Head and eye movements and the role of memory limitations in a visual search paradigm

Gregor Hardiess  
Sabine Gillner  
Hanspeter A. Mallot

The image information guiding visual behavior is acquired and maintained in an interplay of gaze shifts and visual short-term memory (VSTM). If storage capacity of VSTM is exhausted, gaze shifts can be used to regain information not currently represented in memory. By varying the separation between relevant image regions, S. Inamdar and M. Pomplun (2003) demonstrated a trade-off between VSTM storage and gaze shifts, which were performed as pure eye movements, that is, without a head movement component. Here we extend this paradigm to larger gaze shifts involving both eye and head movements. We use a comparative visual search paradigm with two relevant image regions and region separation as independent variable. Image regions were defined by two cupboards displaying colored geometrical objects in roughly equal arrangements. Subjects were asked to find differences in the arrangement of the objects in the two cupboards. Cupboard separation was varied between 30° and 120°. Images were presented with two projectors on a 150° × 70° curved screen. Head and eye movements were simultaneously recorded with an ART head tracker and an ASL mobile eye tracker, respectively. In the large separation conditions, the number of gaze shifts between the two cupboards was reduced, while fixation duration increased. Furthermore, the head movement proportions negatively correlated with the number of gaze shifts and positively correlated with fixation duration. We conclude that the visual system uses increased VSTM involvement to avoid gaze movements and in particular movements of the head. Scan path analysis revealed two subject-specific strategies (encode left, compare right, and vice versa), which were consistently used in all separation conditions.

Keywords: working memory, VSTM, eye movements, head movements, gaze, comparative visual search


Introduction

Visual processing of large-field scenes is performed in an interplay of gaze shifts (gaze = eye-in-head + head-in-space) and memorization of local visual stimuli in a visual short-term memory (VSTM). VSTM is considered a subdivision within the visual part of the working memory (Baddeley, 1978, 1992, 2003) and characterized by a limited storage capacity (Phillips, 1974). The upper limit of this storage capacity was identified as three to four objects, independently of the number of feature dimensions (e.g., color, shape, or orientation) probed for each object (e.g., Irwin & Andrews, 1996; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). In addition, the total information capacity of VSTM is also limited (Alvarez & Cavanagh, 2004; Xu, 2002). Alvarez and Cavanagh (2004) suggested that there is an upper limit on storage that is set in terms of the total amount of information and that there is a functional dependence between the complexity of the objects and the total number of objects that can be stored. Limits in the storage capacity of the VSTM could be impressively shown by research on change detection, revealing the phenomenon of change blindness (Simons, 2000; Simons & Levin, 1997). In these experiments, two almost identical images are presented in alternation with short blanks separating them in time. Subjects are strikingly insensitive even to large changes between both images. For the case of eye movements, Ballard, Hayhoe, and Pelz (1995) and Hayhoe, Bensinger, and Ballard (1998) investigated VSTM usage in a block-copying task. During copying, a presented pattern of colored blocks from a model area to the workspace, participants made only minimal use of VSTM. The authors phrased this as “just-in-time” processing strategy, where observers acquire the specific information they need just at the point at which it is required in the task. This makes sense because the information is easily accessible by the sensors—the eyes. This “just-in-time” processing strategy could also be observed in everyday activities (e.g., Hayhoe, Ballard, Triesch, & Shinoda, 2002; Land, Mennie, & Rusted, 1999). Ballard et al. and Hayhoe et al. showed that this strategy cannot be attributed to hard capacity...
limitations. When the distance between the model and the workspace area was increased (70°), forcing subjects to perform larger eye and head movements, the frequency of the “memoryless” pattern decreased gradually and eye movements occurred about half as frequently (Ballard et al., 1995). The trade-off between memory load and saccadic behavior thus seems to be controlled by the minimization of combined “costs.” On the memory side, such costs could be associated with the amount of stored information, whereas on the eye and head movement side, costs may arise from duration, energy consumptions, or an increased need for correction saccades at larger saccade amplitudes. A general discussion of costs in terms of time requirements has been given by Gray, Sims, Fu, and Schoelles (2006). To investigate the trade-off between the usage of VSTM and the execution of eye movements, Inamdar and Pomplun (2003) used a comparative visual search paradigm developed by Pomplun et al. (2001). In their study, two identical columns of simple geometrical objects (each column containing 20 objects) had to be compared to detect the single difference (target). To vary the costs of eye movements, three distances between the columns were used (15°, 30°, and 45°). For the larger hemifield separations, fewer inter-hemifield saccades was found. Furthermore, the processing time, that is, the time spent within a hemifield before jumping to the other side, was prolonged for larger hemifield separation. To further increase the time needed for inter-hemifield saccades, the attended hemifield was masked when the saccade started and was unmasked only after a delay of 0, 500, and 1000 ms. Delayed unmasking enhanced the original effects, that is, it decreased the number of gaze shifts (i.e., inter-hemifield saccades) between the two hemifields and it increased the time for processing intervals. Inamdar and Pomplun argued that VSTM usage can be flexibly adapted to optimize task performance as long as the creation of internal representations does not take longer than roughly one second. These findings together with the experiments by Ballard et al. provide strong evidence for the VSTM versus eye movement trade-off in visual tasks. In natural, large-field environments, head movements interact with eye movements to carry out gaze shifts. Within the functional range of eye movements (up to ±22°; Stahl, 1999), head movements are unlikely, but their frequency increases as the gaze saccade amplitude grows (Freedman & Sparks, 2000; Hanes & McCollum, 2006; Land, 2004). Because head movements are more costly than movements of the eye, both in terms of time and energy consumption, the VSTM versus gaze shift trade-off should be shifted toward increased memory use for tasks involving large gaze saccades. Indeed, Ballard et al. presented initial evidence showing that subjects use more memorization in a block-copying task requiring large head movements when making gaze changes.

