Eye Movements During Visualizations of Pictures and Verbal Descriptions

An Experimental Investigation of Eye Movements and Mental Images

Master's Thesis, Cognitive Science, VT 2004-05-19

By Roger Johansson

Supervisor: Kenneth Holmqvist

Lund University Cognitive Science

ABSTRACT

In this paper it is hypothesized that the spatial locations of objects both in a verbal description and from a complex picture are reflected in the eye movements during a visualization of these objects. In four 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 screen 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, and that we indeed have mental images or at least procedures that make us experience mental images. It is also argued that tacit knowledge can not be used as an argument against the results of the experiments.

1. Introduction

Eye movements are essential to visual perception and create the illusion, when shifted, that we see entire scenes. The eye movements can be said to be used to organize the parts of a scene that are fixated and then integrate them into an entire scene. But what about eye movements during mental imagery? Is it possible that the eyes scan a visualized internal image in a similar way as when you look at an external scene?

As early as in 1968 Donald O. Hebb proposed that the eye movement scanpaths that occur when viewing objects and scenes are automatically triggered when the same object or scene is later imagined (Hebb, 1968).

With the eye tracking equipment of today it is possible to monitor and record a person's eye movements during the scanning of an external scene and during a mental recreation of that scene. Experiments of this type have been done and there is strong evidence that the eye movements during the scanning of a scene are reflected by the eye movements that occur during a mental visualization of that scene (Brandt & Stark, 1997, Laeng & Teodorescu, 2002). It is even possible that the eye movements are functionally involved in the mental imagery process.

An area that has not been studied in the same extent is how the eye movements are reflected during a mental visualization of a verbal description. But a couple of experiments have been done that show the same tendencies in eye movements as for mental visualizations of perceived scenes (Demerais & Cohen, 1996, Spivey & Geng, 2001). How can this be? Is it possible that verbal descriptions can generate internal images that are of a similar nature as those generated by external pictures and scenes? Philip N. Johnson-Laird (1981) has created a theory of how language is understood and argues that discourse sometimes is represented in a form akin to that of perception and internal images. He states in his theory that (Johnson-Laird, 1981, p. 353):

> "First utterances are translated into a mental code that provides a direct linguistic representation of them. This stage concerns the identification of speech sounds, the recognition of words, and the recovery of superficial syntactic structure. Secondly, the linguistic code may be used as part of the basis for the inferential construction of a *mental model* of the state of affairs that the utterances describe."

A common argument against the possibility that eye movements during a visualization do reflect an internal image is that they occur because of *tacit knowledge* (Pylyshyn, 2003, p.113), i.e. that they occur because we mimic the behavior we have during perception. But if we can construct mental models from a verbal description, and if the spatial locations in the description are proved to be reflected in the eye movements, i.e. when no actual perceptual process has taken place, this argument seems unlikely.

This paper describes a study of eye movements both during the visualization of a complex picture and during a verbal description. The goal is to examine if the eye movements reflect the spatial locations of objects from the picture and the description, and, if so, if these reflections are similar or different. If the reflections are found to be similar for the eye movements generated by pictures and verbal description this will be used as an argument against the possibility that the eye movements occur because of tacit knowledge.

2. Theoretical Background

2.1 Vision

The visual system that allows us to perceive visual information is physiologically dependent on our eyes. The most important area of the eyes is the *fovea*, which is a very small area at the center of the retina. It is within this area that we can perceive detailed information of high acuity. The fovea extends a visual angle of about 2°, which is about the width of one's thumb held out at arm's length or the width of an average word held at normal reading distance (Glenstrup, Engell-Nielsen, 1995, *3.1:Eye structure*). The remaining part of the retina, outside of the fovea, does not have the same acuity (only 15-50% of the acuity of the fovea) and is therefore said to offer *peripheral vision*. By moving our eye, i.e. changing the location of the fovea and our peripheral vision, we perceive and experience the visual information around us. The eye movements that occur when the eyes are moving from one location to another is called *saccades*. Between the saccades, when the eyes foveate an object, the so-called *fixations* occurs. It is during the fixations that the main processing of a retinal image takes place. A saccade lasts for about 10-100 ms and a fixation for about 150-600 ms (Duchowski, 2003, pp. 44-49). Besides fixations and saccades there exist a number of other eye movements. The most common is *pursuit motion*, which occurs when the eyes are following a moving object, i.e. when the eyes keep a moving object foveated (Glenstrup, Engell-Nielsen, 1995, *3.3:Eye movements*). However, in this study only saccades and fixations are of interest.[1]

