Journal of Experimental Psychology: Human Perception and Performance (in press).

Attentional Capture by Abrupt Onsets: New Perceptual Objects or Visual Masking?

Steven Yantis
Johns Hopkins University

John Jonides
University of Michigan

Preparation of this article was supported by National Institute of Mental Health grant R01-MH43924 to S. Yantis and by a grant from the Office of Naval Research to John Jonides. D. Burke provided valuable technical assistance. We thank Bill Bacon, Vince Di Lollo, Howard Egeth, Jim Enns, Brad Gibson, Anne Hillstrom, Doug Johnson, Jim Johnston, Peter Kramer, Joe Lappin, Gordon Logan, and Cathleen Moore for helpful comments and discussion. Correspondence may be sent to Steven Yantis, Department of Psychology, Johns Hopkins University, Baltimore, MD 21218-2685 (email: yantis@jhu.edu).

In this article, we review the claims that we and our colleagues have made about attentional capture by the onset of new perceptual objects, and we examine Gibson's (1996) challenge to these claims. We present some new data that replicate Gibson's findings but that support a different interpretation of his data. We conclude that our claims concerning attentional capture by the appearance of a new perceptual object remain the best account of the relevant data.

Attentional capture by abrupt visual onset

In a series of studies, we and our colleagues have shown that an object appearing abruptly in a previously blank location captures attention in visual search (Yantis, 1993a, in press). Yantis and Jonides (1984) demonstrated this first in an experiment in which subjects searched for a prespecified letter on each trial of a visual search task. An array of figure-eight placeholders was presented for 1 s immediately preceding the appearance of the letter array. One of the letters in the array appeared in a previously blank location while the remaining letters appeared in locations previously occupied by figure eight placeholders. Thus, each of these remaining letters was present during the placeholder display but camouflaged by irrelevant line segments; such stimuli were first used by Todd and Van Gelder (1979), who called them *no-onset stimuli*. In our experiments, the target was present on half the trials, and when it was present, it had an abrupt onset on $1/n$ of the trials (where n is display size) and was a no-onset letter on the remaining trials. Therefore, there was no incentive built into the design of the experiment encouraging subjects to attend to the onset stimulus.

In this and many similar experiments (e.g., Jonides & Yantis, 1988; Yantis & Johnson, 1990), we found that when the target did happen to be the abrupt onset stimulus, mean response time (RT) did not increase significantly with display size. In contrast, when the target was one of the no-onset stimuli, RT increased systematically and approximately linearly with display size. The slope of the latter function is typically about half the slope of the target-absent function. This pattern of response times has led us to conclude that onset stimuli capture attention in a purely stimulus-driven fashion (i.e., without deliberate top-down or strategic control).

It is possible, of course, that abrupt visual onset is not unique in capturing visual attention; perhaps any visually distinct stimulus would capture attention as well. To test this possibility, Jonides and Yantis (1988) conducted experiments in which a unique and visually salient stimulus was embedded in an array of other, homogenous, stimuli. Specifically, subjects searched for a target letter in an array consisting of one red letter and n -1 green letters, or in an array consisting of one bright letter and n-1 dim letters. In both cases, the target was unique (i.e., red or bright) on $1/n$ of the trials, so, as in the onset experiments, there was no incentive to attend to the unique stimulus. In these experiments, we found that salience or uniqueness per se was not sufficient to yield attentional capture. This finding has been corroborated and extended in several subsequent studies in which unique color, brightness, and motion failed to capture attention in visual search (e.g., Folk & Annett, 1994; Hillstrom & Yantis, 1994; Yantis & Egeth, 1994). Based on these findings, we have concluded that abrupt visual onset may have a privileged status in its ability to capture attention in a stimulus-driven fashion compared to certain other salient visual properties.

We have argued that visual onsets capture attention by virtue of the fact that they coincide with the appearance of a new perceptual object, and not because of the luminance increment that typically accompanies abrupt onset. According to this account, when a new object appears in the visual field, it is necessary to create a new perceptual representation for it, analogous to what Kahneman and Treisman (1984; Kahneman, Treisman, & Gibbs, 1992) have called an *object file*. When a representation is created for a new object, attention is typically directed to that object. This makes adaptive sense in that new objects are likely to have behavioral significance and to require rapid and accurate identification and/or response. This hypothesis has been supported by data from several studies in which attentional capture occurs when a new object appears, even when that object is not accompanied by a luminance increment (Yantis & Hillstrom, 1994), and attentional capture does not occur when a new object does not appear, even in the presence of a luminance increment (Yantis & Hillstrom, 1994) or motion (Hillstrom & Yantis, 1994).

There are exceptions: for example, Yantis and Jonides (1990) found that when attention is directed in advance to another spatial location, then the appearance of a new object does not capture attention (see also Folk, Remington, & Johnston, 1992, 1993; Yantis, 1993). Nevertheless, attention typically is captured by the creation of a new perceptual object representation. Note also that a new perceptual object representation might be created as a result of any number of visual events. Ongoing perceptual organization of a dynamic scene may require new object representations to be created as objects are segmented from their background (Hillstrom & Yantis, 1994). Abrupt onset is likely to require a new perceptual object representation, but this need not always be the case (e.g., when an existing object reappears after having been temporarily occluded). Thus abrupt onsets are by no means unique in their ability to capture attention.

As described above, our studies of this phenomenon have included a procedure that permits us to present new objects and "old" objects at the same time by camouflaging the old objects as figure eights and removing the camouflaging line segments at the moment the letters are to appear. The rationale is that when the figure eights appear 1 s before the beginning of the letter display, new object representations are created for each of them (leading to a temporary tagging of these elements as being high in attentional priority; see Yantis & Johnson, 1990; Yantis & Jones, 1991). By the time 1 s has elapsed and the onset letter appears, the priority tags associated with the figure eights have decayed (Yantis & Jones, 1991). Thus when the onset letter appears, it is the highest priority letter in the display and will therefore enjoy an attentional advantage over the remaining letters in the display. Behaviorally, this amounts to attentional capture.

This procedure leaves open the possibility, of course, that the no-onset stimuli--the "old" elements--manifest a sensory, rather than an attentional, deficit relative to the onset letters. Put differently, it is possible in principle that there is a (sensory) deficit in processing the no-onset stimuli rather than an (attentional) advantage for the onset stimulus. In fact, we were concerned in our original experiments that the figure eights might render the no-onset letters difficult to encode (Yantis & Jonides, 1984). We therefore ran a control experiment in which we measured response times to identify onset and no-onset letters that were attended in advance. The idea was to remove any possible attentional difference between onset and no-onset letters by inducing subjects to focus attention on them; if there remained a residual disadvantage for the no-onset letters, it would have to be attributed to non-attentional, perhaps sensory, factors. We found no difference (indeed, in Yantis & Jonides, 1984, the no-onset stimuli were identified slightly but nonsignificantly faster than the onset letters when they were attended in advance). The same control experiment has been conducted in subsequent studies, and no difference between onset and no-onset identification times has been found (e.g., Yantis & Hillstrom, 1994).

