Memory & Cognition
1984, 12 (6), 607-612

Enumeration of dots: An eye movement analysis

MICHIEL P. VAN OEFFELEN and PETER G. VOS
University of Nijmegen, Nijmegen, The Netherlands

The present study reports the measurement of response latencies and the recording of eye movements in a task in which adults had to enumerate dots in figures that differed in number of dots ($n_d = 19$-23) and grouping of dots. The functional relationship between latencies per dot and mean group size was in agreement with earlier findings (van Oeffelen & Vos, 1982). Temporal information from eye movement data indicated that the relative contribution of fixation durations to overall latency was far larger than the contribution of saccades, which superseded the contribution from eyeblinks. Spatial information in the form of eye movement trajectories indicated that, in general, there occurred one or two fixations at the starting position. From this position onward, eye movements were directed toward areas of dots rather than to each dot in particular. Scanning behavior was sometimes reiterative, in the sense that groups of dots were visited more than once. The results are discussed with respect to the nature of strategies employed during a dot-enumeration task.

In a previous study, van Oeffelen and Vos (1982) chronometrically investigated the interactive effect of number of dots and their pattern upon the processing of visual numerosity. With dot figures consisting of 14-23 dots, it was found that the dots were not counted one by one, but in groups of 2 or more dots. More precisely, when the number of dots within a proximity-related (sub)group of dots did not exceed 5, the number was established by subitizing, a very rapid and accurate perceptual process (Kaufman, Lord, Reese, & Volkman, 1949), the numerical result being transiently stored for further processing. The analysis of reaction times (RTs) for those stimuli indeed showed that the main contributor to overall latencies was the time needed to sum up the various partial results of subitizing. However, when a stimulus field of dots could not be segmented into small groups on the basis of proximity cues, the strategy of numerosity processing was not as clear. In discussing the data, van Oeffelen and Vos concluded that the most plausible strategy was counting in twos and threes, rather than that of counting one by one. Yet, how subjective grouping of twos and threes took place in the absence of proximity cues remained unclear.

One way to acquire an objective picture of the perceiver's strategy of subjective grouping and the counting strategies based thereon is to analyze visual scanning patterns, which are considered as overt behavioral correlates of ongoing internal processes. Since it is the function of the eye to gather information, it seems reasonable to assume that the eye generally is directed toward regions of space that contain the most information. Thus, with respect to the counting process under study, we expect scanning trajectories to show relatively large saccades between subitizable groups of dots and only a few fixations located at successive positions within a group, whereas scanning trajectories are expected to show many fixations and small saccades for dot figures consisting of large groups of dots. In addition, eye movement trajectories could demonstrate whether subjects had restarted counting somewhere during the task. It was believed (Jensen, Reese, & Reese, 1950; Klahr, 1973) that this latter phenomenon was responsible for the slightly positively accelerated function for RT versus number of dots.

The present study reports the measurement of response latencies and the recording of eye movements in a dot-enumeration task. Eye movements were recorded using the pupil-center corneal-reflection method.

METHOD

Subjects

Seven undergraduate psychology students at the University of Nijmegen (five males and two females) were paid to participate in the experiment. All subjects were naive with respect to the experimental task.

Stimuli

Thirty-seven dot figures were constructed. Each figure differed both in number ($n_d = 19$-23) and in arrangement of dots. Seven different arrangements were used, except for $n_d = 22$ and $n_d = 23$, for which there were eight arrangements. For each n_d, there was one configuration consisting of one large group, one configuration of two groups, and so on, up to one configuration properly segmented into seven (or eight for $n_d = 22$ and $n_d = 23$) distinctly different groups. Care was taken that different groups within one configuration contained about the same number of dots. Objective criteria for grouping dots within a dot figure were established by CODE, a cluster algorithm the purpose of which is to formalize

---

This research was supported by a grant to P.G. Vos from the Netherlands Organization for the Advancement of Pure Research (Zuiver Wetenschappelijk Onderzoek). The authors wish to thank L. H. Bavinck for his helpful comments on this manuscript. The authors' mailing address is: Psychology Department, K.U.N., Postbus 9104, 6500 HE Nijmegen, Nijmegen, The Netherlands.

the Gestalt rule of relative proximity (van Oeffelen & Vos, 1982, 1983). All dot figures were subjected to the algorithm, which yielded a description of their groupability in terms of contours around groups of dots. Figure 1 depicts seven of the stimuli with the same number of dots ($n_d = 21$) but with different configurations. Applying CODE to the dot figures resulted in perceptually relevant boundaries (contours). These contours are also drawn in Figure 1.

