Cognitive Science 30 (2006) 1053–1079

Pictures and Spoken Descriptions Elicit Similar Eye Movements During Mental Imagery, Both in Light and in Complete Darkness

Roger Johansson, Jana Holsanova, Kenneth Holmqvist

Department of Cognitive Science, Lund University

Received 7 December 2004; received in revised form 28 April 2006; accepted 5 June 2006

Abstract

This study provides evidence that eye movements reflect the positions of objects while participants listen to a spoken description, retell a previously heard spoken description, and describe a previously seen picture. This effect is equally strong in retelling from memory, irrespective of whether the original elicitation was spoken or visual. In addition, this effect occurs both while watching a blank white board and while sitting in complete darkness.

This study includes 4 experiments. The first 2 experiments measured eye movements of participants looking at a blank white board. Experiment 1 monitors eye movements of participants on 2 occasions: first, when participants listened to a prerecorded spoken scene description; second, when participants were later retelling it from memory. Experiment 2 first monitored eye movements of participants as they studied a complex picture visually, and then later as they described it from memory. The second pair of experiments (Experiments 3 and 4) replicated Experiments 1 and 2 with the only difference being that they were executed in complete darkness. This method of analysis differentiated between eye movements that are categorically correct relative to the positions of the whole eye gaze pattern (*global correspondence*) and eye movements that are only locally correct (*local correspondence*). The discussion relates the findings to the current debate on mental imagery.

Keywords: Imagery; Attention; Perception; Visual and verbal elicitation; Darkness; Light; Eye movements; Visual deixis

1. Introduction

It has been proposed that we use imagery when we mentally invent or recreate personal experience (Finke, 1989; Kosslyn, 1980), when we retrieve information about the physical properties of objects or about physical relations among objects (Finke, 1989; Kosslyn, 1980), when we plan future events, when we imagine transformations by mental rotation and mental anima-

Correspondence should be addressed to Kenneth Holmqvist, Humanistlaboratoriet, Språk och Litteraturcentrum, Lund University, Box 201, S-221 00 Lund, Sweden 22100. E-mail: kenneth@lucs.lu.se

tion (e.g., Hegarty, 1992), and when we solve problems (e.g., Huber & Kirst, 2004; Yoon & Narayanan, 2004).

Several recent studies have found a close relation between eye movements and mental imagery. Brandt and Stark (1997) and Laeng and Teodorescu (2002) showed that spontaneous eye movements occur in participants asked to do visual imagery and that these eye movements closely reflect the content and spatial relations of the original picture or scene. A similar eye movement effect has also been found for spatial relations that are verbally described. Demarais and Cohen (1998) demonstrated that participants tend to move their eyes in the same direction as simple relational terms; and Spivey and Geng (2001) and Spivey, Tyler, Richardson, and Young (2000) showed that participants listening to a spatial scene description make eye movements that correspond to directions indicated in the described scene. In the absence of a visual input, participants have to construct a spatial mental model of an object from linguistic input (e.g., Bower & Morrow, 1990; Johnson-Laird, 1981). Although it appears that similar eye movements occur when visualizing verbal descriptions as when visualizing previously seen pictures, it is still not known whether the effect is equally strong.

Brandt and Stark (1997) and Laeng and Teodorescu (2002) interpreted their results as eye movements that reflect an *internal image* (Finke, 1989; Kosslyn, 1994) that is used to remember and visualize the locations and spatiality of the imagined scene. The existence of internal images is, however, an issue of debate. Pylyshyn (2002, 2003), for example, argued that there are no such things as internal images in the brain. Instead, all of our mental representations are propositional (Pylyshyn, 2002, 2003).

Spivey and Geng (2001) and Spivey et al. (2000) had a different approach than Brandt and Stark (1997) and Laeng and Teodorescu (2002). They interpreted their results in an "embodied" view of the mind (e.g., Ballard, Hayhoe, Pook, & Rao, 1997; Spivey, Richardson, & Fitneva, 2004) whereby motor processes, such as eye movements, are naturally and tightly coupled with the perceptual and cognitive processes that subserve mental representations.

We present four studies of eye movements of participants during three description tasks: while participants listen to a spoken description, retell a previously heard spoken description, and describe a previously seen picture. These tasks were executed both while looking at a blank white board and in complete darkness. The first hypothesis is that the eye movements in all three cases indicate the spatial locations of objects from the picture and from the description. The second hypothesis is that this effect is equally strong in retelling from memory, irrespective of whether the original elicitation was spoken or visual. The third hypothesis is that the eye movements in all three cases also indicate the spatial locations of objects from the picture and the description in complete darkness.

2. Experiment 1: Listening to and then retelling a spoken description

Our first experiment concerned the issue of imagery and scenes described in speech. Demarais and Cohen (1998) demonstrated that participants who solved spoken syllogisms containing the words "left" and "right" exhibited more horizontal eye movements, and syllogisms containing "above" and "below" exhibited more vertical eye movements. Spivey and Geng (2001) extended Demarais and Cohen's experiments and showed that participants who listened to a spatial scene description, and who were instructed to imagine it, tended to make eye movements in the same directions as indicated in the described scene. The descriptions were of the following type:

> Imagine a train extending outwards to the left. It is pointed to the right, and you are facing the side of the engine. It is not moving. Five cars down is a cargo holder with pink graffiti sprayed on its side. Another six cars down is a flat car. The train begins to move. Further down the train you see the caboose coming around a corner.

This experiment was also followed by another that demonstrated that participants tended to make eye movements in the same directions as in a description even when their eyes were closed and when they had no instructions to imagine anything (Spivey et al., 2000).

The goal of Experiment 1 was to extend the previous studies (Demarais & Cohen 1998; Spivey & Geng 2001; Spivey et al., 2000) in two respects: First, instead of only studying simple directions, we focused on complex spatial relations (expressions like *at the center, at the top, between, above, in front of, to the far right, on top of, below, to the left of*). Second, apart from measuring eye movements during the listening phase, we added a retelling phase in which the participants were asked to retell the described scene from memory. Eye movements were measured during both phases. To our knowledge, these aspects had not been studied before. In addition, we collected ratings of the vividness of imagery during both the listening phase and the retelling phase, and asked participants whether they usually imagine things in pictures or words.

We hypothesized that spatial positions of objects would be reflected by the eye movements of participants, both while participants listened to the description and while they retold it.

2.1. Participants

Twelve students at the University of Lund (6 women and 6 men) volunteered to participate in an experiment in cognitive science. All participants reported normal vision or vision corrected to normal (with contact lenses or glasses).

2.2. Apparatus and stimuli

The eye tracker used was an SensoMotoric Instruments (SMI) iView X infrared pupil and corneal reflex imaging system, sampling data at 50 Hz. The eye tracker consisted of a bicycle helmet with a scene camera and an eye camera. With magnetic head tracking, the helmet allowed the participant freedom of motion of the head. Glasses and lenses typically presented no problem (whereas mascara did). The outputs of the system were an MPEG video with sound and overlaid gaze mark, as well as a file with eye movement coordinates for each participant.

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

> Imagine a two-dimensional picture. There is a large green spruce growing at the center of the picture. There is a bird sitting at the top of the spruce. To the left of the spruce, and at 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 is sitting. There is a tree growing to the right of the large spruce and to the far right in the picture. It is as tall as the spruce. The leaves of the tree are yellow and red. A bird is flying a bit above the tree at the top of the picture. Between the spruce and the tree there is a man standing in blue overalls 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 that runs from the picture's left edge to the picture's right edge. At the left edge of the picture, there is a bike leaning up against the fence, and just to the right of the bike there is a yellow mailbox. There is a cat sleeping on top of the mailbox. In front of the fence, i.e. below the fence in the picture, there is a road that goes from the picture's left edge to the picture's right edge. On the road, to the right of the mailbox and the bike, there is a black-haired girl bouncing a ball. To the right of the girl a boy wearing a red cap is sitting and watching her. To the far right on the road, a lady wearing a big red hat and carrying books under her arm is walking. To the left of her, on the road, a bird is eating a worm.