Masking hypothesis revisited

Gibson (1996) challenges our claim that abrupt onsets capture attention by virtue of their status as new perceptual objects. Instead, he argues, the advantage accruing to onset letters is due exclusively to the no-onset letters having been masked by the preceding figure-eight placeholders. This assertion is based on two empirical observations. First, Gibson found that visual search for targets in displays with all no-onset stimuli was slower than visual search in displays with all onset stimuli. Second, he found that the speed of visual search was systematically influenced by the intensity of the figure eight placeholders (bright placeholders produced more slowing than dim placeholders).

It is worth emphasizing where we and Gibson agree and disagree. We all agree that an onset stimulus appearing among no-onset stimuli enjoys an attentional advantage in visual search. We disagree with Gibson about the mechanism that is responsible for this advantage. We claim that the appearance of a new perceptual object triggers the creation of a new object representation, and this in turn attracts attention. Gibson asserts instead that when the no-onset and onset letters are first physically present, the camouflaging line segments persist beyond their physical offsets; visual integration of each figure eight placeholder with the subsequent no-onset letter delays the availability of that letter. The observer therefore directs attention to the one letter that is available--the onset letter. From this point forward, the interpretation is the same as in Yantis and Jonides (1984): if the onset letter is the target, it is identified as such and a response is made immediately. The number of nontargets in the display is irrelevant, and a flat display-size function results. If the onset letter is not the target, then search proceeds to the remaining (no-onset) letters in the display, yielding display size functions with positive slopes.

The key idea in Gibson's critique is that the figure eight placeholders (in particular, the segments of the placeholders that are removed) visibly persist into the interval occupied by a letter, and are therefore integrated with the letter (e.g., Eriksen, 1966), which delays perception of that letter. In the remainder of this article, we lay out several arguments and some data from new experiments that call into question a sensory masking account.

It is known that the duration of visible persistence depends on the duration and intensity of the stimulus (Coltheart, 1980). First, the duration of persistence decreases as the duration of the stimulus increases (the *inverse-duration effect*). For example, Di Lollo (1980) found that the degree of successful visual integration (and hence the duration of visible persistence) was greatest for very short duration stimuli and fell to near zero for stimulus durations of 200 ms or more. The duration of the figure-eight placeholders in our experiments and those of Gibson was always at least 1000 ms, well beyond the maximum stimulus duration for which persistence is observed.

In his Experiment 4, Di Lollo (1980) provides direct evidence concerning the degree of masking produced by a temporally leading stimulus with no interstimulus interval (ISI). The test stimulus was an alphabetic character that had to be identified by the subject. The mask was composed of randomly arranged letter fragments appearing in the same spatial region as the target. When the subject pressed a button, the mask appeared for 20, 40, 80, 160, 320, or 640 ms, and was followed immediately (ISI=0) by the target, which always had a duration of 20 ms. The results from four subjects revealed monotonically increasing performance as the mask duration increased, from 20-30% correct with a mask duration of 20 ms, to over 90% correct for mask durations of 320 ms or more. This experiment provides the most direct evidence that forward masking by a temporally leading stimulus is virtually nonexistent for long mask durations such as those used in our experiments. Di Lollo (1980) concluded that "as the duration of the mask was increased, its visual persistence diminished while leaving the persistence of the test stimulus unaffected . . . [at long durations,] the mask's persistence was exhausted, and hence it could not merge with (and degrade) the representation of the trailing test stimulus" (pp. 86-87). We conclude, therefore, that the 1000-ms placeholders in our experiments yielded little or no visible persistence.

Second, the duration of visible persistence decreases as the intensity of the stimulus increases (the *inverse-intensity effect* ; see Di Lollo & Bischof, 1995). According to Gibson, there should be more masking (i.e., longer visible persistence duration) for bright placeholders than for dim ones (and his experiments showed a larger effect for bright than for dimmer placeholders). The durations of the figure eight placeholders in all likelihood eliminated any persistence, but even if persistence was possible at those durations, the inverse-intensity effect dictates that persistence would be longer in duration (and therefore more effective as a mask) for dimmer placeholders. The inverse-intensity effect also calls into question a masking account of Gibson's brightness manipulation.[1]

The most direct way to determine whether some form of sensory masking is an important determinant of response times in our task would be to measure RT to identify a single letter that is and is not preceded by a figure eight. As Gibson notes, we have carried out experiments (described earlier) in which we made just such measurements. Based on them, we ruled out a masking account.

Gibson questions these experiments by suggesting that the attentional manipulation we used may have altered the manifestation of masking that was actually present. He notes that focused attention can influence temporal order judgements (attended events are judged as occurring earlier in time than unattended events, as shown by Stelmach & Herdman, 1991, and others) which could, he implies, differentially affect response times. However, this account does not explain why the temporal advantage for attended items would accrue only to no-onset stimuli and not to onset stimuli, in exactly the quantity required to eliminate any observable difference in RT. Furthermore, Gibson states that the putative masking effect occurs prior to the operation of visual selective attention (p. 12). If the masking effect slows encoding before attention can operate, the putative disadvantage produced by masking should persist even when the stimuli are attended. Our control experiments show that it does not: Focused attention eliminates the difference in identification time for onset and no-onset displays, which shows that the difference cannot have a sensory or preattentive basis.

The effect of the distribution of attention on the magnitude of the putative masking effect itself suggests that the effect is not sensory in origin as suggested by Gibson. A standard metacontrast mask, which is known to be sensory in origin, produces the same effect on performance whether it is attended or not. In contrast, the effect of the figure eight placeholders in our experiments depends upon whether attention is focused in advance (no effect; Yantis & Jonides, 1984) or distributed (the placeholder produces slower responses; Gibson, 1996). Enns and Di Lollo (1995) have recently reported a new phenomenon termed "attentional masking" in which a masking stimulus that failed to produce any performance decrement when the mask-target pair was attended nevertheless produced a significant performance decrement when

---

[1]Gibson cites Turvey's classic 1973 paper to support his contention that dimmer masks produce smaller masking effects. However, there are several reasons why this principle is not applicable to the present case. First, Turvey used pattern masks containing many randomly oriented line segments; the figure eight placeholders are not pattern masks of this type. Furthermore, the vast majority of the target and mask durations tested by Turvey were substantially less than 100 ms, and observed masking effects were generally quite small for the longest-duration stimuli used by Turvey; in the present case, the placeholder duration was always 1000 ms; it is unknown what magnitude of masking would have been observed if Turvey had used such durations. Finally, the measure of masking used by Turvey was detection accuracy, not response time; the mapping between these two functions is not straightforward.

attention was distributed over the display. This pattern parallels the findings reported by us and by Gibson. Enns and Di Lollo concluded that their attentional mask operates at a higher level of visual representation, one in which object representations are formed and interact. As we argue later in this article, the slowing produced by placeholders also operates at a post-sensory level of representation.