Procedure

All subjects were tested individually in a quiet laboratory room that contained the complete Whittaker 1998-S Eye View Monitor and TV-Pupillometer System (EVM), stimulus presentation screen, and computer monitor, both connected to a PDP-11/45 computer system situated in a neighboring room. Figure 2 shows a schematic representation of the experimental setup. The illumination in the room was dimmed during the actual experimental session. A subject was seated in an adjustable chair, and her/his head was held steady by a headstand with headrests in the back. The stimuli were presented on a 35 × 35 cm projection screen (Vector General). The subject viewed the screen at eye level at approximately 1 m (visual angle about 20°) while a TV camera photographed the subject's left eye. In this way, reflections were recorded from an infrared (IR) source light that was continuously directed at the eye. A filter on the IR light absorbed thermal radiation, which could have been harmful to the eye. The subject's eye rotation, and, consequently, the point of fixation, was determined by measuring the center of the pupil with respect to the center of the corneal reflection. Because the center of the pupil and the center of the corneal reflection moved together with eye rotation, the difference between their positions was indicative of the eye's point of fixation. Thus, the eye position was independent of the head position as long as the pupil image was contained within the field of view of the TV camera. The continuous flow of eye-position information was presented as a spot superimposed on the video monitor scene available to the experimenter. The digitized output of the microcomputer was passed on to the PDP-11/45 computer. Eye position was calculated in terms of horizontal and vertical coordinates in the EVM-system representation. A third output was the pupil diameter measured in numbers of scan lines that intersected the image of the pupil on the experimenter's TV-monitor screen. These three signals were delivered at a rate of 50 data points per second.

At the beginning of each experimental session, a calibration procedure was started to match the EVM-coordinate system with that of the field of stimulus presentation. Calibration trials consisted of a subject's fixating each position of a grid of nine calibration points. The nine points were situated in a 3 × 3 matrix that covered almost the entire presentation screen.

Once the calibration had been carried out and the calibration measures had been stored in a data file together with the subject information, the experiment started. The subject had to fixate a point at the upper left corner of the screen before the presentation of each scene. Presentation order of the stimuli was random. The subject was instructed to attend to the number of dots in each stimulus and to report this number orally. The task was self-paced; a stimulus appeared on the screen immediately after the participant had pushed a button. The stimulus remained visible until the response, mediated by a microphone (Sennheiser headset), had surpassed a previously selected critical level. Latencies were registered automatically. In the mean time, EVM data were gathered and stored in a data file. The experimenter, who had a list of stimulus specifications, scored each response according to whether it was correct or incorrect and then stored it in the computer. Whenever a subject committed an error, she/he received immediate feedback ("Wrong") from the experimenter. Each subject completed the session within .5 h.

RESULTS

Not more than 5% of the 259 responses appeared to be errors; these were discarded from further analysis.

To begin with, the mean and standard deviation of the measured latencies were computed for each stimulus and over all subjects. The standard deviations were on the order of 8%-15% of the means. According to earlier theory (van Oeffelen & Vos, 1982), response latencies for correct number responses should be a function of both n_d, the number of dots, and n_g, the number of groups; the relation was formally expressed as follows:

$$RT(n_d,n_g) = b_0 + n_d \times b_1 + (n_g - 1) \times b_2 + b_3(n_d,n_g), \quad (1)$$

where b_0 represents global perception and motor response time, b_1 stands for the time per dot consumed by the subitizing process for naming dots within small groups, b_2 stands for the time to switch from one group to another and for computing the running sum, and $b_3(n_d,n_g)$ stands for the time to segment large proximity-related groups into smaller subgroups *during* the counting process. In the situation in which, after the initial global inspection, the stimulus field of dots can easily be divided into n_g groups, each of subitizable size, the term $b_3(n_d,n_g)$ can be omitted. In that case, n_g, the number of groups, equals $n_d/<g_n>$, where $<g_n>$ is the mean of dots per group. Dividing RT by n_d gives us the RT per single dot:

