Vision Res. Vol. 26, No. 7, pp. 1111–1117, 1986
Printed in Great Britain

0042-6989/86 $3.00 + 0.00
Pergamon Journals Ltd

THE IMPORTANCE OF EYE MOVEMENTS IN THE ANALYSIS OF SIMPLE PATTERNS

K.-H. Schlingensiepen,[1]* F. W. Campbell,[2] G. E. Legge[3] and T. D. Walker[3]

[1]Max-Planck-Institute of Biophysical Chemistry, Abt. 11 Am Fassberg, 3400 Gottingen, F.R.G., [2]The Physiological Laboratory, Cambridge CB2 3EG, England and [3]Department of Psychology, University of Minnesota, Minneapolis, MN 55455, U.S.A.

(*Received* 1 *July* 1985; *in revised form* 14 *January* 1986)

Abstract—How important are eye movements to visual pattern analysis? Previous findings indicate that at least one visual task (counting) is seriously impaired without them. We asked whether a comparable limitation applies to pattern recognition. Subjects were presented with pairs of randomly generated arrays composed of black and white pixels. The subjects indicated whether the arrays were identical or differed by one pixel. In one experiment, they were instructed to use normal eye movements, in another they were required to fixate on a point between the arrays. When eye movements were permitted, subjects' performance showed evidence for a search in which the discrepant pixel was eventually found, given adequate inspection time. When fixation was required, search was less efficient and the discrepant pixel was sometimes not found, despite prolonged inspection time. These results were independent of target size over a wide range. Our findings indicate that eye movements play a crucial role in pattern analysis that is not related to resolution.

Eye movements Pattern recognition Visual search Visual resolution

INTRODUCTION

Jevons (1871) demonstrated a surprising limitation on visual information processing. He threw beans into an open box and attempted to count them at a glance. He found that when there were more than four beans in the box, he began to make errors. Atkinson, Campbell and Francis (1976) verified and extended Jevons' original observation by examining the speed and accuracy with which subjects could count rows of dots. When there were fewer than five dots, subjects performed perfectly with response times averaging 0.35 sec. For five or more dots, subjects made errors. Miller (1956) has summarized many studies that show that subjects are limited in their ability to count or identify stimuli along a variety of perceptual dimensions.

On the other hand, we know that the visual system can perform complex pattern-recognition tasks, such as the identification of a face in a crowd, and can work at high speeds, as in reading. How do we reconcile these feats with the severe constraints that the counting tasks appear to reveal?

Subjects can undoubtedly count more than four items accurately if allowed unlimited time. But, is it the prolonged inspection time or the eye movements that is critical to the improved performance? Kowler and Steinman (1977) have shown that accurate counting of arrays of from 7 to 16 dots cannot be achieved without the aid of eye movements. Atkinson *et al.* (1976) used an afterimage technique to show that more than four dots could not be counted accurately in the absence of eye movements, despite inspection times of up to 60 sec.

Here we ask whether pattern recognition, like counting, requires eye movements or whether adequate inspection time *per se* is all that is required for efficient performance.

METHODS

Our stimuli consisted of pairs of $N \times N$ arrays of white and black pixels, displayed on the white screen of a T.V. monitor. See Fig. 1 for a demonstration, but read the caption for instructions before examining the figure. Except where specified, 7×7 square arrays were used. Each array subtended 40 min arc on a side. Each pixel subtended $6' \times 6'$. In one experiment, we varied angular pixel (and array) size over a range of 100 to 1 by varying viewing distance. The white areas of the screen had a luminance of 160 cd/m^2. All experiments were conducted with the corner-to-corner arrangement shown in

*To whom reprint requests should be sent.

Fig. 1, except for one control experiment with side-by-side arrays in which no significant difference in results was found.