The goal of this present study is to examine the trade-off between gaze movements (involving both eye and head movements) and VSTM usage for large-field stimuli. For that purpose, we developed a comparative visual search task, similar to those used by Gajewski and Henderson (2005), Inamdar and Pomplun (2003), and Pomplun et al. (2001), and we increased the distance between the two stimulus regions up to 120°. As comparative visual search stimulus, we used two cupboards filled with geometrical objects distributed over four shelves. The task in this paradigm was to identify differences between the object configurations in the left and right cupboards. For large gaze saccades, we anticipate a strong shift of the trade-off toward VSTM usage due to larger eye and head movement costs. The trade-off will be described quantitatively by a simple model based on the total time required to solve the task, cf. Gray et al. (2006). Further points include the possible additional costs of head movements in the gaze shift versus VSTM trade-off and the identification of scanning strategies used by individual subjects.

Methods

Experimental setup

All experiments are performed using a virtual reality environment displayed on a large, curved projection screen shown in Figure 1. This screen provided a horizontal field of view of 150° and a vertical one of 70° to the subject. The geometrical shape of the projection screen was that of a conic shell with a vertical axis, an upper radius of 1.83 m, and a lower one of 1.29 m.

Figure 1. Schematic view of the experimental setup. Two video projectors illuminate the complete conical screen. The horizontal field of view at eye level is ±75°. The upper vertical elevation angle (βu) is +25 and the lower one (βl) −45°. Small picture: ASL501 eye tracker with fixed gauge object for head tracking, comprising four light reflecting balls.
Subjects were seated upright with the back tightly at the chair and with their head in the axis of the conical screen (eye level at 1.2 m with 1.62 m screen distance). Two video projectors (SANYO PLC-XU46 with 1024 × 768 pixel resolution) were used to illuminate the whole screen.

The setup was running on a 2.6-GHz PC under Linux RedHat 9.0 as operating system (graphic card: NVIDIA Quadro 4 980XGL with dual video projector connection). The spatial resolution was 2048 × 768 pixels with a frame rate of 60 Hz. Experimental procedures and rendering of the virtual environment was programmed in the SGI OpenGL Performer™. Compensation for image distortion generated by the curved screen was programmed in C++. Soft edge blending was done in hardware using two partial occluders in front of the projector lenses.

Eye-in-head movements were recorded with a head mounted, infrared light-based eye tracker (bright pupil type, model 501 from Applied Science Laboratories, Bedford, USA) with approximately 2° accuracy and a real-time delay of 50 ms. To record head-in-space movements, an infrared light-based tracker system (ARTtrack/DTrack from A.R.T. GmbH, Weilheim, Germany) with 6° of freedom, 0.1° accuracy, and a real-time delay of 40 ms was used. This device tracks a rigid body (configuration of four light reflecting balls) fixed to the eye tracker (see Figure 1, small picture). Both trackers had a temporal resolution of 60 Hz.