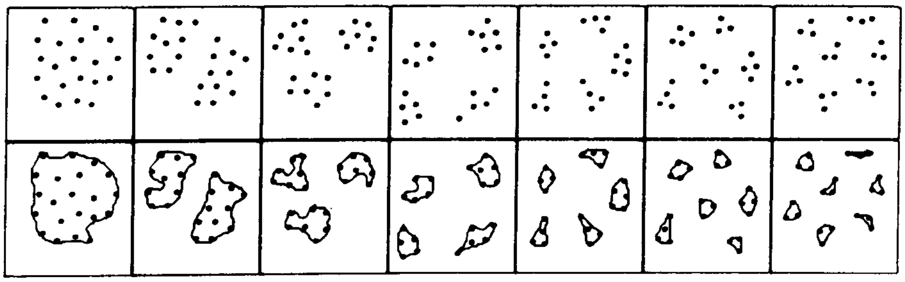

Figure 1. Seven of the stimuli with the same number of dots (n_d = 21) but with different configurations. The perceptually relevant boundaries (contours) that were the result of CODE applied to the dots are also shown.

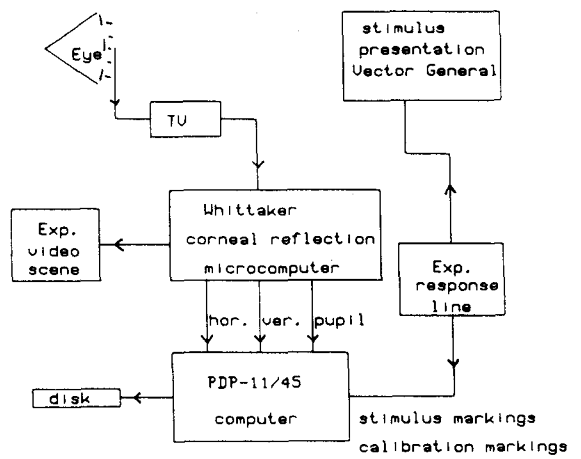

Figure 2. Schematic representation of the experimental setup.

$$RT(n_d, \langle g_n \rangle)/n_d$$
$$= b_1 + [(b_0 - b_2)/n_d] + b_2/\langle g_n \rangle, \qquad (2)$$

where $\langle g_n \rangle$ is small. Van Oeffelen and Vos (1982) argued that the contribution from the second term can be ignored in cases when n_d is much larger than $\langle g_n \rangle$. Hence, we may approximate Equation 2 by an expression that is a function of $\langle g_n \rangle$ only:

$$RT\langle g_n \rangle/n_d = b_1' + b_2/\langle g_n \rangle, \qquad (3)$$

where b_1' is a constant primarily representing timing aspects attributed to the subitizing process. Equation 3 says that, for dot figures containing groups of subitizable size only, RT/n_d is a hyperbolic function of mean group size.

In the case in which the stimulus field of dots could not easily be divided into subitizable groups, it was assumed (van Oeffelen & Vos, 1982) that subjects perform some segmentation of large proximity-related groups during the counting process. However, we do not know how large those subgroups are, and therefore we do not know how many subgroups there are. The same authors argued that, in this case, response latency is a purely linear function of n_d, as follows:

$$RT(n_d)/n_d = A\langle g_n \rangle + B, \qquad (4)$$

where $\langle g_n \rangle$ is large, and where B primarily represents motor response time and A is attributed to temporal aspects of grouping, subitizing, and adding operations.

With respect to the chronometric analysis of the present data, the mean latencies were divided by the corresponding values of n_d. In this way, the overall means of processing time per dot were obtained. Figure 3 presents the mean latencies per dot ($\langle RT \rangle_{tot}/n_d$) as a function of mean group size, $\langle g_n \rangle$. The successive curve fittings applied to the mean latencies per dot yielded the following results: $\langle RT \rangle_{tot}/n_d = 368.4/\langle g_n \rangle + 149.3$, $r = .711$, for $\langle g_n \rangle \leq 5$, and $\langle RT \rangle_{tot}/n_d = 3.4\langle g_n \rangle + 272.5$, $r = .953$, for $\langle g_n \rangle \geq 6$.

