Psychonomic Bulletin & Review
2004, 11 (5), 876-881

Searching for stimulus-driven shifts of attention

STEVEN L. FRANCONERI
Harvard University, Cambridge, Massachusetts

DANIEL J. SIMONS
University of Illinois at Urbana-Champaign, Urbana, Illinois

and

JUSTIN A. JUNGE
Harvard University, Cambridge, Massachusetts

Several types of dynamic cues (e.g., abrupt onsets, motion) draw attention in visual search tasks even when they are irrelevant. Although these stimuli appear to capture attention in a stimulus-driven fashion, typical visual search tasks might induce an intentional strategy to focus on dynamic events. Because observers can only begin their search when the search display suddenly appears, they might orient to any dynamic display change (Folk, Remington, & Johnston, 1992; Gibson & Kelsey, 1998). If so, the appearance of capture might result from task-induced biases rather than from the properties of the stimulus. In fact, such biases can even create the appearance of stimulus-driven capture by stimuli that typically do not capture attention (Gibson & Kelsey, 1998). The possibility of task-induced, top-down biases plagues the interpretation of all previous studies claiming stimulus-driven attention capture by dynamic stimuli. In two experiments, we attempt to eliminate potential task-induced biases by removing any need to monitor for display changes. In the first experiment, search displays did not change on most trials. In the second experiment, although new search displays appeared on each trial, we ensured that observers never saw the changes, by making them during large saccades. In both cases, dynamic events still received search priority, suggesting that some dynamic stimuli capture attention in a stimulus-driven fashion.

Imagine you are in a crowded airport terminal to meet a friend. When you spot her, you wave your arm wildly, and she sees you. Did you capture her attention? Or, did she notice your arm only because she was searching for someone waving? Would she have noticed you had she not been expecting to meet you? More broadly, can any visual stimulus capture our attention, even when we are not looking for it?

Some dynamic events do seem to draw attention during visual search tasks. For example, objects that appear abruptly are inspected with priority over other items (e.g., Jonides & Yantis, 1988). But is this attentional priority truly driven by the properties of the stimulus, or does the search task itself induce a top-down incentive to intentionally prioritize dynamic events (Folk et al., 1992; Gibson & Kelsey, 1998)? Here we provide new evidence that some dynamic visual events capture attention despite efforts to remove all potential top-down incentives.

The *irrelevant feature search* task is thought to measure stimulus-driven shifts of attention in the absence of top-down interference (Jonides & Yantis, 1988; Yantis & Jonides, 1984). In this task, the observer's only goal is to locate a target letter among distractor letters. On each trial, a randomly chosen letter has a special property (e.g., late appearance, unique color). This property does not predict the target location, and observers have no reason to give it search priority. If observers still inspect the special letter with priority, we can conclude that the property draws attention.

Surprisingly, features that "pop out" in traditional visual search tasks are not searched with any priority in the irrelevant feature search task. For example, observers are not drawn to salient color singletons (e.g., a red item among gray items) (Jonides & Yantis, 1988; but see Turatto & Galfano, 2001), form singletons (Theeuwes, 1990), luminance singletons (Jonides & Yantis, 1988), or luminance changes (Enns, Austen, Di Lollo, Rauschenberger, & Yantis, 2001; Yantis & Hillstrom, 1994). Only a subset of dynamic events appear to receive priority in this task, including the abrupt appearance of a new object (Jonides & Yantis, 1988), sudden motion (Abrams & Christ, 2003; Franconeri & Simons, 2003; Thomas & Luck, 2000), looming (Franconeri & Simons, 2003), and concurrent

This research was supported by National Institutes of Health/National Institute of Mental Health Grant R01 MH63773-02 to D.J.S. He was also supported by a fellowship from the Alfred P. Sloan Foundation. S.L.F. was supported by a National Defense Science and Engineering Graduate fellowship. Thanks to Corey Ip, Matt Kamen, Sarah Scudder, and Damien Williamson for their assistance in data collection. We are also grateful to Chip Folk, Brad Gibson, Todd Horowitz, Stephen Mitroff, Steve Most, Robert Rauschenberger, and Michael Silverman for their valuable comments. Correspondence concerning this article should be addressed to S. L. Franconeri, Department of Psychology, Harvard University, 33 Kirkland St., 7th Floor, Cambridge, MA 02138 (e-mail: francon@wjh.harvard.edu).

changes in luminance contrast and contrast polarity (Enns et al., 2001).