White and black values were assigned with equal probability at random to each pixel of one of the arrays. On half of the trials, the second array was identical to the first. On the other half, one black pixel in the second array was selected at random and replaced with a white one. The observer's task was to determine whether the two arrays were identical or different. Most of the errors resulted from the failure of the observer to detect the difference in the arrays. Very few errors (< 10%) were made when the patterns were the same. Subjects heard a buzzing noise following incorrect responses.

In some conditions, subjects were required to maintain careful fixation. Occasionally, they were aware that their eyes had deviated during a trial. By pressing a key, these trials were deleted from the data analysis. Despite this precaution, small, fixational eye movements undoubtedly occurred. The effect of these would be to reduce the difference between our fixation and eye-movement conditions.

Three subjects with corrected-to-normal acuity participated. K.H.S., one of the authors, was highly practiced, having completed several thousand preliminary trials before final data collection. K.B. had much less practice, about 100 trials each with fixation and eye movements. Nevertheless, he showed very similar results to K.H.S. From this, we infer that long-term learning did not play an important role. T.W. was one of the authors and was also highly practiced.

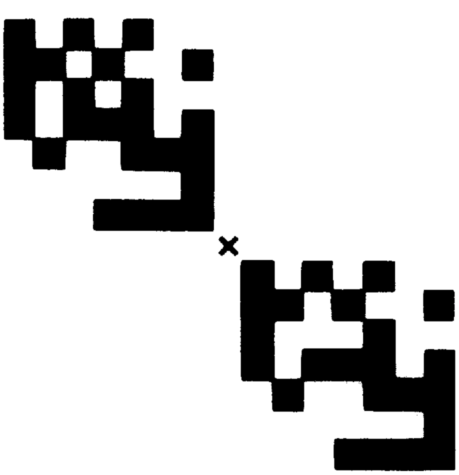

Fig. 1. Example of the stimuli used in the experiments. The two 7 × 7 arrays are identical except for one pixel. Before examining the arrays, fixate the cross between them and try to find the difference. After a few seconds, let your eyes search through the arrays until you find the discrepancy.

RESULTS AND DISCUSSION

Percent correct was obtained in blocks of 50 trials. These percentages were transformed to an equivalent number of pixel comparisons as follows. Let the number of pixels in each array be A. Assume that across a series of trials, the subject examines N pairs of corresponding pixels, one from each array. If the target pair is found within this subgroup of examined pairs, a correct "difference" response is generated. If not, the subject guesses. The proportion correct (P) in the forced-choice procedure would then be given by

$$P = 0.5 + 0.5\,(N/A)$$

and the number of pixel comparisons (N) is

$$N = A(2P - 1).$$

Mean values of N (left ordinate), and corresponding values of percent correct (right ordinate) are plotted in Fig. 2.

Performance was measured as a function of target duration. In the first experiment the observers were instructed to use normal voluntary eye movements to scan the arrays. (Presumably these were saccadic eye movements.) Figure 2(A) shows results for subject K.H.S. Performance for durations up to 3 sec was well fit by a rising straight line. It intersects the vertical axis at a value of 4.5 pixel comparisons. We call this value the Jevons constant. It represents the number of pixel comparisons that can be made at a glance. The slope of the rising straight line was 10.5 pixel comparisons/sec. Since 2 pixels are inspected per comparison, the subject inspected 21 pixels/sec, yielding a search rate of about 50 msec/pixel. This rate corresponds closely to values found by Bergen and Julesz (1983a) and Treisman and Gelade (1980) for serial search tasks.

