enhancement effect were due to categorical priming, it should be strongly affected by categorical similarity between the two targets.

A First Pass Categorical Discrimination

A key finding is that the effects of the LT on the report of CT are obtained whether or not the LT itself was reported by the participants. In fact, in every experiment, CT enhancement was greater on trials in which the LT was not reported.

At one level, these data corroborate the findings of both McCormick (1997) and Ivanoff and Klein (2003), who demonstrated that cues that were not consciously reportable due to their

brief presentation and strong masking could still produce an attentional benefit (reduced reaction time) at the cued location. In these experiments, cues were defined by onset.

However, this crucial finding goes beyond a replication of cueing by an unreported stimulus. In our experiments, the LT had to be successfully categorized as a digit or letter in order to trigger attention. Thus, this categorization process must have occurred preattentively by detecting a match between a stimulus and a categorically specified set of targets that cannot be readily discriminated by gross visual features. This detection of a probable target seems to precede the subsequent encoding of the item as a durable conscious representation.

Other work has found a similar distinction between categorization and subsequent identification. In a series of experiments, Grill-Spector and Kanwisher (2005) presented subjects with brief, masked, natural images and asked them to report (a) whether they had seen an object or no object, (b) the basic-level category of the object (e.g., one of 10 responses such as fish, car, or flower), or (c) the specific object (e.g., a German Shepherd versus another breed of dog). As exposure duration was varied, the probability of correct report of the basic-level category of the object matched the ability to tell whether an object had been presented, whereas the ability to report the specific identity of the object required a longer exposure duration. That is, subjects could report that a picture was a flower without being able to specify that it was a rose. Given an additional 20–30 ms of viewing time, the same subjects could resolve the flower's identity. Thus, these results suggest that the initial processing of an object coincides with its basic-level categorization, but specific identification requires additional time.

A number of theories of visual processing have described a process of initial target detection followed by attentional deployment. For example, the biased competition theory of Desimone & Duncan (1995) suggests that visual processing initially occurs in parallel across multiple objects of a visual display, but attention resolves the location of a target within about 100 ms and biases processing towards cells that respond to the target. To accommodate this rapid target detection, it is suggested that even complex objects can be processed as wholes prior to attentional binding. Formally specified models, such as the theory of visual attention (Bundesen, 1990) and the neuronal theory of visual attention (Bundesen, Habekost, & Kyllingsbæk 2005) are computational depictions of how the visual system might use an initial wave of unselective processing to detect pertinent visual stimuli in order to adjust the deployment of attention before subsequently encoding or pigeonholing the identity of relevant objects. These models describe the processing only of simultaneously presented stimuli, but the present set of experimental results can be described in the same terminology. Asking subjects to report the identity of digits configures their attentional mechanism to filter stimuli according to the property of being a digit. A stimulus that matches this filter setting causes a rapid adjustment of attentional weights to focus processing on its spatial location. A following target in the same location benefits from this weighting and is thereby more rapidly pigeonholed (identified) for report.

The demonstration of transient attention as triggered by letter or digit targets validates a number of existing computational accounts of RSVP processing. In a number of recent models, attention acts as a gate that is opened by detection of a target to allow encoding and then quickly closes (Bowman & Wyble, 2007; Nieuwenhuis et

al., 2005; Reeves & Sperling, 1986; Shih, 2008; Weichselgartner & Sperling, 1987). In these models, a rapidly deployed attentional window readily accounts for Lag 1 sparing in the attentional blink effect, which we describe below.

Neurophysiological Support for Contingent Attentional Deployment

Neurophysiological studies of brain function support the hypothesis that processing of visual stimuli up to the level of complex form discrimination proceeds rapidly after onset. Recent experiments recording from IT neurons in macaque monkeys suggest that the ventral pathway can perform this computation rapidly. Within 125 ms of the onset of a complex shape stimulus, these neurons produce a firing pattern with sufficient information to specify the category, identity, and approximate location of that stimulus (Hung, Kreiman, Poggio, & DiCarlo, 2005). Given that it takes parvocellular inputs, which are thought to be critical to object identification, 50–75 ms to reach V1 from the retina (Schmolesky et al., 1998), the ventral stream would have only 50–75 ms to compute this coarse representation through feed-forward processing from V1 to IT cortex.