**Experimental task**

Two cupboards equally filled with simple geometrical objects in four geometrical shapes (triangles, circles, diamonds, and squares) and in four different colors (green, blue, yellow, and black) were used as stimuli (see Figure 2). Each cupboard included 20 objects in four shelves. Each shelf included five objects in a row and one cupboard subtended 30° of the subjects’ horizontal field of view. The diameter of an object was 3°, the horizontal separation between two objects was 5°, and the vertical one was 11°.

For the comparative visual search task, the object configurations in the two cupboards were either completely equal (zero target condition) or differed at one or two positions (one and two target conditions, respectively). Target objects differed only in shape whereas all other objects pairs had identical features (distractors). A maximum number of two targets was introduced to avoid premature trial completion. Because subjects did not know the number of targets, they should not terminate the comparative search after detecting the first target. Four different cupboard distance conditions were used in the experiment. The horizontal distance between the centers of both cupboards was 30°, 60°, 90°, or 120°. These larger cupboard distances should induce the need for head movements and therefore higher costs for more time consuming gaze shifts. A session consisted of 36 trials in random order (4 distance conditions × 3 target conditions × 3 repetitions for each target condition). Object configuration for both targets and distractors was randomized for each trial.

In the beginning of the experiment, the eye tracker was calibrated by displaying a 9-point calibration pattern on the screen. For this procedure, head movements were prevented with a chin rest and all nine points had to be fixated. Also the head tracking target was calibrated with fixed head. Additionally, each experimental trial started with a 5-s fixation phase during which a fixation cross was displayed at eye level (1.2 m elevation) in the center of the screen (point of origin, cf. Figure 2). During this phase, the subject had to rotate the head to align the naso-occipital axis to the fixation cross followed by fixating the cross with the eyes. Thus, the subjects’ gaze offset with respect to the calibrated systems was measured for each trial. After this fixation phase, the cross disappeared and the two cupboards became visible. The subjects were free to move their head and eyes to find the number of targets (i.e., zero, one, or two) as quickly and reliably as possible. The subject terminated the trial by pressing a button and reported verbally the number of targets. The next trial started with the fixation phase to the fixation cross after pressing the button. Participants were free to take breaks in between trials if desired.

Twenty-five subjects participated in this study (age: 23–34 years). Subjects were students at the University of Tübingen with normal or corrected to normal (only contact lenses were allowed) vision and were naive to the purpose of the experiment. They received monetary reimbursement for their participation. Because of the wide spread stimulus size, satisfactory eye tracking was obtained only for 12 of 25 subjects. Eye tracking loss was indicated by the tracker.

![Figure 2](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932849/)
by a zero value for the pupil diameter. This was resulting from eye blink or a loss of the Purkinje reflex or of the pupil. Subjects were excluded if percentage of eye tracking loss exceeded 15% of all time. Only data from the 12 subjects were analyzed for the recent study.

Data analysis

To analyze the recorded data, the MATLAB® software (The MathWorks Company, Natick, USA) was used. Based on head and eye tracking data, the gaze vector was calculated in angles with an azimuth and an elevation component (\( \alpha \) and \( \beta \), respectively) relative to the point of origin (see Figure 2). Thus, the gaze vector includes both the head-in-space and the eye-in-head vectors. Gaze fixations were defined in the following way: For each time step \( t \), we consider a gliding window of length 120 ms centered at \( t \). Let \( v_{\text{min}} \) and \( v_{\text{max}} \) denote minimal and maximal gaze velocities obtained within the window. The instant \( t \) is classified as belonging to a fixation, if \( v_{\text{max}} - v_{\text{min}} < 100 \text{ deg/s} \). This procedure is iterated through all time steps. Adjacent instants in time satisfying the condition are combined to fixational events.

Results

Stimulus-related search performance

As error rate, we defined the proportion of subjects’ incorrect responses in terms of the number of targets. Similar to Inamdar and Pomplun (2003), we found relatively low error rates of less than 7% for all cupboard distance conditions (Figure 3), indicating for subjects’ accurate task performance. The error rates did not vary significantly between the four cupboard distance conditions, \( F(3,33) = 1.416, MSE = 9.84, p = .256 \).

Furthermore, we analyzed the response time, that is, the time needed to detect the number of targets and finishing one trial. We found a significant increase with increasing cupboard distance, \( F(3,33) = 33.186, MSE = 4.754, p < .001, \eta^2 = .751 \) (Figure 3). The response time increased linearly by about 2 s for each 30° step of cupboard distance.

