What Do Eye Movements Reveal About Mental Imagery? Evidence From Visual And Verbal Elicitations

Roger Johansson (roger_json@hotmail.com)
Department of cognitive science, Lund University
Kungshuset, Lundagård, 222 22 Lund

Jana Holsanova (jana.holsanova@lucs.lu.se)
Department of cognitive science, Lund University
Kungshuset, Lundagård, 222 22 Lund

Kenneth Holmqvist (kenneth@lucs.lu.se)
Department of cognitive science, Lund University
Kungshuset, Lundagård, 222 22 Lund

Abstract

This paper shows evidence that eye movements reflect the positions of objects during the description of a previously seen picture; while listening to a spoken description, and during the retelling of a previously heard spoken description. This effect is equally strong in retelling from memory irrespective of whether the original elicitation was spoken or visual. In two experiments, eye movements were recorded while test subjects recalled objects that were either previously observed in a complex picture or presented in a verbal description. In both cases, the subjects spontaneously looked at regions on a blank board that reflected the spatial locations of the objects they recalled. These results contribute to evidence that the eyes are connected with the cognitive processes that occur during imagery. In the discussion the results are related to the current debate on mental imagery.

Keywords: Eye-movements; Mental-imagery; Perception; Descriptions; Attention

Introduction

There is strong evidence that the eye movements during visual scanning of a scene reappear in the eye movements that occur during a mental visualization of the same scene. Brandt and Stark (1997) showed that spontaneous eye movements occur during visual imagery and that these eye movements closely reflect the content and spatial relations from the original picture or scene. In this study the subjects were first introduced to a simple visual grid pattern that they should memorize, and shortly afterwards they were asked to imagine the pattern. Holsanova et al (1998) found similar results as Brandt and Stark (1997), with the difference that the original picture showed a natural, real life scene. Laeng and Teodorescu (2001) replicated and extended Brandt and Stark's experiment and showed that subjects who fixed their gaze centrally during a scene perception did the same, spontaneously, during imagery. They also showed that subjects free to explore a pattern during perception, when required to maintain central fixation during imagery, exhibited a decreased ability to recall the pattern.

There is also indication that eye movements reflect verbally constructed scenes. Demarais and Cohen (1998) demonstrated that subjects that solved auditory presented syllogisms containing the words "left" and "right" elicited more horizontal eye movements, and syllogisms containing "above" and "below" elicited more vertical eye movements. Spivey and Geng (2001) extended Demarais and Cohens experiments and showed that subjects listening to a spatial scene description tend to make eye movements in the same directions as in the described scene. It appears that similar eye movements appear during visualization from verbal descriptions as from pictures. But it is still not known whether the effect is equally strong.

An explanation to phenomena of these types could be that eye movements reflect an internal mental image that are constructed in a "visual buffer" (e.g. Kosslyn, 1994) of the working memory. In this visual buffer it is possible to shift attention to certain parts or aspects of the mental image. Eye movements would thus somehow be connected with these attention shifts. Mast and Kosslyn (2002) argues that eye movements are stored as spatial indexes that are used to arrange the different parts of the mental image correctly. However, the mental image does not necessarily have *all* the properties (for instance, detail) as a real picture, only *some* (for instance, spatial extension) (Finke, 1989).

Pylyshyn denies the existence of a visual buffer and suggests that there are no similar properties between perception of an object and the mental representation of this object, and claims that all our mental representations are of the same functional nature (e.g. Pylyshyn, 2002).

The purpose of the present study is to extend the experiments by Brandt and Stark (1997), and Laeng and Teodorescu (2001) by studying eye movements when subjects visualize a more complex picture then the simple grids they used. The study also attempts to extend the experiments by Spivey and Geng (2001) by studying eye movements when subjects listen to a complex description that they are to mentally visualize.

We present two studies of eye movements in three description tasks: During the description of a previously seen picture; while listening to a spoken description, and during the retelling of a previously heard spoken description. The hypothesis under examination is that the eye movements in all three cases indicate the spatial locations of objects from the picture and the description, respectively. We also test the hypothesis whether the effect is stronger in either condition.

