Perception, 2004, volume 33, pages 485 – 494

DOI:10.1068/p5130

Eye movements in a simple spatial reasoning task

Christof Körner¶, Iain D Gilchrist
Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, UK
Received 30 July 2003, in revised form 19 December 2003

Abstract. We report two experiments in which participants read a question about the spatial relationship between two letters, then viewed a visual display containing the letters and were required to respond to the question. The format of the question influenced the nature of the eye movements generated to the visual display. Participants also had a tendency to make additional eye movements in order to generate a fixation sequence that corresponded to the order of the letters in the question. This demonstrates an influence of stored information on eye movement generation, and suggests that the scanpath plays a role in structuring the visual information to facilitate reasoning.

1 Introduction

Patterns of eye movements reflect both the content of the visual scene and the task that the observer is carrying out (eg Yarbus 1967). In addition, a number of recent studies have shown that eye-movement patterns systematically reflect the spatial layout when participants simply imagine the scene. For example, Brandt and Stark (1997) presented a chessboard-like display with black and white squares to their participants. Participants were then shown a blank board and were asked to imagine the previously seen spatial arrangement of black squares. The resulting eye movements were very similar to those generated when participants viewed the original display. A visual display (like the chessboard) need not necessarily be present for eye movements to coincide with stored spatial information. Richardson and Spivey (2000) and Spivey and Geng (2001) presented objects and animations in one of four locations on the screen to their participants. When later asked questions referring to one of the events while sat in front of a blank screen, participants were more likely to move their eyes to the location where that event had occurred. Together, these findings show that eye movements can emanate from stored visual representations.

Additional support for the idea that eye-movement patterns can be influenced by internal representations comes from studies in which linguistic information affects the way information is extracted from a visual scene. Altmann and Kamide (1999) showed participants a visual scene depicting, for example, a boy, a cake, and various distractor objects. The cake was the only edible object in the scene. When the participants heard the sentence "the boy will eat the cake" they moved their eyes faster to the target object (the cake) than when they heard the sentence "the boy will move the cake". This is evidence for a direct and on-line interaction between the linguistic and visual information. Such effects can even occur in the absence of any useful visual information. For example, Demarais and Cohen (1998) had their participants solve verbally presented spatial inference problems. When the problems involved the relational terms *above* or *below*, participants made more vertical eye movements than when the problems involved the terms *left* or *right*.

Together, these results demonstrate how visual and non-visual representations can shape the eye movements that are generated. In addition, these results suggest that eye movements can emanate from such representations even in the absence of a visual display.

¶ Address for correspondence: Institut für Psychologie, Karl-Franzens-Universität Graz, Universitätsplatz 2/III, A-8010 Graz, Austria; e-mail: christof.koerner@uni-graz.at

We report here two experiments in which we investigated how eye movements can be shaped by stored information in a simple reasoning task. This new paradigm allows us to demonstrate how the eye movements mediate between the visual display and the stored non-visual representation. In addition, these experiments give an insight into the structure of the stored representation, an issue we will return to at the end of this paper.

In our experiments, participants were required to reason about a two-dimensional spatial arrangement of four letters, two of which were presented in the top row and two in the bottom row of the display. Before the letter display was presented, we displayed linguistic information and asked the participants questions about the spatial relation of two of the letters (target letters). For example the question could be "Is a better than c?" which in this context means, Is *a* drawn above *c*? ('better'-type question) or the question could be "Is a less than c?" meaning: Is *a* drawn below *c*? ('less'-type question).

In analysing the data we looked for evidence that the question type and structure influences the sequence of eye movements that occur during the presentation of the letter display. This in turn provides evidence that the stored question information shapes the eye movement sequence.

2 Experiment 1

2.1 *Method*

2.1.1 *Overview of the task.* The task required the participants to read a visually presented question of the form "Is a better than c?" and then to view a display that consisted of four letters and decide whether the letter *a* was presented above the letter *c* in the display. Participants then responded with a manual button press. An example trial sequence is shown in figure 1.