If these dynamic events are irrelevant to the search task and observers have no goal-directed reason to search the cued items with priority, any prioritization may reflect a stimulus-driven shift of attention. However, all prior studies of capture that have used search tasks are vulnerable to a powerful criticism—that dynamic events *are* relevant. Although the dynamic singleton does not predict the target location, the requirements of the task itself might make dynamic events important. In search tasks, observers must await the appearance of a search display in order to begin the search. Monitoring for the display appearance, regardless of the contents of the search display itself, might induce observers to monitor more generally for any visual change. If the display appearance induces observers to actively prioritize dynamic events, dynamic events become a relevant feature in the irrelevant feature search.

In fact, by manipulating the properties that signal the start of the search task, other features that typically do not capture attention in this task can be made to capture. For example, red singletons typically do not capture attention in the irrelevant feature search task. However, when the search display itself is made red, the appearance of red signals the start of the search. Under these conditions, a red precue captures attention even though the color of the search array is irrelevant to the search task itself (find an H or a U) (Gibson & Kelsey, 1998). The red search display causes observers to prioritize red objects because the appearance of the red display items signals the start of the task. If correct, this task-induced priority explanation could undermine all evidence for stimulus-driven attention capture by dynamic events in the irrelevant feature search task.

To date, no demonstration of stimulus-driven attention capture by dynamic events has escaped the possibility of being driven by a task-induced bias. Therefore, in two experiments, we altered the irrelevant feature search such that the start of search was not signaled by a dynamic visual event. If dynamic events still captured attention, task-induced biases might not entirely explain attention capture. In Experiment 1, participants viewed an unchanging display of letters but could not begin their search until a voice prompt identified the target. Shortly before or after the start of the prompt, a randomly chosen letter was cued with a large contrast change. In Experiment 2, displays changed only when the participants made a large saccade to a location offscreen. Thus, participants never saw new displays appear. In both experiments, participants should have no reason to anticipate a change signal accompanying a new search display. Search priority for letters cued by an irrelevant dynamic event would suggest a true stimulus-driven shift of attention.

EXPERIMENT 1

In Experiment 1, search displays did not change on most trials. Although search displays were visible, participants could not respond until a voice prompt specified the target. When search displays did change, the new display appeared immediately after the last response, so that the voice cue, and not a display change, signaled the start of the next trial.

Method

Participants. Twenty-eight undergraduates at Harvard University participated in the study in exchange for $9 or class credit.

Stimuli. Stimuli were displayed on an iMac 15-in. CRT monitor using custom software created with the VisionShell C libraries (http://www.visionshell.com). From an approximate viewing distance of 50 cm, the entire display was 31.28º wide × 23.46º high. Figure 1 depicts a typical search display. The background was gray (17.7 cd/m^2) with a white (107 cd/m^2) fixation point. Three, five, seven, or nine letters (each 1.3º high × 0.9º wide, Geneva bold font, with randomly assigned luminances of 6.9, 10.2, 27.1, or 54.7 cd/m^2) were equally spaced on an imaginary circle around the fixation point at a distance of 5.9º. On 50% of the trials, one letter underwent a concurrent luminance contrast and contrast polarity change from 6.9 to 27.1 cd/m^2 (or 27.1 to 6.9) or 10.2 to 54.7 cd/m^2 (or 54.7 to 10.2) (see Enns et al., 2001). Possible letters were B, C, E, F, G, K, L, P, and R. Approximately half of the letters on each trial were mirror reversed.

Procedure. Although search displays were visible as soon as the trial began, participants could not start their search until a voice prompt specified the name of the target letter (prompts were brief: $M = 387$ msec, $SD = 59$ msec). In order to minimize the incentive to monitor for display changes, the same search array was used for four consecutive trials and then was changed to a new display immediately after the response on the fourth trial. After a response, participants pressed a key to begin the next trial, and the next voice prompt began 1,200–1,550 msec later.