Stimulus-related eye and head movements

The subjects’ initial gaze fixation was directed to the point of origin (i.e., \( \alpha = 0 \) and \( \beta = 0^\circ \)). This was ensured by the calibration cross fixation phase before starting an experimental trial. After both cupboards became visible, all investigated subjects shifted their gaze to the upper left part of the left cupboard (cf. Figure 2 and Movie 1). This was followed by oscillating gaze changes between the left and right hemifield including shelf level shifts down to the lowest part of the cupboards. This general search pattern could be observed for all participants.

The smooth oscillatory shape of horizontal head movements is clearly visible in Figure 4. Maximum head movement velocities, averaged over all participants, of about 165 deg/s have been observed for the largest cupboard distance condition (120°), whereas the averaged velocity for 60° distance was only 48 deg/s. These
maximum velocities are reached during the head’s left/right and right/left shifts, respectively. To enable a stable fixation during head movements, subjects performed compensatory eye movements in the direction opposite to the head movement under the control of the vestibular–ocular reflex (cf. Figure 4C).

Figure 5 shows the head and eye components of all recorded gaze saccades pooled over all subjects. Due to the cupboard design we used, gaze saccades were mainly performed in horizontal directions (cf. Figures 2 and 9). Under these circumstances, head and eye movements occur in parallel. Thus, eye saccade amplitudes can directly be obtained by subtracting head from gaze amplitudes. For 30° gaze saccades, eye movement amplitudes dominated the gaze and reached about 27.3°. For larger gaze amplitudes, the proportion of the head component increased up to 71° for gaze saccade amplitudes of 130°. This corresponds to approximately ±35.5° for the head’s movement range and ±29.5° for the eye’s movement range. This result (±29.5°) is well below the anatomical limit of the ocular motor range estimated to be ±53° (Guitton & Volle, 1987). The relation between gaze amplitude and head’s proportion was quadratic with a correlation coefficient \( r = .879 \) (see Figure 5). The clustered distribution of the data points in this figure reflects the four cupboard distance conditions.
Stimulus-related gaze behavior

The most important objective of the experiment was to analyze the trade-off between gaze movements and VSTM for large gaze shifts. As compared to the results of Inamdar and Pomplun (2003) for eye saccades, we expected increased preferences for memory use resulting from high head movement costs. In the experiment, this should show up by a reduced number of inter-hemifield gaze saccades and longer fixation durations.

Figure 6A2 shows the number of inter-hemifield gaze shifts per trial, averaged over all subjects and depending on the cupboard distance. This number of inter-hemifield saccades per trial was significantly reduced for larger hemifield distances, \( F(3,33) = 75.341, \text{MSE} = 3.714, p < .001, \text{eta}^2_p = .873 \). Subjects performed approximately 10 gaze shifts for the largest cupboard distance condition. This value is roughly half compared to the 30° condition.

The decrease of inter-hemifield gaze shifts was approximately the same for all 12 participants (cf. Figure 6A1). The fixation duration for each level of distance averaged over all subjects is shown in Figure 6B2. We found a significant increase for the fixation duration with increasing hemifield distance, \( F(3,33) = 9.331, \text{MSE} = 137.01, p < .01, \text{eta}^2_p = .459 \). On average, subjects fixated about 25 ms longer for the most “expensive” condition in this experiment compared to the relatively “inexpensive” 30° cupboard distance condition. Figure 6B1 shows the fixation duration averaged over all trials per distance condition separately for each subject. All subjects performed longer lasting fixations related to the increased distance between both cupboards. Both the fixation duration and the number of inter-hemifield saccades over all subjects tend to saturate for large hemifield separations.

The fixation number per trial, averaged over all subjects (Figure 6C2), did not vary with the cupboard distance and was 39.8 fixations per trial, \( F(3,33) = 0.974, \text{MSE} = 11.606, p = .417 \). The same holds for the individual subjects, but with variation across subjects in large (Figure 6C1). Subjects performed the search task with a minimal fixation number of about 25 and a maximal number of about 65. Still, each subject showed a constant number of fixations over all distance conditions.

For large distances between the two cupboards, the time needed for a gaze shift was increased. Related to this increased time costs, the number of inter-hemifield saccades was decreased. This relation is shown in Figure 7. Per trial, subjects needed on average 4.97 s for all inter-hemifield gaze shifts in the 120° distance condition. For a single gaze shift, this amounts to a duration of about 0.5 s. For the 30° distance condition, the cumulative duration for all gaze shifts within one trial was only 2.15 s. Subjects performed on average 17.5 gaze shifts between the hemifields. This leads to 0.12 s duration for a single inter-hemifield gaze saccade.