2.1.2 *Stimuli and procedure.* Before each trial, the participant fixated a small black disc in the centre of the screen that was used for drift correction of the eye tracker (see below). After the drift correction was completed, a question was presented in the centre of the screen for 2000 ms. The question was set in Arial font and subtended 1.1 deg vertically and 9.6 deg ('better' question) and 8.6 deg ('less' question) horizontally. The question was followed by a centred fixation stimulus (a small cross) for 1000 ms. Then the screen went blank for 100 ms before a display was presented for a maximum of 5000 ms. Participants were instructed to read the question, fixate the fixation stimulus, and respond, after the display had been presented, as accurately and rapidly as possible. Participants were instructed to answer a question by checking the spatial relationship between the target letters in the display; they were told to press the right button on a button box for a "yes" response and the left button for a "no" response. The participant's button-press ended a trial and cleared the display.

The question always asked for the spatial relationship between the two target letters *a* and *c*. The block of trials was preceded by a cover story in which it was explained that 'better' was denoted in the display by being drawn above; this corresponds to the phrasing of questions in similar experiments (see, eg, Körner 2004; Körner and Albert 2002).

A question could be either "Is a better than c?", meaning: Is *a* drawn above *c*? ('better'-type question) or "Is a less than c?" ('less'-type question). The combination of the two types of questions ('better' versus 'less') and the two letter orders resulted in four different questions. Note that for each question there exists a corresponding question in a different structural form eg "Is a better than c?" is asking for identical information to "Is c less than a?".

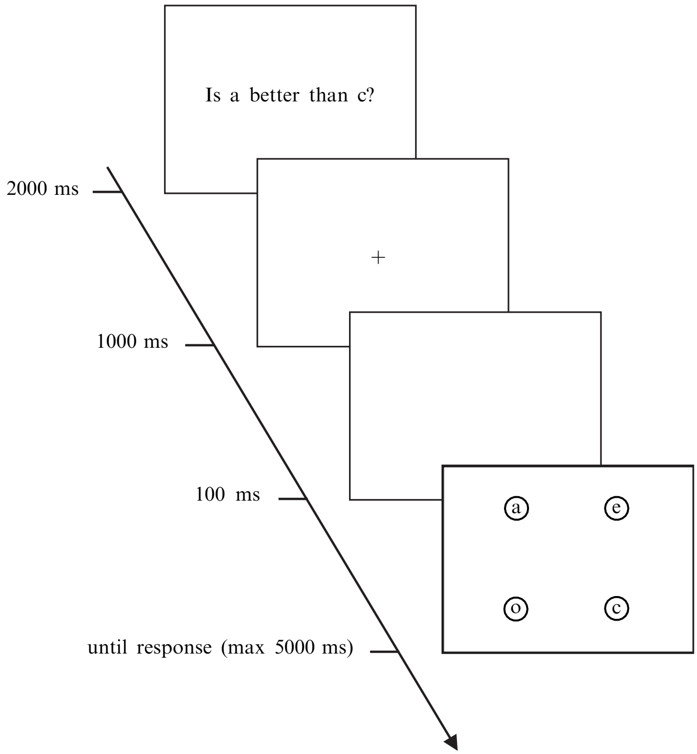

Figure 1. Display sequence in a trial.

The display consisted of an arrangement of four letters (a, c, e, o) located at the corners of an imaginary square with an edge length of 12 deg. A black circle surrounded each letter. The circle served two purposes: the first was that it acted to reduce the ability to identify the letter without fixation (cf Bouma 1970) and the second was that it provided a clear target for the saccadic orienting. The outer diameter of an encircled letter was 1.5 deg (see figure 1). We used all 24 possible display arrangements. In 8 of the displays, the letter *a* was drawn in the top row and above the letter *c*, in 8 of the displays the letter *c* was drawn above the letter *a*. In the remaining 8 displays the two target letters were drawn at the same height, either in the top row or in the bottom row.