The next step in the analysis of the experimental results concerned the eye movement recordings. To achieve visual scanning patterns in terms of fixations and saccades, the raw EVM data were subjected to a cluster algorithm. The algorithm was a slightly modified version of the one developed by Spaninks (1978). The algorithm yielded listings of positions and durations of fixations and saccades, and periods of disturbances (primarily eyeblinks). In the analysis of the temporal information, mean durations of fixations, $\langle RT \rangle_{fix}$, saccades, $\langle RT \rangle_{sac}$, and eyeblinks, $\langle RT \rangle_{bli}$, were determined over all subjects. According to Equations 3 and 4, curve fittings to the partial data were applied, yielding the following results: fixations: $\langle RT \rangle_{fix}/n_d = 185.5/\langle g_n \rangle + 130.6$, $r = .500$, for $\langle g_n \rangle \leq 5$, $\langle RT \rangle_{fix}/n_d = 3.0\langle g_n \rangle + 196.5$, $r = .888$, for $\langle g_n \rangle \geq 6$; saccades: $\langle RT \rangle_{sac}/n_d = 128.7/\langle g_n \rangle + 16.2$, $r = .744$, for $\langle g_n \rangle \leq 5$, $\langle RT \rangle_{sac}/n_d = .8\langle g_n \rangle + 54.0$, $r = .614$, for $\langle g_n \rangle \geq 6$; and eye blinks: $\langle RT \rangle_{bli}/n_d = 54.2/\langle g_n \rangle + 2.4$, $r = .281$, for $\langle g_n \rangle \leq 5$, $\langle RT \rangle_{bli}/n_d = -.3\langle g_n \rangle + 20.7$, $r = -.358$, for $\langle g_n \rangle \geq 6$.

From the fact that $RT_{tot} = RT_{fix} + RT_{sac} + RT_{bli}$, it is evident that, for both conditions of $\langle g_n \rangle$, addition of the partial-fit results should satisfy the curve fitting to the overall results. In Figure 3 are plotted the mean latencies per dot, $\langle RT \rangle_{tot}/n_d$, and the mean contributions from fixations, $\langle RT \rangle_{fix}/n_d$, saccades, $\langle RT \rangle_{sac}/n_d$, and eyeblinks, $\langle RT \rangle_{bli}/n_d$, all as a function of mean group size, $\langle g_n \rangle$. In addition, the best-fitting hyperbolic curves for $\langle g_n \rangle \leq 5$ and best-fitting straight lines for $\langle g_n \rangle \geq 6$ are drawn for both the overall results, $\langle RT \rangle_{tot}/n_d$, and for the partial results, $\langle RT \rangle_{fix}/n_d$, $\langle RT \rangle_{sac}/n_d$, and $\langle RT \rangle_{bli}/n_d$. From Figure 3, one can see that the durations of fixations exceeded those of saccades to a considerable degree, whereas durations of eyeblinks contributed little to the overall latency. Based upon the successive curve fittings, Figure 4 depicts the *relative* contributions of durations of fixations, saccades, and eyeblinks to overall latency (i.e., the percentage of RT_{tot}).

Finally, spatial information from the eye movement data was considered. With the use of appropriate calibration parameters, positions of fixations and saccades in the EVM-system representation were transformed into positions in the stimulus-presentation system. Figure 5 represents some eye movement trajectories that were typically recorded during the counting task. Inspection of the trajectories revealed the following regularities. Corresponding to the prescribed starting position, the subject's eye position at the moment of stimulus onset was fixed at the upper left position of the stimulus field. Immediately after stimulus onset, the eye barely moved to other positions; instead, one or two eye fixations were situated near the starting point. A mean duration of 305 ± 45 msec was found, a value that was independent of