2.3. Procedure

Throughout, participants were told that the experiment concerned pupil dilation during the retelling of descriptions held in memory. It was explained to them that we would be filming their eyes, but nothing was said about our knowing in which directions they were looking. They were asked to keep their eyes open so that we could film their pupils, and to look only at the white board in front of them so that varying light conditions beyond the board would not disturb the pupil dilation measurements. The eye tracker was then calibrated; participants looked at 13 successive positions across the board as indicated by a laser pointer. Calibration typically took less than 20 sec.

The experiment consisted of two main phases: one *listening* phase in which the participants listened to the verbal description and one *retelling* phase in which the participants retold in their own words the description they had listened to. Eye movements were recorded both while participants listened to the spoken description and while they retold it. At the beginning of the description phase, participants received the following instructions:

> You will soon hear a pre-recorded, spoken description. The description will describe a two-dimensional picture. We want you to listen to the description as carefully as possible and to imagine it as thoroughly as possible. During this description we will measure your pupil size. It is important that you do not close your eyes, but you may look wherever you want on the white board.

When the prerecorded description ended, the participant was asked to describe the scene in his or her own words. Participants were again specifically asked to keep their eyes open during this phase, but they were free to look wherever they wanted on the white board. They were also informed that their pupil size would again be measured while they spoke. These instructions took about 40 sec. After the participants had finished their retelling, which lasted about 1 to 2 min, they were asked the following questions in a short interview:

1. What do you think the objective of this experiment was?
2. Rate the vividness of your visualization during the listening phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).

3. Rate the vividness of your visualization during the retelling phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).
4. Assess whether you usually imagine things in pictures or in words.

Before they left, we promised to send each participant a description of the true nature of the experiment as soon as all participants had been recorded.

2.4. *Method of analysis*

The analysis of the eye data was done with an eye-tracking analysis program developed by SMI (iView for Windows), which can trace the saccades and fixations of a participant's eyes over time. Correspondences of the eye movements were analyzed for all objects that were mentioned in the prerecorded description (the fence and the road were analyzed both in the condition "below" and "from the left to the right"). Spatial schematics for these objects can be seen in Fig. 1.

It is possible that participants could imagine a scene either using the whole white board or only a certain part of it. Therefore, it would be unsound to define actual physical coordinates of objects on the white board (as Brandt & Stark, 1997, and Laeng & Teodorescu, 2002, had done). Instead, we developed a method to analyze the relative position of an eye movement compared to the overall structure of the scanpath.

Fig. 2 shows four examples of how the eye movements of one participant are represented in iView over four successive points in time during the description (circles represent fixations; lines represent saccades). Apart from measuring spatial correspondence, we also needed temporal criteria for the occurrence of the fixation to ensure that it concerned the same object that was mentioned verbally. In a study of simultaneous descriptions of the same stimulus picture, Holsanova (2001, p. 104f) found that eye–voice latencies (i.e., the time from when an object is mentioned until the eye moves to the corresponding location) typically range between 2 and 4 sec. Similar results were found by Richardson and Dale (2005), who showed that participants who listen to a speaker talking about present picture elements are most likely to look at the

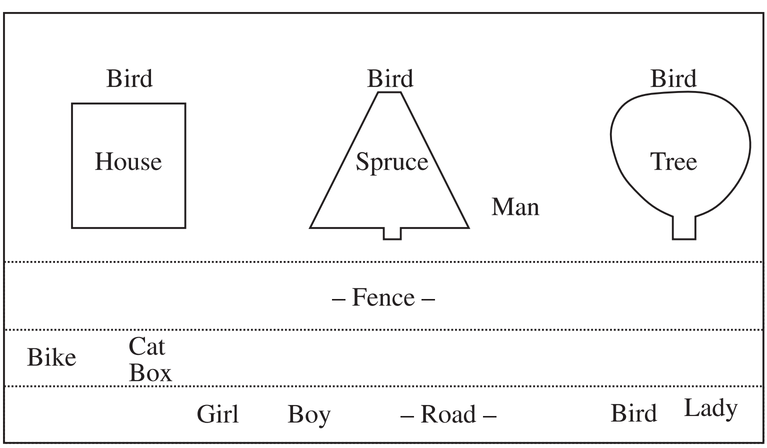

Fig. 1. Spatial schematics for the objects in the prerecorded description.

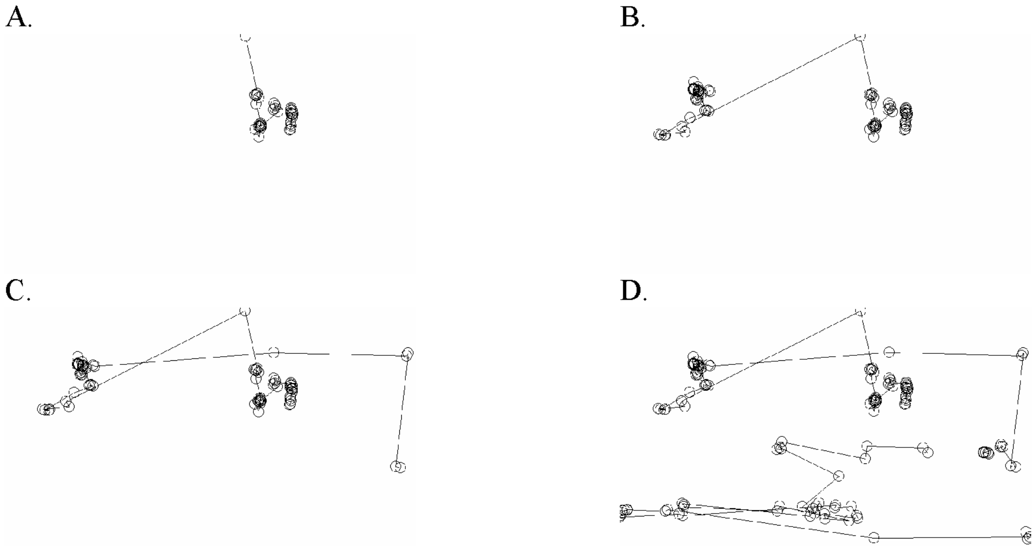

Fig. 2. iView analysis of the first 67 sec for one participant. (A) 0 to 19 sec; spruce and bird in top. (B) 19 to 32 sec; the house to the left of the spruce with bird on top of chimney. (C) 32 to 52 sec; the tree to the right of the house and the spruce. (D) 52 to 67 sec; the man between the spruce and the tree, and the fence in front of them running from left to right.

same picture elements with a delay of 2 sec. Whereas eye movements in the listening phase only can occur *after* the mentioning of an object, eye movements in the retelling phase may *precede* the mentioning of an object (i.e., some participants first move their eyes to a new position and then start the retelling of that object, whereas others start the retelling of an object and then move their eyes to the new location).

On average, the voice–eye latency was 2.1 sec while listening to the description and 0.29 sec during the retelling of it. The maximum value over all participants was 5 sec during both the description and the retelling of it. Therefore, a 5-sec limit was chosen for both before and after the verbal onset of an object.

In sum, the eye movements of the participants were scored as *global correspondence, local correspondence,* and *no correspondence.* Eye movements to objects were considered correct in global correspondence when fulfilling the following spatial and temporal criteria:

1. When an eye movement shifts from one object to another it must finish in a position that is spatially correct relative to the participant's eye gaze pattern over the entire description or retelling.
2. In the listening phase, the eye movement from one position to another must appear within 5 sec after the object is mentioned in the description.
3. In the retelling phase, the eye movement from one position to another must appear within 5 sec before or after the participant mentions the object.

