

Visual Search Has No Memory

Humans spend a lot of time searching for things, such as roadside traffic signs, soccer balls, or tumors in mammograms. These tasks involve the deployment of attention from one item in the visual field to the next. Common sense suggests that rejected items should be noted in some fashion so that effort is not expended reexamining items that have been attended and rejected. However, common sense is wrong. We had human observers search for a letter "T" among letters "L". This search demands visual attention and normally proceeds at the cost of 20-30 milliseconds per item. In the critical condition, we randomly relocated all letters every 111 msec. This made it impossible to keep track of the progress of the search. Nevertheless, the efficiency of the search was unchanged. Theories of visual search have uniformly assumed that search relies on accumulating information about the identity of objects over time". Such theories must predict that search efficiency will be drastically reduced if the scene is continually shuffled while the observer is trying to search through it. Since efficiency is not impaired, the standard theories must be revised.

When a target item differs from distractors on a simple visual feature, such as a red bar among green bars, the target automatically grabs one's attention and can be detected independently of the number of distractor items present. When targets and distractors differ only in their spatial arrangement, however, the search becomes attention-demanding and the reaction time increases by 20-30 milliseconds per item. Theories of visual search explain this phenomenon in one of two ways. "Serial" models propose that attention can process the identity of only one item at a time. Once an item has been identified and rejected as a distractor, an inhibitory "tagging" mechanism prevents that item from being revisited. As a result, a successful search for a target will require subjects to examine, on average, only half the items in the display⁵. "Parallel" theories assume that identity is computed in parallel for each item, and that an item's identity becomes gradually more certain over the course of a trial. A response is issued either when sufficient information confirms one item as the target, or when all of the items have proven to be distractors⁶. Both theories have in common the assumption that efficient search is based on accumulating information about the contents of the scene over the course of the trial; we refer to this as memory-driven search. We propose an alternative, that visual search processes are amnesic: They act on neural representations which are continually rewritten and have no permanent existence beyond the time span of visual persistence.

To test the hypothesis that visual search relies on memory-driven mechanisms, we designed our stimuli so that, during a trial, the scene would be constantly changing, yet the meaning of the scene (as defined by the required response) would remain constant. The task was to report as quickly as possible whether or not the target letter, T, was present in the display. In order to measure the increase in reaction time when extra items were present in the display, we varied the number of letters in the display (the set size) between 8, 12, and 16. The slope of the target-present reaction-time X set-size function measures the efficiency of search through the display. This slope represents the added cost of each additional item. We concentrate on target-present slopes because their interpretation is more straightforward; the question of when to stop

searching when you have not yet found a target is more complicated than the question of when to respond once you have found a target. In half of the trials, all the letters were Ls, and these trials demanded a "no" response. On the remaining trials, which required a "yes" response, one of the letters was a T. Both Ts and Ls could appear, randomly, in any of four orientations: 0°, 90°, 180° and 270° (see Fig. 1).

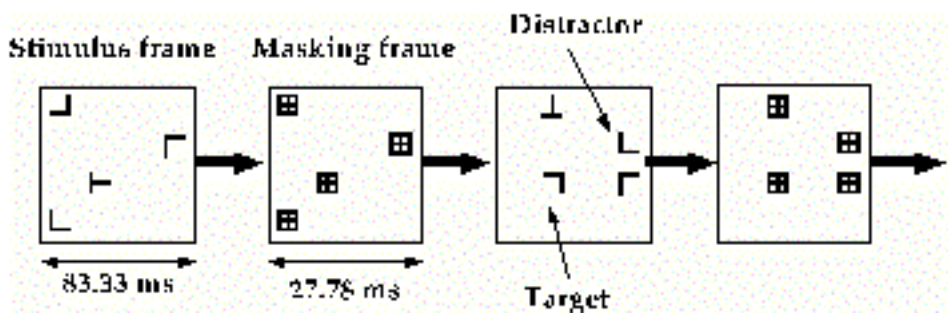


Figure 1: Two example stimulus frames from experiment 1, each followed by its corresponding masking frame. An actual trial in experiment 1 had four stimulus frames, repeated through five cycles (see text). In experiment 2, the masking frames were eliminated and each stimulus frame was presented for 106.7 msec.

There were two stimulus conditions in the experiments: random and static. In the random conditions, the stimulus locations were changed every 111 ms (see Fig. 1). For any memory-driven search mechanism, this manipulation would be disastrous. It would cripple parallel accumulation of information about the identity of a particular letter. A serial model would be unable to keep track of where it had been, and would be forced to resample already searched locations. For a given rate of serial sampling of items, Monte Carlo simulations show that serial sampling with replacement should result in mean reaction time $\times$ set size slopes that are twice as steep as those resulting from the normally assumed serial sampling without replacement.

