Visual marking of old objects

JAN THEEUWES
TNO Human Factors Research Institute, Soesterberg, The Netherlands

and

ARTHUR F. KRAMER and PAUL ATCHLEY
University of Illinois at Urbana-Champaign, Urbana, Illinois

Watson and Humphreys (1997) presented evidence that selection of new elements can be prioritized by on-line top–down attentional inhibition of old stimuli already in the visual field (visual marking). The experiments on which this evidence was based always presented old elements in green and new elements in blue. Because of this, instead of prioritizing new objects by inhibiting old objects, selection could have been based on color. The present experiment, which does not contain this confound, showed that visual marking is a strong and robust process that enables subjects to visually mark at least 15 old elements even when these elements have the same color as the new elements. The results indicate that preview of the elements is critical—not the fact that those elements contained a common feature.

On the basis of a series of innovative experiments, Watson and Humphreys (1997) recently proposed a new mechanism for prioritizing the selection of new events or objects, which they referred to as "visual marking." Visual marking occurs because observers actively inhibit the locations of old or previously processed objects. Because the old objects are visually marked, they no longer compete as strongly for selection. Visual marking has important theoretical implications for theories of visual search; it is also ecologically important because it may be applied flexibly to optimize selection in a variety of real-world situations.

Visual marking was proposed on the basis of evidence obtained in nine experiments using a modified visual search task. A single-feature search task (e.g., a search for a blue H among blue A distractors) and a conjunction search task (search for a blue H among green H's and blue A's) served as baseline conditions for the critical "gap" conditions, in which there was a 1,000-msec gap between the distractors in the two colors. In other words, green H's were presented first for 1,000 msec, after which the blue letters were added to the display. The results showed that the search functions in the gap condition were similar to those in the feature condition, suggesting that in the gap condition, subjects searched only among the blue letters and ignored the initially presented green letters. Search in the gap condition was as efficient as in the feature condition, indicating that subjects could limit their search to the blue subset of elements.

On the basis of these findings, Watson and Humphreys (1997) proposed visual marking: Subjects actively inhib-

ited the initial "old" green distractors, enabling a higher priority to be assigned to the "new" blue elements that were presented later. Although the data are consistent with the visual marking interpretation, a confound was present in all the experiments: Oldness and newness was perfectly correlated with color. All old elements were always green and all new elements were always blue. Previous research has shown that in conjunction search, subjects can limit their search to items of one color and ignore items in a different color (Egeth, Virzi, & Garbart, 1984; Kaptein, Theeuwes, & Van der Heijden, 1995). For example, when searching for a red vertical line segment between red tilted and green vertical line segments, Kaptein et al. showed that subjects serially searched among the red elements while ignoring the green elements. Given these findings, it is conceivable that the findings of Watson and Humphreys do not imply visual marking of old objects but simply entail another demonstration of subset selective search on the basis of color.

Watson and Humphreys (1997) recognized this alternative account but refuted it on the basis of their results in the standard conjunction baseline condition, in which both blue and green elements were presented at the same time. In this condition, there was no evidence for selective search among the relevant blue letters, indicating that the figure–ground segmentation was not strong enough to render subset selective search on the basis of color. Yet, this argument is not convincing when it is assumed that it may take some time to set up an adequate figure–ground segmentation to allow color subset selective search. This may very well be the case when the colors used are not very discriminable (i.e., blue and green), as in the experiments of Watson and Humphreys.

Since visual marking has important theoretical implications, it is crucial to determine whether visual marking is a genuine phenomenon or whether the data can be accounted for by selective search on the basis of color. Fur-

thermore, rather than relying on evidence for visual marking on the basis of comparing feature search with search in the gap condition, it would appear important to obtain independent evidence that subjects do indeed limit their search to only the new elements. Instead of a typical conjunction search task, we used a serial search task in which subjects had to search for a target letter (the letter H) among a variable number of different other letters. In these kinds of letter search tasks, there is little information from the parallel process that can be used to guide search. This is the case because of the high degree of heterogeneity of the distractor letters. Therefore, search is expected to be serial, self-terminating through the entire set of elements (Wolfe, 1994). Finally, to determine the limits of visual marking, we used display sizes that were almost twice as large as those employed by Watson and Humphreys (1997).