A critical test of Gibson's hypothesis would be an experiment similar to the ones carried out by Yantis and Jonides (1984) in which visual quality is manipulated by some means other than masking by figure eights. Again, such a critical experiment has been carried out. Jonides and Yantis (1988) had subjects search for a target letter in an array of *n* elements in which one letter was bright and *n*-1 letters were dim; Yantis and Hillstrom (1994) carried out a similar experiment. The position of the target was uncorrelated with the position of the bright letter. Gibson's masking account makes the strong prediction that the bright letter should capture attention because it is available earlier to attention than the dim letters are. This did not occur; instead, both Jonides and Yantis (1988) and Yantis and Hillstrom (1994) failed to observe attentional capture by the bright letter. These results cannot be explained with Gibson's masking account. Gibson (1996) states that the effect "may have been obscured by the specific combination of experimental variables used in those experiments" (p. 12), but this is difficult to assess without a more specific claim.

The duration of the figure eight placeholders in our experiments is too great to have yielded effective visible persistence, making masking by integration highly unlikely. The effect of Gibson's brightness manipulation would be opposite to that predicted by the inverse-intensity effect. The results of our single-element control experiments (e.g., Yantis & Hillstrom, 1994; Yantis & Jonides, 1984) and the brightness experiments (Jonides & Yantis, 1988; Yantis & Hillstrom, 1994) provide strong evidence that the no-onset letters were not masked by the figure-eight placeholders. Together, these considerations undermine Gibson's masking account.

However, this still leaves us with an empirical effect that must be explained: In Gibson's (1996) experiments, RT to all no-onset displays was slower than RT to all onset displays. To identify the source of this effect, we first conducted a replication of Gibson's experiment to ensure that the effect could be obtained with our equipment and procedures.

Experiment 1: Replication of Gibson (1996)

Method

Fifteen Johns Hopkins University students each participated in a 50-minute session. The visual search task was to determine whether an E or an H was presented on each trial of the experiment and to press a corresponding button as quickly as possible.

Stimuli were displayed on a 21" Taxan color monitor driven by an Artist Graphics TI34020-based graphics accelerator board. A chinrest was used to stabilize head position with the observer's eyes 55 cm from the screen. Letters were .86° high and .43° wide, and appeared on the circumference of an imaginary circle 4.0° in diameter. The minimum center-to-center distance between letters was 1.25°. The luminance of a uniform patch the same intensity as the letter contours was 67.4 cd/m^2 and the background luminance was 0.12 cd/m^2. The letters were thus similar in contrast to Gibson's bright condition.

Two main conditions were randomly intermixed within blocks. In the onset condition, a fixation point appeared in the center of a blank screen for 1 s, followed by 3 or 7 letters. In the no-onset condition, an array of figure eight placeholders surrounding a central fixation point appeared for 1 s, at which time a subset of the line segments in the figure eights was removed to reveal 3 or 7 letters. Subjects were instructed to determine as quickly as possible whether an E or an H was present (and one of these two target letters was presented on every trial), and to press the right button if E was present and the left button if H was present.

Each subject completed 6 blocks of 64 trials. Display size was 3 or 7 and display type was onset or no-onset. These two factors were completely crossed within subjects.

Results and Discussion

Mean RT for onset and no-onset displays of size 3 and 7 are shown in Figure 1. There was a significant display size effect, $F(1, 14) = 48.9$, $p < .001$, and a significant effect of display type, $F(1, 14) = 20.4$, $p < .001$, but no interaction, $F(1,14) = 3.1$, $p > .05$. These results are very similar to those reported by Gibson (1996), although the main effect of display type was substantially smaller in our Experiment 1 (27 ms) than in Gibson's Experiment 1 (82 ms). Gibson asserts that these results implicate a form of masking as the mechanism for slowing the no-onset identification times.

An alternative account

We propose a different interpretation of this result based on the principle that the appearance of a new perceptual object triggers the creation of a new representation for that object, and this representation captures attention, thereby speeding identification relative to objects that do not enjoy an attentional advantage.

In the all onset condition, all the elements in the display are new, and perceptual object representations must be created for each of them. All of the object representations so created are tagged as high priority (Yantis & Johnson, 1990). We assume that this is a fast and efficient process, in part because rapid identification of new perceptual objects is highly adaptive, and in part because this is an extremely common perceptual event. Whenever we walk into a room, whenever we turn a corner, we are confronted with new perceptual objects that must be identified as quickly as possible. The creation of an episodic object representation is assumed to occur in parallel across the visual field, although this may be a limited-capacity process, and identification may take place in parallel or serially.

Quite a different state of affairs holds for the no-onset condition. Here, according to our account, episodic perceptual object representations are created at the onset of the figure eight placeholders, 1 s before the trial actually begins. At the end of the 1-s placeholder interval, each of the figure eights changes form slightly; identities change from "8" to "E," "8" to "P," and so forth. We suppose that the process of updating a continuing perceptual object representation is not a rapid process. After all, it is rare for an object to change its identity abruptly (gradual changes in attributes such as shading or perspective--due to observer motion for example--are common and typically unimportant, but identity changes are not). Partly for this reason, there does not seem to be any urgency to updating a changing perceptual object; once an object has been identified and a valence has been assigned to it (e.g., dangerous), that object is stable and attentional capacity can be reserved for the identification of new, as-yet-unidentified perceptual objects. Therefore, the updating of a representation may be slow relative to the creation of a new representation. We assume that the updating process occurs in parallel for all the existing perceptual object representations. This account explains the observed response slowing for all no-onset displays relative to all-onset displays.

Gibson (1996, Experiment 2) found that RT to no-onset displays with dim placeholders was faster than with bright placeholders but slower than with no placeholders. According to our account, what matters is whether the letters are experienced as new perceptual objects rather than as continuations of "old" perceptual objects. The luminance of the bright figure eights in Gibson's experiments was almost two log units greater than the luminance of the dim figure eights. Although Gibson stated that both types of figure eights were "easily perceivable," this is not the relevant criterion. The critical issue is whether the degree of change from the placeholder display to the letter display is experienced as the appearance of a new perceptual object or not. A change from a dim, near-threshold stimulus to a much brighter, well above-threshold stimulus could easily be experienced as the appearance of a new perceptual object replacing the previous dim object; Gibson's data are consistent with this conclusion. This is an empirical question that will not be resolved here. However, we assert that this explanation provides a better account than does masking, which if anything predicts greater slowing for dim than bright placeholders because of the inverse intensity effect (as discussed earlier), contrary to what Gibson observed.

This account of the onset advantage for all onset and all no-onset displays differs from Gibson's account in the emphasis it places on perceptual object representations. This property provides a way to test the two accounts directly against one another. Our account states that the advantage for onset over no-onset stimuli is object-based, while Gibson's account states that it is location-based. To test these predictions, we carried out another experiment.

Experiment 2: Masking or New Objects?