To prevent participants from using the contrast change itself as a signal to start the search, half of the trials did not contain a change. In the remaining trials, participants understood that the change was unpredictive of the target location; for trials with n items in the display, the target was cued on $1/n$ trials, and a distractor was cued on $(n-1)/n$ trials. The contrast change occurred at a stimulus onset

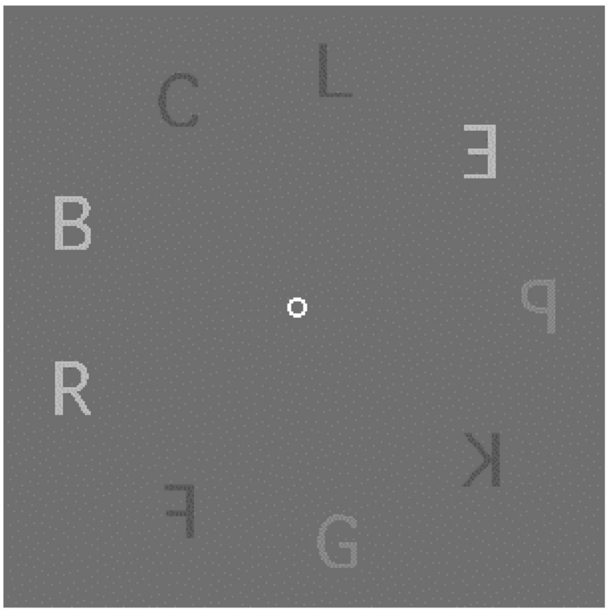

Figure 1. Sample search display for Experiment 1.

asynchrony (SOA) of 50 msec before, at the same time as, or 50 or 100 msec after the start of the voice prompt. After hearing the voice prompt, participants pressed the quote key as quickly as possible if the target was a normal letter, or the A key if the target letter was mirror reversed. Errors resulted in a 2-sec time-out.

Participants completed 18 practice trials, followed by five test blocks of 192 trials each (96 with a contrast change and 96 without). The 96 change trials were equally divided among the four onset timings. Each of these sets of 24 trials included 1 valid cue trial and $n-1$ invalid cue trials for each set size (n = 3, 5, 7, or 9 items). Pilot studies suggested that participants often tried to memorize the smallest display sizes to speed their search. Therefore, we included trials with three letters as a foil and did not include those trials in our analyses. All trials within a block were randomly ordered and then grouped so that set sizes were the same for 4 trials in a row, enabling the search array to remain exactly the same for 4 consecutive trials.

Results and Discussion

Error rates were low (M = 1.7%) and not significantly dependent on set size, cue validity, or cue-voice SOA. Error trials and trials with response times less than 300 msec or greater than 2,500 msec (less than 2% of all responses) were excluded from all analyses.

Attention capture in this task is indicated by increased search efficiency on trials for which the target happens to be the item with a contrast change relative to trials for which a distractor changes. Efficient search is indicated by minimal increases in response latency as a function of the number of items in the display (a shallow search slope). If the contrast changes capture attention, search should be more efficient when the target happens to change luminance than when a distractor happens to change luminance. This effect would be revealed by a significant interaction between cue type and set size.

Response time data (see Figure 2) were entered into a two-way repeated measures analysis of variance (ANOVA) with cue type (no cue, valid cue, or invalid cue) and set size (5, 7, or 9) as factors. Main effects of set size and cue validity were significant [both Fs(2,54) > 227, p < .001], as was their interaction [$F(4,108) = 25, p < .001$]. Because a separate ANOVA on only the invalid cue trials revealed no differences between SOAs, for clarity Figure 2 shows response times collapsed over all SOAs for invalid cue trials. Most important for present purposes, for each SOA considered separately, there were significant main effects of cue type (all Fs > 80, all ps < .001), set size (all Fs > 21, all ps < .001), and interactions between set size and cue type (all Fs > 4.1, all ps < .02) for each SOA.

Slopes for the no-cue (M = 54 msec/item) and invalid cue (M = 52 msec/item, collapsed across SOA) trials were fairly steep. For valid cue trials, however, slopes were near zero when the cue preceded the start of the voice prompt (−50 msec SOA, M = 7 msec/item, CI = ±15 msec/item), suggesting that the concurrent contrast and contrast polarity changes captured attention. Slopes at SOAs of 0 msec (M = 16), 50 msec (M = 25), and 100 msec (M = 24) were low, but not near 0, suggesting that attention was captured only partially or on a subset of trials at those SOAs. The differing levels of priority across SOAs are not unexpected. Because exogenously cued targets only receive processing priority for a short time after the cue (Nakayama & Mackeben, 1989), cues