The theoretical curves in Fig. 2 were derived from two serial search models (Engel, 1977). In the directed search model, a subject searches systematically through the arrays without skipping or repeating any elements. In the random search model, the subject examines corresponding pairs of pixels in random sequence. For a directed-serial-search with rate R, the number of pixel comparisons in time t is Rt. If the Jevons constant J is added, this model predicts

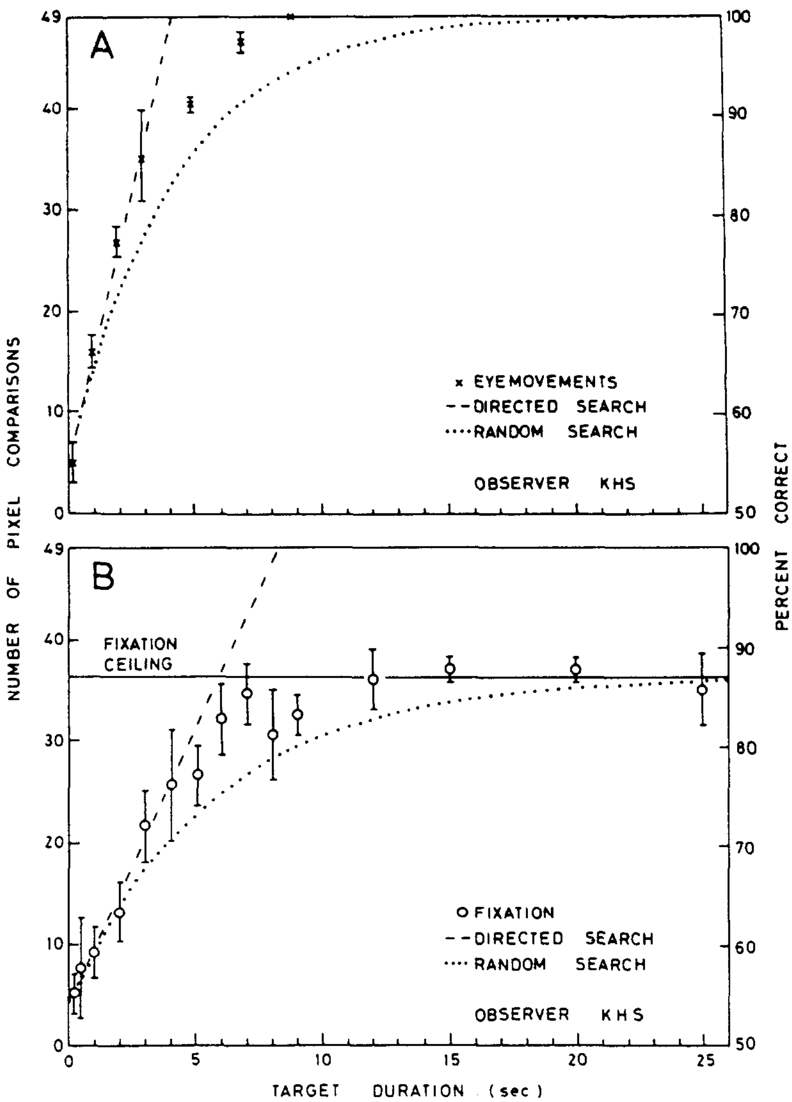

Fig. 2. The number of pixel comparisons (left ordinate) and percent correct (right ordinate) are plotted as a function of target duration. Each point is based on 3–10 blocks of 50 trials each. The data are for observer K.H.S. In (A), he used normal eye movements. In (B), he fixated carefully. Error bars represent ± 1 SE. The theoretical curves represent performance to be expected from two models that are described in the text. In panel (B), the suboptimal ceiling value of 36.5 is taken as the value for A to which the curve asymptotes.

that $J + Rt$ pixel comparisons will be made after time t. In a random-serial-search at rate R through arrays with A elements, the curve exponentially approaches the ceiling value A. For a Jevons constant J, the number of pixel comparisons in time t is $N = J + A'[1 - (1 - 1/A')^{Rt}]$ where $A' = A - J$. With a Jevons constant of 4.5 and rate of 10.5/sec the directed-serial-search reaches its maximum of 49 (7 × 7) after 4.24 sec. Instead, the data in Fig. 2(A) fall below the straight line and only reach the maximum value of 49 after 9 sec. The data, however, lie above the curve predicted by a random-serial-search model. Although subjects felt they used serial search, it is possible that a parallel-processing model could be constructed to account for the data of Fig. 2.