The logic of this experiment is simple. We used apparent motion to create a visual search display containing pre-existing episodic object representations that could not be masked by preceding figure eight placeholders. According to the new-object account, these elements should exhibit a slowing relative to onset elements, because the relevant factor is whether an object is old or new, not whether the preceding contours appeared in the same retinal or spatial location. According to the masking account, in contrast, there should be no difference between the onset and no-onset arrays because there is no sensory masking in either case.

Method

A new group of 15 subjects participated in Experiment 2. The displays in Experiment 2 were identical to those in Experiment 1, with one exception (see Figure 2). The fixation point (in the onset condition) or the array of figure-eight placeholders together with the fixation point (in the no-onset condition) initially appeared in the periphery of the display and moved toward the center of the screen in a sequence of jumps that were timed to produce a vivid perception of apparent motion. Each step of the motion was greater than the height of one letter, so no successive stimuli overlapped spatially. The center-to-center distance between successive stimulus locations was 1.1°; the edge-to-edge distance depended on whether the motion was vertical (producing an edge-to-edge distance of .24°), horizontal (producing an edge-to-edge distance of .67°), or oblique (producing intermediate edge-to-edge distances). The duration of each frame was 100 ms, and there was a blank 50-ms interstimulus interval between pairs of successive frames, yielding a stimulus-onset asynchrony of 150 ms. On each trial, 15 stimulus frames were presented. All but the final frame consisted of figure eights (for the no-onset condition) or the fixation point alone (for the onset condition); the final frame consisted of the array of letters. Thus, the motion in each trial lasted 2100 ms before the letter display appeared. The letter display remained until a response was made or 2000 ms elapsed. In all other respects, this experiment was identical to Experiment 1.

The perceptual object representations in Experiments 1 and 2 are identical: they are created at the onset of the figure eights in the no-onset condition, and they are created at the onset of the letters in the onset condition. The possibility of sensory masking has been eliminated in Experiment 2 by virtue of the fact that the letters in the no-onset condition appeared in a location that was previously blank. If there remains an RT advantage for the onset condition as compared to the no-onset condition in Experiment 2, it must be attributed to differences in the time required to update the object representations.[2]

Results and Discussion

Mean RT is plotted in Figure 3. The data are remarkably similar to those obtained in Experiment 1, with main effects of display size, $F(1, 14) = 80.5$, $p < .001$, display type, $F(1, 14) = 21.6$, $p < .001$, and no interaction, $F(1, 14) < 1$. Thus, we find that the advantage in processing onsets versus no-onsets obtains both when the camouflaging figure eights and the letters themselves appear in the same spatial locations (Experiment 1) and when they appear in different locations (Experiment 2). If masking were the proper account for this effect, one would expect the magnitude of the onset vs. no-onset difference to decrease substantially. Instead, this difference decreased only slightly: the mean difference between the all onset and all no-onset conditions was 27 ms for Experiment 1 and 24 ms for Experiment 2. We must conclude, as we have in the past, that masking plays little or no role in producing the difference between onsets and no-onsets.

The pattern of results in Experiment 2 depends critically on the extent to which the stimuli appearing in successive frames of the display are perceived as manifestations of continuously present objects that are moving through space. One might have conducted a very similar experiment with different spatiotemporal parameters that could have given quite different results: for example, if the timing of the successive frames did not produce optimal apparent motion, then we would not expect to observe a difference between the onset and no-onset conditions, because the “no-onset condition” in this case would really be a degenerate onset condition. What is important is perceived continuity of the “old” objects.

Additional evidence

Bacon and Remington (1995) have recently reported results from an independently conceived experiment that corroborates the conclusions we drew on the basis of Experiment 2. The logic of their experiment was similar to ours in that the figure-eight placeholders in their experiment exhibited apparent motion that eliminated their ability to perceptually mask the letters. In the Bacon and Remington experiment, the display contained several no-onset letters

---

[2]One could argue that a form of paracontrast masking is operating here (one could also argue that retinotopic forward masking across eye movements plays a role; this possibility is discussed in the next section). Paracontrast refers to a mask that appears spatially adjacent to and temporally before the target (Breitmeyer, 1984). In paracontrast masking experiments, the target is typically a disk and the mask is an annulus, or the target is a line and the mask is a pair of flanking lines. This possibility is unlikely given the timecourse of paracontrast masking: masking peaks for SOAs of 30-70 ms and dissipates to near zero for SOAs of 100 ms or more for spatial separations of 0.3° or more (e.g., Growney, Weisstein, & Cox, 1977). In the present experiment the SOA was 150 ms. Furthermore, paracontrast masking typically interferes with stimulus detection accuracy or with judgements of stimulus brightness, not with stimulus identification latency (see Footnote 1). Finally, the paracontrast account holds that two very different mechanisms produced RT effects that were remarkable similar in Experiments 1 and 2. For these reasons, paracontrast is unlikely to be a major factor here.

and a single onset letter (similar to the design of Yantis and Jonides, 1984). The placeholders revolved in a clockwise direction around a central fixation point, maintaining their eccentricity throughout movement. There were 9 frames of movement; the figure eights turned to letters and the onset letter appeared in frame 7. All of the letters in this display appeared in a previously blank location; the onset letter differed from the other letters in the display in that it did not correspond with any object that was present during the motion of the figure-eight placeholders.

According to the masking account advanced by Gibson (1996), visual search time in the Bacon and Remington (1995) experiment should not depend on whether the target was an onset letter or a moving no-onset letter, because no sensory masking is possible in any location. The new-object account predicts instead that the onset letter should capture attention because it is a new perceptual object. In fact, Bacon and Remington found that onset targets produced a significant attenuation of the display-size effect relative to no-onset targets, even though sensory masking could not have occurred in their experiment.

The Bacon and Remington (1995) experiment also helps to rule out a possible masking-based interpretation of the results of the present Experiment 2. If subjects accurately maintained fixation on the fixation point as it moved during the 1-s placeholder interval, then in principle the figure-eight placeholders might stimulate the same retinal locus throughout this interval, and one might argue that this could lead to masking. There are two reasons to doubt this argument, however. First, it is unlikely that subjects could follow the fixation point as accurately as would be required for this explanation to hold. Second, the Bacon and Remington (1995) experiment is not subject to this possibility because the figure eights always stimulated different retinal locations as they moved.

Experiment 3: Even “masked” letters can capture attention

The masking hypothesis holds that stimuli preceded by figure eight placeholders are perceptually delayed relative to stimuli that appear in previously blank display locations, hence the latter stimuli exhibit temporal precedence and thereby capture attention. We tested this hypothesis directly by modifying our standard task such that all the letters were preceded by figure eight placeholders. However, all but one of the figure eights had durations of 1000 ms, and the remaining figure eight had a duration of only 200 ms (i.e., it was presented 800 ms after the onset of the other figure eights). In this case, the degree of masking should be equal for all the letters, or, if anything, persistence should be greater for the delayed placeholder (due to the inverse-duration effect; Di Lollo, 1980), producing more masking. Gibson's masking hypothesis thus predicts no difference (or perhaps a slowing) when the target happens to appear in the delayed placeholder location relative to when it appears in one of the early placeholder positions. In contrast, our new-object account predicts that the delayed placeholder will capture attention, because it is a new perceptual object, and this will yield an attentional advantage for targets that appear there.