Experiment 1

In this experiment subjects viewed a picture which was later to be orally described. The chosen picture included many objects with rich detail and clear spatial relations. Based on previous research (Brandt & Stark, 1997; Holsanova, et al, 1998; Laeng and Teodorescu, 2001) it is hypothesized that the spatial positions of objects in a naturalistic picture are reflected by the eye movements of subjects who describe it while not seeing it.

Participants

Twelve students at the University of Lund, 6 females and 6 males volunteered to participate in an experiment in cognitive science. All subjects reported normal vision, or corrected to normal (with contact lenses or glasses). The participants were told that their pupil size was being measured during a visualization task. At the end of each session, participants were questioned about their beliefs about the purpose of the experiment. It was confirmed that all participants were naive about the fact that their eye movements were recorded and that they had no specific knowledge about the experimenters' expectations.

Apparatus and procedure

The eye tracker that was used is an SMI iView X 50 Hz pupil and corneal reflex imaging system. The eye tracker consists of a headset, with magnetic head-tracking, which allows the subject freedom of motion of the head. The outputs of the system were an MPEG video and a file with eye movement coordinates.

The visual stimuli used in the experiment consisted of a complex picture (500mm × 700mm) (figure 1) and a white board. The participants were seated in front of the picture or the white board at a distance of 150 cm.

The experiment consisted of two main phases, one perception phase and one retelling phase. Eye movements were recorded both in the perception phase and in the retelling phase.

The picture was shown for about 30 seconds. When the description finished the subject was told to describe the picture freely with his or her own words. The subjects where also specifically told to keep their eyes open during this phase, but that they were free to look where ever they wanted on the white board.

Figure 1: The picture

Analysis

To analyze the data the test subjects' descriptions were first transcribed so it was possible to analyze *when* certain objects are mentioned. The analysis of the eye data was done with an eye-tracking analysis program, *iView for Windows*, which can trace the saccades and fixations of the subject's eyes over time (an example of this can be seen in experiment 2).

While it is impossible to define an actual physical coordinate of an area of interest on the white board – it is e.g. possible to imagine the scene on the whole white board or on a certain part of it – it is not useful to analyze the actual physical location of eye movements like Brandt and Stark (1997), and Laeng and Teodorescu (2001) did. Instead a method analyzing the relative position of an eye movement compared to the overall structure of the scanpath was developed. To achieve this method of analysis eye movements of the test subjects were scored as either *high correspondence, low correspondence* or *no correspondence*. Eye movements to objects were considered correct in high correspondence when fulfilling the following criteria:

> <u>High correspondence</u>
> The eye movement to an area of interest must *finish in a position that is spatially correct relatively to the subject's eye tracking pattern over the entire description*.

While several experiments have shown that subjects rotate, change size, change shape, change colour, and reorganize and reinterpret mental images (Finke, 1989). Such image transformations may affect our results, in particular if they take place in the midst of the descriptions. Therefore we devised an alternative low correspondence measure:

> <u>Low correspondence</u>
> When an eye movement is moving from one area of interest to another during the description it must *move in the correct direction*.

The key difference between high and low correspondence is that high correspondence requires fixations to take place at the categorically *correct spatial position* relative to the whole eye tracking pattern. Low correspondence only requires that the eye move in the *correct direction* between two consecutive objects in the description. Schematics of this can be seen in figure 2 (the low correspondence eye movement is illustrated in grey colour).

No correspondence was considered if neither the criteria for low correspondence or high correspondence is fulfilled.