It is known that retrieved information about physical relations among objects can undergo several changes (cf. Barsalou, 1999; Finke, 1989; Kosslyn, 1980). Several experiments have shown that participants rotate, change size, change shape, change color, reorganize, and reinterpret mental images (Finke, 1989). We found similar tendencies in our data. Such image transformations may affect our results, in particular, if they take place in the midst of the description or the retelling phase. Therefore, we devised an alternative local correspondence measure.

Eye movements were considered correct in local correspondence when fulfilling the following spatial and temporal criteria:

1. When an eye movement shifts from one object to another during the description or the retelling it must move in the correct direction.
2. In the listening phase, the eye movement from one position to another must appear within 5 sec after the object is mentioned in the description.
3. In the retelling phase, the eye movement from one position to another must appear within 5 sec before or after the participant mentions the object.

The key difference between global and local correspondence was that global correspondence required fixations to take place at the categorically correct spatial position relative to the whole eye-tracking pattern. Local correspondence only required that the eye move in the correct direction between two consecutive objects in the description. Examples and schematics of this can be seen in Figs. 3 and 4.

No correspondence was considered to exist if neither the criteria for local correspondence nor those for global correspondence were fulfilled (typically, when the eyes did not move or moved in the wrong direction).

As a consequence of applying this spatial criterion, a binominal distribution in the data was obtained: The spatial relations were either correct or not (for each coding). We then defined the possibility that a test participant would move his or her eyes to the correct position by chance.

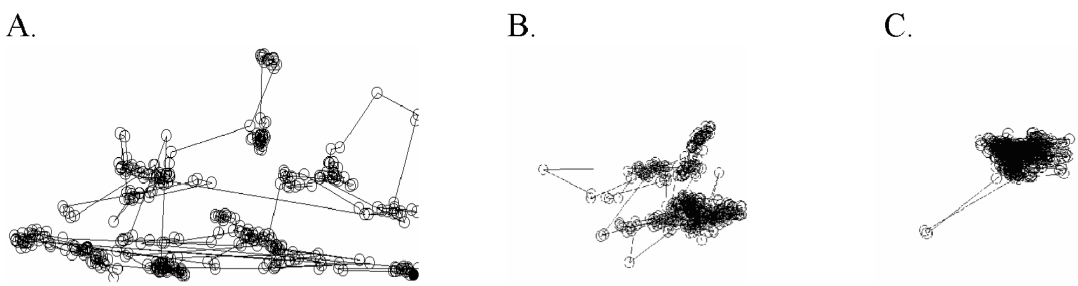

Fig. 3. (A) Example of mostly global correspondences. (B) Example of mostly local correspondences. (C) Example of no correspondences at all.

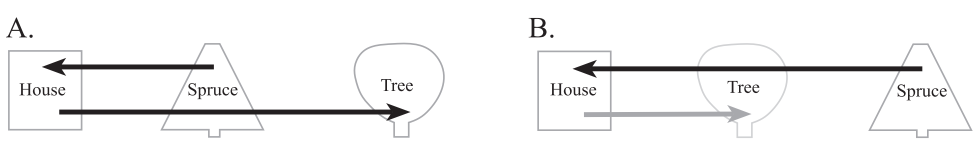

Fig. 4. Schematics of global (A) and local (B) correspondence.

For global correspondence coding, both the direction and the distance of the movement had to be correct. There were many possible movements. A conservative estimate was that the eyes could move in at least four directions (up, down, left, right). For each direction, they could move to at least two different locations (full and half distance). In addition to these eight possibilities, the eye could stand still. For global correspondence, the probability that the eyes moved to the correct position at the correct time was thus definitely less than 1/9 (11%). For local correspondence coding, we only required correct direction; thus, the local correspondence probability was 1/5 (20%). We used the Wilcoxon Signed-Ranks test (Wilcoxon, 1945) for significance between the number of correct eye movements and the expected number of correct movements by chance.

2.5. Results and discussion

Data for correct and incorrect eye movements during the listening phase and retelling phase in local and global correspondence coding are presented in Tables 1 and 2. As the tables show, more than one half of all objects mentioned had correct eye movements, according to the conservative global correspondence criteria. Allowing for recentering and resizing of the image—as with local correspondence—almost ¾ of all objects have correct eye movements. The participants' spatial pattern of eye movements was highly consistent with the original spatial arrangement.

The goal in Experiment 1 was to extend the previous studies (Demarais & Cohen 1998; Spivey & Geng 2001; Spivey et al., 2000) in two respects: (a) to focus on the complexity of the spatial relations instead of simple directions and (b) to add a retelling phase in which the participants were asked to retell the described scene from memory. Our data indicate that even for complex relations, spatial locations are, to a large extent, preserved during both the listening

Table 1
Listening to the description

Coding	% Objects With Correct Eye Movements	Direction Significance
Local correspondence	64.3	$W = 72$, $z = 2.8$, $p = .0026$
Global correspondence	54.8	$W = 72$, $z = 2.8$, $p = .0026$

Note. W = Wilcoxon signed rank statistic.

Table 2
Retelling the description

Coding	% Objects With Correct Eye Movement	Direction Significance
Local correspondence	74.8	$W = 78$, $z = 3.04$, $p = .0012$
Global correspondence	55.2	$W = 68$, $z = 2.65$, $p = 0.004$

Note. W = Wilcoxon signed rank statistic.

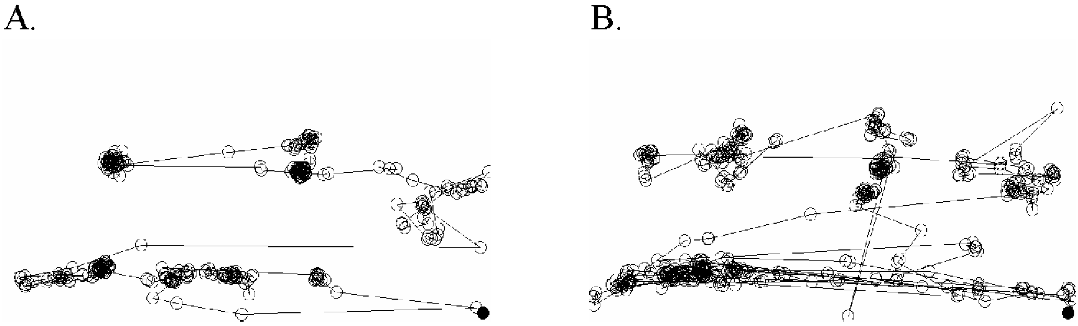

Fig. 5. Same participant: listening phase (A) and retelling phase (B).

and the retelling phase. The effect we measured is strong. Therefore, we concluded that participants make eye movements to appropriate spatial locations while retelling a previously heard description of a scene that was never seen in the first place. To the best of our knowledge, this is the first time that such an effect has been observed.

For a few participants, some eye movements were recentered and shrunk into a smaller area during the retelling phase (thus yielding more local correspondence). However, the majority of eye movements had the same proportions during the listening phase as during the retelling phase. An example of this can be seen in Fig. 5.

Posttest interviews showed that no participant had comprehended the true nature of the experiment. Neither imagery nor eye movements were part of their guesses as to what we were studying. Typically, our participants would guess that it was a memory test or a mental workload test measured by pupil dilation. A simple correlation analysis of how well participants' rating of their own visualizations correlated to the degree of correct eye movements showed no correlation for either the listening phase (–0.1615) or the retelling phase (–0.1393). The participants' assessments as to whether they usually thought in words or pictures were proportionally distributed across four possibilities: (a) words, (b) pictures, (c) a combination of words and pictures, or (d) no guesses. Again, a simple correlation analysis showed no correlation between these assessments and the degree of correct eye movements, either for the listening phase (–0.0330) or for the retelling phase (–0.0119).