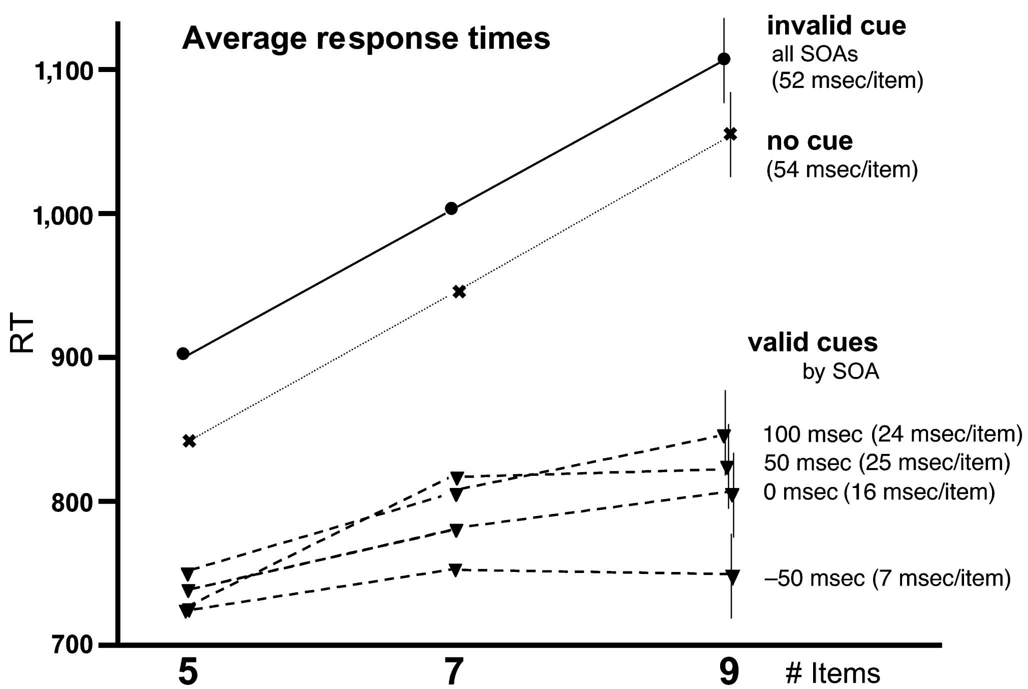

Figure 2. Average response times (RTs) in each condition for Experiment 1.

that appear too soon or too late should not capture attention as strongly as cues that appear at the optimal time.

Given that the appearance of the display in our task did not signal the start of the search task, participants should have had no reason to adopt an attention set for onsets. Thus, no obvious goal-directed process can account for attention capture in this experiment.[1] However, there is an alternative explanation—waiting for an auditory event (the voice prompt) might induce an attention set for transients *in general*, across both auditory and visual modalities. To rule out this alternative, in Experiment 2, we eliminated the voice prompt and modified the traditional irrelevant feature search task in such a way that participants did not *see* the displays change. New search displays appeared after every trial, but the change to a new display occurred while the participant made a large saccade to a location offscreen.

EXPERIMENT 2

In Experiment 1, we minimized the possible influence of task-induced biases by eliminating display changes on most trials. In Experiment 2, the search displays changed on every trial, but participants never *saw* the changes. At the end of each trial, participants saccaded to a point 33.4º below the center of the search array. Display changes occurred when the eyes were below the monitor and saccade velocity was near its peak, so that the transient signal from the display change was suppressed (see Shioiri & Cavanagh, 1989). If onset targets are still processed with priority over non-onset targets, attention capture by the onset is unlikely to be due to biases induced by display changes.

Method

Participants. Twenty undergraduates at Harvard University participated in the study in exchange for $8 or class credit. Data from 3 additional participants were excluded from the analyses because of software errors.