Each of the 24 displays was combined with the four questions yielding 96 different problems. As a result, 32 of the problems with target letters in different rows required a "yes" response (yes problems) while 32 problems required a "no" response (no problems). Participants were instructed to respond "no" to the 32 problems involving target letters in the same row. To compensate for these additional 'no' problems the 32 'yes' problems with target letters in different rows were repeated. This resulted in 128 problems that were presented in a random order. There were 12 practice trials followed by the 128 experimental trials.

2.1.3 *Apparatus.* We recorded two-dimensional eye movements using an SMI Eye-Link eye tracker (SensoMotoric Instruments GmbH, Germany), which is an infrared video system with a 250 Hz sampling rate and a head-movement compensation mechanism. We recorded from both eyes, and analysed the data from the eye that produced the best spatial resolution, which was typically better than 0.3 deg. Displays were presented on a personal computer, while a second PC recorded the eye-position data online. The display monitor was a 17 inch SVGA monitor with a resolution of 800×600 pixels. The viewing distance was 57 cm, and a chin-rest was used to minimise head movement.

2.1.4 *Participants.* Sixteen undergraduate students (nine female, seven male) from the Department of Experimental Psychology, University of Bristol, served for class credit. Their average age was 21.1 years, ranging from 18 to 35 years. They had normal or corrected-to-normal vision.

2.2 *Results and discussion*

The overall percentage of correct responses was 95.9%: only these correct trials were considered for further analysis. In addition, only trials in which participants fixate the fixation stimulus after the question display were included. Following this additional criterion, 92.6% of trials were analysed.

The mean response time from letter display onset was 1616 ms. A 2×2 repeated-measures ANOVA was carried out on the mean response times with question type ('better' versus 'less') and question sequence (*a* followed by *c* or vice versa). There was a reliable effect of question type with 'better' questions being answered 131 ms faster than 'less' questions ($F_{1,15} = 49.39$, $p < 0.01$).

For the eye-movement analysis, we used gaze areas of 6 deg × 6 deg visual angle (corresponding to an entire quadrant) surrounding a letter. Fixations outside these defined areas were deleted from the fixation sequence (< 2%). Consecutive fixations in the same area were combined into a single inspection and inspection duration. The use of an inspection or *gaze measure*, respectively, that combines the primarily orienting saccade with the corrective saccade to gain exact target fixation is a recognised method in the reading literature (eg Rayner et al 1998). The total number of inspections analysed was 7214.

The mean number of inspections per trial was 3.80. There were 0.12 additional inspections for 'less' questions and this difference was statistically reliable ($F_{1,15} = 9.47$, $p < 0.01$).

2.2.1 *First inspection was modulated by question type.* We first investigated the relationship between the first inspection and question type ('better' versus 'less'). Most of the first inspections were directed to letters above the central fixation point (60.3% top left, 17.3% top right) as opposed to bottom letters (18.1% bottom left, 4.3% bottom right). This may reflect a kind of reading strategy adopted by participants. However, over and above this effect, first inspections were strongly dependent on the type of question: 'Better' questions produced more first inspections in the upper locations than 'less' questions; conversely, 'less' questions produced more inspections in the lower locations than 'better' questions. This effect is illustrated in figure 2. Since all the 'better' questions account for 50% of the data and the 'less' questions for the remaining 50%, the percentages are not independent of each other. Therefore, we compared the difference of the percentages between upper and lower locations for 'better' questions (44% on average) with the respective difference for 'less' questions (11%) to confirm the pattern of results displayed in figure 2 ($t_{1,15} = 4.34$, $p < 0.01$).