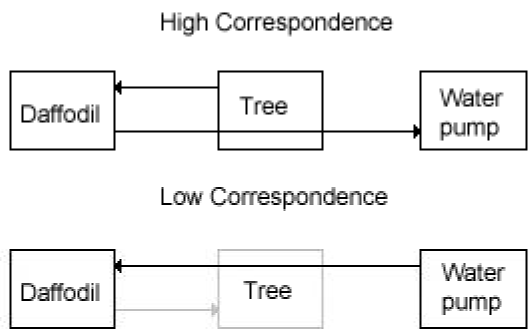

Figure 2: High and low correspondence

As a consequence of applying this spatial criterion a binominal distribution in the data is obtained: the spatial relations are either correct or not (for each coding). We then defined the possibility that a test subject would move his or her eyes to the correct position by chance. For high correspondence coding, both the direction and the distance of the movement must be correct. There are many possible movements. A conservative estimate is that the eyes can move in at least 4 directions (up, down, left, right). For each direction, they can move at least to two different locations (full and half distance). In addition to these eight possibilities, the eye can stand still. For high correspondence, the probability that the eyes move to the correct position at the correct time is thus definitely less than 1/9. For low correspondence coding, we only require correct direction, and thus the low correspondence probability is 1/5. We used the Wilcoxon Signed-Ranks test for significance between the number of correct eye movements and the expected number of correct movements by chance.

Experiment 2

In this experiment subjects listened to a prerecorded verbal description which was later to be orally retold. The description was designed to include objects with clear spatial relations. On the basis of previous research on imagery and verbal descriptions (Demarais & Cohen, 1998; Spivey & Geng, 2001) it is hypothesized that subjects' eye movements should reflect the positions of objects in the description, both while listening to the description and while retelling it.

Participants

Twelve new students at the University of Lund, 6 females and 6 males, volunteered to participate in an experiment in cognitive science. All subjects reported normal vision, or corrected to normal (with contact lenses or glasses). The participants were told that their pupil size was being measured during a visualization task. At the end of each session, participants were questioned about their beliefs about the meaning of the experiment. It was confirmed that all participants were naive about the fact that their eye movements were recorded and that they had no specific knowledge about the experimenters' expectations.

Apparatus and procedure

The eye tracker and its output were the same as in experiment 1.

The visual stimulus used in the experiment consisted of a white board (657mm × 960mm), and the auditory stimulus used in the experiment consisted of a prerecorded description (2 minutes and 6 seconds). The participants were seated in front of the white board at a distance of 150 cm. The prerecorded description was the following (here translated to English):[1]

> "Imagine a two dimensional picture. At the center of the picture a large green spruce grows. In the top of the spruce a bird is sitting. To the left of the spruce and to the far left in the picture there is a yellow house with a black tin roof and white corners. The house has a chimney on which a bird sits. To the right of the large spruce and to the far right in the picture a tree grows, which is as high as the spruce. The leaves of the tree are colored in yellow and red. A bit above the tree at the top of the picture a bird flies. Between the spruce and the tree there stands a man in a blue overall, who is raking leaves. In front of the spruce, the house, the tree and the man, i.e. below them in the picture, there is a long red fence, which runs from the pictures left edge to the pictures right edge. At the left edge of the picture, a bike is leaning towards the fence, and just to the right of the bike there is a yellow mailbox. On top of the mailbox a cat is sleeping. In front of the fence, i.e. below the fence in the picture there is a road, which goes from the pictures left edge to the pictures right edge. On the road, to the right of the mailbox and the bike, a black haired girl is bouncing a ball. To the right of the girl a boy who wears a red cap is sitting and watching her. To the far right on the road walks a lady who is wearing a big red hat and who has books under her arm. To the left of her, on the road, a bird is eating a worm."

The experiment consisted of two main phases, one description phase in which the participants listened to the verbal description and one retelling phase in which the participants with own words retold the description they had listened to. Eye movements were recorded both while

[1] The initial Swedish verb was "Föreställ dig…" which is neutral to the modality (image or word) of thinking. "Föreställ dig…", like its German equivalent "Stell dich vor…" is not as visual as the English "Imagine…".

subjects listened to the verbal description and while they retold it.

When the description finished the subject was told with own words freely to describe the scene. The subjects where also specifically told to keep their eyes open during this phase, but that they were free to look wherever they wanted on the white board.

Analysis

To analyze the data the test subjects' descriptions were as in experiment 1 transcribed so it was possible to analyze when certain objects are mentioned.

The eye movements of all test subjects were also as in experiment 1 scored as (and with the same criteria) *high correspondence*, *low correspondence* and *no correspondence*. This was done for both the description and the retelling of it.

The analysis of the eye data was also analyzed in the same way as in experiment 1, and again done with the eye-tracking analysis program, *iView for Windows*. An example of this is shown in figure 3, where the fixations (rings) and saccades (lines) are present one minute and seven seconds into the description – the spruce in the centre, the house to the left, the tree to the right, the man between the spruce and the tree, and below them the fence from the left to the right edge.