In contrast, an amnesic search mechanism would be oblivious to the randomization manipulation. For simplicity, we will describe only the serial version of an amnesic system, though an equivalent parallel interpretation can also be developed. Our model assumes that the visual system generates a priority ranking of each item in the field according to the salience of the item. This neural representation is somewhat noisy and fluctuates dynamically. Under normal circumstances, this priority ranking would reflect important feature differences in the scene (such as color and size) and thus allow attention to be efficiently guided to the most likely target locations, effectively gating out stimuli which are unlikely to be targets. The stimuli in these experiments were specifically designed so as not to allow such guidance.

As the representation is assumed to be noisy, there will be spurious differences in salience even between identical items. As a consequence, the priority assigned to each item will change over time, even in the static case. An amnesic search would proceed by determining the most likely (salient) target at the moment and directing attention to that location. If that item were to be identified as a non-target, the next item would be selected according to the same criterion: its momentary salience. From the point of view of an amnesic mechanism, there are n items at any given moment. Thus with every sample, the system would have a $1/n$ chance of picking out the target. When the stimuli are shuffled about between samples, a memory-driven model cannot

keep track of where it has been and loses information. An amnesic mechanism, by contrast, was not keeping track in the static condition and, thus, does not lose anything when the stimuli are shuffled.

[Figure 2](#) shows results of correct target-present trials in two experiments; in the first experiment masked stimuli were used, in the second unmasked stimuli were used (see [Methods](#)). Subjects are slower and slightly less accurate in the random conditions. However, slopes of the random and static target-present reaction-time X set-size functions ([Table 1](#)) are statistically indistinguishable for both experiments ($t_{\text{masked}}(8) = .13$, $p < .50$, $t_{\text{unmasked}}(8) = 1.52$, $p > .15$). This follows the prediction of the amnesic model. In contrast to the predictions of the memory-driven search theories, there is no evidence that subjects are searching half as efficiently in the random conditions as in the static conditions. In fact, shallower slopes are produced under random conditions than under static conditions. Data for the masked and unmasked conditions are comparable, which argues that the flickering masks in experiment 1 did not noticeably affect the search.