To investigate the role of head proportions, the maximum head amplitude occurring in each trial within the two largest separation conditions (i.e., 90° and 120°) was considered. For each subject, two correlations were analyzed, (i) maximal head amplitude versus fixation duration and (ii) versus number of gaze shifts. The tests performed over the correlation coefficients of all 12 subjects revealed significant differences from zero correlation, for details see Figure 8. Correlations of maximum head amplitude with the number of inter-hemifield saccades were negative for 10 of 12 subjects in both separation conditions (Figure 8, left). Correlations of maximum head amplitude with the fixation duration were positive for 11 of 12 subjects in both separation conditions (Figure 8, right). These data indicate that for a fixed hemifield separation, larger head proportions correlate with smaller number of inter-hemifield saccades and longer fixation duration.

Stimulus-induced search strategies

It turned out that the number of fixations (on average, 40 per trial over all subjects and conditions) was not evenly distributed over both stimulus hemifields. For example, in the gaze scan path shown in Figure 2, most fixations are located in the left cupboard. We analyzed the number of fixations with respect to the left and the right hemifield. The results indicate strong differences between the subjects (Figure 9). Eight participants showed more frequent fixations into the left and three participants more frequent fixations into the right hemifield. In the sequel, the first group is called “left hemifield subjects” and the second one is called “right hemifield subjects.” Only one
subject performed about 50% in either hemifield over all cupboard distances. The dependence of the proportion of fixations in the left hemifield on subjects was significant, $F(11,384) = 52.52, MSE = 46.8, p < .001, \eta^2_p = .6$. The post hoc analysis produced a significant difference between the right hemifield subject group and the left hemifield subject group ($p < .01$).

Interestingly, for the left hemifield subjects, the duration of individual fixations in the left hemifield was longer than fixation durations in the right hemifield. Conversely,
right hemifield subjects’ fixations to the right hemifield lasted longer than their fixations to the left. This phenomenon is illustrated in Figure 10. There the previously shown fixation duration averaged over all 12 subjects (cf. Figure 6B2) was split up for the left and the right hemifield subjects’ fixation duration to the left and to the right hemifields. Two sample t tests revealed significant differences between the fixation duration in the left and in the right hemifield for all cupboard distance conditions except for the one case marked in the Figure 10. The main effect for increased fixation duration with increased cupboard distance level was still visible in all four distance conditions. In addition, the left hemifield subjects’ fixation durations were always longer than those performed by the right hemifield subjects.

Figure 7. Trade-off between the sum of all inter-hemifield saccade durations ($T_\phi$) and the number of inter-hemifield saccades ($n(\phi)$) for each trial averaged over all subjects for each cupboard distance ($\phi$). Regression indicates a power function relation with $r = .99$. Error bars indicate standard error of the mean.

Figure 8. Box-whisker plots (median with 10th and 90th percentiles) of the correlation coefficients of the maximum head movement amplitude with the number of inter-hemifield saccades (left side) and the fixation duration (right side) in the 90° and the 120° cupboard distance conditions. Correlations were calculated over trials and separately per subject. Significant differences from zero correlation are given for each box-plot (*$p < .05$; **$p < .01$).

Figure 9. Hemifield distribution of all fixations for each subject. Subjects with more than 50% fixations over all distance conditions into the left cupboard are called left hemifield subjects and these with less than 50% right hemifield subjects. Only one subject showed no preference for the left or the right side. Error bars indicate standard error of the mean.

Figure 10. Fixation duration for the left and the right hemifield subjects divided for the two hemifields for all four cupboard distance conditions. Significant differences are calculated for the left and the right hemifield subjects between their fixation durations to the left and to the right hemifield for each cupboard distance (*$p < .05$; **$p < .01$; ***$p < .001$). Error bars indicate standard error of the mean.
As an additional parameter characterizing the search strategy, we analyzed the level shifts between shelves in the gaze scan path. Because the search task is carried out in a sequential manner, information encoded from one cupboard shelf has to be compared to information on the same shelf level of the opposing hemifield (cupboard). Therefore, shelf level shifts should occur only after one comparison has been completed and novel information is to be encoded. We divided the occurring shelf level shifts into two groups according to their goal, that is, level shifts within or into the left hemifield (cf. Figure 11A) and level shifts within or into the right hemifield (cf. Figure 11B).