3. Experiment 2: Viewing and then describing a picture

The results in Experiment 1 supported the hypothesis that participants' eye movements reflect the positions of objects during a verbal description and during the retelling of it. Experiment 2 examines whether the effect is equally strong when describing a visual scene from memory.

A growing body of research suggests that there is a clear correspondence between eye movements during the examination of a picture or a scene, and eye movements during a visualization of that same picture or scene. Brandt and Stark (1997) showed that spontaneous eye movements occurred during visual imagery and that these eye movements closely reflected the content and spatial relations from the original picture or scene. The participants were first introduced to a simple visual grid pattern that they were to memorize; shortly afterward, they were asked to imagine the pattern. Their eye movements were recorded during this procedure, and it was possible to show that the unique scanpaths established during the viewing of a pattern spontaneously reappeared when the participants later imagined the same pattern. These results were in line with two studies by Noton and Stark (1971a, 1971b), who showed that participants tended to fixate regions of special interest according to certain scanpaths. Holsanova, Hedberg, and Nilsson (1998) found tendencies similar to those observed by Brandt and Stark (1997), with the difference being that the original picture shown was a natural, real-life scene. Laeng and Teodorescu (2002) showed that participants who fixed their gaze centrally during a scene perception did the same, spontaneously, during the imagery phase. They also showed that participants free to explore a pattern during perception, when required to maintain central fixation during the imagery phase, exhibited a decreased ability to recall the pattern.

In Experiment 2, we monitored eye movements of participants who viewed a picture that was later to be described orally while participants looked at a white board. The goal of this experiment was twofold: first, to extend the findings of Brandt and Stark (1997) and Laeng and Teodorescu (2002), who used relatively simple visual stimuli, by introducing a picture of high complexity. Because it was possible that the imagery ability depended on the complexity of the scene, we chose a picture that included many objects with rich detail and clear spatial relations (see Fig. 6). We hypothesized that spatial positions of objects in a complex, naturalistic picture

Fig. 6. The picture that the participants observed. From *Kackel i trädgårdslandet* [Festus and mercury: Ruckus in the Garden] by S. Nordqvist, 1990, Bromma, Sweden: Bokförlaget Opal. Copyright © 1990 by Bokförlaget Opal. Reprinted with permission.

were to be reflected in the eye movements of participants who described them from memory, in the absence of the stimulus picture.

Second, we examined whether this eye movement effect was equally strong, irrespective of whether the original elicitation was spoken or visual, by comparing results from Experiments 1 and 2. To our knowledge, this had not previously been investigated.

In addition, we again collected ratings of the vividness of imagery during the description of the picture, and examined whether the participants usually imagined things in pictures or words.

3.1. Participants

Twelve new students at the University of Lund (6 women and 6 men) volunteered to participate in an experiment in cognitive science. All participants reported normal vision or vision corrected to normal (with contact lenses or glasses).

3.2. Apparatus and stimuli

The same eye tracker was used as in Experiment 1. The visual stimulus was a picture from Nordqvist (1990). It was chosen because of its rich detail and its playful mood.

3.3. Procedure

Throughout, the participants were told that our study concerns pupil dilation during the description of pictures. It was explained to them that we would be filming their eyes, but nothing was said about our knowing where they were looking. They were asked to keep their eyes open so that we could film their pupils, and to look only at the picture or board in front of them so that varying light conditions beyond the board would not disturb the pupil dilation measurements. The eye tracker was then calibrated in the same way as in Experiment 1.

Experiment 2 consisted of two main phases: a *viewing* phase in which the participants inspected the stimulus picture and a *description* phase in which the participants described this picture in their own words from memory. Eye movements were recorded during both phases. At the beginning of the viewing phase, each participant received the following instructions:

> You will soon see a picture. We want you to study the picture as thoroughly as possible. While you study the picture we will measure your pupil size.

The picture was shown for about 30 sec and was then covered by the same white board used in Experiment 1. The participant was then asked to describe the picture in his or her own words. Participants were again specifically told to keep their eyes open during this phase, but that they were free to look wherever they wanted to on the white board. They were also informed that their pupil size would be measured during their description as well. These instructions took about 40 sec. Then, the description phase began, usually taking 1 to 2 min.

After the participants had finished their descriptions, eye movement recording was stopped, and they were asked the following questions in a short interview:

1. What do you think the objective of this experiment was?
2. Rate the vividness of your visualization during the description phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).
3. Assess whether you usually think in pictures or in words.

Before they left, we promised to send them a description of the true nature of the experiment via e-mail as soon as all participants had been recorded.

3.4. Method of analysis

Because the participants themselves were describing the picture, their eye movements were analyzed according to objects derived from their descriptions. These objects can be mentioned in a number of ways, and can be divided either into smaller units (*verbal foci*) or into larger units (*verbal superfoci*; Holsanova, 1999, p. 16f; 2001, p. 15f). The participants' descriptions were first transcribed to analyze *when* certain picture elements were mentioned. The transcripts were then segmented as in the examples that follow. Each line represents a verbal focus expressing a new "idea" formulated at a particular point in time. The numbers indicate the time in minutes and seconds when a new focus starts. The examples are again translated into English:

00:42—And there is a tree in the middle.
00:45—where small animals live and stuff like that.

A verbal superfocus is a coherent chunk of speech consisting of several foci that are, in turn, connected by the same thematic aspect. In the following superfocus, the flowers to the left in the picture are described by the direction left, categorized as large daffodils, and described as having animals in them:

01:20—and ehhh to the *left* in the picture.
01:23—There are large *daffodils*.
01:26—It looks like there were also some *animals sitting there* perhaps.

During this superfocus, one would expect the participant to move his or her eyes toward the left part of the white board during the first focus. Then, it would be plausible to inspect the referent of the second focus (the daffodil). Finally, we could expect the participant to dwell for some time within the daffodil area (on the white board), searching for the animals (three birds, in fact) that were sitting there in the stimulus picture.

An analysis of what elements in the picture were actually mentioned by participants led us to define the objects of interest (e.g., "the tree", "cat with water pump," and "dragonfly"). Because it was common to summarize several versions of the man and the cat ("there were four versions of this man," "there were four versions of the cat"), those four as a unit were also considered as one object. To get correspondence for one of these units, the participant's eyes had to move in a way similar to the way in which the unit was positioned in the vertical plane.

As in Experiment 1, we distinguish between global correspondence and local correspondence. Eye movements were considered correct in global correspondence when they fulfilled the following spatial and temporal criteria:

1. When an eye movement shifts from one position to another it must finish in a position that is spatially correct relative to the participant's entire eye gaze pattern.
2. The eye movement from one position to another must appear within 5 sec before or after the participant mentions an object.

Eye movements were considered correct in local correspondence when the following criteria were fulfilled:

1. When an eye movement shifts from one position to another it must move in the correct direction.
2. The eye movement from one position to another must appear within 5 sec before or after the participant mentions an object.

The time limit of 5 sec was used again. Average latencies in this experiment were –0.75, and the maximum latency was 4 sec. With the same criteria as in Experiment 1, we again used the Wilcoxon Signed-Ranks test (Wilcoxon, 1945) for significance between the number of correct eye movements and the expected number of correct movements by chance.

3.5. Results and discussion

Data for correct and incorrect eye movements during the picture description in local and global correspondence coding are presented in Table 3. As the table shows, the results are significant both in the local and in the global correspondence coding. This suggests that participants visualize the spatial configuration of the scene as a support for their descriptions from memory.

The first goal in Experiment 1 was to extend the findings of Brand and Stark (1997) and Laeng and Teodorescu (2002), who used relatively simple visual stimuli, by introducing a picture of high complexity. Our data indicate that even for a complex picture, spatial locations are preserved to a high degree when described from memory. Again, the effect we measured is strong. More than one half of all picture elements mentioned had correct eye movements, according to the conservative global correspondence criteria. Allowing for the recentering and resizing of the image—as with local correspondence—almost ¾ of all picture elements have correct eye movements.