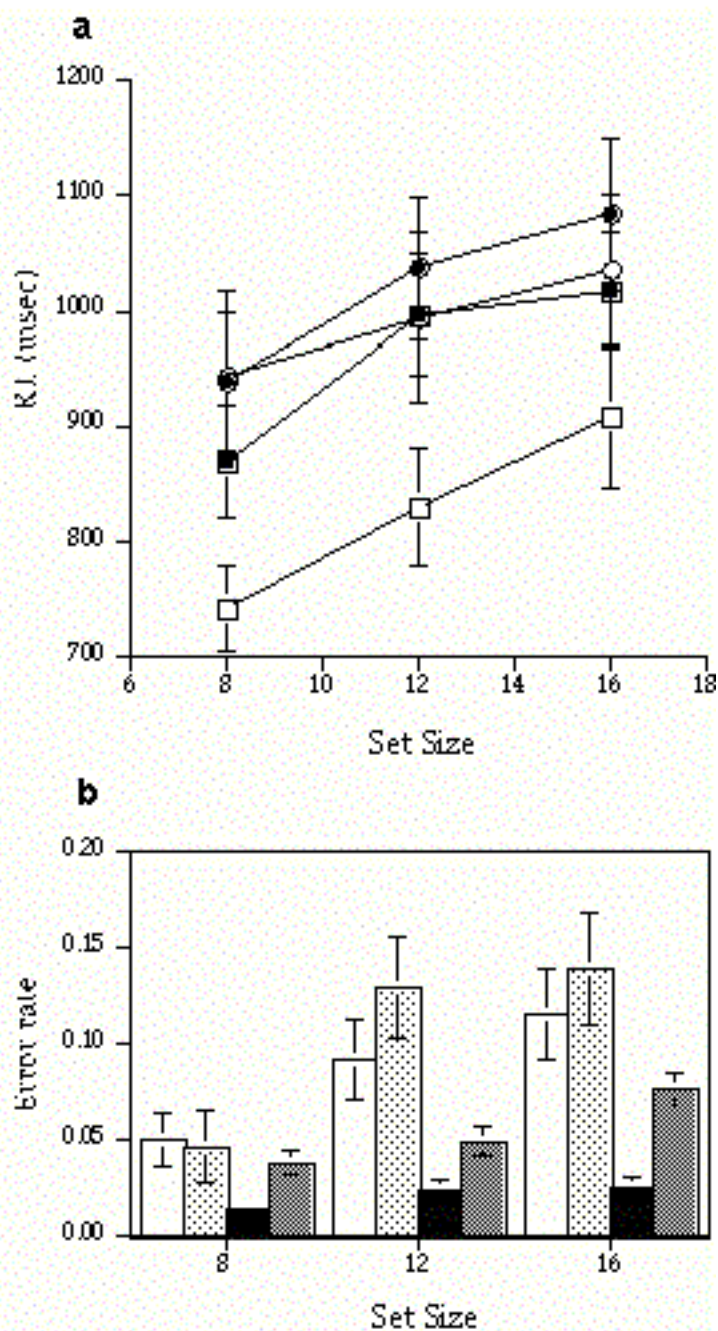


Figure 2

Figure 2 Results of experiments 1 and 2 **a**, Mean correct target-present reaction times (RTs) plotted against set size for the random and static conditions from experiments 1 and 2. Squares denote the static condition and circles the random condition. Filled symbols represent experiment 1 (masked) and open symbols experiment 2 (unmasked). Error bars indicate the

s.e.m. The main finding is that changing the location of items every 111 ms (random condition) does not alter the efficiency of visual search (the slopes of the lines). **b**, Error rates by set size, stimulus condition, and experiment. For each set size, bars from left to right indicate the masked static condition, the masked random condition, the unmasked static condition, and the unmasked random condition. Error bars indicate the s.e.m. More errors are committed under the more difficult, random conditions. However, subjects are not trading off accuracy for speed.

Although the slopes of the reaction-time x set size functions are no steeper in the random conditions, the mean reaction times do appear to be longer; however, the reaction-time cost is reliable in only experiment 2 ($F(1,8)=18.81$, $p < .005$). We suspect that the increased mean reaction times reflect subjects' decreased confidence in their responses. Consider a subject who believes she has found a target. In the static case, the physical stimulus is still available for confirmation, while in the random case, it is not. The slope data show that, contrary to predictions of any memory-driven account, search efficiency is similar under random and static conditions. The mean reaction time data merely suggest that subjects may be less confident in the Random condition.

Error rates are higher under random conditions. This is not surprising because the random conditions are more difficult than the static conditions. However, it does raise the possibility of a speed-accuracy trade-off. Artificially shallow slopes might occur if subjects guess early in a trial rather than waiting to confirm the presence of a target. Half of the time their guesses will be correct and will contribute to a shallow slope. Half of the time they will be wrong, producing "false alarms". Conversely, when the target is difficult to find, subjects may give up and inaccurately respond "no", thus unfairly taking long reaction times out of the distribution.

In a third experiment, we eliminated the option to respond "no" by having subjects respond to target identity, rather than target presence. A target letter "E" or "N" was present on each trial, embedded in distractors selected from the remaining letters of the alphabet (except for "I" and "J"). Subjects identified the target letter. Otherwise, the procedure was identical to that of the unmasked experiment.

Once again, the slopes (29.53 ms per item for the random condition and 34.67 ms per item for the static) were statistically indistinguishable. As expected, errors were substantially lower in this experiment (5.6% errors overall for the random condition and 2.8% overall for the static) than in previous experiments. There were still more errors under random conditions than under static conditions, but not enough to make the memory-driven story plausible. For instance, if we take the subjects with the smallest differences between random and static error rates, we still find that random and static slopes are essentially the same. Space constraints preclude a complete discussion of the speed-accuracy trade-off issue. For more details, see

http://search.bwh.harvard.edu/recent_projects/nature/speed_error.htm.

These experiments provide dramatic evidence that the visual system does not accumulate information about object identity over time during a search episode. Instead, the visual system seems to exist in a sort of eternal present. There is little integration of visual information across saccadic eye movements, and observers are remarkably oblivious to dramatic scene changes when the moment of change is obscured by a brief flicker or an intervening object. Although subjects in such a "change blindness" experiment may suffer momentary embarrassment, an amnesic visual system may be a handicap only in the laboratory. The structure of the world

makes it unnecessary to build fully elaborated visual representations in the head. If an observer knows that an elephant was present a moment before, she can be quite sure that it will be an elephant when it is attended to again. She can guide attention to the elephant on the basis of basic features such as color, and higher-level mnemonic strategies ("I already looked for elephants near the trees.") can prevent attention from retracing its steps too often. Amnesia may be an efficient strategy for way to build a visual system.