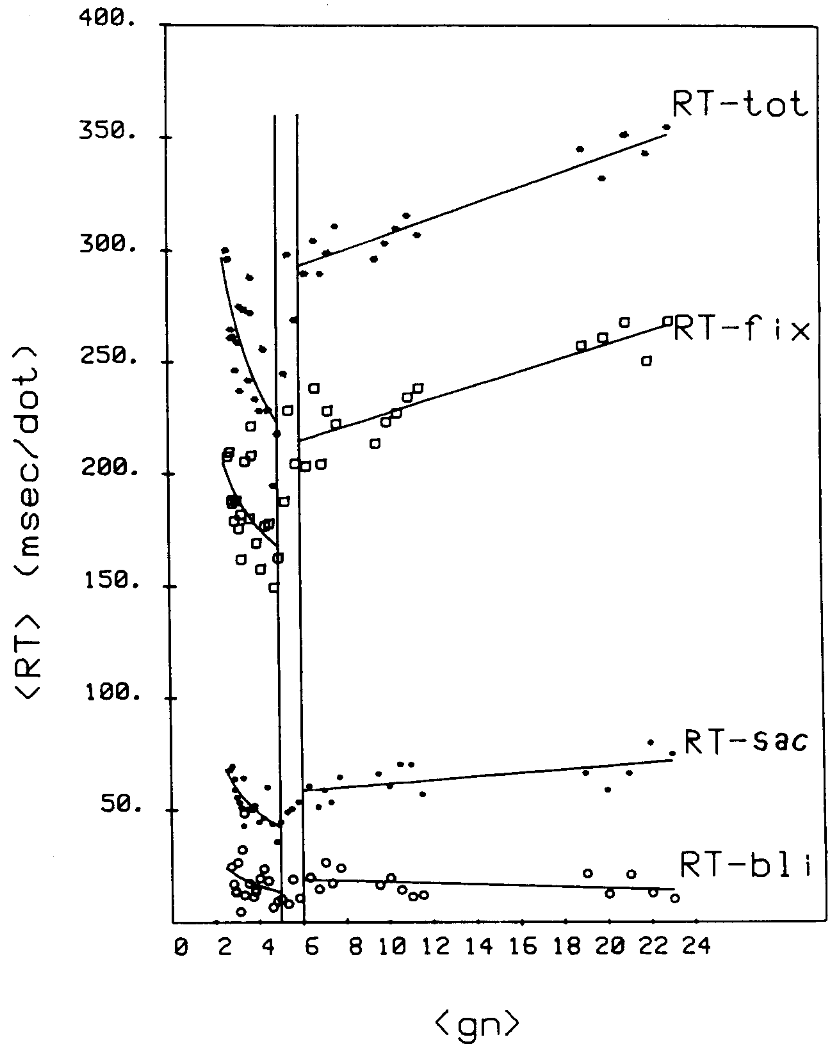

Figure 3. The mean latencies per dot, $<RT>_{tot}/n_d$, and the mean contributions from fixations, $<RT>_{fix}/n_d$, saccades, $<RT>_{sac}/n_d$, and eyeblinks, $<RT>_{bli}/n_d$, are plotted as a function of mean group size, $<g_n>$. In addition, the best-fitting hyperbolic curves for $<g_n> \leq 5$ and best-fitting straight lines for $<g_n> \geq 6$ are drawn for both the overall results and the partial results.

the number of dots and of the number of groups. From the starting position onward, eye movements were guided toward dots, or groups of dots that were relatively near. Only in a few cases (6%) were movements directed to the midpoint of the stimulus figure. In general, it was found that the scanning of the figure proceeded in a clockwise fashion. Not only with small groups, $<g_n> \leq 5$, but also with large groups, fixations were directed to areas of dots rather than to individual dots. With small groups, trajectories sometimes (11%) showed iterative scan behavior, in the sense that an already scanned group of dots was revisited a second or even a third time. This behavior was not found with trajectories on large groups.

DISCUSSION

The chronometric analysis of the overall latencies yielded results that were in agreement with earlier findings (van Oeffelen & Vos, 1982): The functional relationship between latency per dot, ($<RT>_{tot}/n_d$), and mean group size, $<g_n>$, appeared to be hyperbolic when $<g_n> \leq 5$ and linear when $<g_n> \geq 6$.

From Figure 3 and Figure 4 it can be seen that the contribution of fixation durations to total latency exceeded that of durations of saccades, whereas durations of eyeblinks contributed little to the overall latency. Under the condition of $<g_n> \leq 5$, goodness of fit measures showed that durations of saccades satisfied a hyperbolic trend better than durations of fixations. The probable reason is that, during saccades, the cognitive operations of adding partial results can be carried out in parallel with the guidance of the eye movements, hence supporting a hyperbolic trend. Within a fixation duration, extra time is demanded to abstract information from the stimulus field and to preguide subsequent eye movements. The extra time might be responsible for disturbing a hyperbolic