2.2 Visual attention

To understand the phenomenon of visual attention it is important to distinguish between "where" and "what". Where we look is not always the same place as what we look at, e.g. it is possible to foveate a certain object but to have the attention at a *peripheral* object. This attentional dichotomy is commonly called *overt* and *covert* attention, where overt attention corresponds to foveal attention and covert attention corresponds to parafoveal attention. This dichotomy is particularly relevant when explaining how we select our attention, and especially in a *bottom-up* explanation (Wolfe, 1998, pp. 43-44). In a bottom-up explanation the visual attention selection mechanism can be said to consist of two stages. First we have an early *pre-attentive* stage ("where" to look) and later an *attentive* stage ("what" to look at). The pre-attentive stage is working in parallel across the entire visual field (parafoveal) and the attentive stage is limited and can only handle one object at a time (foveal). When an object is processed from the pre-attentive to the attentive stage it is considered to be selected. This attention selection also means that visual attention is shifted to a new location *before* the saccade occurs. This selection process of attention has been explained by a number of different metaphors. The most common metaphors are Posner's "spotlight", Treisman's "glue", and Kosslyn's "window" (Duchowski, 2003, pp. 9-12). The spotlight-metaphor suggests that the attentional mechanism moves in the same manner as a spotlight and that the object in the spot is what we attend to. The glue-metaphor means that attention provides some sort of glue that integrates separated features into a conjunction that makes it possible to

[1] Since nothing moves in the stimulus that are used in the experiments.

perceive an object in its entirety. With this explanation attention can be said to select features from a “master map” (that shows “where” the features are but not “what” they are) and glue them together. The window-metaphor means that we have some sort of window that is responsible for the selection from a “visual buffer”. This window is needed because while some information must be filtered out some must be taken in. This window has the ability to adjust itself incrementally, i.e. it is scalable.

To sum all this up one can say that when given a stimulus, like an image or a scene, this is first mostly seen in parallel and through peripheral vision. At this phase of attention certain features that attract the viewer may “pop out” in the field of view. This pop-out effect then directs the attention towards these features for further and more detailed inspection. When this occurs the attention is disengaged from the foveal location and is repositioned to the feature that has attracted the attention. When the eyes have completed this movement the fovea is directed towards the interesting feature, and the attention is now engaged so that the feature might be inspected at high resolution.

This is, however, a bottom-up model of how attention is directed during an inspection of a scene. But it is also possible that the attention is directed *top-down*, i.e. that the interesting features are voluntarily chosen by a certain interest, i.e. user-driven, and not by some sort of pop-up effect (Wolfe, 1998, pp. 44-45). This voluntary and task dependent attention has been showed when people look at pictures, i.e. they look at them differently depending on what they are looking for and what their interests are (Yarbus, 1967, pp. 171-196).

However, the nature of the representations formed during the viewing of pictures and scenes is a wildly debated question. There are two main theories that are competing. The first, the *localist-minimalist approach* (Henderson & Hollingwoth, 2000, p. 5), suggests that coherent visual representations falls apart as soon as attention is withdrawn from an object, which means that the visual representation of a scene is wiped out (Rensink, 2002, Rensink, O’Regan & Clark, 1997). The second suggests that visual representations do not necessarily disappear when attention is withdrawn, but can instead be stored in a visual memory and be integrated with previous attended regions of a scene (Henderson & Hollingworth, 2000, pp. 6-8). The localist-minimalist theory could serve as a good explanation to why we sometimes have change blindness, i.e. that we sometimes fail to apprehend changes in pictures and scenes. But it has quite convincingly been shown that relatively detailed visual representation of objects from a scene are retained over time and across several eye fixations, which gives evidence to the theory with a visual working memory (Henderson & Hollingworth, 2003; Hollingworth & Henderson, 2002; Hollingworth, Williams & Henderson, 2001; Irwin & Zelinsky 2002; Ishai & Sagi, 1995).