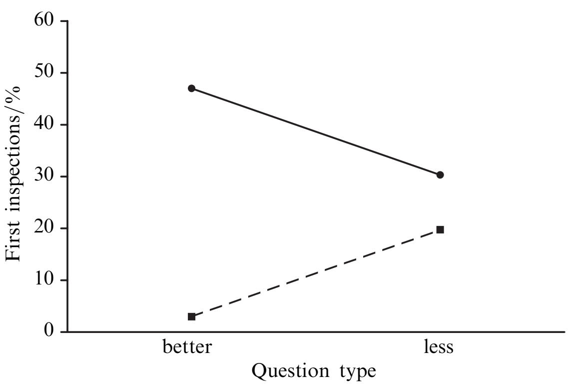

Figure 2. Percentage of first inspections on the upper letters in the display (solid line) versus lower letters (dashed line) plotted by question type.

2.2.2 *Reinspections on the question letters.* Target-letter inspections accounted for about 60% of all inspections. For 37.0% of the trials participants fixated both targets exactly once. For 16.1% of the trials participants fixated only one of the target letters before responding and for 2.3% of the trials participants responded without a target inspection. For the remaining 846 trials (44.6%), reinspections (of targets or distractors) occurred after both targets had been fixated at least once. Consequently, participants fixated both targets at least once before responding in the majority of 1548 trials (81.6%).

For a more detailed description of inspection paths we use the following string notation. For each trial a target letter inspection is denoted by T. Single and multiple subsequent distractor-letter inspections are denoted by a single D. The 12 most common inspection patterns coded in this manner are depicted in figure 3 against their relative frequency. These 12 patterns accounted for more than 75% of all patterns obtained. Amongst these patterns, the three most common patterns involved the inspection of both target letters and one or two distractor-letter inspections (31.9%). Four patterns (DT, TD, DTD, and T) involved a single target letter inspection (14.5%). Target-letter reinspections occurred in 32.9% of the patterns.

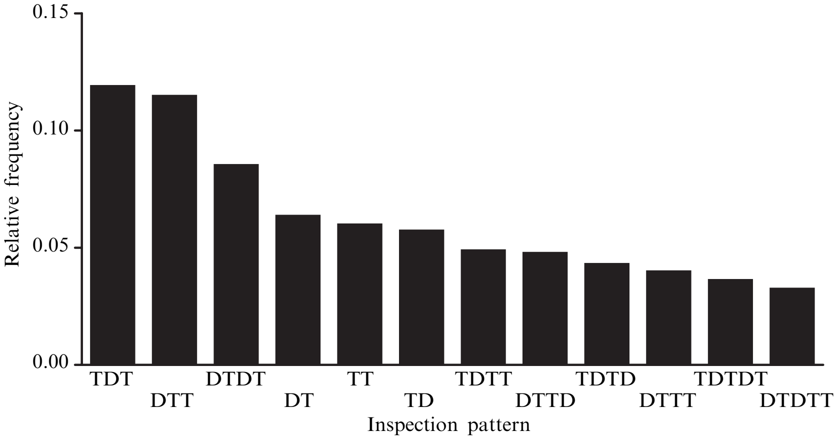

Figure 3. Relative frequency of the 12 most common inspection patterns. Patterns are represented by sequences of letters; a T represents inspection of the target, and a D represents inspection of distractors.

2.2.3 *Reinspections on target letters resulted in inspection – question alignment.* To further clarify how inspection patterns were influenced by the question we categorised the scanpath patterns by question sequence. If a participant fixated the first letter in a question before the second letter, we called this an aligned pattern; if the second letter was fixated first, this was called a non-aligned pattern.

Of the inspection patterns, 823 were aligned and 725 were non-aligned. If the inspection pattern was aligned, participants responded without a subsequent target reinspection in 71.0% of the trials. If the inspection pattern was non-aligned, they responded without target reinspection on 61.1% of the trials. This result is illustrated in figure 4. Participants responded more readily when they had fixated the target letters in the sequence provided by the question. If that was not the case (non-aligned patterns), first-target reinspections were increased by 9.9% ($F_{1,15} = 7.29$, $p < 0.05$). This is equivalent to a tendency to make additional inspections to align the target inspections with the question.