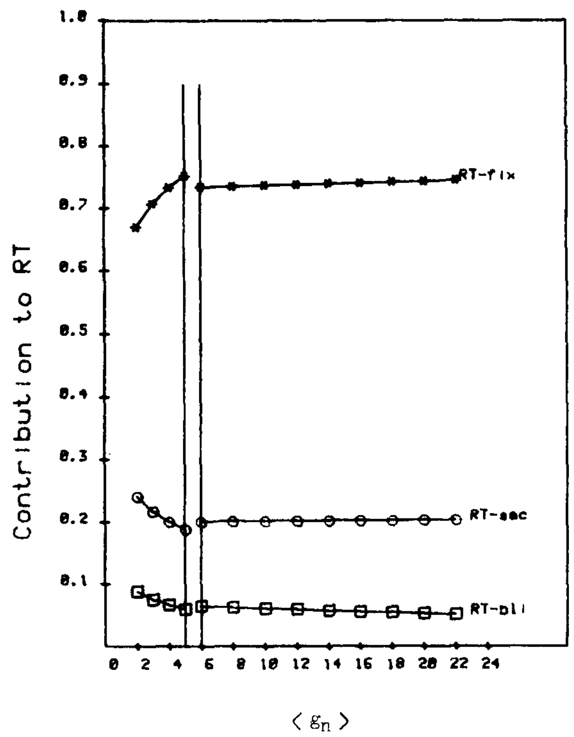

Figure 4. Relative contributions of durations of fixations, saccades, and eyeblinks to overall RT, as a function of mean group size (g_n).

trend. A noisy occurrence of eyeblinks probably is responsible for the low goodness of fit in question.

It is interesting to observe from Figure 4 that the relative contribution of fixations to overall RT continuously increases from 67%, when $<g_n> = 2$, to 75%, when $<g_n> = 5$. The contribution remains at that value for larger group sizes. The reason for this effect probably has to do with increasing amounts of information that have to be processed instantaneously. It is known (Mackworth & Bruner, 1970) that fixation duration is partly dependent upon the amount of information to be processed. With increasing group sizes, the duration of each particular fixation is enlarged and reaches its maximum value when $<g_n> = 5$. A further increase of group size, $<g_n>$, therefore, could affect only the number of fixations and not their particular durations.

As has already been mentioned, a positive increase in RT was believed to result from an increase in the number of restarts. Evidence against the proposition of an increase in the number of restarts comes from inspection of the eye movement trajectories. Neither with an increase in n_d nor with an increase in $<g_n>$ was there an increase in the number of restarts. A possible explanation for the increase in RT is that, with a larger number of dots, peripheral information progressively is needed to discriminate between what has and has not been counted. Hence, more fixations are needed.

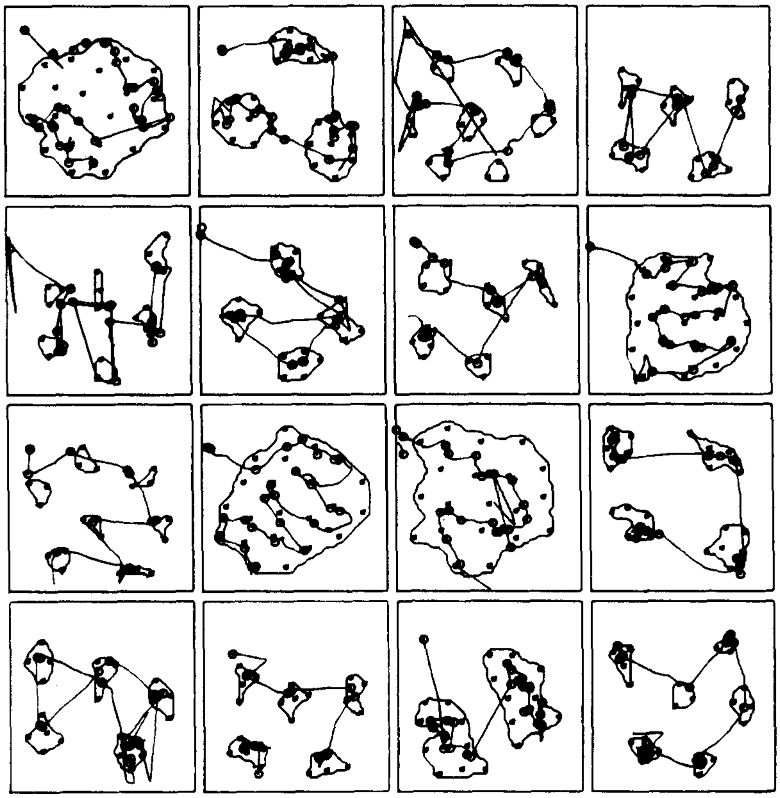

Figure 5. Some eye movement trajectories that were typically recorded during the counting task.