Method

The displays used in Experiment 3 are depicted in Figure 4. Two main conditions were run in alternating blocks of trials. In the Delayed Placeholder condition, each trial began with a display of 6 figure eight placeholders. After 800 ms, an additional placeholder was added to the display. The 7 placeholders remained for 200 ms, after which all the placeholders changed to letters (on display size 7 trials), or 3 placeholders changed to letters and 4 disappeared (on display size 3 trials). According to the masking account, there should be no difference in the perceived time of appearance of the letter in the delayed placeholder location relative to the other letter stimuli; if anything, the inverse duration effect (e.g., Di Lollo, 1980) predicts longer persistence duration for the delayed figure eight, with its 200 ms duration, than for the figure eights in the other locations, with their 1000 ms durations. The masking account therefore predicts that RTs to detect targets in the delayed placeholder location should be equal to or slower than RTs to detect targets in the other locations.

In contrast, our new object account holds that the delayed placeholder is a new object and therefore captures attention when it appears; the increased priority for this object is maintained for the 200 ms that elapses until the figure eight changes into a letter (Yantis & Jones, 1991). Our account therefore predicts that when the target appears in the delayed placeholder location, RT should be significantly faster, and the display size effect significantly smaller, than when the target appears in one of the other placeholder locations.

We might also predict the display size effect in the delayed placeholder location to be greater than it would be if the letter had appeared in a completely blank location (as in the standard onset task) because of the 200 ms interval during which the priority tag has time to decay (Yantis & Jones, 1991). To assess this possibility, we included a second condition, termed the Onset condition, in which there was no delayed placeholder; instead, a letter appeared in a previously blank location, just as in our standard onset task.

Fifteen naive Johns Hopkins University undergraduates participated in a one-hour session. The other details of the experiment were as before.

Results and Discussion

The results are shown in Figure 5. The bottom panel shows the mean RT for the onset condition. We observed the usual interaction of target type with display size: responses when the target was the onset letter were almost completely unaffected by display size (slope=2.0 ms/element) but were significantly affected for no-onset targets (slope=13.4 ms/element). This interaction was also significant for the Delayed Placeholder condition, shown in the top panel of Figure 5: the slope for delayed onset targets (4.4 ms/element) was significantly lower than for no-onset targets (18.1 ms/element). A repeated-measures ANOVA was conducted with condition, target type, and display size as factors. There were main effects of condition, $F(1,14) = 21.8$, $p < .001$, target type, $F(1,14) = 65.0$, $p < .001$, and display size, $F(1,14) = 20.9$, $p < .001$. The only interaction to reach significance was the display size by target type interaction, $F(1,14) = 14.3$, $p < .01$. None of the interactions involving condition were significant, indicating that the magnitude of the interaction between display size and target type was not different for the Delayed Placeholder and Onset conditions, respectively.

The results of Experiment 3 are at odds with the predictions of Gibson's masking account and are consistent with our new object account. In particular, a stimulus that was preceded by a figure eight placeholder with a relatively short duration received high priority in visual search, yielding virtually no display size effect. This directly contradicts the prediction that this stimulus should have been masked as much or more than the other letters in the display due to the relatively short duration of the placeholder that preceded it. If masking and temporal precedence were the sole mechanisms underlying attentional capture, we would not have observed these results.

Concluding Remarks

There is an advantage in searching a display of all onset characters compared to a display of all no-onset characters, as Gibson (1996) documented and we confirmed in Experiment 1. We disagree about why this occurs. Gibson argues that no-onset stimuli are masked by their prior camouflage. We show that this account cannot be sustained in light of our analysis of Gibson's argument and in light of experiments that are not subject to a putative sensory masking effect (i.e., our Experiment 2 and that of Bacon & Remington, 1995). Rather, we argue, onsets enjoy an advantage over no-onsets in that the perceptual representations for onset stimuli are created rapidly and efficiently, while those for no-onset stimuli must emerge via the updating of an existing perceptual object representation. Experiment 3 provides further corroboration for this conclusion: a letter that was preceded by a figure eight placeholder captured attention by virtue of its late appearance within the display; a masking account predicts instead either a disadvantage for a letter preceded by a brief placeholder, or (at most) no effect of such a manipulation relative to long-duration placeholders. Sensory masking cannot satisfactorily account for this constellation of results.

We noted earlier that there is substantial agreement between Gibson's position and our own about the phenomenon in question. In particular, we agree that onset letters enjoy an attentional advantage, relative to no-onset letters, in visual search. The agreement may run even deeper than this. Gibson's primary assertion is that onset letters exhibit temporal precedence relative to no-onset letters by virtue of the no-onset letters having been visually masked and therefore delayed by the figure eight placeholders, and this temporal precedence causes observers to direct attention to the onset stimuli preferentially. We have shown, however, that what Gibson refers to as forward masking is not masking at all, at least not as that term is standardly used in the literature. The influence of preceding placeholders on object identification may be related to what Enns and Di Lollo (1995) term "attentional masking," which is explicitly nonsensory and concerns how perceptual object representations are established, maintained, and updated.

Our account emphasizes the distinction between the rapid, high priority process of creating new perceptual object representations and the sluggish, lower priority process of updating existing object representations. Further investigation of the temporal dynamics of object formation and attention in early visual processing will help clarify and ultimately resolve these theoretical issues.

References

Bacon, W. F., & Remington, R. W. (1995, November). Attentional capture by abrupt onsets in moving displays. Poster presented at the 36th Annual Meeting of the Psychonomic Society, Los Angeles.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Coltheart, M. (1980). Iconic memory and visible persistence. *Perception & Psychophysics, 27,* 183-228.

Di Lollo, V. (1980). Temporal integration in vision. *Journal of Experimental Psychology: General, 109,* 75-97.

Di Lollo, V., & Bischof, W. F. (1995). Inverse-intensity effect in duration of visible persistence. *Psychological Bulletin, 118,* 223-237.

Enns, J. T., & Di Lollo, V. (1995). Attentional masking: Visual interference by object substitution. Submitted for publication.

Eriksen, C. W. (1966). Temporal luminance summation effects in backward and forward masking. *Perception & Psychophysics, 1,* 87-92.

Folk, C. L., & Annett, S. (1994). Do locally defined feature discontinuities capture attention? *Perception & Psychophysics, 56,* 277-287.

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

Folk, C. L., Remington, R. W., & Johnston, J. C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance, 19,* 682-685.

Gibson, B. S. (1996). Visual quality and attentional capture: A challenge to the special role of abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance, 22,* xxx-xxx.

Growney, R., Weisstein, N., & Cox, S. I. (1977). Metacontrast as a function of spatial separation with narrow line targets and masks. *Vision Research, 17,* 1205-1210.

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics, 55,* 399-411.

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

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology, 24,* 175-219.

Stelmach, L. B., & Herdman, C. M. (1991). Directed attention and perception of temporal order. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 539-550.