In the present experiment, subjects searched for a white H among a variable number of other white letters. In the gap condition, the number of elements before the gap (5, 10, or 15) was independently varied from the number of elements presented after the gap (5, 10, or 15 elements). In other words, the number of old and new elements was orthogonally manipulated (see Kaptein et al., 1995, for a similar manipulation). If subjects did not search among the old elements, as suggested by the visual marking account, it was expected that the number of old elements would have no effect on the time to find the target. Search times should depend only on the number of new elements (elements presented after the gap) and should be independent of the number of old elements (elements presented before the gap). In addition to this manipulation, we also included two baseline conditions, similar to the feature and conjunction baselines in Watson and Humphreys (1997). In the half-elements baseline condition, display size was 5, 10, or 15 elements, comparable to the number of elements presented after the gap in the gap condition. In the all-elements baseline condition, display sizes were 10, 20, and 30, comparable to the display size if subjects had to search all elements in the gap condition.

To ensure that old and new elements had physically the same color, the letters were presented in white against a gray background. Previous research has shown that because of local chromatic adaptation, initially presented elements may be perceived as having a color or brightness that is slightly different from elements that are added later to the display (Theeuwes & Lucassen, 1993). Again, since this may introduce a confound between the appearance of the letters and their newness, we used white letters against a gray background, which minimizes the effect of chromatic adaptation. This manipulation basically guarantees that old and new letters are perceived as physically similar in terms of brightness and color.

METHOD

Subjects
Twelve young adults ranging in age from 19 to 28 years participated in the study as paid volunteers. All had self-reported normal or corrected-to-normal vision.

Apparatus
A personal computer (Amdek AM/815E) with an SVGA color monitor controlled the timing of the events, generated stimuli, and recorded reaction times (RTs). The "/" key and the "z" key of the computer keyboard were used as response buttons. Each subject was tested in a sound-attenuated, dimly lit room, with his/her head resting on a chinrest. The CRT was located at eye level, 95 cm from the chinrest.

Stimuli
The target was a white capital letter H, and the distractors were white capital letters randomly drawn (with replacement) from a subset of letters (A, B, C, E, F, G, J, K, M, N, P, R, S, T, U, V). The letters (0.54° × 0.30°) were randomly positioned within a stimulus field of 11.3° × 9.6°. In the gap condition, 5, 10, or 15 elements were presented before the gap and 5, 10, or 15 elements were presented after the gap. There were two baseline conditions in which just one display was presented with no added elements. In the all-elements baseline condition, 10, 20, or 30 elements were presented. In the half-elements baseline condition, 5, 10, or 15 elements were presented. The target was present in 50% of the trials. The letters were presented in white (8.0 cd/m²) and the background was presented in gray (0.9 cd/m²).

Design and Procedure
There were three main conditions: In the all-elements baseline condition, a white fixation cross was presented for 1,500 msec; this was followed by the search display, consisting of 10, 20, or 30 letters, which remained on until the subject responded with a maximum of 6,000 msec. Half of the subjects pressed the z key when the target letter H was present and pressed the / key when it was absent. This assignment was reversed for the other half of the subjects. The half-elements baseline condition was essentially the same except that 5, 10, or 15 letters were presented. In the gap condition following the fixation cross, 5, 10, or 15 elements were first displayed for 1,000 msec, after which 5, 10, or 15 elements were added to the display. In all three conditions, the fixation cross remained on throughout a trial. Subjects were asked to remain fixated until the search display came on. In case of the gap condition, they were asked to remain fixated until the second set of letters was added to the display. Subjects were explicitly told that if the target letter were present, it would be present among the set of elements added to the display after the gap.

All subjects performed all three conditions in separate blocks of trials. The order of these blocks of trials was counterbalanced across subjects. Each block of trials started with 20 practice trials. Each subject performed a total of 882 trials, consisting of 216 all-elements trials, 216 half-elements trials, and 450 gap-condition trials. Display size (for the gap condition, both number of elements presented before and after the gap) was random across trials. Subjects received feedback about their performance (in terms of RT and error rates) after each block of 50 trials.

RESULTS

To compare the results across the three main conditions, display sizes 10, 20, and 30 were used even though in the half-elements condition the true display size was only half of this (see Watson & Humphreys, 1997).[1] In the gap condition, in which the number of elements before and after the gap were independently varied, the data as a function of the number of elements presented after the gap was used. The data were averaged over the number of elements before the gap.