2.4 Eye movements and mental imagery

In the early days of eye tracking the pioneer Yarbus (1967) showed that while subjects view pictures and scenes eye movements are not random. The eye movements are in some way related to the content of the picture or scene that the subjects are watching. The pattern of the eye movements can also be altered depending on the instructed search task, e.g. if you are instructed to observe the picture freely or with the task to study certain aspects of the picture.

Noton & Stark extended Yarbus’ results and showed that subjects tend to fixate regions of special interest according to certain “scanpaths” (Noton & Stark, 1971a, 1971b). These scanpaths are however very individual and differ from viewer to viewer and even when the same scene is viewed by the same individual at different occasions. Noton and Stark (1971a, 171b) suggested that the internal representation or memory of the scene is an alternating sequence of sensory and motor memory traces, recording alternately a feature of the scene and the eye movement required to reach the next feature.

Another pioneer, Hebb, analyzed imagery and proposed that eye movements have an important and necessary function in visual imagery (Hebb, 1968). He suggested that, as in perception, eye movements are necessary to put together and organize the “part-images” to construct a whole visualized image. This means, according to Hebb, that the eye movements during scene perception are automatically triggered during imagery of that scene.

Kosslyn found, in 1973, that the time during a scanning was linear with the spatial distance between points in a mental image (Finke, 1989, pp. 62-65). He showed that when subjects mentally focused on one end of a previously observed drawing and then “looked” for a designated feature of the drawing on their image the time increased as the distance between the feature and the initial point of focus increased (spatial equivalence as mentioned above). This finding with other similar experiments supported Kosslyn’s theory of a visual working memory with two-dimensional properties (Finke, 1989, pp. 62-65). Further experiments suggest that this visual working memory also may have a three-dimensional structure (Finke, 1989, pp. 65-68).

A study by Brandt and Stark (1997) has shown 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. 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 subjects later imagined the same pattern.

These findings suggest that there is a clear correspondence between the eye movements during an examination of a picture, or a scene, and the eye movements during imagery of the same picture or scene. Brandt and Stark (1997) propose that eye movements play a significant and functional role in visual imagery and suggest that the scanpaths are linked to the arranging of part images into their correct locations.

But although eye movements during perception and imagery show significant similarity (and perhaps are constituting imagery) there are differences. It has been found that the fixations during imagery are longer and that the amplitude of the saccades is smaller (Brandt & Stark, 1997, p. 33). The longer fixations can be explained by that a construction of a mental image is more difficult than in perception, i.e. it is harder to arrange the sub-features into a whole scene. The smaller amplitude can be explained by that in imagery there is not the same need for full range eye movements as in perception, where the fixations are necessary to identify certain sub-features of a scene.

Laeng and Teodorescu (2001) have with a recent study given new evidence to the idea that eye movements play a functional and important role in visual imagery. They 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, got a decreased ability to recall the pattern. According to these results Laeng and Teodorescu proposed that the eye movements during perception are stored along with the visual representation and are then used as a spatial index for the parts of the image during the imagery.

Although the results by Brandt and Stark (1997), and Laeng and Teodorescu (2001) are very interesting, they used quite simple visual stimuli in their experiments (grids with black and white squares). You can not make the assumption that these results would be the same for a more complex stimulus, like a real picture or a real scene.

While the correspondence of eye movements during perception of pictures, or scenes, and during imagery have been quite well studied, less work has been done in the area concerning eye movements during a verbally constructed scene and during the imagery of that scene.

The difference with imagery during verbally constructed scenes is the absence of a visual input, i.e. when imagining an object we have to develop a mental representation of that object that has a spatial structure. During the construction of such a spatial mental model cognition often uses linguistic input to activate memory representations (Bower & Morrow, 1990; Johanson-Laird, 1981). These memory representations may then be used in imagery to partially activate perceptual representations (Spivey, Tyler, Richardson, & Young, 2000).

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, when listening to a spatial scene description, tend to make eye movements in the same directions as in the described scene. The descriptions where of the following type:

> *“There is 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.”*

In this experiment the subjects were instructed to imagine the scene. This experiment was also followed by another that demonstrated that subjects tend to make eye movements in the same directions as in a description even when their eyes where closed and when they had no instructions to imagine anything (Spivey, Tyler, Richardson, & Young, 2000).