Todd, J. T., & Van Gelder, P. (1979). Implications of a transient-sustained dichotomy for the measurement of human performance. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 625-638.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80,* 1-52.

Yantis, S. (1993a). Stimulus-driven attentional capture. *Current Directions in Psychological Science, 2,* 156-161.

Yantis, S. (1993b). Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 676-681.

Yantis, S. (in press). Objects, attention, and perceptual experience. In R. Wright (Ed.), *Attention*. New York: Oxford University Press.

Yantis, S., & Egeth, H. E. (1994). Visual salience and stimulus-driven attentional capture. *Investigative Ophthalmology and Visual Science, 35,* 1619.

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

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 812-825.

Yantis, S., & Jones, E. (1991). Mechanisms of attentional selection: Temporally-modulated priority tags. *Perception & Psychophysics, 50,* 166-178.

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.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 121-134.

Figure Captions

Figure 1. Results from Experiment 1 with stationary placeholder displays. Response time is plotted as a function of display size for the all onset and all no-onset conditions, respectively.

Figure 2. Stimulus displays used in Experiment 2. Top: No-onset condition. Bottom: Onset condition. Each frame shows a snapshot of the display at three points in time. Frame 1 is the initial frame of motion, Frame 8 is approximately 1050 ms into motion, and Frame 15 is the stationary search display, which appeared at the end of motion. Motion lasted 2250 ms. The starting point and direction of motion varied randomly from trial to trial. Stimuli not drawn to scale.

Figure 3. Results from Experiment 2 with moving placeholder displays as shown in Figure 2. Response time is plotted as a function of display size for the all onset and all no-onset conditions, respectively.

Figure 4. Stimulus displays used in Experiment 3. Top: Delayed Placeholder condition. Bottom: Onset condition.

Figure 5. Results from Experiment 3. Top: Delayed Placeholder condition. Bottom: Onset condition.

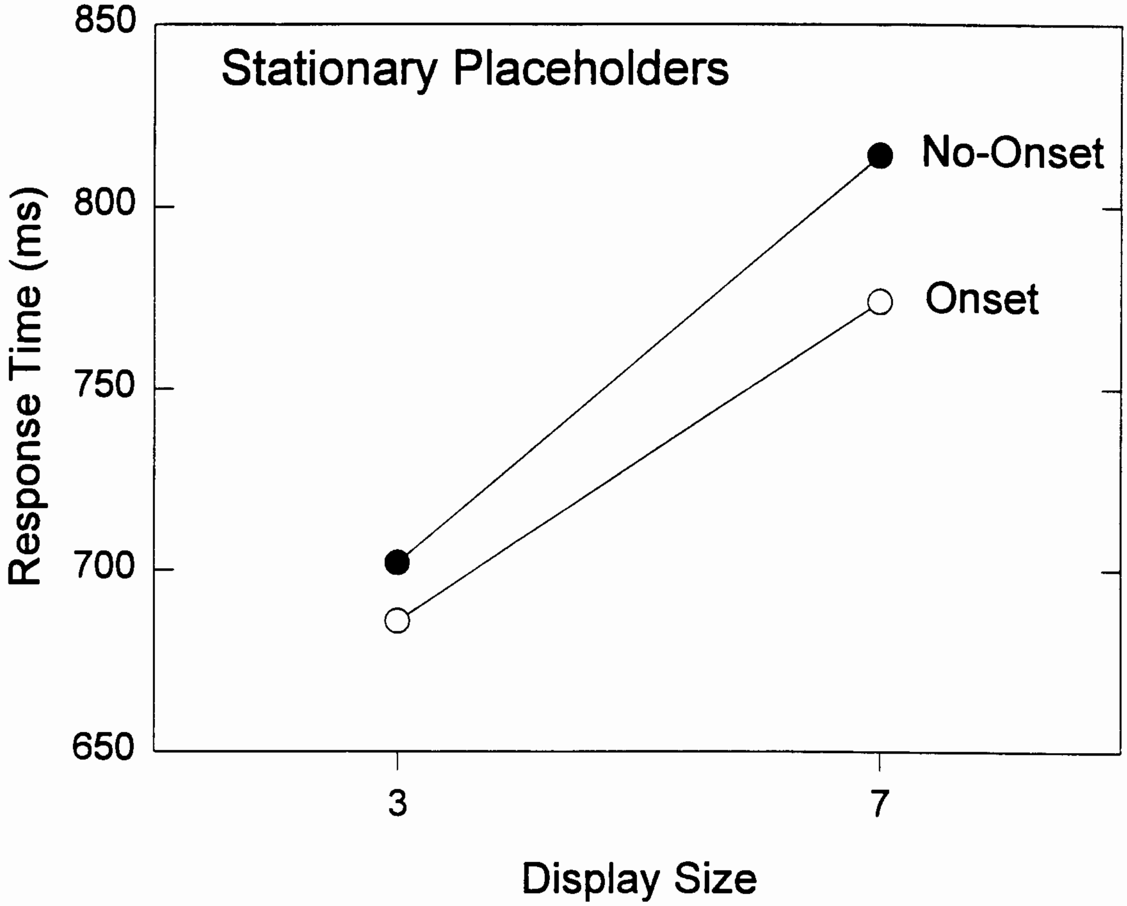
Stationary Placeholders
No-Onset
Onset
850
800
750
700
650
Response Time (ms)
3
7
Display Size