The current experiment provides evidence for a relationship between the stored question information and the fixation pattern that occurs in the letter display. First, the type of question influenced whether the first eye movement was to the upper or lower part of the display, and, second, participants made additional inspections which resulted in an alignment of the fixation order to the order of the letters in the question.

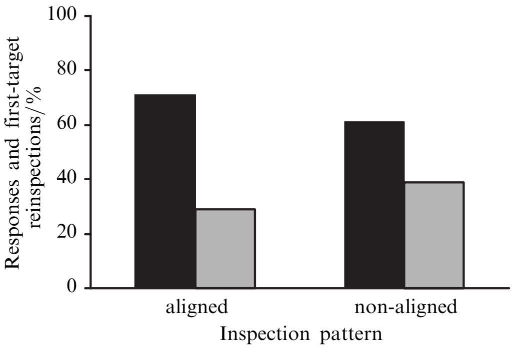

Figure 4. Percentage of responses (black bars) and first-target reinspections (grey bars) for aligned and non-aligned inspection patterns.

However, it is clear that in a significant minority of trials (16.1%) participants were able to answer the question following inspection of only one of the target items. Because the displays always included both target letters, it is possible for participants to adopt a specific but efficient problem-solving strategy. For example, if the question was "Is a better than c?" and the letter *a* was found by the first inspection in the bottom row, the question could be denied immediately. Likewise, the question "Is a less than c?" could be denied readily by finding letter *a* in the top row. If the relevant target letter is not found with the first inspection, the question can still be answered after inspecting the other letter in the same row. Thus, trying to deny rather than to verify the question is the most economic strategy as measured by the number of inspections necessary. However, participants inspected both targets in the majority of trials, but the presence of a small but substantial amount of responses based on a single target letter inspection suggests that this was a strategy adopted by some participants on some trials.

To reduce the use of such a strategy, in experiment 2 we replicated experiment 1, but inserted trials in which there was only one target letter and three distractor letters present in the display. Note that reasoning based on a single target letter inspection is still possible; however, the a priori probability of finding the critical target letter with the first inspection is reduced.

3 Experiment 2

3.1 *Method*

The design and procedure were almost identical to experiment 1. The only change was the introduction of displays in which only one of the target items was present.

3.1.1 *Stimuli and procedure.* The four letters for the display were selected from a set of five letters (*a*, *c*, *e*, *o*, and *u*). We randomly sampled 8 displays from the set of all displays with target letter *a* drawn above target letter *c*. The two distractor letters were taken from amongst *e*, *o*, and *u*. Likewise, 8 displays with letter *c* drawn above *a* were sampled. Combining the displays with four question types as before resulted in 64 problems requiring an equal amount of "yes" and "no" responses. Displays with both target letters present in the same row (as used in experiment 1) were replaced by displays with only one target letter (either *a* or *c*) and three distractor letters. To accomplish this, 8 displays with a target letter in the top row and 8 displays with a target letter in the bottom row were selected. They were each combined with two questions. Participants were instructed to respond "no" to the 32 resulting problems for which only a single letter from the question was present. To compensate for these 'no' problems a further 16 displays with both targets present in different rows were sampled and each combined with two questions in such a way that 'yes' problems resulted. This was done for each participant separately.

3.1.2 *Participants.* Sixteen undergraduate students (eight female, eight male) with normal or corrected-to-normal vision served for class credit. Their average age was 19.3 years, ranging from 18 to 22 years.

3.2 *Results and discussion*

The percentage of correct responses was 97.2% overall, with two responses given after the timeout coded as incorrect. As before, data for which fixation was not maintained between the question display and the letter display were also excluded. The resulting correct data (92.3%) were analysed. To allow a direct comparison with experiment 1, we analysed only the trials in which both target letters were present.