Subject K.B. showed a similar pattern of results. His Jevons constant was 4.5, and his rate was 9 pixel-comparisons/sec. K.H.S. showed a similar pattern of results in an experiment with 9 × 9 arrays: Jevons constant = 4; search rate = 10 comparisons/sec. His performance reached the ceiling of 81 pixels for a duration of 12 sec.

In the second experiment, the conditions were the same, except that the observer was required to maintain steady fixation on a cross lying between the two arrays. (See Fig. 1.) Data for K.H.S. are shown in Fig. 2(B). Once again, the short-duration (up to 3 sec) data can be fit by a straight line. Its intercept is 4, about the same as that for the voluntary eye-movement data. This is to be expected since very short durations don't allow time for eye movements. However, the fixation data differ in two important ways from the eye movement data. First, the rate of search for short durations up to 3 sec is about 5.5 comparisons/sec, almost a factor of two slower than the rate obtained with eye movements. Although some form of search is apparently possible, even in the absence of eye movements (cf. Sperling and Reeves, 1981), it is less effective. Either the search was conducted comparatively slowly, or it was conducted at the same rate but less accurately. Second, unlike the eye-movement data, the fixation data approach a ceiling at long durations that is less than the number of pixels (49) in the array. The ceiling in Fig. 2(B) is 36.5. This is a surprising and puzzling result. Despite very long viewing times, up to 25 sec, performance remains suboptimal. About the same results were obtained with subject K.B. He showed a Jevons constant of 5, a search rate of 6 pixels/sec and a ceiling of 36.5 pixels.

We considered four possible explanations for the differences between fixation and eye-movement results.

One possibility is that the targets fade (Troxler's effect) when fixated, limiting the effective time available for inspection. However, subjects noticed no fading of the arrays. Moreover, in a control experiment, the subjects moved their eyes along a line orthogonal to the axis joining the centres of the arrays so that the images of the two arrays would be in motion on the retina and would not fade. Performance was no better than in the fixation experiment. (We checked that the most remote pixels in the arrays remained within the resolution limit of the eyes when they reached the endpoints of the orthogonal trajectory.) This experiment also shows that the presence or absence of eye movements *per se* does not account for the difference between fixation and eye movement data.

A second possible explanation is that individual pixels were not resolved in peripheral vision during fixation. We tested the effect of pixel size in the fixating condition by conducting 3-sec measurements for viewing distances ranging from 20 cm to 20 m (corresponding to pixels subtending 100 × 100 to 1 × 1 min arc). Three blocks of trials were run at each viewing distance. Mean percent correct and standard deviations are given in Table 1. Performance deteriorated for the 1′ × 1′ pixels at the acuity limit. Elsewhere, performance did not vary systematically with viewing distance. Similar data were collected for 10-sec durations and pixels subtending 2.7 × 2.7 deg and 3.3 × 3.3 min arc. No significant difference was found.

A third possibility is that neighbouring pixels might mask target pixels. This possibility is suggested by the observation that acuity for isolated letters in peripheral vision is substantially greater than acuity for letters flanked by other letters (Bouma, 1970). Moreover, Engel (1974) has shown that detection of a disk target in peripheral vision is impaired when disks that are similar in size or luminance are presented in nearby regions of the field. If such lateral interference exists, we would expect performance to be better when target pixels occupied one of the 24 edge positions than when they occupied one of the 25 interior positions, because on average, edge positions have fewer neighbouring black pixels. We would expect that the effect of lateral interference would be diminished when eye movements are permitted because lateral interference plays a much smaller role in central vision. K.H.S. and T.W. participated in an experiment in which the array locations of discrepant pixels were recorded so that they could be analyzed separately for edge and interior positions. The upper panels of Fig. 3

Table 1. Effect of viewing distance (fixation conditions, 3-sec duration)[a]

	Viewing distance (m)		0.2	0.5	1.0	2.0	5.0	10.0	20.0
	Pixel size (min arc)		100	40	20	10	4	2	1
Subjects	K.H.S.	Mean %	73	66	64	72	73	71	55
		SD	3	6	2	7	5	2	12
	T.W.	Mean%	68	67	70	68	60	64	47
		SD	2	6	5	3	1	2	3

[a]Means and standard deviations are based on measurements of percent correct in 3 blocks of 50 trials.