The Wilcoxon Signed-Ranks test was as in experiment 1 used to test direction significance both during the description and the retelling of it.

Figure 3: iView analysis

Results and discussion

Average values (based on every single eye movement for all of the subjects) of how well the eye movements corresponded - either they did or they did not - in both low correspondence and high correspondence coding, for both experiments, are presented in table 1.

The direction significance for both low correspondence and high correspondence coding for both experiments are presented in table 2.

As can be seen in table 2 the eye movements during the retelling of a picture (experiment 1), during a description and during the retelling of that description (experiment 2) did show significant direction correspondence for both the low correspondence coding and for the high correspondence coding. Table 1 indicates that the effect we measured is strong. More than half of all objects mentioned – in both experiment 1 and 2 – had correct eye movements, according to the conservative high correspondence criteria. Allowing for re-centering and resizing of the image – as with low correspondence – makes almost three quarters of all objects have correct eye movements.

An interesting observation of table 1 is that the low correspondence coding results were better when the subjects retold the verbal description than when they listened to it (64.3% and 74.9%). A possible explanation to this could be that not all eye movements that appeared were related to image scanning. It has been found that eye movements that appear during verbal tasks may be caused by general arousal, orienting reactions, and/or in cognitive change (Demerais & Cohen, 1998). It is possible that eye movements appeared because of these effects and sometimes were included in the low correspondence coding. However, in the high correspondence coding, with the higher demand on the results, this difference between listening to the description and retelling it was not found. As can be seen in table 1 the results during the verbal description and the retelling of it are almost identical (54.8% and 55.2%).

Although, the most interesting observation from table 1 is that the results of the retelling of what could be seen in the picture (experiment 1) and the retelling of the verbal description (experiment 2) were also almost identical (74.8% and 74.9% for low correspondence coding, and 54.4% and 55.2% for high correspondence coding). These results consequently indicate that the effect is equally strong for eye movements generated by the complex picture as for eye movements generated by the verbal description.

Table 1: Average values

	Exp1: Retelling	Exp2: Descript.	Exp2: Retelling
Low corr.	74,8%	64,3%	74,9%
High corr.	54,4%	54,8%	55,2%

Table 2: Direction significance

	Exp1: Retelling	Exp2: Descript.	Exp2: Retelling
Low corr.	p = .0015	p = .0026	p = .0012
High corr.	p = .0051	p = .0026	p = .0040

General discussion

Our results could be interpreted as further evidence that eye movements play a functional role in visual mental imagery and that eye movements indeed are stored as spatial indexes that are used to arrange the different parts correctly when a mental image is generated in a visual buffer.

Pylyshyn (2002) disagrees with this interpretation. He argues that studies that support internal mental images occur because of *tacit knowledge*. The knowledge of *what things would look like* to subjects in situations like the ones in which they are to imagine themselves, i.e. when subjects are asked to "imagine x" they use their knowledge of what "seeing x" would be like and they *simulate as many of these effects as they can* (Pylyshyn, 2002). Eye movements then merely mimic the behavior we have during perception.

Our subjects were not asked to imagine anything, we argue, only to "föreställa sig" in one condition in one of the two experiments. The major instruction was to describe. Yet it is obvious that in the listening phase of our Experiment 2, subjects must have had knowledge about what houses and various trees look like, and what "between" and "a bit above" is. But our subjects also incorporated these sub-meanings into a larger whole – the scene – that allowed them to make later eye movements to correct places relative to earlier established positions. High correspondence results can only occur if there is a working memory holding these various spatial positions of different parts active. As for the content of that working memory, it appears considerably simpler to store spatial scene information as one image, as suggested by Laeng and Teodorescu (2002), than as a large collection of propositional statements. Additionally, it seems very unlikely that we are able to mimic so precisely a behavior in eye movements that entire scenes of objects with correct spatial locations are built up, as the high correspondence results indicate. The number of points and the precision of the eye movements to them are too high to be remembered without a support (like an internal image) to tie them together in a context.