The effects of resizing (i.e., the fact that participants shrank, enlarged, and stretched the image) were quite common during picture description. One possible explanation for enlargement

Table 3
Picture description

Coding	% Objects With Correct Eye Movements	Direction Significance
Local correspondence	74.8	$W = 76$, $z = 2.96$, $p = .0015$
Global correspondence	54.4	$W = 66$, $z = 2.57$, $p = .0051$

Note. W = Wilcoxon rank statistic.

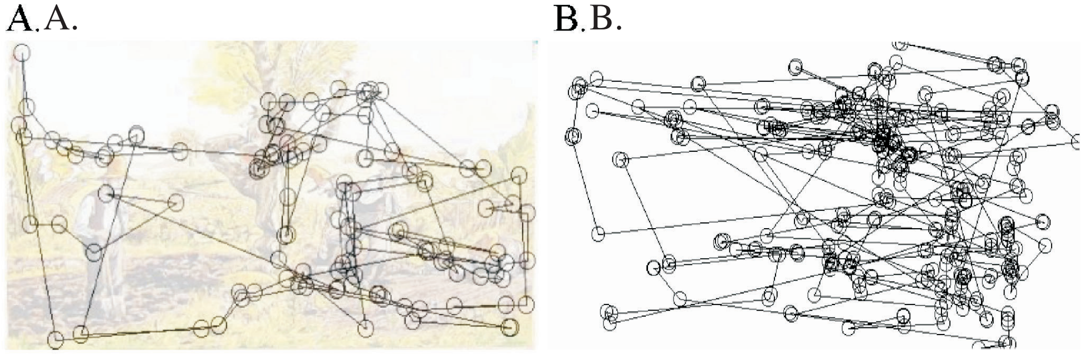

Fig. 7. Same participant: viewing phase (A) and description phase (B).

and stretching could be that the white board was slightly larger than the picture. It was also common that participants recentered the image from time to time, thus yielding local correspondence. Overall, there was a good similarity between data from the viewing and the description phases, as can be seen in Fig. 7.

The second goal of Experiment 2 was to examine whether this eye movement effect is equally strong, irrespective of whether the original elicitation was spoken or visual. Taking the results of Experiment 1 and Experiment 2 together, there is a convincing similarity in the percentages of correct eye movements when participants describe a previously seen picture and when they retell a previously heard, spoken description (cf. Table 4). The data indicate that this effect is equally strong, irrespective of whether the original elicitation was spoken or visual.

The results, although strong, are in fact methodologically conservative. A more fine-grained method of analysis, potent in counting "local" global correspondences between resizings and recenterings, could make the results stronger yet.

In addition, eye movements yielding no correspondence do not necessarily contradict our hypothesis. Because the saccade amplitudes tend to be smaller during mental imagery (Brandt & Stark, 1997), some of them could have been so small that they were not counted with our method of analysis. Demarais and Cohen (1998) argued that certain individuals develop a tendency to suppress large eye movements while inspecting details of an image (much like learning to relax jaw tension to breathe more quietly when listening intently).

Another possibility, as suggested by Johansson, Holsanova, and Holmqvist (2005), is that participants shrink the picture, or parts of it, so much that they are able to covertly "scan" most of their mental image (i.e., shifting their inner attention without eye movements).

Table 4
Percentages of objects with correct eye movements during description and retelling

Coding	Pictorial Elicitation—Later Description (%)	Spoken Elicitation—Retelling (%)
Global correspondence	54.4	55.2
Local correspondence	74.8	74.9

The posttest interviews showed that no participant had comprehended the true nature of the experiment. Neither eye movements nor imagery were part of their guesses as to what we were studying. The answers were about the same as those in Experiment 1. In a procedure similar to that of the previous experiment, we conducted a correlation test for the following two conditions: (a) between correctness in eye movements and the participants' vividness ratings and (b) between correctness in eye movements and the participants' assessments as to whether they think in pictures or words. The participants' vividness ratings yielded no result (–0.2202), nor did their assessments as to whether they think in pictures or words (0.3163). Despite the fact that the majority of participants in Experiments 1 and 2 had appropriate imagery patterns, no such correlation was found. One possible interpretation might be that people, in general, are not aware of which mental modality they are thinking in.

4. Experiment 3: Listening to and then retelling a spoken description in complete darkness

The main goal of Experiments 1 and 2 was to test whether the reflection in eye movements was equally strong in retelling from memory, irrespective of whether the original elicitation was spoken or visual. The objective in Experiments 3 and 4 was somewhat different.

Within a number of situated approaches it has been suggested that we can use our environment as an external memory store (e.g., O'Regan, 1992; O'Regan & Noë, 2001). The eyes could then "leave behind" a deictic pointer or spatial index in the external memory (Altmann, 2004; Ballard et al., 1997; Spivey et al., 2004). In this view, the "scanning" of an image is accomplished by binding the imagined objects onto the actual visual features in the world. A visual feature is then used as a sort of demonstrative reference. Pylyshyn (2000, 2001, 2002) called the mechanism for binding imagined objects to perceived ones *visual index.*

If the visual environment were used as an external memory store, then each object in participants' memories in Experiments 1 and 2 would be associated with a position in the actual visual environment (i.e., the white board). Slight visual features on the board and in the surroundings could serve as *visual indexes* that bind the spatial locations of objects from the working memory (Pylyshyn, 2000, 2001, 2002). Thinking that something is at a certain location is then no more than thinking, "This is where I imagine X to be located" (Pylyshyn, 2002). Such an association would assume no pictorial properties of the memory, only the binding of active memory objects to real objects. Visual indexes would make internal pictorial representations of spatiality unnecessary as far as eye movements were concerned. Consequently, the eyes would just move to that position in the real world that was associated to the current propositional object.

In complete darkness, associations from memory objects to positions in the visual environment would be impossible. The goal of Experiment 3 was to replicate Experiment 1 in complete darkness. If the results were similar to Experiment 1, visual indexes would not be a plausible explanation for eye movement patterns during visualization of a described scene and the retelling of this scene. To the best of our knowledge, no such study has previously been conducted.

4.1. Participants

Twenty-eight students at the University of Lund (14 women and 14 men) volunteered to participate in an experiment in cognitive science. All participants reported normal vision or vision corrected to normal (with contact lenses or glasses).

4.2. Apparatus and stimuli

The same eye tracker was used as in Experiments 1 and 2, with the adjustment that we installed a 950 nm infrared diode that is completely invisible to the human eye. The eye tracker was also tested in complete darkness for about 30 min to assure that no light from the diode could be seen even after adaptation to darkness.

4.3. Procedure

Throughout, participants were told that the experiment concerned pupil dilation during retelling of descriptions held in memory. It was explained to them that we would be filming their eyes, but nothing was said about our knowing where they were looking. They were asked to keep their eyes open so that we could film their pupils, and to look wherever they wanted. The eye tracker was then calibrated in the same way as in Experiments 1 and 2.

The experiment consisted of two main phases: one *listening* phase in which the participants listened to the spoken description and one *retelling* phase in which the participants retold in their own words the description they had listened to. Eye movements were recorded both while participants listened to the spoken description and while they retold it. At the beginning of the description phase, participants received the following instructions:

> You will soon hear a pre-recorded, spoken description. The description will describe a two dimensional picture. We want you to listen to the description as carefully as possible and to imagine it as thoroughly as possible. Before the description starts, the lights will be turned out and you will be sitting in complete darkness. During this description we will measure your pupil size. It is important that you do not close your eyes, but you are free to look wherever you want.

When the prerecorded description ended, the participant was asked—while still sitting in complete darkness—to describe the scene in his or her own words. Participants were also specifically asked to keep their eyes open during this phase, but were told that they were free to look wherever they wanted. They were also informed that their pupil size would again be measured while they spoke. After the participants had finished their retelling (lasting for about 1–2 min), they were asked the following questions in a short interview:

1. What do you think the objective of this experiment was?
2. Rate the vividness of your visualization during the listening phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).