The mean response time per trial was 1468 ms on average. Response times were 78 ms faster for 'better' questions compared to 'less' questions: $F_{1,15} = 11.32$, $p < 0.01$. We found no other effects.

The mean number of inspections per trial was 3.93. Here we found no significant effect of question type.

3.2.1 *First inspection is modulated by question type.* As in experiment 1 most of the first gaze shifts were made to top letters (56.1% top left, 17.6% top right), as opposed to bottom letters (21.8% bottom left, 4.4% bottom right). The first gaze shift strongly depended on the type of question (see figure 5). The percentage of first inspections on bottom letters is even slightly higher (1%) for 'less' questions than for 'better' questions. Thus, comparing the difference of the percentages between top letters and bottom letters for 'better' questions (45% on average) with the aforementioned 1% difference for 'less' questions confirmed the pattern of results ($t_{1,15} = 6.41$, $p < 0.01$).

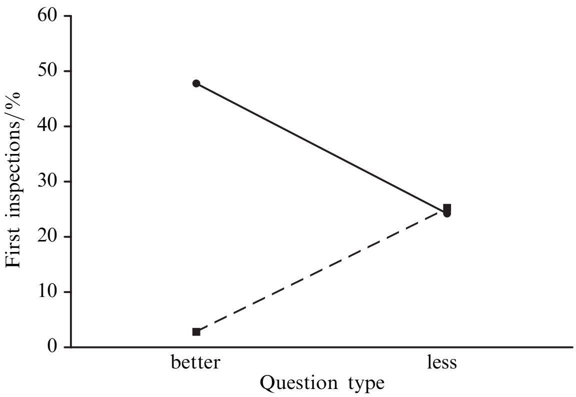

Figure 5. Percentage of first inspections on the upper letters in the display (solid line) versus lower letters (dashed line) plotted by question type.

3.2.2 *Reinspections on the question letters.* Target-letter inspections accounted for about 57.1% of all inspections, while the remaining inspections were equally distributed over distractor letters.

Participants fixated one target in 8.6% of all trials. Trials without a target inspection no longer occurred. For 43.8% of the trials participants fixated both targets exactly once. Reinspections (of targets or distractors) after both targets had been fixated at least once occurred in 675 trials (47.6%). Participants fixated both targets at least once before responding in 91.4% of all trials.

The 12 most common gaze inspection patterns and their relative frequencies obtained in this experiment are depicted in figure 6. These 12 patterns accounted for 83.9% of all trials. The 5 most common patterns which account for 59.1% of all trials, all involve the inspection of both target letters and one or two distractor-letter inspections. Only 5.8% of trials involved patterns in which a single target-letter inspection occurred. 25.0% of trials involved patterns in which target-letter reinspections occurred.

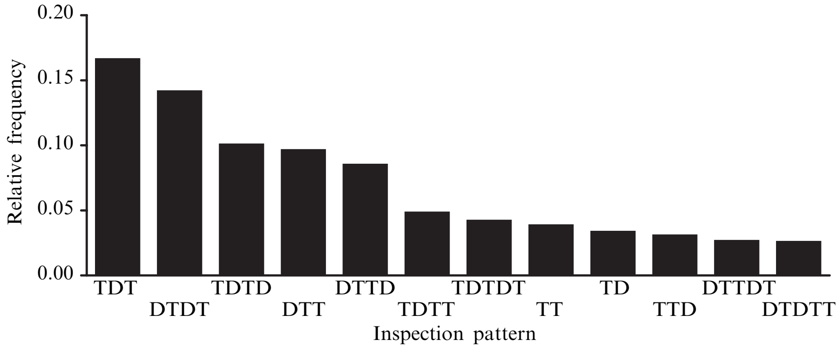

Figure 6. Relative frequency of the 12 most common inspection patterns. Patterns are represented by sequences of letters; a T represents inspection of the target, and a D represents inspection of distractors.