Stimuli. Stimuli were displayed on a 21-in. Sony Multiscan G500 monitor using custom software created with the VisionShell C libraries (http://www.visionshell.com). The display was 48º wide × 36.9º high from a viewing distance of 45 cm (fixed via chinrest). Four, six, or eight block letters (2.64º high × 2.64º wide with 1-pixel-thick lines) were arranged on an imaginary circle with a radius of 13.34º. The possible letters, A, C, E, F, H, L, P, S, and U, were constructed of subsets of the seven segments that make up a block figure 8. The letters were gray (22.8 cd/m^2) on a black background (0.02 cd/m^2). On trials with an onset, one letter was cued by a set of four gray (22.8 cd/m^2, 0.28º high × 0.28º wide) dots positioned in a diamond configuration 2.64º from the center of the letter. The dots appeared for 59 msec, disappeared for 59 msec, reappeared for 59 msec, and then disappeared.

Procedure. Participants wore an SMI EyeLink eye tracker controlled by an Acer Pentium Pro PC. The PC transmitted eye position data via an Ethernet cable to a Power Macintosh G4 computer, which controlled the stimuli. Eye position was calibrated before the experiment and again during the experiment, if necessary. Participants made about 20 practice saccades to a black X written on a yellow sticky note placed 15.2º below the bottom edge of the screen (33.4º below the center of the search array). Using realtime graphs of eye position and velocity, the experimenter estimated the fastest eye velocity that reliably occurred for each participant (M = 357º/sec, SD = 48º/sec). Individual thresholds were set so that search displays would not change unless the participant's eye speed exceeded this threshold speed, and the participant's eye position was at least 6º below the bottom edge of the screen.[2]

Figure 3 depicts a typical trial sequence. Participants completed 24 practice and 426 experimental trials, half of which had a cue. On cue trials, flickering dots appeared around one letter, beginning at a randomly chosen SOA of 0, 50, or 100 msec after eye position was within the search display and eye speed dropped below 93º/sec. Each display contained either a U or an H, and participants pressed keys as quickly and accurately as possible to indicate which letter was present. As in Experiment 1, the location of the cue was not correlated with the position of the target. After each response, participants saccaded to the X below the monitor, and then back to the new search display.[3] Participants were instructed to take breaks if desired by postponing the downward saccade.

Results

Error trials (M = 1.6%) and trials with response times less than 300 msec or greater than 2,500 msec (less than 2% of all responses) were excluded from the analyses. Error rates varied only as a function of SOA [$F(2,18)$ = 4, p = .037], due to a slightly lower error rate for the 0-msec SOA (M = .07%) than for the 50- and 100-msec SOAs (M = 2%).

Figure 4 shows average response times for each combination of cue type and SOA (0, 50, or 100 msec). To de-

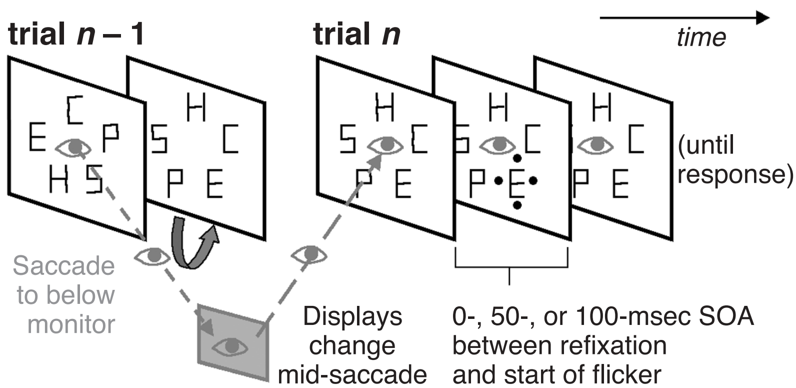

Figure 3. Sample trial sequence for Experiment 2.

termine whether search was faster on valid cue trials, average response times were entered into a 3 (cue type: no cue, invalid cue, valid cue) × 3 (set size: 4, 6, or 8) repeated measures ANOVA. There were main effects of cue type [$F(2,38) = 38, p < .001$] and set size [$F(2,38) = 98, p < .001$], and an interaction between the two [$F(4,76) = 14.7, p < .001$]. This interaction reflects shallower search slopes on valid cue trials ($M = 17$ msec/item) than on no cue ($M = 50$ msec/item) and invalid cue trials ($M = 55$ msec/item). Slopes for no-cue and invalid cue trials did not differ significantly [$t(19) = 1.1, p = .28$]. For clarity, Figure 4 shows the response times for invalid cue trials collapsed across SOAs [for invalid trials, slopes at the 0-msec ($M = 50$ msec/item) and 50-msec ($M = 48$ msec/item) SOAs did not differ, but slopes at the 100-msec SOA ($M = 67$ msec/item) were slightly steeper than the slopes at the 0- and 50-msec SOAs (both ts > 2.2, both ps < .04)].