3. Rate the vividness of your visualization during the retelling phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).
4. Assess whether you usually imagine things in pictures or in words.

Before our participants left, we promised to send them a description of the true nature of the experiment as soon as all participants had been recorded.

4.4. Method of analysis

The methods of analysis were the same as in Experiment 1. The time limit of 5 sec was used again. Average latencies in this experiment were 1.94 sec during the description and –0.26 sec during the retelling of it. The maximum latency was 5 sec during both the description and the retelling of it. With the same criteria as in Experiment 1, we again used the Wilcoxon Signed-Ranks test (Wilcoxon, 1945) for significance between the number of correct eye movements and the expected number of correct movements by chance.

4.5. Results and discussion

Data for correct and incorrect eye movements during the listening phase and retelling phase in complete darkness are presented in Tables 5 and 6. As these tables show, about one third of all objects received correct eye movements according to the conservative global correspondence criteria, and about two thirds of all objects received correct eye movements according to the local correspondence criteria. The data suggest that participants' spatial patterns of eye movements were consistent with the original spatial arrangement.

Table 5
Listening to the Description

Coding	% Objects With Correct Eye Movement	Direction Significance
Local correspondence	60.8	$W = 366, z = 4.39, p < .0001$
Global correspondence	27.6	$W = 218, z = 2.61, p = .0045$

Note. W = Wilcoxon signed rank statistic.

Table 6
Retelling the Description

Coding	% Objects With Correct Eye Movement	Direction Significance
Local correspondence	64.9	$W = 400, z = 4.55, p < .0001$
Global correspondence	35.1	$W = 340, z = 3.87, p = .0001$

Note. W = Wilcoxon signed rank statistic.

The goal of Experiment 3 was to replicate Experiment 1 in complete darkness. The difference between watching a white board, as in Experiment 1, and sitting in complete darkness, as in Experiment 3, was that eye movements to a higher degree were recentered and shrunk into a smaller area; thus yielding more cases of local and no correspondence. Some participants also tended to frequently blink when in darkness. This made data disappear and harder to analyze. Despite this, our data indicate that even for complete darkness, spatial locations are to a high degree preserved during both the listening and the retelling phase.

The posttest interview showed similar results to those in Experiments 1 and 2. No participants had seen through the nature of the experiment. Neither eye movements nor imagery were part of their guesses as to what we were studying. The answers were about the same as in Experiments 1 and 2. In a procedure similar to the previous two experiments, we conducted a correlation test for the following two conditions: (a) between correctness in eye movements and the participants' vividness ratings and (b) between correctness in eye movements and the participants' assessments as to whether they think in pictures or words. The participants' vividness ratings yielded no result for either the listening phase (–0.10797) or the retelling phase (0.15441). No correlation was found either between their assessments as to whether they think in pictures or words and the degree of correct eye movements, whether for the listening phase (0.135531) or for the retelling phase (–0.18285).

5. Experiment 4: Viewing and then describing a picture in complete darkness

The goal of Experiment 4 was to replicate Experiment 2 in complete darkness. If the results were to be similar to Experiment 2, visual indexes would not be a plausible explanation for eye movement patterns during visualization of a picture. To our knowledge, no such study has been previously conducted.

5.1. Participants

Twenty-eight new students at the University of Lund (14 women and 14 men) volunteered to participate in an experiment in cognitive science. All participants reported normal vision or vision corrected to normal (with contact lenses or glasses).

5.2. Apparatus and stimuli

The same eye tracker was used as in Experiments 1 and 2, and with the same adjustment as in Experiment 3. The visual stimulus was the same as in Experiment 2.

5.3. Procedure

Participants were initially told that the experiment concerned pupil dilation during the description of pictures. It was explained to them that we would be filming their eyes, but nothing was said about our knowing where they were looking. They were asked to keep their eyes open so that we could film their pupils, and to look at the picture in front of them. The eye tracker was then calibrated in the same way as in Experiments 1, 2, and 3.

Experiment 4 consisted of a viewing phase and a description phase. At the beginning of the viewing phase, each participant received the following instructions:

> You will soon see a picture. We want you to study the picture as thoroughly as possible. While you study the picture we will measure your pupil size. When you have studied the picture for a while the lights will be turned out and you will be sitting in complete darkness.

The picture was shown for about 30 sec, then the lights were turned out. The participant was then asked to describe the picture in his or her own words. Participants were also specifically told to keep their eyes open during this phase, but that they were free to look wherever they wanted. They were also informed that their pupil size would again be measured during their description. These instructions took about 40 sec. Then, the description phase took place, typically lasting for about 1 to 2 min.

After participants had finished their descriptions, eye movement recording was stopped, and they were asked the following questions in a short interview:

1. What do you think the objective of this experiment was?
2. Rate the vividness of your visualization during the description phase on a scale ranging from 1 (*not very vivid*) to 5 (*extremely vivid*).
3. Assess whether you usually think in pictures or in words.

Before our participants left, we promised to send them a description of the true nature of the experiment via e-mail as soon as all participants had been recorded.

5.4. Method of analysis

The methods of analysis were the same as in Experiment 2. The time limit of 5 sec was used again. Average latencies in this experiment were –0.11 sec, and the maximum latency was 4 sec. With the same criteria as in Experiment 1, we again used the Wilcoxon Signed-Ranks (Wilcoxon, 1945) test for significance between the number of correct eye movements and the expected number of correct movements by chance.

5.5. Results and discussion

Data for correct and incorrect eye movements during the picture description in complete darkness are presented in Table 7. As Table 7 shows, more than one third of all objects received correct eye movements according to the conservative global correspondence criteria, and about two thirds of all objects received correct eye movements according to the local correspondence criteria. The data suggest that participants' spatial pattern of eye movements were consistent with the original spatial arrangement.

The goal of Experiment 4 was to replicate Experiment 2 in complete darkness. Our data indicate that even for complete darkness, spatial locations are to a high degree preserved during picture description.

Table 7
Describing the Picture

Coding	% Objects With Correct Eye Movement	Direction Significance
Local correspondence	69.0	$W = 370$, $z = 4.44$, $p < .0001$
Global correspondence	40.5	$W = 290$, $z = 3.48$, $p = 0.0003$

Note. W = Wilcoxon signed rank statistic.

The posttest interview showed that no participants had seen through the nature of the experiment. Neither eye movements nor imagery were part of their guesses as to what we were studying. A correlation test between correctness in eye movements and the participants' vividness ratings yielded no result (0.181268), nor did a similar test for their rating of whether they think in pictures or words (0.113699).

Taking results from Experiments 3 and 4 together, the data indicate that the effect we measured is strong also in complete darkness. However, the results might actually have been stronger had it been possible to eliminate the sometimes excessive blinking that occurred in some participants. In addition, if we had been able to allow global correspondence in periods between resizings and recenterings, the results would possibly have been even stronger.

The similarity in the percentage of correct eye movements when participants describe a previously seen picture and when they retell a previously heard, spoken description in complete darkness is not as good as when they retell while watching the white board (see Table 8). To compare the results from Experiments 1 and 2 to Experiments 3 and 4, a *t* test for independent samples was employed. The proportions of correct eye movements for every participant from Experiments 1 and 2 were compared to the proportions of correct eye movements for every participant from Experiments 3 and 4. It was expected that there would not be any differences between the experiments. As can be seen in Table 9, however, the proportion of correct eye movements is significantly lower in the darkness condition, but only for the global correspondence coding. For the local correspondence coding, the effect is not significant. We take this to be further support for the interpretation that the more sensitive global correspondence coding is more heavily effected by the darkness condition that causes frequent blinking and increases resizings and recenterings.