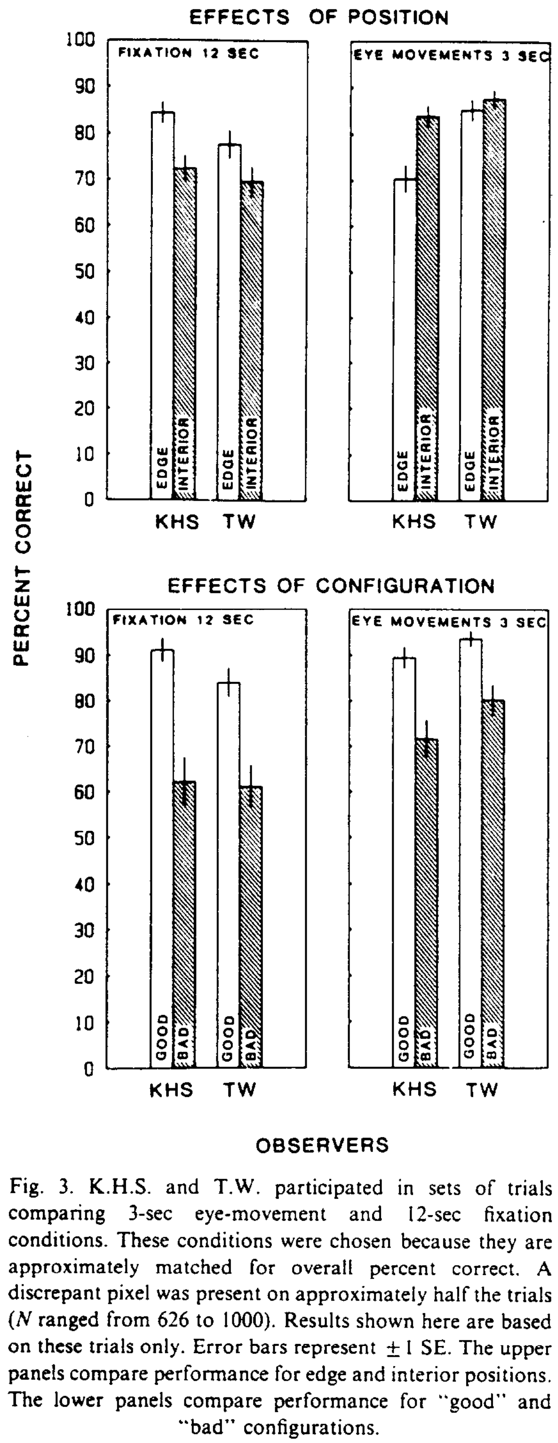

Fig. 3. K.H.S. and T.W. participated in sets of trials comparing 3-sec eye-movement and 12-sec fixation conditions. These conditions were chosen because they are approximately matched for overall percent correct. A discrepant pixel was present on approximately half the trials (N ranged from 626 to 1000). Results shown here are based on these trials only. Error bars represent ± 1 SE. The upper panels compare performance for edge and interior positions. The lower panels compare performance for "good" and "bad" configurations.

summarize the results. For the fixating eye, performance was better when the missing pixel was positioned on an edge rather than the interior. This was not true for eye movements. These findings suggest that lateral interference limits performance in peripheral vision, not only for letter acuity but for pattern analysis more generally.