2.3 Theories of Mental Imagery

But what does it mean that reflection in eye movements appear during mental imagery? During imagery, there is no visual input to drive eye movements bottom-up and no scene or picture to be inspected top-down. So what can eye movements say about the nature of mental imagery?

Mental images are not always an exact copy of the things they depict. Most people experience mental images that merely resemble the things they depict. Mental images are also subjective, i.e. they are not observable for others than those who produce them. Mental images are also very elusive, e.g. they can appear at one moment and fade away at the next moment. Because mental images are subjective, not observable and elusive it is very hard to define and study them.

Finke defines "Mental imagery" as *the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event, either in conjunction with, or in the absence of, direct sensory stimulation* (Finke, 1989, p. 2). This definition is inspired by the quasi-pictorial theory that has been developed by Stephen M. Kosslyn (Kosslyn, 1980, Kosslyn, 1994). The quasi-pictorial theory suggests that the mental representations of a mental image have *some* properties that are the same as pictures, but not necessarily *all* the properties. This theory is also in line with the attention theory that we have some sort of visual working memory (or buffer).

Finke also identifies five major principles of imagery that are intended to provide a general description of the fundamental characteristics of mental images (Finke, 1989). These principles are: the principle of *implicit encoding*, the principle of *perceptual equivalence*, the principle of *spatial equivalence*, the principle of *transformational equivalence*, and the principle of *structural equivalence*. These principles are based on experiments that have been done, but are not to be seen as absolute laws. They are more like hints to how today it is possible to describe mental images based on the collected evidence from the experiments and research that have been done on the subject.

The principle of implicit encoding means that *mental imagery is instrumental in retrieving information about the physical properties of objects, or about physical relationships among objects that was not explicitly encoded at any previous time* (Finke, 1989, p. 7). The point is that imagery is particularly useful when recalling information about spatial relations and objects whenever the information has not been explicitly encoded.

The principle of perceptual equivalence means that *imagery if functionally equivalent to perception to the extent that similar mechanisms in the visual system are activated when objects or events are imagined as when the same objects or events are actually perceived* (Finke, 1989, p. 41). The point is that mental images have many visual characteristics in common with perceived objects and events. For example constraints on resolution for mental images do in a way correspond to those in visual perception, and changes in visual-motor coordination for mental images do resemble those resulting when one adapts to actual vision distortions. But there are also important differences. The principle of perceptual equivalence seems to be limited down to the levels in the visual system where visual associations occur. For example there does not seem to be any retinal or precortical involvement in mental imagery (which are responsible for chromatic aftereffects), and it does not seem to be any involvement in the initial information process stages in the visual cortex (where simple features are analyzed).

The principle of spatial equivalence means that *the spatial arrangement of the elements of a mental image correspond to the way objects or their parts are arranged on actual physical surfaces or in an actual physical space* (Finke, 1989, p. 61). The point is that spatial relations among objects are preserved in images. But it is possible that they are distorted.

The principle of transformation equivalence means that *imagined transformations and physical transformations exhibit corresponding dynamic characteristics and are governed by the same laws of motion* (Finke, 1989, p. 93). The point is that imagined transformations are like physical transformations holistic and continuous, e.g. in size and shape.

The principle of structural equivalence means that *the structure of mental images corresponds to that of actual perceived objects, in the sense that the structure is coherent, well organized, and can be reorganized and reinterpreted* (Finke, 1989, p. 120). The point is that mental images possess structural characteristics corresponding to those of physical objects. For example the relationships among an object's parts can be reinterpreted and preserved. But this principle seems very limited when it comes to detect "hidden" parts of a pattern, or when it comes to interpretations of ambiguous figures.

Applied to eye movements during mental imagery the quasi-pictorial theory would say that these reflections in eye movements support the quasi-pictorial theory because perception and imagery shares the same patterns in eye movements. For example Kosslyn and Mast have interpreted the results of Laeng and Teodorescu (2001) quasi-pictorial, and argue that eye movements play a functional role in mental imagery and that the eye movements are stored as spatial indexes that are used to arrange the parts of the image correctly during mental imagery (Kosslyn & Mast, 2002).