3.2.3 *Reinspections on targets resulted in inspection – question alignment.* Target-letter inspections were aligned with the question in 722 cases; 499 patterns were non-aligned. If the inspection pattern was aligned, participants responded, ie terminated the search, without a target reinspection in 88.0% of the trials; in 12.0% of the trials they refixated the first target. If the inspection pattern was non-aligned, responses were made in 65.8% of the trials, while the first target was refixated in 34.2% of the trials (see figure 7).

This result shows that participants responded more readily when they had fixated the target letters in the sequence provided by the question. If that was not the case (non-aligned patterns), first-target reinspections were increased by 22.2% ($F_{1,15} = 30.05$, $p < 0.01$).

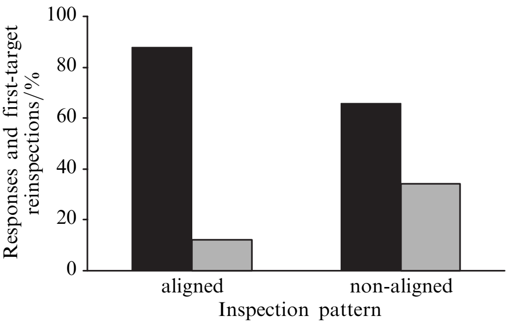

Figure 7. Percentage of responses (black bars) and first-target reinspections (grey bars) for aligned and non-aligned patterns.

In the current experiment we replicated the two basic results from experiment 1: the type of question again influenced whether the first eye movement was to the upper or lower part of the display and participants made additional inspections which resulted in an alignment of the fixations to the order of the question. These results provide evidence for a relationship between the stored question information and the fixation pattern that occurs in the letter display. In this experiment, by introducing trials with a single target item, we eliminated the benefit of searching for a single item and this, in turn, reduced the number of trials in which a single target was fixated prior to responding. It is, of course, still possible to answer the question in this way but it is no longer an optimal strategy across trials.

4 General discussion

In two experiments we investigated the role of eye movements in a spatial reasoning task. We have found effects of the stored linguistic information from the question on where participants look first and on the pattern of their inspections.

The number of inspections might appear lavish, given that the task can be carried out with only a few. One explanation for this may come from the relative effort associated with moving the eyes and reasoning. Multiple eye movements can be carried out relatively easily while the cognitive cost associated with inferences made on the basis of partial information about target location might be substantial. The number of eye movements then simply reflects a payoff between the effort associated with the manipulation and storage of display information and the effort of moving the eyes to pick up information directly. Participants clearly shifted the load of the reasoning task away from abstract cognitive manipulations (inferences and integration) of target and distractor information. Rather, they 'acted out' with their eye movements the information that was provided with the question (see Ballard 1991; Brooks 1986; Findlay and Gilchrist 2003). The claim is supported by the following findings from the current experiments.

First, target letters were inspected more frequently than distractors even though distractor fixation can support reasoning in this context. This suggests that participants generally avoided inferences from distractor locations about the target-letter locations. Instead, they preferred verifying the target-letter relation directly by moving their eyes to both targets.

Second, we have shown that participants' first inspections are strongly influenced by the type of question asked. For certain display – question combinations, the task could be accomplished with a single target inspection if the participants tried to deny rather than to verify the assertion implied by a question. This strategy would imply first saccades to the bottom row of the display for 'better' questions and top-row first saccades for 'less' questions. However, this was not what participants did. In both experiments, their first inspections depended substantially on the question type but the other way round: If we presented participants a 'less' question, they were more likely to start their search in the bottom row of the display as compared with the presentation of a 'better' question. It is, of course, possible that participants planned this first eye movement on the basis of the question type alone, during the 1000 ms fixation period prior to the display of the spatial array. However, this possibility does not affect our basic conclusion for a link between the eye movement and the question structure. This finding suggests that participants may try to verify rather than to deny the assertion implied by a question. Denying a question on the basis of one or two inspections in the row not suggested by the question may demand a costly reasoning effort. Making more eye movements in order to inspect both target letters and reason about their relation without inferences from information provided from just one or two inspections might be less costly.