We now come to the question of why one or two fixations were needed before scanning could begin. Probably, the subject initially is structuring the stimulus field to inventory the task-relevant problems. From a perceptual point of view, the subject is building up a "primal sketch" (Marr, 1975) of a stimulus, which yields a global picture of the arrangement of dots. From such a primal sketch, a subject could decide which strategy she/he should follow. The question then arises of why the initial fixations were not directed to the "center of gravity" of the dot figure. Such a result has been found (van Oeffelen & Vos, 1984) in an experiment in which children of about 5 years of age had to count a small number of dots. In 60% of the trials, the children's eye movements were directed to the midpoint of the stimulus field immediately after stimulus onset. The difference in this respect between adults and children is probably due to the fact that adults' peripheral vision provides sufficient information to structure the visual field. Moreover, in order to minimize mnemonic difficulties due to lateral inference of what has and has not been counted, it seems easier to start at a well-defined beginning position (upper left part of the screen). It could also explain why further counting occurs in clockwise fashion.

It should be interesting to investigate whether the eye is directed to the centers of gravity of the various groups of dots. Will the eye jump to the centers of gravity of, for instance, a triangular or a quadrilateral (sub)figure? Some evidence for this was found by Findlay (1980), who used a scanning task with only two differently colored spots to be scanned in one or the other prescribed order. He found that, as a rule, the eye moved to the midpoint between the two targets before performing the prescribed scanning path. Findley interpreted the jumping to the midpoint as an indication of an initial, elementary processing of the global characteristics of the stimulus.

At this point, some criticism must be made regarding the usefulness of the registration of eye movements. First, cognitive operations as part of a specific counting strategy can, in principle, be carried out with eyes closed. The continuous flow of eye movement information does not provide us with the precise moments of cognitive action. Second, records of eye movements can show only the succession of eye fixations: They cannot show precisely what information is being processed at each moment. Third, the point of fixation need not match the center of the field of attention. It is, for instance, entirely possible to shift attention to points in the periphery of the visual field without actually moving the eyes. All points of criticism are closely related and concern the problem of the status of attention. In addition to other methods, such as the measurement of latencies and threshold values, the registration of eye movements is one more method for clarifying this problem further.

REFERENCES

FINDLAY, J. M. (1980). The visual stimulus for saccadic eye movements in human observers. *Perception*, **4**, 361-377.

JENSEN, E. M., REESE, E. P., & REESE, T. W. (1950). The subitizing and counting of visually presented fields of dots. *Journal of Psychology*, **30**, 363-392.

KAUFMAN, E. L., LORD, M. E., REESE, T. W., & VOLKMAN, J. (1949). The discrimination of visual number. *American Journal of Psychology*, **62**, 498-525.

KLAHR, D. A. (1973). A production system for counting, subitizing, and adding. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

MACKWORTH, N. H., & BRUNER, J. S. (1970). How adults and children search to recognize pictures. *Human Development*, **13**, 149-177.

MARR, D. (1975). *Analyzing natural images, a computational theory of vision* (AI Memo No. 334). Cambridge, MA: M.I.T., Artificial Intelligence Laboratory.

VAN OEFFELEN, M. P., & VOS, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition*, **10**, 396-404.

VAN OEFFELEN, M. P., & VOS, P. G. (1983). An algorithm for the description of relative proximity. *Pattern Recognition*, **16**, 341-348.

VAN OEFFELEN, M. P. & VOS, P. G. (1984). The young child's processing of dot patterns: A chronometric and eye movement analysis. *International Journal of Behavioral Development*, **7**, 53-66.

SPANINKS, H. (1978). *Analysis of eye movement recordings* (Internal Report). Tilburg: Katholieke Hoge School.

(Manuscript received December 5, 1983;
revision accepted for publication September 5, 1984.)