Figure 1 shows the mean correct RTs as a function of display size for target-present (left panel) and target-absent (right panel) trials. The mean RTs for correct trials were entered into an analysis of variance (ANOVA) with condition (gap, half, and all elements), display size (10, 20, and 30), and target (present–absent) as factors.

All three main effects were significant: condition $[F(2,22) = 27.6, p < .001]$, target $[F(1,11) = 440.5, p <$
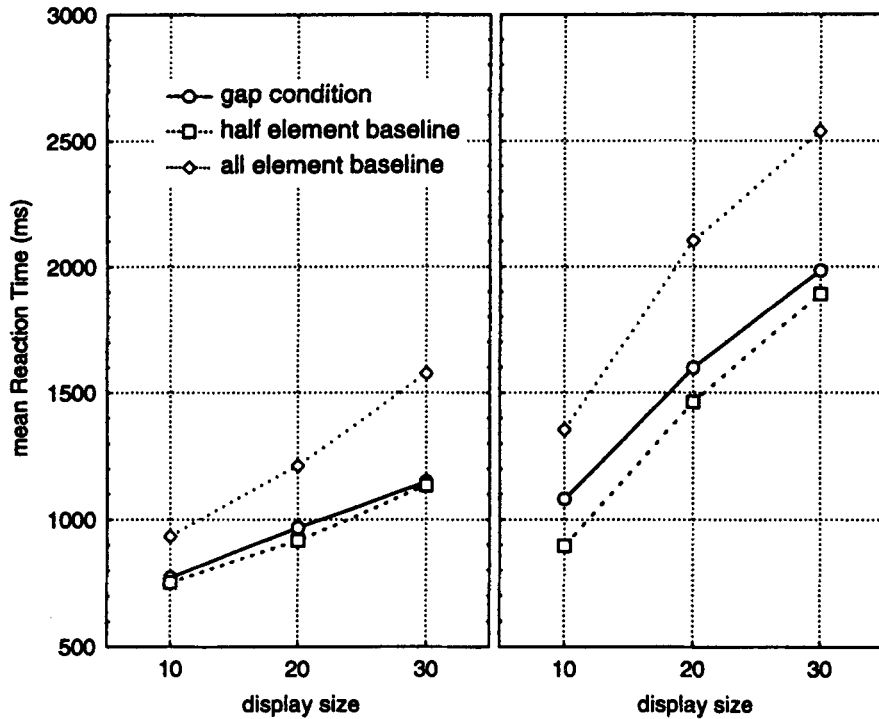
Figure 1. Mean reaction time as a function of condition for target-present (left panel) and target-absent (right panel) trials.

.001], and display size [$F(2,22)$ = 347.8, $p < .001$]. All two-way interactions were also significant: condition × target [$F(2,22)$ = 11.4, $p < .001$], condition × display size [$F(4,44)$ = 7.7, $p < .001$], and target × display size [$F(2,22)$ = 227.7, $p < .001$]. Finally, the three-way interaction was also significant [$F(4,44)$ = 2.9, $p < .05$]. The error rates and search slopes are given in Tables 1 and 2. The speed with which the elements were searched was about 60 msec per item (the all-elements target-absent condition; see Table 1).

Planned comparisons among the three main conditions for target-present and target-absent responses showed that there were no reliable differences between the half-elements and gap condition for either target-present (p = .50) or target-absent trials (p = .19). The difference between the gap condition and the all-elements condition was significant for target-present [$F(1,11)$ = 43.3, $p < .001$] and target-absent trials [$F(1,11)$ = 25.3, $p < .001$]. This analysis clearly indicates that there were no reliable

differences between the gap and the half-elements conditions, suggesting that subjects were able to ignore the old distractor elements from search even when they did not have the opportunity to filter on the basis of color, as was the case in Watson and Humphreys (1997).