Neurons within IT cortex have spatially confined receptive fields, corresponding to a distribution over 3° of visual angle (DiCarlo & Maunsell, 2003). A coarse-coded representation over a population of these broadly tuned cells would be able to specify the location of a visual stimulus. Accordingly, a feed-forward representation of target identity in IT cortex would also specify the target's location in the visual field, allowing attention to be redirected to that location.

Electrophysiological evidence in humans finds an electrical signature of the process of detection of a categorically specified target (e.g., animal, vehicle) that begins as early as 125 ms after onset of a novel picture (VanRullen & Thorpe, 2001). VanRullen and Thorpe's results also suggest an earlier (e.g., 80 ms) electrical potential that is specific to a given category (vehicles or animals) and independent of task. These results, along with a wealth of behavioral studies, support the hypothesis that initial detection or categorization occurs in a feed-forward sweep: A visual stimulus induces a rapid cascade of processing through the ventral visual pathway, which provides a coarse approximation of its neural representation (VanRullen, 2007).

This rapid stimulus-categorization process is well suited to control the deployment of transient attention. If top-down control signals were able to selectively bias the activity of IT neurons that correspond to stimuli within the target category, those neurons would receive convergent top-down and bottom-up input whenever a target stimulus was presented in the visual field. Thus, these cells would be able to signal that a probable target exists in the visual field and where it is likely to be. Mechanisms associated with rapid and reflexive attentional deployment, such as the temporoparietal junction (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Serences et al., 2005) could then be recruited to deploy attention reflexively to the relevant location without specific mediation by volitional processes. This conceptual model of contingent, reactive orienting of visual attention is functionally similar to that described by Folk et al. (1992).

Transient Attention, RSVP, and Lag 1 Sparing

In RSVP, subjects view a continuous stream of stimuli in a single location with each stimulus replacing the previous one at a regular SOA, typically 100 ms. In studies of the attentional blink (Raymond, Shapiro, & Arnell, 1992), one or more target items defined by some feature, such as luminance, color, case, or categorical identity, appear in the stream among distractors. If participants are shown two targets in rapid succession at SOAs in the range of 200–500 ms, perception of the second target is impaired: the attentional blink. In many studies of the attentional blink, there is little or no blink when the SOA between targets is less than about 150 ms. This effect is commonly known as Lag 1 sparing because, at the typical rate of 100 ms/item, the spared target follows immediately after the first target (Potter, Chun, Banks, & Muckenhoupt, 1998).

The present work suggests that Lag 1 sparing is the result of transient attention triggered by the first target (T1) that lasts long enough to enhance detection of a following second target (T2) if it occurs closely in time (i.e., within an SOA of about 100 ms or less). For T2s at slightly longer SOAs (200–500 ms), the attentional enhancement triggered by the T1 abates and the T2 is left vulnerable to the attentional blink.

The idea that sparing is related to the deployment of transient attention predicts that the effect should be found only if the targets occur in the same location. In attentional blink studies that have varied the spatial location of targets, sparing is rarely found when T1 and T2 occur in different locations (Breitmeyer, Ehrenstein, Pritchard, Hiscock, & Crisan, 1999; Visser, Bischof, & Di Lollo, 1999; but see Potter, Staub, & O'Connor, 2002, for sparing in proximal streams when viewers know the targets will be in different locations).

If sparing is the result of transient attention of the kind studied here, then sparing should be critically dependent on the temporal relationship between T1 and T2, rather than simply a consequence of T2s directly following T1. That is, sparing should be only incidentally related to the Lag 1 serial position but, rather, should be found within any SOA shorter than about 100 ms, whether or not there is an intervening distractor. In all of the experiments reported here, at least one distractor intervened between the two targets, and yet the attentional enhancement was still observed. Similarly, using a single stream and a presentation duration of 53 ms per item, Potter et al. (2002) found sparing of T2 at Lag 2, corresponding to a 107-ms SOA between the targets. Bowman and Wyble (2007) replicated this finding using digit–letter RSVP streams. Potter et al. also found that T2 has a competitive advantage over T1 at a very short SOA (53 ms), even though both targets were presented for the same duration. Both of these findings suggest an attentional effect that benefits T2 if it occurs within approximately 150 ms of the onset of T1 in an RSVP stream.