Finally, our results show that reinspections of target letters can be partly explained by an alignment of inspections and questions. Although target reinspections still occurred when participants had inspected the targets in the sequence provided by the question, reinspections were more likely when question and inspection sequence were non-aligned. Conversely, participants responded more readily, ie without reinspection, when inspection sequence and question were aligned. The first (non-aligned) target fixation may perhaps only serve to establish the location of the target, while the actual spatial reasoning is carried out with the reinspection which aligns the inspection sequence with the question. Again, the additional eye movements necessary for alignment may be less costly than additional reasoning.

In our experiments, participants showed a strong tendency not to rely on cognitively sophisticated spatial reasoning strategies but instead make multiple eye movements to reduce the cognitive load (see also Hayhoe et al 1998). The number and pattern of these eye movements were on the surface very similar to the structure of the question.

This suggests, in turn, that the question was encoded in a relatively figurative manner that retained the order of the letters in the question and the type of question rather than extracting the underlying logical meaning of the question.

In summary, the current results provide further support for the ability of stored, non-visual, information to influence the pattern of eye movements generated (eg Altmann and Kamide 1999; Brandt and Stark 1997; Demarais and Cohen 1998; Richardson and Spivey 2000; Spivey and Geng 2001). In addition, the current results support a view in which perception and eye movements are an integral part of cognition (eg Barsalou 1999). Sequences of eye movements can deliver visual information in a manner that allows an efficient mapping with the stored information. Organising the visual information at input may be more efficient than the more costly 'mental' manipulation.

Acknowledgment. This work was supported by a Marie Curie Fellowship grant (HPMF-CT-2000-00986) from the European Commission and by The Wellcome Trust.

References

Altmann G T M, Kamide Y, 1999 "Incremental interpretation at verbs: restricting the domain of subsequent reference" *Cognition* **73** 247–264

Ballard D H, 1991 "Animate vision: An evolutionary step in computational vision" *Journal of the Institute of Electronics, Information, and Communication Engineers* **74** 343–348

Barsalou L W, 1999 "Perceptual symbol systems" *Behavioral and Brain Sciences* **22** 577–660

Bouma H, 1970 "Interaction effects in parafoveal letter recognition" *Nature* **226** 177–178

Brandt S A, Stark L W, 1997 "Spontaneous eye movements during visual imagery reflect the content of the visual scene" *Journal of Cognitive Neuroscience* **9** 27–38

Brooks R, 1986 "A robust layered control system for a mobile robot" *IEEE Journal of Robotics and Automation* **2** 14–22

Demarais A M, Cohen B H, 1998 "Evidence for image-scanning eye movements during transitive inference" *Biological Psychology* **49** 229–247

Findlay J, Gilchrist I D, 2003 *Active Vision: The Psychology of Looking and Seeing* (Oxford: Oxford University Press)

Hayhoe M M, Bensinger D G, Ballard D H, 1998 "Task constraints in visual working memory" *Vision Research* **38** 125–137

Körner C, 2004 "Sequential processing in comprehension of hierarchical graphs" *Applied Cognitive Psychology* **18** 467–480

Körner C, Albert D, 2002 "Speed of comprehension of visualized ordered sets" *Journal of Experimental Psychology: Applied* **8** 57–71

Rayner K, Sereno S C, Morris R K, Schmauder A R, Clifton C, 1989 "Eye movements and on-line language comprehension processes" *Language and Cognitive Processes* **4** 21–49

Richardson D C, Spivey M J, 2000 "Representation, space and Hollywood Squares: Looking at things that aren't there anymore" *Cognition* **76** 269–295

Spivey M J, Geng J J, 2001 "Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects" *Psychological Research* **65** 235–241

Yarbus A L, 1967 *Eye Movements and Vision* (New York: Plenum)

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 33 2004

www.perceptionweb.com