Most important, 2 (cue type) × 3 (set size) ANOVAs conducted separately for each SOA showed that search slopes were shallower for valid cue trials than for invalid cue trials.[4] At each SOA, there were significant effects of cue type [all Fs(1,19) > 19, all ps < .001], set size [all Fs(2,38) > 10.1, all ps < .002], and their interaction [all Fs(2,38) > 4.1, all ps < .03]. These significant interactions indicate that at each SOA, search slopes were shallower for trials in which the onset dots appeared at the same location as the target than for trials in which the dots appeared at a distractor location. When the onset dots began to flicker at the same time as, or slightly after, eye speed dropped below 93°/sec (the 0- and 50-msec SOAs), valid cue search slopes were nearly flat ($M = 5$ msec/item, CI = ±16 msec/item; $M = 13$ msec/item, CI = ±18 msec/item). Valid slopes were lower, but not flat, at the 100-msec SOA ($M = 22$ msec/item). As in Experiment 1, capture effects may be stronger at some SOAs because of the relatively transient time course of exogenously cued attention (Nakayama & Mackeben, 1989).

CONCLUSIONS

As our goals change, our need for different types of visual information does as well. Although we can adjust our attentional priorities according to these goals (e.g., Folk et al., 1992), some stimuli may demand attention even when we know that they are irrelevant. Our attempts to exclude top-down biases suggest that at least one property of the visual world—a dynamic event—attracts attention even in the absence of top-down prioritization. Prior claims for stimulus-driven prioritization of dynamic events required observers to await the dynamic appearance of a search array, a task that may itself induce a bias to prioritize dynamic events. In eliminating this confound, we showed that irrelevant dynamic cues capture attention even in the absence of any obvious reason to prioritize them. In Experiment 1, dynamic events captured attention when the start of the trial was signaled by an auditory cue rather than by a dynamic visual event. In Experiment 2, dynamic events captured attention even when participants never saw the search displays change.

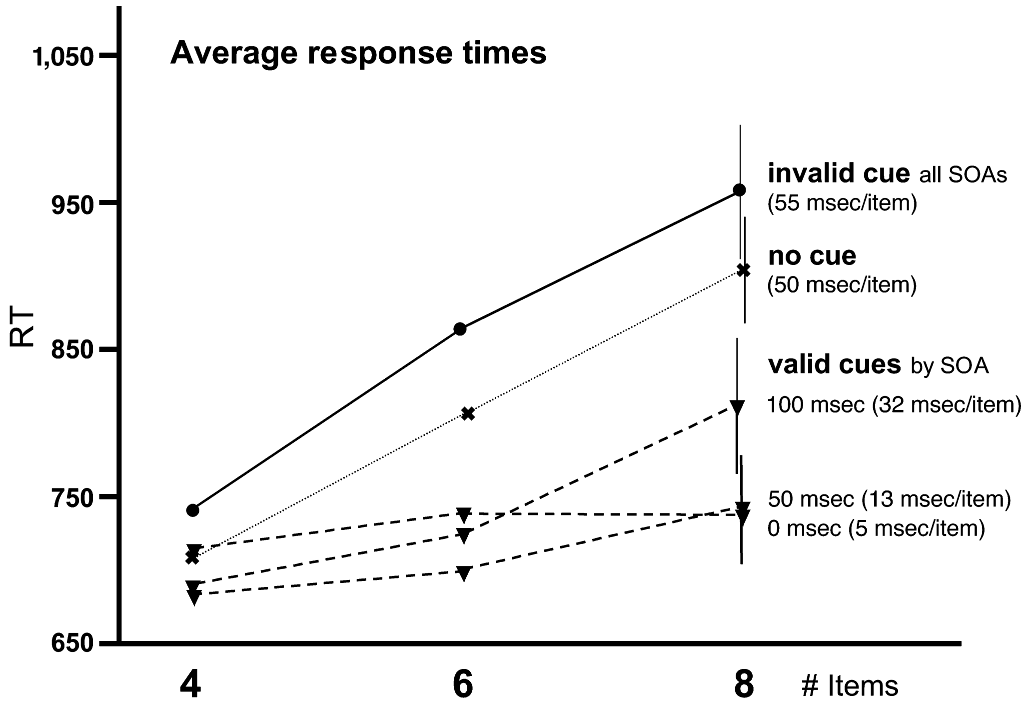

Figure 4. Average response times (RTs) in each condition for Experiment 2.