The error rates were low (4.77%) and generally followed the same trends as the RT data. An ANOVA on mean error rates showed main effects of target [$F(1,11)$ = 62.3, $p < .001$], indicating that subjects tended to respond "absent" when in fact the target was present. There was also a main effect of display size [$F(2,22)$ = 10.7, $p < .001$], indicating that more errors were made when display size increased. Two interactions were significant: condition × target [$F(2,22)$ = 7.1, $p < .01$] and display size × condition [$F(2,22)$ = 4.5, $p < .05$]. More importantly, planned comparison showed no difference between the error rates in the gap condition and the half-elements condition for target-present trials (p = .097) or for target-absent trials (p = .21). Also, there were no differences

Table 1
Search Slopes for Each of the Conditions

| | Condition | | | | | |
| | Gap | | Half-Elements | | All-Elements | |
Measure	Present	Absent	Present	Absent	Present	Absent
Slope (msec)	18.8	45.5	19.2	49.8	31.2	59.2
Intercept (msec)	587	647	605	814	551	422
Absent:present ratio	2.4		2.6		1.9	

Table 2
Mean Percentage Error Rates

Condition	Display Size		
	10	20	30
Gap			
Present	5.8	7.0	11.2
Absent	1.0	1.0	5.6
Half-elements			
Present	5.6	6.0	7.6
Absent	1.0	0.8	3.5
All-elements			
Present	6.1	10.6	13.3
Absent	0.2	0.7	1.5

between the error rates of the all-elements and gap condition for target-present ($p = .20$) or target-absent trials ($p = .61$). Overall, in line with the RT data, the error data indicated no differences between the half-elements and the gap condition.

In the gap condition, the number of elements before and after the gap were factorially manipulated, enabling us to determine whether subjects were able to completely ignore the old elements. An ANOVA on mean RT with number of elements before the gap (display size before) and the number of elements after the gap (display size after) and target (present, absent) showed a main effect of target [$F(1,11) = 129.5, p < .001$], number of elements before the gap [$F(2,22) = 10.3, p < .001$], and

number of elements after the gap [$F(2,22) = 97.5, p < .001$]. Also, the interaction between (1) the number of elements before the gap and target and (2) the interaction between the number of elements after the gap and target were significant [$F(2,22) = 7.8, p < .01$, and $F(2,22) = 97.6, p < .001$, respectively].

Additional analyses performed separately for target-present and target-absent trials indicated that the effect of number of elements before the gap was not reliable for target-present trials [$F(2,22) = 2.0, p = .15$], but the effect was reliable for target-absent trials [$F(2,22) = 11.1, p < .001$].

This analysis shows, as is clear from Figure 2, that subjects were able to completely ignore the old elements on the target-present trials. In line with the predictions of visual marking, the results show that search time depends only on the number of elements presented after the gap and is independent of the number of elements presented before the gap. Yet, for target-absent trials, there was a reliable effect of the number of elements presented before the gap, suggesting that subjects did search among the old elements. The results seem to suggest that subjects first search the new elements. If the target is not found among the new elements (e.g., in case of a target-absent trial), subjects do search among a few old elements. If 60 msec/item is an accurate estimate of the search rate (the absent search slope for the all-elements condition; see Table 1), the data indicate that for target-absent
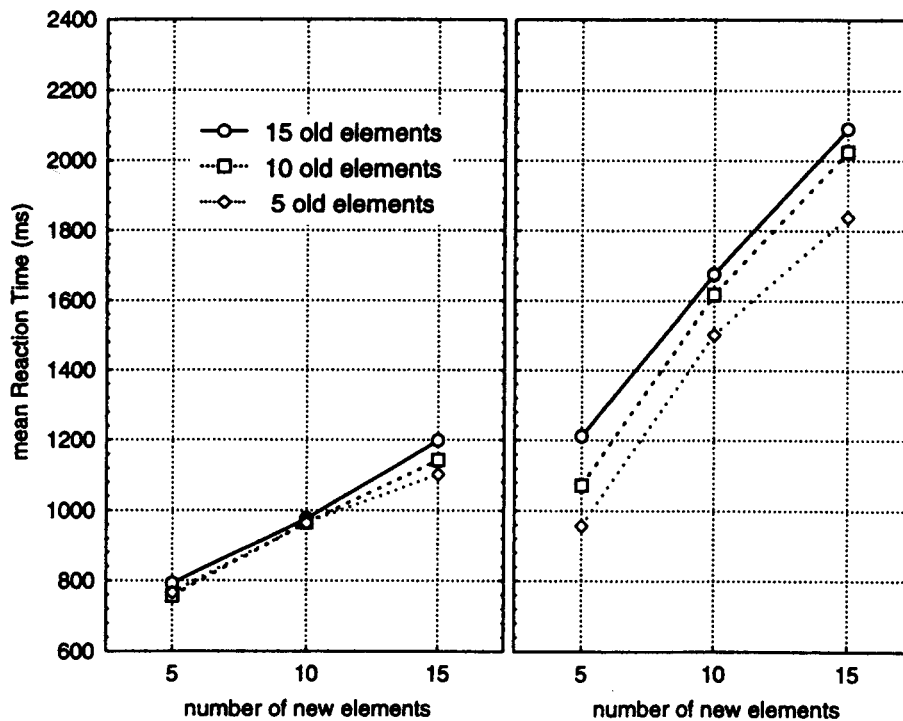

Figure 2. Mean reaction time as a function of the number of new elements, separately for each number of old elements for target-present (left panel) and target-absent (right panel) trials.

trials, on average about four old elements are searched before subjects decide that the target is not present.