The competing theory to the quasi-pictorial theory is the description theory, often represented by Zenon W. Pylyshyn, who proposes instead that the mental representations that we experience of images do not share any properties with the perception of a scene or a picture, i.e. there are no equal properties between perception of an object and the mental representation of this object. This theory also claims that all our mental representations are of the same functional nature, i.e. whether we have mental representations of for example a scene or a verbal sentence these mental representations have the same functional nature. The consequences of this theory is that the cognitive processes that are used during perception are absent during imagery (Pylyshyn, 2000). This theory is also in line with the localist-minimalist theory of attention and means that there is no such thing as a visual working memory (or buffer).

Pylyshyn also suggests that the reason that many experimental findings support the theory that we have internal pictures is because of *tacit knowledge* (Pylyshyn, 2003, p.113). This tacit knowledge means that when people imagine something they *simulate* many of the aspects that would happen when they perceive something, i.e. findings of similarities in cognitive processes between perception and imagery do not appear because of a functional connection.

Another approach to the question of mental images is that we use the world as an outside memory, i.e. that our environment can be considered as a kind of external memory store that is used as an index for spatiality for both perception and imagery (Pylyshyn, 2000, O'Regan, 1992). This approach is also compatible with the description theory because this external memory could be so strong that internal pictorial representations are unnecessary.

Applied to eye movements during mental imagery the description theory would say that these eye movements occur because of tacit knowledge, task-induced demand or that the eye movements are a by-product that has no functional role in mental imagery.

Although the quasi-pictorial theory and the description theory are the main opponents in the imagery debate there is an interesting contender called the perceptual activity (PA) theory. The PA theory has a proceduralist approach to memory and suggests that the nature of perceptual learning does not involve storing of pictures, or descriptions, of what we perceive, but instead as a continual updating and refining of procedures that specify *how* to direct our attention in different situations, i.e. how to examine and interpret scenes and objects (Thomas, 1999, pp. 218-219). This means that neither inner pictures nor descriptions are created, i.e. no thing in the brain *is* the image. Perceptual experience is instead an ongoing exploration of the environment guided by certain procedures. An image in the PA theory is something we experience when a procedure, that is not necessarily relevant to the exploration of the environment, takes control of the exploratory apparatus (Thomas, 1999, p. 218).

Applied to eye movements during mental imagery the PA theory would say that the reflections of eye movements during mental imagery happen in order with the procedures that take control of the exploratory apparatus during the experience of an image, i.e. the eyes constitute a fundamental part of the perceptual experience of an image.

3. Purpose

The purpose of this 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, Tyler, Richardson and Young (2000), and Spivey and Geng (2001) by studying eye movements when subjects listen to a complex description that they are to mentally visualize. The results will then be analyzed, i.e. if, and how, the eye movements reflect the spatial locations of objects during these visualizations. In line with Johnson-Laird’s (1981) theory that language and discourse sometimes are represented in a form akin to that of perception and internal images the results will be compared and similarities or differences in eye movements in these two situations will be studied. If there are no significant differences the results will be used as an argument against the tacit-knowledge explanation, because tacit-knowledge is hard to apply to reflections in eye movements that are generated by a verbal description, i.e. when no actual perception has occurred.

4. Experiment 1

One group of subjects was instructed to listen to and visualize a prerecorded verbal description, which was later to be imagined when they were questioned about its objects and spatial relations. The description described a two dimensional picture.

Based on the results by Spivey and Geng (2001), Spivey, Tyler, Richardson and Young (2000), and Demarais and Cohen (1998) the hypothesis is that the positions of objects in the description are reflected by the eye movements during the description and during the questions about it.

This means that when the subject listens to the verbal description the eye movements should follow the spatial relations of the objects in the description. These eye movements shall also appear later when the subject is questioned about the objects from the description, i.e. when the subject is asked questions about the position of an object relative to another from the verbal description, the eye movements should be similar to the eye movements that occurred when these objects first were presented in the verbal description.

4.1 Method

4.1.1 Participants

Twelve students at the University of Lund, 6 females and 6 males volunteered to participate to an experiment in cognition 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 what had been studied in the experiment. It was confirmed that eleven participants were naive about the fact that there eye movements were recorded and that they had no specific knowledge about the experimenters' expectations. One of the participants thought that the eye movements were recorded in some way, but it was pretty vague and it was not considered that this participant had seen through the nature of the experiment.