Finally, we considered the possibility that some configurations of pixels might be easier for peripheral vision to analyze than others. Beck and Ambler (1973) have shown that a tilted "T" is more easily found than an "L" in an array of peripherally viewed vertical "T"'s. Banks and Prinzmetal (1976) showed that a target letter can be made more difficult to detect by including it as part of a perceptual grouping. The Gestalt psychologists were the first to explore the idea of figural goodness. Although their laws usually lacked predictive power, their demonstrations showed convincingly that some patterns are perceived more easily or remembered better than others. Garner and colleagues have attempted to quantify figural goodness in terms of the symmetry of patterns (cf. Garner, 1974, Chap. 1). Consider the set of eight patterns that can be derived from a given pattern by rotations of 0, 90, 180 and 270 deg, and by reflections about vertical, horizontal and two diagonal axes. Garner's hypothesis is that the number of patterns in the set of eight that are identical to the original pattern provides a measure of figural goodness. Palmer (1982) has shown that the figural goodness of local regions of complex patterns can also be described in terms of their symmetry properties.

We asked whether Garner's hypothesis could be used to predict when discrepant pixels are easy or difficult to detect in our paradigm. Since each pixel has eight adjacent neighbours, four located at the sides and four located at the corners, there are $2^8 = 256$ "local" configurations. We examined the rotation and reflection symmetries of each of these configurations. Only four of the 256 configurations have eight-fold symmetry, that is, the original pattern is the same for each of Garner's eight transformations. Eight of the 256 configurations have four-fold symmetry, 100 have two-fold symmetry, and 144 were different from the original pattern for all of the eight transformations. We grouped the 256 configurations into these four categories to see whether performance was better for the more symmetric patterns. Using the same data in which position effects were examined, we computed percent correct for each of the four symmetry categories. There were no significant differences across categories for either fixation or eye-movement conditions. Apparently, Garner's measure of figural goodness does not predict the results of our experiment.

However, our observers reported that the target pixel seemed easier to detect when it was missing from a solid region of black such as a row or column. It seemed harder to detect in the presence of black pixels located at the corners or in the absence of any neighbouring black pixels. To evaluate these reports, we divided the set of 256 local configurations into three categories: (1) "good" configurations, in which two or more black pixels were located on the sides, and in which the number of these exceeded the number of black pixels located at the corners, (2) "bad" configurations, in which the number of pixels located at the corners exceeded the number of black pixels located on the sides, and (3) all the rest. Figure 1 shows a "bad" configuration in which there are three black pixels on corners and only one black pixel on a side. Once again, using the data in which position effects were examined, we analyzed the results according to category. The results are shown in the lower panels of Fig. 3. There is a clear difference; performance was substantially better for the "good" configurations than for the "bad" ones. The difference was more pronounced for fixation than for eye movements. Moreover, configuration appeared to play a greater role than position in determining performance (compare upper and lower panels in Fig. 3).

These findings suggest that the inferior performance during fixation is due to both positional and configurational effects. Targets located in "bad" configurations or in the interior of large arrays are hardest to detect, particularly during fixation. The idea that certain pattern features are more easily processed than others has recently been developed in some detail as part of Julesz's texton theory (see Bergen and Julesz, 1983a, b) and in the feature-integration theory (Treisman and Gelade, 1980).

Our findings suggest that the advantages of eye movements in pattern analysis are related to superior capacities of central vision above and beyond acuity. The argument rests on the assumption that eye movements bring portions of patterns to central vision where critical discriminations are made. Because we did not measure eye movements in our experiments, it is logically possible that subjects moved their eyes but did not use central vision. However, studies of visual search in which eye movements were recorded have shown that targets must be brought within some critical angular distance of a point of fixation before recognition is possible. The critical distance depends on characteristics of the target and background patterns and upon their similarity (Engel, 1977; Kundel and Nodine, 1978; Prinz and Kehrer, 1982).

The superiority of central vision may be sensory in origin. It may be related to the increased density or overlap of receptive fields at some level in the visual pathway (cf. Levi and Klein, 1985). Alternatively, attentional processes may underlie the difference between central and peripheral vision. Beck and Ambler (1973) showed that certain pattern discriminations were possible in peripheral vision only when explicit information was available to guide focal attention. Perhaps, in the absence of such information, as in our experiments, focal attention cannot be employed for systematic search through a complex pattern.