However, the percentages indicate a similarity in complete darkness also—especially for the local correspondence coding. The reason that the percentages are higher during the description of the picture (Experiment 3) than during the retelling of the description (Experiment 4) could be that the description includes more objects than the picture does. This increases the

Table 8
Percentages of Objects With Correct Eye Movements During Description and Retelling in Complete Darkness

Coding	Pictorial Elicitation—Later Description (%)	Spoken Elicitation—Retelling (%)
Local correspondence	69.0	64.9
Global correspondence	40.5	35.1

Table 9
Comparing the results of Experiments 1 and 2 to the results of Experiments 3 and 4

Coding	Light Versus Darkness
Local correspondence	$p = .088775$
Global correspondence	$p = .000543$

possibility of reorganizing and recentering the eye movements (something that was much more common in complete darkness than while watching the white board), and thus yields more local and no correspondence.

The goal of Experiments 3 and 4 was to replicate Experiments 1 and 2 in complete darkness. Our data indicate that even in complete darkness, spatial locations are to a high degree preserved while listening to a scene description, during the retelling of the same scene description, and during the description of a picture. On the basis of these results, we conclude that visual indexes in the strong sense that Pylyshyn proposes (2000, 2001, 2002) are not a plausible explanation for eye movement patterns during mental imagery. However, it is possible that spatial indexes or deictic pointers in a weaker interpretation could partly contribute to this eye movement effect. It seems intuitive to assume that the existence of a visual environment and the white board in Experiments 1 and 2 makes it easier to avoid the frequent recenterings and resizings found in complete darkness. If spatial indexes, instead of being the sole explanation, are said to merely serve as an aid to preserve the center and to maintain a reference frame, they could in fact be the main reason to why the results for global correspondence are significantly stronger in light than in complete darkness (Table 9).

6. General discussion

In four experiments we have presented evidence that eye movements reflect the positions of objects while participants are listening to a spoken description; during the retelling of a previously heard, spoken description; and during the description of a previously seen picture. The participants' spatial pattern of eye movements was highly consistent with the original spatial arrangement. This effect was found to be equally strong in retelling from memory, irrespective of whether the original elicitation was spoken or visual. In addition, this effect occurs both while watching a blank white board and while sitting in complete darkness.

Do these results show that our participants had engaged internal image representations to remember better and support retelling?

6.1. Eye movements and internal images

There is much empirical evidence in favor of internal image representations (e.g., Finke, 1989). It should, however, be noted that an internal image is hardly an exact copy of the scene it depicts, and it does not necessarily have *all* the properties (e.g., detail) of a real picture; only *some* properties (e.g., spatial extension; Finke, 1989). It is often claimed that the internal image is constructed in a "visual buffer" of the working memory (Kosslyn, 1994). Distance, location, and orientation of the internal image can be represented in this visual buffer, and it is possible to *shift attention* to certain parts or aspects of it (Kosslyn, 1994). Neural findings that visual imagery and visual perception draw on most of the same neural machinery (e.g., Ganis, Thompson, & Kosslyn, 2004), and the fact that areas of primary visual cortex are topographically organized and preserve spatiality from the retina (e.g., Kosslyn, Thompson, Kim, & Alpert, 1995), have been interpreted as support for the existence of a visual buffer. Eye movements during imagery would, thus, somehow be connected with the internal attention shifts in the visual buffer. Hebb (1968) suggested a functional role for eye movements and proposed that, as in perception, eye movements during imagery are necessary to put together and organize the "part images" to construct a whole visualized image. This suggestion is supported by Laeng and Teodorescu (2002), who interpreted their results as a confirmation that eye movements play a functional role during image generation. Mast and Kosslyn (2002) agreed with this interpretation and proposed, in a way similar to Hebb, that eye movements are stored as spatial indexes that are used to arrange the parts of the image correctly.

A large amount of criticism against this interpretation comes from propositional accounts (e.g., Pylyshyn, 2002, 2003), which claim that there are no such things as internal images or visual buffers. Instead, all our mental representations are propositional and have the same functional nature. A common criticism made by propositional accounts against mental imagery studies focuses on the danger of revealing the study objective in the instructions to participants and, in doing so, producing the desired behavior (e.g., Intons-Peterson, 1983). We argue that this study cannot be a target of this criticism. First, several measures were taken to conceal the study's real objective. Second, after participating, when asked about the purpose of our study, participants did not guess that we were measuring their eye movements or that we were studying imagery. However, the most "classic" objection to imagery experiments that support internal image representations is that the results only appear because of the participant's "tacit knowledge" (e.g., Pylyshyn, 2002, 2003); the knowledge of what things would look like to persons in situations like the ones in which they are to imagine themselves. This implies that all interlocutors acquire tacit knowledge about how they are supposed to look at scenes and objects. Consequently, when participants 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). We argue that tacit knowledge cannot be an explanation for our results. The major instruction in the four experiments was to "describe." Only one third of the instructions included *föreställ dig* (cf. *stell dir vor* in German and translated into English as *imagine*). However, it is obvious that in the listening phase of our Experiments 1 and 3, participants must have had previous knowledge about what houses and various trees look like and what "between" and "a bit above" are. However, our participants also incorporated these submeanings into a larger whole—the scene—which allowed them to make later eye movements to correct places relative to previously established positions. It seems very unlikely that participants are able to mimic a behavior so precisely in their eye movements. The number of points and the precision of the eye movements to each point are too high to be remembered without a support to tie them together in a context (such as an internal image). In addition, it appears considerably simpler to store spatial scene information as one image than as a large collection of propositional statements. Therefore, we argue that internal images are a plausible explanation for our results, and perhaps eye movements are stored as spatial indexes that help to arrange parts of an image.

6.2. *Eye movements and indexes to the external world*

A different approach to our results can be found within the external world account. O'Regan and Noë (2001) suggested that the outside world serves as its own, external representation. Instead of relying on an internal image and a visual buffer, features in the external environment are used. An imagined scene can then be projected over those external features. Storing the whole scene internally would thus be unnecessary. Instead, as Ballard et al. (1997) suggested, participants leave behind "deictic pointers" to locations of the scene in the environment that may be perceptually accessed later when they are needed. In this approach, there is a moment-by-moment trade-off between the visual information maintained in the working memory and information acquired by eye fixations. In effect, it would be possible to encode very little visual information in the working memory and thus minimize internal storing. Spivey and Geng (2001) supported this "embodied" view and suggested that eye movements during imagery resemble a situation in which one must decide where to hang a picture on a wall. In this situation, one tends to look at blank regions and imagine how the picture would fit in that particular location.

Pylyshyn (2000, 2001, 2002) developed a somewhat similar approach to support his propositional representations. He argued that imagined objects and spatial locations are bound to visual features in the external world. These bindings are called *visual indexes* (Pylyshyn, 2000, 2001, 2002). Thinking that something is in a certain location then amounts to no more than thinking, "This is where I imagine X to be located" (Pylyshyn, 2002). This approach is, however, not equivalent to that of Ballard et al. (1997) because it assumes no pictorial properties whatsoever of the "projected image"; only the *binding* of imagined objects to real, perceived ones. Consequently, Pylyshyn (2002) argued that eye movements that reflect spatial locations during imagery occur because the eyes simply move to the visual feature that is associated with the current imagined propositional object.