The idea that transient attention plays a role in RSVP perception can also explain why the relative duration of T1 and T2 is of critical importance in determining whether sparing is observed. Kristjansson and Nakayama (2002) reported results from an experiment with eight simultaneous RSVP streams of letters at an SOA of 140 ms. These eight streams contained two luminance-marked digit targets at different lags and spatial offsets. In striking contrast to the results reported here, those investigators found that a second target was reported less often when it directly followed the first target in the same location relative to other locations. However, in RSVP at a rate of 140 ms per item, an attentional transient with a duration of 150 ms would benefit T1 more than the following T2, allowing the T1 to compete more effectively than the LT does in the experiments reported here.

Categorical Distinctiveness and Familiarity

Experience clearly determines the ability to categorize rapidly. Consequently, familiarity and categorical distinctiveness, due to a lifetime of exposure, must affect the ability of a particular stimulus to trigger the deployment of attention. For example, in the data of Schneider and Shiffrin (1977), we can observe the development of an initially arbitrary categorical boundary over the course of hundreds of trials. One can ask whether letters and digits would attract attention in a similar paradigm without being treated as targets by the subject. The categorical distinctiveness of digits presented among a different category of distractors may contribute to their ability to trigger attention. Differences in the familiarity of the shapes of targets and distractors may play a role as well, suggesting individual differences that result from each subject's reading history. Another critical issue is to what extent the categories established by a subject's preexperimental experience serve as boundaries in establishing task set. In other words, does attending to a specific subset of digits tend to produce attention to the entire set of digits irrespective of their featural similarity? Experiments that carefully explore categorical distinctiveness, category boundaries, and familiarity as factors that affect the deployment of attention are necessary to understand how visual attention reacts to categorically defined task sets.

Conclusion

The present experiments demonstrate that exogenously driven visual attention can be selective for categorical identity. Like RSVP, the paradigm we have used de-emphasizes the salience of onsets with the intent of isolating the system's ability to respond to object identity. Under these conditions, visual attention reacts to targets specified by category affecting the identification of a second target arriving within about 100 ms. The present results add to a growing body of convergent evidence that visual stimuli undergo rapid processing by the ventral visual stream, producing an evaluation of category prior to the development of a consciously accessible representation.

At a functional level, these results suggest that reactively deployed attention may be useful in a broader array of cognitive functions than previously thought. Rather than serving primarily as a mechanism for reacting to unexpected or arresting visual events, reactive covert attention may play an important role in directed visual search of displays that contain an otherwise overwhelming amount of information, as occurs, for example, when paging through a book rapidly in search of a particular word or scrolling through material on the Web. The time course of deployment of spatiotemporally selective attention, on the order of 100 ms, is well suited to allow a target to trigger a deployment of attention and then be processed, all within a single fixation.

References

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 834–846.

Barnard, P., Scott, S., Taylor, J., May, J., & Knightley, W. (2004). Paying attention to meaning. *Psychological Science, 15,* 179–186.

Bowman, H., & Wyble, B. (2007). The simultaneous type, serial token model of temporal attention and working memory. *Psychological Review 114,* 38–70.

Brainard, D. H. (1997). The psychophysics toolbox, *Spatial Vision, 10,* 433–436.

Breitmeyer, B. G., Ehrenstein, A., Pritchard, K., Hiscock, M., & Crisan, J. (1999). The roles of location specificity and masking mechanisms in the attentional blink. *Perception & Psychophysics, 61,* 798–809.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291–328.

Chiarello, C., Liu, S., Shears, C., Quan, N., & Kacinik, N. (2003). Priming of strong semantic relations in the left and right visual fields: A time-course investigation. *Neuropsychologia, 41,* 721–732.

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.