A possible support could, however, be the external world, i.e. that we are able to use our environment as an external memory store (O'Regan, 1992).Visual features in the external world are then used as *visual indexes*, i.e. the eyes move to external features in the actual world that are used to bind the spatial locations of the "mental image" (Pylyshyn, 2001; 2002). The white board (that was used in the imagery phase of all the experiments) was plain and completely uniform in color, i.e. it did not possess any visual features that could be used as visual indexes. However, it could be argued that during the visualization the frame or very slight features on the board were used as visual indexes. For example, the dragonfly was positioned about ten percent from the right edge of the frame and about twenty-five percent from the top edge of the frame, and therefore I move my eyes to that position during the visualization. To test if visual indexes could be the explanation to our results the experiments have to be replicated in complete darkness.

In recent years – inspired by embodied cognition – other approaches to mental imagery have been developed. Thomas (1999) "perceptual activity (PA) theory" suggests that perception is "active" in a similar way as active vision systems in robotics. Perception is then not about storing mental images or shifting attention in a visual buffer. No thing in the brain *is* the image. Instead we store a continually updated and refined set of procedures or schemas that *specify how to direct out attention* in different situations (Thomas, 1999). A perceptual experience consists in the ongoing activity of a schema-guided perceptual exploration of the environment. Imagery is than the *reenactment* of the specific exploratory perceptual behavior that would be appropriate for exploring the imagined object if it were actually present. In this reenactment the procedure or schema sends some of its "orders" to lower level motor processes, like eye movements. In this approach we always encode *how* to direct our attention and eye movements thus happens when we "act out" how we would visually explore a scene. Although the perceptual activity theory seems to have a plausible way to explain eye movement effects of this type it does not explain how the procedures that generate the eye movements actually work.

A somewhat similar approach is favored by Barsalou and his theory of perceptual symbol systems (1999). A perceptual symbol is *not* a mental image but a record of the neural activation that arises during perception. Imagery is then the re-enactment or *simulation* of the neural activity. These simulations do not contain only sensory states but also motor (e.g. eye movements) and mental states, but might contain distortions and are never complete re-enactments of the originally neural activity. Remembering something that occurred in a specific spatial location would thus during the re-enactment make the eyes more likely to revisit that location than others.

Summary

To summarize, this research has provided new evidence that the eye movements that occur during the visualization, both for a complex verbal description and for a complex picture, do reflect the spatial locations of objects that appear in the description and in the picture. These results were just as strong for eye movements generated by the complex picture as for eye movements generated by the complex verbal description. It is hard to explain these results without internal images and a visual buffer.

We have argued that our results can not be explained in terms of tacit knowledge. However, it could be argued that the eye movements occur because of visual indexes in the external world, or that they are a product of procedures or neural re-enactments that make us experience mental images.

Acknowledgements

The authors wish to thank Juliane Steinberg and Halfdan Koot-Fryd for discussions on methodology and theory.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-660.

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.

Demarais, A., & Cohen, B.H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology*, *49*, 229-247.

Finke, R.A. (1989). Principles of Mental Imagery, Cambridge, MA: MIT Press.7

Holsanova, J. , Hedberg, B. & Nilsson, N. (1998). Visual and Verbal Focus Patterns when Describing Pictures. In Becker, Deubel & Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects*. Plenum: New York, London, Moscow.

Kosslyn, S. M. (1994). *Image and brain.* Cambridge, Mass.: The MIT Press.

Laeng, B., & Teodorescu, D-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science*, 26, 207-231.

Mast, F. W., & Kosslyn, S. M. (2002). Eye movements during visual mental imagery, *TRENDS in Cognitive Science,* Vol. 6, No. 7.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory, *Canadian Journal of Psychology,* 46:3, 461-488.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision, *Cognition*, 80 (1/2), 127-158.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory, *Behavioral and brain sciences*, 25 (2), 157-238.

Spivey, M., & Geng, J., (2001). Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects, *Psychological Research,* 65, 235-241.

Thomas, N. J. T., (1999). Are Theories of Imagery Theories of Imagination? An Active Perception Approach to Conscious Mental Content. *Cognitive Science, vol. 23 (2).*