The view that eye movements are used as pointers or indexes to relieve or replace working memory load during mental imagery seems very attractive and intuitive, and several experiments seem to support it (cf. Ballard et al., 1997; Spivey et al., 2004). However, Experiments 3 and 4 in this study were carried out in complete darkness (i.e., without any possible visual features) and still yielded eye movements that reflected objects from both the description and the picture. Therefore, we argue that visual indexes that only assume the binding of propositional objects to real ones (Pylyshyn, 2000, 2001, 2002) cannot explain eye movements during mental imagery. However, with a more liberal interpretation than Pylyshyns (2000, 2001, 2002) visual indexes, like the deictic pointers suggested by Ballard et al., the situation is somewhat different. Deictic pointers are not supposed to do all the work during mental imagery and do not rule out the possibility that pictorial properties (such as location and spatiality) are stored in memory. Deictic pointers would, in this sense, only be used to relieve the working memory. It is likely that in the experiments executed in light, the white board could be used to maintain a reference frame and serve as some workload relief. In complete darkness, there would not be any visual information present to relieve the working memory. The only possibility would be if the physical act of moving the eyes in itself could serve as some support. Consequently, if deictic pointers are not the sole explanation to our results, but have a supportive role, then they could in fact be a likely explanation to why the recentring and resizing effects increase in complete darkness and thus make the global correspondence coding results significantly weaker (Table 9).

6.3. Eye movements as automatic reenactments

There are, however, other embodied approaches that do not rely on pointers to the external world but which may explain our results. Thomas (1999) founded the *perceptual activity theory,* and argued that perception is "active" in a way similar to active vision systems in robotics. Perception is then not about storing mental images or shifting attention in a visual buffer. Nothing in the brain *is* the image. Instead, we store a continually updated and refined set of procedures or schemas that specify how to direct our attention in different situations (Thomas, 1999). A perceptual experience consists of the ongoing activity of a schema-guided, perceptual exploration of the environment. Imagery is then the reenactment of the specific exploratory perceptual behavior that would be appropriate for exploring the imagined object as if it were actually present. In this reenactment, the procedure or schema sends some of its "orders" to lower-level motor processes to help the exploration. Eye movements during mental imagery would thus occur because of the procedures that take control of the exploratory apparatus during the experience of a "mental image." In this approach, we always encode *how* to direct our attention. Eye movements thus happen when we "act out" the way in which we would visually explore a scene.

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

6.4. Eye movements and language production

Because our participants had the task of verbally describing a scene or picture, it is interesting to mention, albeit briefly, the relation between eye movements and language production. By asking *when* and *why* speakers gaze, Griffin (2004) summarized the role of eye movements in language production, thought, and communication. It turns out that speakers gaze at objects while selecting their object names, whereas they rarely gaze at them when articulating the names. Also, eye movements reflect message planning at certain units of discourse (cf. also Holsanova, 2001) and seem to play a role in monitoring speech. The question of why we gaze is answered by a number of hypotheses: (a) Gazes reflect attention and mental effort, (b) gazes support our memory of objects, and (c) gazes aid object name production. Apart from that, when studying more complex meaningful units of discourse, we can find similarities between the temporal and semantic patterns of eye movement data and spoken language data (Holsanova, 2001). For example, before uttering plural nouns (e.g., 3 birds in the tree) or mentioning objects related by a common activity (e.g., flying objects), speakers show counting-like eye movements and fixate each of these multiple objects.

However, eye–voice latencies in our data differ considerably from language production studies with visible pictures: Latencies are much longer, up to 5 sec. Also, when participants retold the scenes, their eyes in many cases moved to an object position *after* they mentioned the object. This might indicate that objects may be remembered without a participant looking at the object position, but looking there makes the object more salient and, thus, easier to describe. Some of these eye movements might thus serve a similar memory-supporting function as so-called *look-backs* in free descriptions of visible pictures (Holsanova, 2001).

6.5. Future research

This article suggests that eye movements may have a functional and necessary role during mental imagery. However, other explanations have been considered. In the future, we aim to develop methods that specifically study the functionality of eye movements during imagery. A first step would be to study how the fixed-gaze method used by Laeng and Teodorescu (2002) applies to pictures and descriptions of high spatial complexity.

Acknowledgments

This research was conducted in the Laboratory for the Humanities at Lund University (http://www.sol.lu.se/humlab/research/humlabResearch.html?expand_menu=7). It has been presented at the Cognitive Science Conference 2005 in Stresa and at the European Conference on Eye Movements in Bern (2005). The work was partly supported by post-doctoral fellowship Grant VR 2002-6308 from the Swedish Research Council (to Jana Holsanova).

We are grateful to Juliane Steinberg and Halfdan Koot-Fryd for discussions on methodology and theory. We also thank the three anonymous reviewers for their valuable suggestions. Many thanks to our two native proofreaders. Last, but not least, we thank the participants for their participation in the experiments.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The “blank screen paradigm”. *Cognition, 93*(2), B79–B87.

Ballard, H. D., Hayhoe, M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral Brain Sciences, 20,* 723–767.

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

Bower, G. H., & Morrow, D. G. (1990). Mental models in narrative comprehension. *Science, 247,* 44–48.

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.

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: An fMRI study. *Cognitive Brain Research, 20,* 226–241.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. Henderson & F. Ferreira (Eds.), *The integration of language, vision, and action: Eye movements and the visual world* (pp. 213–247). New York: Psychology Press.

Hebb, D. O. (1968). Concerning imagery. *Psychological Review, 75,* 466–477.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1084–1102.

Holsanova, J. (1999). På tal om bilder. Om fokusering av uppmärksamhet i och strukturering av talad beskrivande diskurs [Speaking of pictures: Focusing attention and structuring spoken descriptive discourse in monological and dialogical settings]. Licentiate dissertation. *Lund University Cognitive Studies, 78.*

Holsanova, J. (2001). *Picture viewing and picture description: Two windows on the mind.* Unpublished doctoral dissertation, Lund University, Sweden.

Holsanova, J., Hedberg, B., & Nilsson, N. (1998). Visual and verbal focus patterns when describing pictures. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research: Physiological and psychological aspects* (pp. 303–304). New York: Plenum.

Huber, S., & Kirst, H. (2004). When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 431–444.

Intons-Peterson, M. J. (1983). Imagery paradigms: How vulverable are they to experimenters' expectations? *Journal of Experimental Psychology–Human Perception and Performance, 9,* 394–412.

Johansson, R., Holsanova, J., & Holmqvist, K. (2005). *What do eye movements reveal about mental imagery? Evidence from visual and verbal elicitations.* Proceedings of the Cognitive Science Conference, Stresa, Italy.

Johnson-Laird, P. N. (1981). Comprehension as the construction of mental models. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 295,* 353–374.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

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

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature, 378,* 496–498.

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, 6,* 271–272.

Nordqvist, S. (1990). *Kackel i trädgårdslandet* [Festus and mercury: Ruckus in the Garden]. Bromma, Sweden: Opal.

Noton, D., & Stark, L. W. (1971a). Scanpaths in eye movements during perception. *Science, 171,* 308–311.

Noton, D., & Stark, L. W. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11,* 929–942.

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

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24,* 939–1031.

Pylyshyn, Z. W. (2000). Situating vision in the world. *TRENDS in Cognitive Science, 4,* 197–207.

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

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25,* 157–238.

Pylyshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the brain? *TRENDS in Cognitive Science, 7,* 113–118.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29,* 39–54.

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

Spivey, M., Richardson, D., & Fitneva, S. (2004). Thinking outside the brain: Spatial indices to visual and linguistic information. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 161–190). New York: Psychology Press.

Spivey, M., Tyler, M., Richardson, D., & Young, E. (2000). Eye movements during comprehension of spoken scene descriptions. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the Cognitive Science Society* (pp. 487–492). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thomas, N. J. T. (1999). Are theories of imagery theories of imagination? An active perception approach to conscious mental content. *Cognitive Science, 23,* 207–245.

Wilcoxon, F. (1945). Individual comparison by ranking methods. *Biometrics, 1,* 80–83.

Yoon, D., & Narayanan, N. H. (2004). Mental imagery in problem solving: An eye tracking study. In A. Duchowski & R. Vertegaal (Eds.), *Proceedings of the 3rd ACM Symposium on Eye Tracking Research and Applications* (pp. 77–83). New York: ACM.