fig 1

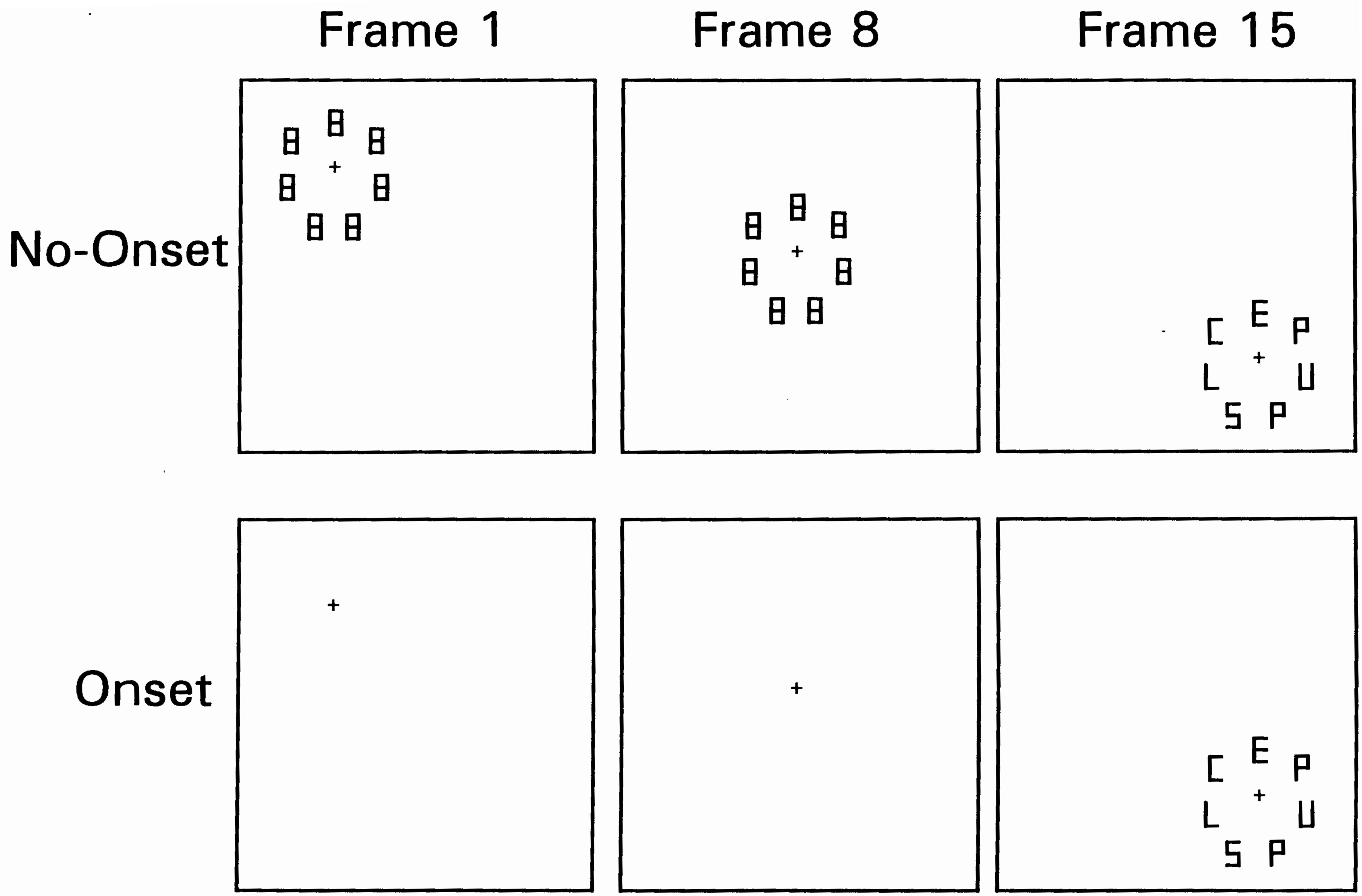

fig 2

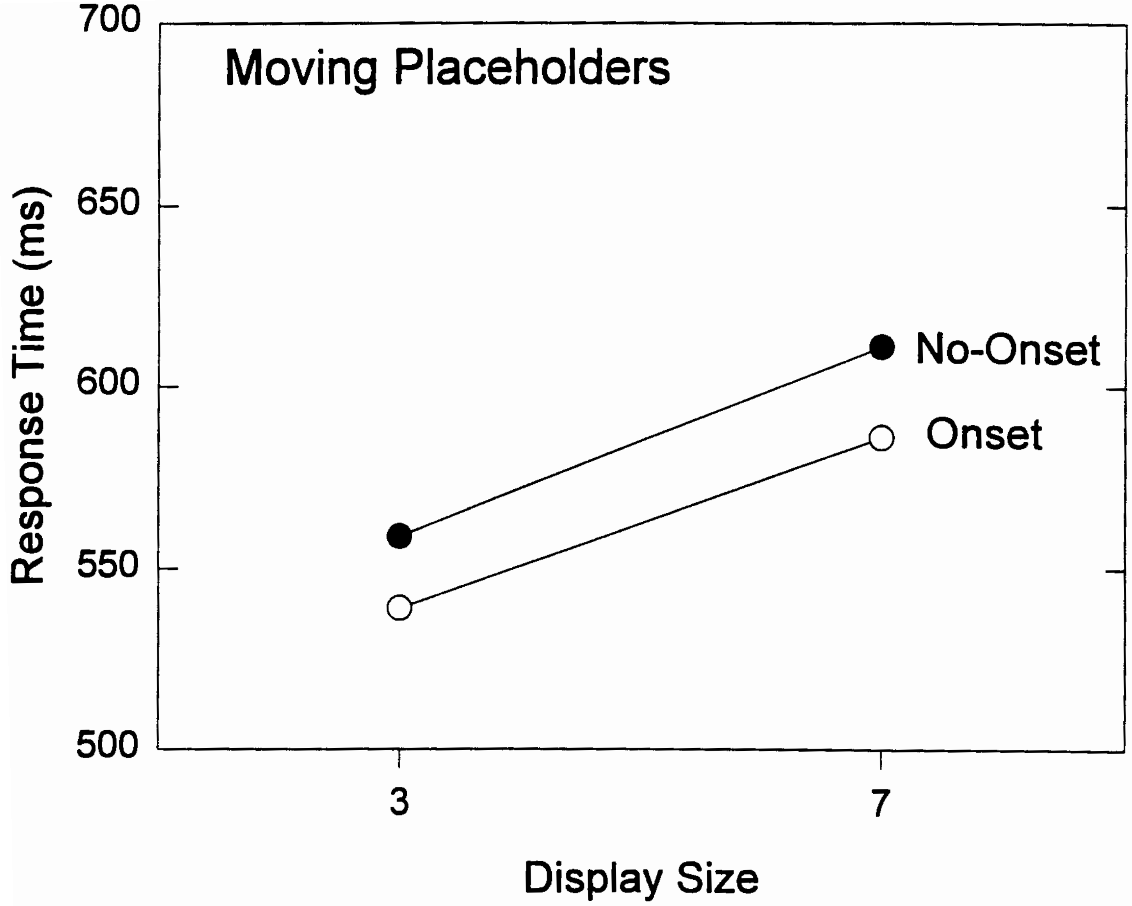
Moving Placeholders
Response Time (ms)
700
650
600
550
500
No-Onset
Onset
3
7
Display Size