Colegate, R., L., Hoffman, J. E., & Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics, 14,* 217–224.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience, 3,* 292–297.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

DiCarlo, J. J., & Maunsell, J. H. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89,* 3264–3278.

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception & Psychophysics, 34*(p. 3), 221–232.

Duncan, J., Ward, R., & Shapiro, K. (1994). Direct measurement of attentional dwell time in human vision. *Nature, 369,* 313–315.

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3,* 674–693.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467.

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Perception & Psychophysics, 64,* 741–753.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030–44.

Gobell, J. L., Tseng, C. H., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research, 44,* 1273–1296.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you see it, you know what it is. *Psychological Science, 16,* 152–160.

Hamilton, J. P., Mirkin, M., & Polk, T. A. (2006). Category-level contributions to the alphanumeric category effect in visual search. *Psychonomic Bulletin and Review, 13,* 1074–1077.

Helmholtz, H. von. (1896) *Handbuch der Physiologischen Optik, Dritter Abschnitt, Zweite Auflage* [Handbook of optical psychology, Vol. 3, 2nd ed.]. Hamburg, Germany: Voss.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Ivanoff, J., & Klein, R. M. (2003). Orienting of attention without awareness is affected by measurement-induced attentional control settings. *Journal of Vision, 3,* 32–40.

Jonides, J., & Gleitman, H. (1972). A conceptual category effect in visual search: "0" as a letter or digit. *Perception & Psychophysics, 12,* 457–460.

Kristjansson, A., & Nakayama, K. (2002). The attentional blink in space and time. *Vision Research, 42,* 2039–2050.

Kyllingsbaek, S., Schneider, W. X., & Bundesen, C. (2001). Automatic attraction of attention to former targets in visual displays of letters. *Perception & Psychophysics, 63,* 85–98.

Liu, T., Pestilli, F., & Carrasco, M. (2005). Transient attention enhances perceptual performance and fMRI response in human visual cortex. *Neuron, 45,* 469–477.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 168–180.

McMains, S. A., & Somers, D. C. (2004). Multiple spotlights of attentional selection in human visual cortex. *Neuron, 42,* 677–686.

Muller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 315–330.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29,* 1631–1647.

Nieuwenhuis, S., Gilzenrat, M. S., Holmes, B. D., & Cohen, J. D. (2005). The role of the locus coeruleus in mediating the attentional blink: A neurocomputational theory. *Journal of Experimental Psychology: General, 34,* 291–307.

Olivers, C. N., & Nieuwenhuis, S. (2005). The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. *Psychological Science, 16,* 265–269.

Pelli, D. G., Palomares, M., Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4,* 1136–1169.

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

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma, & D. Bowhuis (Eds.) *Attention & performance X: Control of language processes* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick and I. J. Saltzman (Eds.), *Modes of perceiving and processing information*. Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

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.

Potter, M. C., Staub, A., & O'Connor, D. H. (2002). The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1149–1162.

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.

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206.

Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S.,

Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79,* 3272–3278.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science, 16,* 114–122.

Shih, S., (2000). Recall of two visual targets embedded in RSVP streams of distractors depends on their temporal and spatial relationship. *Perception & Psychophysics, 62,* 1348–1355.

Shih, S. (2008). The attention cascade model and attention blink. *Cognitive Psychology, 56,* 210–236.

Theeuwes, J. (1994a). Endogenous and exogenous control of visual selection. *Perception, 23,* 429–440.

Theeuwes, J. (1994b). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 799–806.

Theeuwes, J., Godijn, R., & Pratt, J. (2004). A new estimation of the duration of attentional dwell time. *Psychonomic Bulletin & Review, 11,* 60–64.

Tipples, J. (2002). Eye gaze is not unique: Automatic orienting in response to uninformative arrows. *Psychonomic Bulletin & Review, 9,* 314–318.

Turk-Browne, N. B., Junge, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564.

VanRullen, R. (2007). The power of the feed-forward sweep. *Advances in Cognitive Psychology, 3,* 167–176.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision making. *Journal of Cognitive Neuroscience, 13,* 454–461.

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

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238,* 778–780.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621.