For the left hemifield subjects, as classified on the basis of fixation frequency and duration, we found a preference for shelf level shifts within or into the left hemifield. These subjects showed a strong preference to begin a new shelf encoding stage on the left side performing about 80% of the shelf level shifts within or into the left hemifield. Vice versa, subjects from the right hemifield subjects group showed a preference for shelf level shifts within or into the right hemifield. They performed about 60% of all shelf level shifts within or into the right hemifield (Figure 12). Statistical analyses showed a significant difference between the left and the right hemifield subjects’ search strategies related to the shelf level shifts for all cupboard distance conditions (see Figure 12).

**Discussion**

**Head movement proportion on gaze amplitude**

In normal scene perception, humans naturally use head and eye movement together to direct their gaze (Hayhoe, 2000; Hayhoe & Ballard, 2005; Hayhoe et al., 2002; Land et al., 1999). To understand the role of head movements and their interaction with eye movements in relation to visual working memory use, we performed a comparative visual search paradigm similar to Inamdar and Pomplun (2003). Our results confirm the findings by Inamdar and Pomplun on eye saccades ranging from 15° to 45°, and these by Ballard et al. (1995) on combined eye and head movements ranging up to 70°. Furthermore, we extend these results toward larger gaze shifts with more pronounced trade-off effects resulting from stronger head movement involvement. To increase the costs for gaze movements, we increased the inter-hemifield distance between two stimuli up to 120°. To perform large gaze amplitudes corresponding to this stimulus size, head movements are required because the full-scale ocular
motor range is limited to ±53° (Guitton & Volle, 1987). Furthermore, Stahl (1999) identified the customary ocular motor range with about ±22°, that is, well below the full-scale ocular motor range. Head movements are performed, Stahl argued, to maintain the eyes within the customary range. Also Becker (1989) reported that when the head is free to move, head movements become a regular feature of gaze saccadic shifts at approximately 20°. Our data support this finding. For the largest cupboard distances, we used 90° and 120°; the averaged eye amplitude was about ±29°. For these large gaze amplitudes, the head was used to maintain the eyes within a range, as described by Stahl. For smaller gaze amplitudes (below 60°), the head’s proportion on gaze saccades showed more variance. With these small distances, the eyes could reach both stimulus hemifields without leaving the customary ocular motor range and no head movements were necessary. This finding differs from other studies (Kowler et al., 1992; Pelz, Hayhoe, & Loeber, 2001) where head movements were a regular feature even for small gaze shifts. In experiments by Pelz et al. (2001), head movements ranged between 1° and 10° for gaze changes of 15°. These different findings about head amplitudes suggest that this property of gaze shifts is not reliable and the authors concluded that the head movement magnitude is probably a function of the experimental constraints (Pelz et al., 2001). The large intersubject variability of head proportion on gaze saccades observed in our study was also reported by Fuller (1992), Pelz et al., and Stahl (1999, 2001). Moreover, even the same subject can alter the movement strategy to better address the demands of a specific task (Oommen, Smith, & Stahl, 2004). Unfortunately, the reason of this variability in head movement tendencies is still unknown.

Trade-off between VSTM and gaze movement behavior

The findings in experiments on scene comparison of Gajewski and Henderson (2005) suggested a strong general bias toward minimal use of VSTM. The authors showed that for complex visual tasks, often only one object is encoded at a time and maintained in visual memory. If subjects make such minimal use of memory also in our experiment, the following optimal scan path can be predicted: Starting from the upper left object, which is usually fixated at the beginning of each trial, fixation should shift to the corresponding position on the right cupboard. If the object at this place is identical, memory could be cleared and the next object right to the last fixated one should be memorized and compared with the corresponding object and so on. In our results, we found similar scan paths, but only for the smallest cupboard distance condition. In this condition, subjects produced on average 4.4 gaze shifts between the two hemifields within one shelf level, 2.2 in each direction. With the five objects per shelf level which have to be compared, a single cycle corresponds to approximately two memorized objects. This hints toward a memoryless search strategy. Also the relatively short fixation duration in the 30° condition, which we take as an indication of processing time, supports the lower memory load (Velichkovsky, 1995). For the largest distance condition (i.e., 120°), we found an approximately halved value for inter-hemifield gaze saccades. Because long lasting fixations are generally taken to indicate extraordinary memory load, the increased number of memorized items together with the much higher processing time hints toward an increased usage of VSTM. The higher the costs for gaze movements, the more memory load will be necessary. The overall fixation durations measured in this study (about 250–270 ms) corresponds well to those reported in other studies (Gajewski, Pearson, Mack, Bartlett, & Henderson, 2005; Henderson, Weeks, & Hollingworth, 1999; Rayner, 1998), which found fixation durations as low as 275 and 247 ms for a visual search task.