fig 3

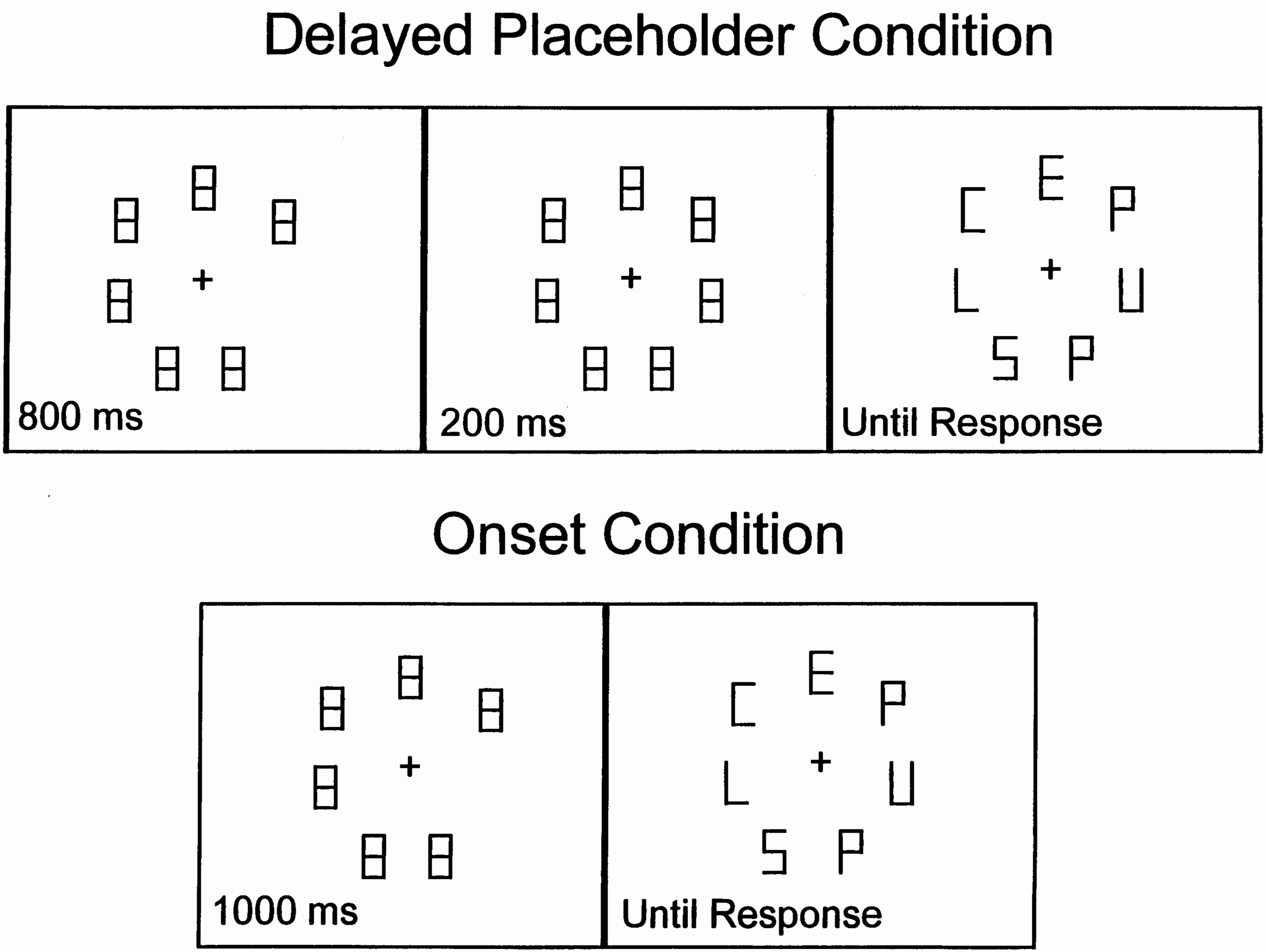
Delayed Placeholder Condition
800 ms
200 ms
C E P
L + U
S P
Until Response
Onset Condition
1000 ms
C E P
L + U
S P
Until Response

fig4

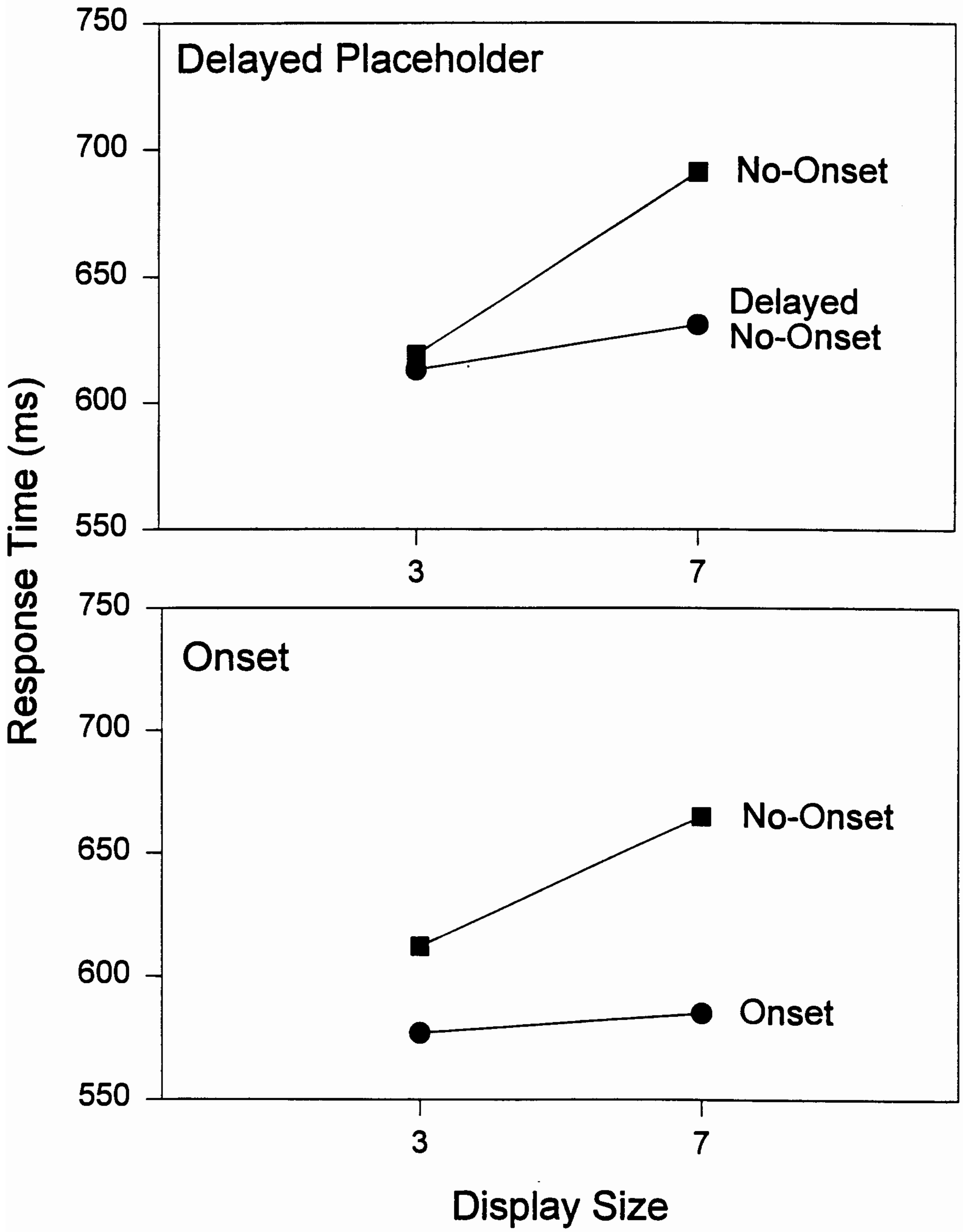
Response Time (ms)
750
700
650
600
550
Delayed Placeholder
No-Onset
Delayed
No-Onset
3
7
750
700
650
600
550
Onset
No-Onset
Onset
3
7
Display Size

fig. 5