Proving that a shift of attention was entirely stimulus-driven may well be impossible because it requires exhaustive elimination of all possible top-down influences. Despite our best attempts, the observer might have lingering goal-directed reasons to attend preferentially to dynamic events. For example, although our dynamic cues appeared on only half of the trials, and usually after the search began, their appearance could still be associated to some degree with the start of the search.[5]

The set of stimuli that appear to capture attention has been described as abrupt onsets (Jonides & Yantis, 1988), new objects (Yantis & Hillstrom, 1994), and a hodgepodge of dynamic events (Abrams & Christ, 2003; Franconeri & Simons, 2003; Thomas & Luck, 2000). Although in these experiments, we tested only flickering dots and concurrent luminance contrast and contrast polarity changes, our results suggest that some dynamic visual events do capture attention in a stimulus-driven fashion.

REFERENCES

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, **14**, 427-432.

Enns, J. T., Austen, E. L., Di Lollo, V., Rauschenberger, R., & Yantis, S. (2001). New objects dominate luminance transients in setting attentional priority. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1287-1302.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 1030-1044.

Franconeri, S. L., & Simons, D. J. (2003). Motion and looming stimuli capture attention. *Perception & Psychophysics*, **65**, 999-1010.

Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for displaywide visual features. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 699-706.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, **43**, 346-354.

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

Shioiri, S., & Cavanagh, P. (1989). Saccadic suppression of low-level motion. *Vision Research*, **29**, 915-928.

Theeuwes, J. (1990). Perceptual selectivity is task dependent: Evidence from selective search. *Acta Psychologica*, **74**, 81-99.

Thomas, S. J., & Luck, S. J. (2000). *Multiple pathways to the automatic capture of attention*. Unpublished manuscript.

Turatto, M., & Galfano, G. (2001). Attentional capture by color without any relevant attentional set. *Perception & Psychophysics*, **63**, 286-297.

Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: Evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 95-107.

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

NOTES

1. An additional 9 participants were run on a color singleton version of Experiment 1. On each trial, one randomly chosen letter was highly saturated red (17.4 cd/m^2), while the others were gray (27 cd/m^2) on a dark gray background (17.7 cd/m^2). Displays remained unchanged for four consecutive trials. Unlike contrast and contrast polarity changes, letters validly cued by color singleton status had slightly higher search slopes (M = 64 msec/item) than did letters cued by nonsingletons [M = 41 msec/item; $t(8) = 2.5$, $p < .04$].

2. At the end of Experiment 2, we conducted a test to ensure that participants could not detect the display changes. The participants were warned that during the downward saccade, their first names would be flashed on the bottom of the screen in large gray letters (3.3º tall, 7.25 cd/m^2). All participants required at least a 35-msec presentation time to discriminate whether the name had been presented. Names flashed for only one frame (11.7 msec) were easily visible to the experimenter.

3. Although there was no fixation point in the display, refixations landed close to the center of the search display with a mean error of 6º (search items and onset cues were located 13º from the center of the search display). The direction of this error was not systematically biased in the direction of the onset cue. In a conservative test using only the 0-msec SOA trials (which demonstrate the strongest capture by the dynamic cue), the angle between the center of the display and the refixation position was not correlated with the angle between the center of the display and the onset location (r = .018). These eye movement analyses are based on data from 7 of the 20 participants, because 13 of the eye movement records were lost due to disk error.

4. We also tested whether static singletons captured attention in the search task used in Experiment 2. Ten additional participants were run in a similar experiment that used color singletons instead of dynamic singletons. On all 426 trials, one randomly chosen letter was red (19 cd/m^2) among green (12.7 cd/m^2) or vice versa. Search slopes for color singleton letters (59 msec/item) did not reflect priority over other letters [M = 62 msec/item, $t(9) < 1$, $p > .6$].

5. Thanks to Brad Gibson for pointing out this potential source of goal-directed bias.

(Manuscript received June 3, 2002;
revision accepted for publication September 16, 2003.)