In the study by Inamdar and Pomplun (2003), without head movements the maximum inter-hemifield distance between the two columns of objects was 45°. The reported number of gaze shifts of about 14.3 matches very well with our findings for comparable distances (see Figure 13). With delayed unmasking of the attended hemifield, Inamdar and Pomplun tried to shift the trade-off between VSTM and eye movements as far as possible toward memory use. Masking time intervals up to 1 s were used. By increasing the time requirements, the authors could show that the employment of VSTM can be flexibly adapted to optimize task performance as long as the creation of internal representations does not take more than about 1 s. The data for the fixation duration (Figures 6B1 and 6B2) and the number of

![Figure 13. Combined presentation of decreasing inter-hemifield gaze saccades with increased stimulus distance for our data and these from Inamdar and Pomplun (2003). Error bars indicate standard error of the mean.](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932849/)
gaze shifts (Figures 6A1 and 6A2) indicate a tendency to approach asymptotically a limit for 120° inter-hemifield distance. For increased hemifield separations, we expect no further change of these variables. Thus, the upper limit of working memory load for the comparative visual search task seems to have been obtained.

The question whether head movements impose additional costs on gaze shifts has been addressed by a correlation analysis of head proportions with the number of gaze shifts and with fixation duration within the 90° and 120° hemifield separation conditions. When pooling over all subjects, the correlations did not reach significance. However, correlation coefficients per subject consistently showed positive signs for fixation duration and negative signs for number of gaze shifts. Across subjects, these effects did reach significance as shown in Figure 8. If gaze saccades with a given amplitude are carried out using a larger head proportion, the costs therefore seem to be increased. These data suggest that costs depend on both gaze shift amplitude and head proportion.

The comparative visual search task requires dealing with a fixed amount of information which does not depend on the hemifield separation. By eye movements, this amount of information is broken up into a number of chunks each of which is processed in a cycle of (i) encoding, maybe including intra-hemifield gaze shifts, (ii) inter-hemifield gaze shift, (iii) comparison, maybe including intra-hemifield gaze shifts, and (iv) inter-hemifield gaze shift back to the encoding side. The total cost can thus be divided into a cost of processing (encoding and comparing; steps i and iii) and a cost for shifting gaze between hemfields (steps ii and iv). The number of cycles, \( n \), needed to solve the task equals half the number of inter-hemifield gaze-shifts. The amount of information processed during each cycle equals \( I = \frac{1}{n} \), where the total information is arbitrarily set to unity. We now assume that the number of cycles, \( n \), is adjusted to minimize the total cost of processing. One possible measure of costs is the time needed to solve the task as suggested, for example, by Gray and Fu (2004) and Gray et al. (2006). This measure is consistent with the instruction given to our subjects, that is, to solve the task as quickly and reliably as possible. Because the error rate was generally low, it seems that time is indeed the most important factor. With respect to the processing cycle, we need to distinguish two time variables. First, the processing time per cycle, \( T_p(I) \), equals the time needed for the encoding and comparison steps; it is assumed to depend on the size \( I = \frac{1}{n} \) of the information chunk, but not on hemifield separation. We choose a power law for \( T_p(I) \), that is, \( T_p(I) = b \cdot I^a \) with constants \( b, a \). Second, shift duration per cycle, \( T_s(\phi) \), is assumed to depend on hemifield separation \( \phi \), but not on chunk size \( I \). Note that the dependence of \( T_s \) on \( \phi \) is taken simply as an empirical fact as reported in Figure 7. It may result from the varying relative contributions of eye and head movements, both in terms of mechanical properties and in terms of neural planning effort for the compound movement.