● Method

Two conditions, static and random, were tested in experiments 1-3. In experiment 1, four stimulus frames were generated for each trial (Fig. 1). If there was a target present in that trial, it was present in all frames. The same number of items was present in each frame of a given trial. In the random condition, the locations and orientations of the targets and distractors were independent from frame to frame. In the static condition, the four frames were identical; items therefore remained fixed and unchanging in their locations, alternating with masks at that location.

For each frame, a corresponding masking frame was generated by placing a mask (a square bisected along both axes) at every location where there was a letter in the stimulus frame, such that all possible line segments of the letters would be masked. Each frame was presented for 83.33 msec, and was followed for 27.78 msec by its masking frame (see Fig. 1). The cycle of four frames repeated every 444.44 msec, and subjects observed 5.25 cycles (21 stimulus frames). Subjects were instructed to respond as soon as possible whether or not the target was present, but were allowed to respond at any time within 5000 ms of stimulus onset. The final frame was always a mask. Subjects were therefore allowed to respond after the final frame, but such responses accounted for less than 2% of target-present responses. Nine subjects were tested for 200 trials in each condition, randomly distributed over three set sizes (8, 12, and 16 items per display).

The second experiment differed in two respects. First, 20 frames were generated for each trial, and no frame was presented more than once, so the total presentation time was shortened to 2140 msec. We also eliminated the masking frames, so each stimulus frame was now shown for 107 ms total. Stimulus exposure time was thus roughly equal across the two experiments. 9 subjects were tested for 480 trials in each condition, using the same set sizes as in experiment 1.

Both experiments were designed to thwart a strategy of attending at one location, waiting for the target to appear in the random case. In experiment 1, the target only appeared at four locations, so such a strategy would lead to failure on 93.75% of trials. In Experiment 2, the target changed location on every trial but remained at the same eccentricity, so a "sit and wait" strategy would still fail on 75% of trials.

In experiment 3, subjects identified which of two targets, an E or N, was presented in each trial. Eleven subjects participated for 480 trials in each condition. Methods were otherwise similar to those used for experiment 2.

Table 1: Reaction-time X set-size slopes for experiments 1, 2, and 3.

Target	Present	Present	Absent	Absent
Condition	Static	Random	Static	Random
Masked Stimuli	18.76 (3.66)	18.13 (4.66)	50.42 (4.25)	23.74 (6.75)
Unmasked Stimuli	20.89 (3.59)	11.51 (3.54)	42.00 (6.07)	12.18 (4.19)
Target Identification	34.67 (3.20)	29.53 (3.02)	n / a	n / a

Data are shown as ms per item, means $\pm$ s.e.m. NA, not applicable.

References

¹ Fairclough, S. & Maternaghan, M. Changes in driver's visual behaviour due to the introduction of complex versus simple route navigation information. In *Visual Search 2* (eds Brogan, D., Gale A. & Carr, K.) 419-431 (Taylor & Francis, London, 1993).

² Helsen, W. & Pauwels, J. M. A cognitive approach to visual search in sport. In *Visual Search 2* (eds Brogan, D., Gale A. & Carr, K.) 379-388 (Taylor & Francis, London, 1993).

³ Gale, A. G. & Walker, G. E. Visual search in breast cancer screening. In *Visual Search 2* (eds Brogan, D., Gale A. & Carr, K.) 231-238 (Taylor & Francis, London, 1993).

⁴ Wolfe, J. M. What can 1 million trials tell us about visual search? *Psychological Science*, **9**, 33-39

(1998).

⁵ Treisman, A. & Gelade, G. A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136 (1980).

⁶ Palmer, J. Attention in visual search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, **4**, 118-123 (1995).

⁷ Grossberg, S., Mingolla, E., & Ross, W. D. A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, **101**, 470-489 (1994).

⁸ Chun, M. M. & Wolfe, J. M. Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology* **30**, 39-78 (1996).

⁹ Wolfe, J. M. Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, **1**, 202-238 (1994).

¹⁰ Irwin, D. E. Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, **5**, 94-100 (1996).

¹¹ Rensink, R. A., O'Regan, J. K., & Clark, J. J. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, **8**, 368-373 (1997).

¹² Simons, D. J., & Levin, D. T. Change blindness. *Trends in Cognitive Sciences*, **1**(7), 261-267 (1997).

Acknowledgments. We thank L. Adams, K. Dahlen, H. Egeth, G. Gancarz, I. Horowitz, J.

Mangels, R. Klein, A. Treisman, R. Ward, and S. Yantis for helpful comments on the manuscript, and N. Klempen for assistance with data collection. Supported by grants from AFOSR , NIH and NSF (to J.M.W.) and an NIH individual NRSA (to T.S.H.).