Whatever the underlying cause, our results suggest that eye movements play a crucial role in pattern analysis that has nothing to do with resolution. When eye movements are prohibited, the observer must rely on slow and imperfect pattern analysis away from the point of fixation. When eye movements are permitted, more rapid and efficient processing is based on sequential comparisons in central vision.

Acknowledgements—This research was supported in part by U.S. Public Health Service Grants EY02857 and EY02934 to Gordon E. Legge. K.-H. Schlingensiepen was supported by the Studienstiftung des deutschen Volkes, the DAAD and the Max-Planck-Gesellschaft. Timothy D. Walker was supported by the Center for Research in Human Learning. Some of the apparatus was provided by the Medical Research Council and the Wellcome Trust. We thank Clive Hood, Andrew Luebker and Gary Rubin for technical help, Anne Griffin for typing and Fiona Hake for artwork. We also thank anonymous referees for their particularly thorough comments.

REFERENCES

Atkinson J., Campbell F. W. and Francis M. R. (1976) The magic number 4 ± 0: A new look at visual numerosity judgments. *Perception* **5**, 327–334.

Banks W. P. and Prinzmetal W. (1976) Configurational effects in visual information processing. *Percept. Psychophys.* **19**, 361–367.

Beck J. and Ambler B. (1973) The effects of concentrated and distributed attention on peripheral acuity. *Percept. Psychophys.* **14**, 225–230.

Bergen J. R. and Julesz B. (1983a) Parallel vs. serial processing in rapid pattern discrimination. *Nature, Lond.* **303**, 696–698.

Bergen J. R. and Julesz B. (1983b) Rapid discrimination of visual patterns. *IEEE Trans. Systems, Man Cybernet* **SMC-13**, 857–863.

Bouma H. (1970) Interaction effects in parafoveal letter recognition. *Nature, Lond.* **226**, 177–178.

Engel F. L. (1974) Visual conspicuity and selective background interference in eccentric vision. *Vision Res.* **14,** 459–471.

Engel F. L. (1977) Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Res.* **17,** 95–108.

Garner W. R. (1974) *The Processing of Information and Structure.* Lawrence Erlbaum, Hillsdale, N.J.

Jevons W. S. (1871) The power of numerical discrimination. *Nature, Lond.* **3,** 281–282.

Kowler E. and Steinman R. M. (1977) The role of small saccades in counting. *Vision Res.* **17,** 141–146.

Kundel H. L. and Nodine C. F. (1978) Studies of eye movements and visual search in radiology. In *Eye Movements and the Higher Psychological Functions* (Edited by Senders J. W., Fisher D. F. and Monty R. A.), pp. 317–328. Lawrence Erlbaum, Hillsdale, N.J.

Levi D. M. and Klein S. A. (1985) Vernier acuity, crowding and amblyopia. *Vision Res.* **25,** 979–989.

Miller G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **63,** 81–97.

Palmer S. E. (1982) Symmetry, transformation and the structure of perceptual systems. In *Organization and Representation in Perception* (Edited by Beck J.), pp. 95–144. Lawrence Erlbaum, Hillsdale, N.J.

Prinz W. and Kehrer L. (1982) Recording detection distances in continuous visual search. In *Cognition and Eye Movements* (Edited by Groner R. and Fraisse P.), pp. 48–56. Elsevier, North-Holland, Amsterdam.

Sperling G. and Reeves A. (1981) Measuring the reaction time of a shift of visual attention. In *Attention and Performance VIII* (Edited by Nickerson R.), pp. 347–360. Lawrence Erlbaum, Hillsdale, N.J.

Treisman A. M. and Gelade G. (1980) A feature-integration theory of attention. *Cogn. Psychol.* **12,** 97–136.