With the above notations, we can compute the total time needed to carry out the comparison in \( n \) steps:

\[
T_{\text{tot}} = n \cdot [T_p(I^{-1}) + T_s(\phi)] = b \cdot n^{(1-a)} + n \cdot T_s(\phi). \tag{1}
\]

The latter equality results from the assumed power law for \( T_p \). The idea of the model is that the trade-off results from minimizing \( T_{\text{tot}} \) with respect to \( n \). Taking the derivative with respect to \( n \) and setting the result to zero yields

\[
(a - 1) \cdot b \cdot n^{-a} = T_s(\phi), \tag{2}
\]

and further

\[
n = c \cdot [T_s(\phi)]^{-(1/a)}, \tag{3}
\]

that is, the relation of the number of cycles to the total time required for gaze shifting follows a power law. Here, \( c \) is a constant depending on \( b \) and \( a \). Figure 7 shows the relation of \( n \) and \( T_s \) from our empirical data together with the theoretical curve given by Equation 3. If \( T_s \) is measured in seconds, the fitting parameters are \( c = 29.5 \) and \( a = 1.47 \).

Why should \( T_p \) depend on the size of the encoded information chunk? One possible explanation for this is that the encoding of additional information may become more error prone as the amount of information already encoded increases. In this situation, additional fixations within one hemifield might be necessary, resulting in a longer processing time \( T_p \). Indeed, the total number of fixations was found to be constant across all separation conditions (see Figure 6C), indicating that the number of intra-hemifield saccades increases as the number of inter-hemifield gaze shifts decreases. We therefore suggest that working memory involvement increases with hemifield separation and that encoding time increases in a nonlinear way with the amount of encoded information.

Subject-specific search strategies

An analysis of the number of fixations divided for the left and right hemifields indicated that subjects use one of two search strategies. On the basis of the percentage of fixations made into the left hemifield, we defined three subject groups. For all four cupboard distance conditions, the left hemifield group (eight subjects) performed more than 50% of all fixations into the left side. The second group, termed right hemifield subjects (three members), made more fixations in the right hemifield. In these two groups, we found no obvious correlation between the cupboard distance and the asymmetry of fixation numbers. Only one subject showed no preference for either side. All tested subjects are right-handed, so this attribute could not explain these fixation asymmetries.
The existence of such a strategy was furthermore supported by the findings on different fixation durations for left and right hemifield subjects for the right and the left sides. We found that the left hemifield subjects not only performed more fixations into the left hemifield; also, their fixation duration to this side was significantly longer. The same effect was true in the other way around for the right hemifield subjects. One could argue for two functional task stages processed by the subjects solving our comparative visual search paradigm. One stage is defined by more and longer fixations related to the preferred hemifield and could be interpreted as “encoding” stage. During the second stage, subjects performed fewer and shorter fixations to the other side. We called this stage the “comparison” stage.

The scheme in Figure 14 summarizes the search strategy process found in our experiment. Within a shelf level (shelf level: both corresponding cupboard shelves together), subjects started the visual search with the encoding stage and performed inter-hemifield gaze shifts in alternation between the comparison and the encoding stage. The encoding stage was identified by increased and the comparison stage by decreased fixation frequency and duration. After processing a shelf level completely, the following between shelf level gaze shift was started with the encoding stage. Left hemifield subjects were left encoders with their first encoding stage for a new shelf level mainly in the left hemifield, and correspondingly right hemifield subjects were right encoders.

Pomplun et al. (2001), using two fields with randomly distributed items, also distinguished two phases of comparative search, based on eye movement characteristics. In a “search and comparison” phase, subjects identified suspicious regions for further inspection. In a “detection and verification” phase, detailed scanning was performed to identify deviating items. In our structured layout, items are grouped on shelf levels that are searched one by one. Therefore, we do not find a “search and comparison” phase in the sense of Pomplun et al. We suggest that our distinction between encoding and comparison stages corresponds to a subdivision of Pomplun et al.’s “detection and verification” phase. Because Pomplun et al. do not analyze their “detection and verification” phase further, a deeper comparison of the search strategies and subject-specific effects is not possible.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (Graduiertenkolleg 778 and grant GI373/1-1 awarded to S.G.) and the European Commission (6th FP NEST-Pathfinder Project “Wayfinding”).

Commercial relationships: none.
Corresponding author: Gregor Hardiess.
Email: gergor.hardiess@uni-tuebingen.de.
Address: Eberhard Karls University, Cognitive Neuroscience, Auf der Morgenstelle 28, 72076 Tuebingen, Germany.

References


Figure 14. Scheme of the search strategy with the encoding and the comparisons stage alternation within and between shelf level gaze shifts.