DISCUSSION

The results indicate that visual marking occurs even when it is ensured that the new objects have physically the same color as the old objects. Although Watson and Humphreys's (1997) findings could be explained in terms of subset selective search among elements in a particular color, the present findings can be explained only by the visual marking hypothesis (Watson & Humphreys, 1997).

The present findings show that visual marking also occurs for display sizes up to 30 elements, suggesting that subjects can mark up to at least 15 old items simultaneously. Even though Watson and Humphreys (1997) claimed that there were basically no limitations in the number of locations that can be marked, they examined visual marking only up to eight old elements. Given the finding that marking up to 15 elements is possible, it is highly unlikely that a mechanism such as FINST (Pylyshyn, 1989) is responsible for the gap effect since FINST assumes that it is possible to tag to only about four locations or objects simultaneously.

In addition, the present findings indicate that visual marking also occurs for a task that is typically considered serial self-terminating through the entire set of elements. The search slope of 60 msec/item is similar to the slope reported by Wolfe (1994) for searching for a T among rotated Ls, a search task typically considered to be serial and self-terminating. The fact that it takes considerable time to finish such a serial self-terminating search (e.g., in our experiment, up to 2 sec) suggests that there is not much decay of the visual marking mechanism.

The manipulation of the number of elements before and after the gap provides converging evidence for visual marking. The results show that subjects first search serially and exhaustively through the set of new elements. If the target is not found among the new elements (in case of target-absent trials), subjects may search among up to four old elements and then decide that the target is not present (i.e., rechecking the old elements; Treisman & Gelade, 1980). Note, however, that the finding that subjects search through a few old elements might also be explained by some decay of the visual marking mechanism.

One aspect of the present data is different from those reported by Watson and Humphreys (1997). Watson and Humphreys reported an intercept difference between the feature search and the gap condition. In the present experiment there was virtually no difference either in terms of search slope or intercept between the gap condition and the feature search condition. Watson and Humphreys explained the intercept effect in terms of resource limitations: Because the inhibition of the old elements requires attentional resources, search for the new objects may have been delayed. It is unclear why there was no intercept effect in the present experiment. If anything, since the old and new letters had iden-

tical colors in the present experiment, one would expect that excluding the old elements would have required more resources. However, because a *serial* search task was used in the present experiment, there could not have been any guidance from the parallel stage. Because parallel processing was ineffective, subjects may have started searching as soon as the new elements were added to the display. This may explain why there was no intercept effect in the present experiment.

The present study provides direct evidence that preview of the elements (i.e., making them "old") is critical—not the fact that those elements contained a common feature. Whether or not marking is due to *inhibition* of the initial distractor set, as suggested by Watson and Humphreys (1997), is open to further examination.

REFERENCES

EGETH, H. E., VIRZI, R. A., & GARBART, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 32-39.

KAPTEIN, N. A., THEEUWES, J., & VAN DER HEIJDEN, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1053-1069.

PYLYSHYN, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, **32**, 65-97.

THEEUWES, J., & LUCASSEN, M. P. (1993). An adaptation-induced pop-out in visual search. *Vision Research*, **16**, 2353-2357.

TREISMAN, A. M., & GELADE, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, **12**, 97-136.

WATSON, D. G., & HUMPHREYS, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attention inhibition of old objects. *Psychological Review*, **104**, 90-122.

WOLFE, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238.

NOTE

1. We realized that the real display sizes were different from those shown here. In the all-elements baseline condition, there were 10, 20, and 30 elements; in the half-elements baseline, there were 5, 10, and 15 elements; in the gap condition, there were 5, 10, or 15 new elements orthogonally combined with 5, 10, and 15 old elements. We choose to present the data in this way to allow a direct comparison among the different conditions, identical to the way Watson and Humphreys (1997) presented their data.