4.1.2 Apparatus and stimuli

The eye tracker that was used is an SMI iView 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 MPEG video and eye movement coordinates.

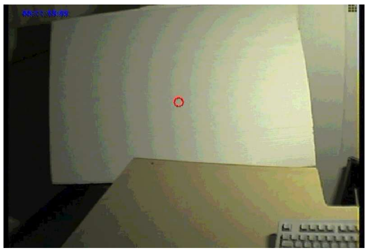
The visual stimulus used in the experiment consisted of a white screen (657mm $\times$ 960mm), and the auditory stimulus used in the experiment consisted of a prerecorded description (2 minutes and 6 seconds). The participants were seated in front of the white screen at a distance of 150 cm (picture 1). The prerecorded description was the following:[2]

> *"Imagine a two dimensional picture. In the center of the picture a large green spruce grows. In the top of the spruce sits a bird. To the left of the spruce and to the far left in the picture is a yellow house with 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 stands a man in 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 goes from the pictures left edge to the pictures right edge. In 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 stands bouncing a ball. To the right of the girl sits a boy who wears a red cap and who is 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."*

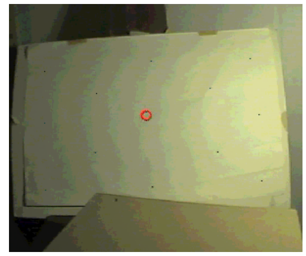
4.1.3 Procedure

[2] In the experiment the description was presented in Swedish, but is here translated to English.

Before the actual experiment, at the beginning of each session, a standard calibration routine was used. The eye position was recorded at nine standard calibration points (appearing as black crosses on a white background), corresponding to the size of the picture (657mm × 960mm). Using a laser pencil the participant was instructed to fixate each point (picture 2).



Picture 1 - Example of the white screen the test subjects were watching during the description (the circle is the position of the eyes).



Picture 2 - Example of the calibration points

The experiment consisted of two main phases, one description phase in which the participants listened to the verbal description and one question phase in which the participants answered 8 questions about the description. Afterwards the subject was asked a few questions about the experiment. Eye movements were recorded both during the verbal description and during the questions.

At the beginning of the description phase the subjects received the following instructions:

> *"You will soon hear a prerecorded verbal description. The description will describe a two dimensional picture. We want you to listen to the description as carefully as possible and to visualize 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 freely look wherever you want on the white screen".*

When the description finished the subject was told to answer, a couple of prerecorded questions about the description. 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 screen. They were also informed that their pupil size again would be measured during these questions. These instructions took about 40 sec, in order to prevent afterimages.[3] After the subjects had answered the questions about the description they were asked to:

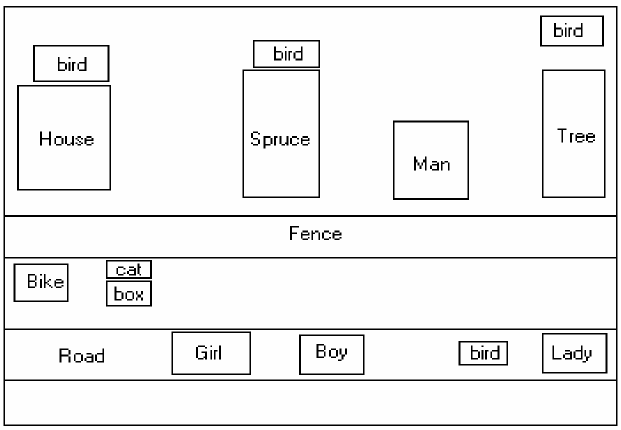
1) Rate the vividness of your visualization during the description phase (a rating scale ranging from 1 to 5).
2) Rate the vividness of your visualization during the question phase (a rating scale ranging from 1 to 5).
3) Make an assumption whether you usually imagine things in pictures or words.
4) What do you think we studied in this experiment (to ensure that the subject was naive about the hypothesis of the experiment)?

[3] When neural mechanisms are activated for prolonged periods, they become fatigued and need time to recover. If they do not get the time to recover perceptual aftereffects can occur (Finke, 1989, p. 44).

4.2 Analysis

The description was divided into eighteen *areas of interest* for which the correspondences of the eye movements were analyzed.

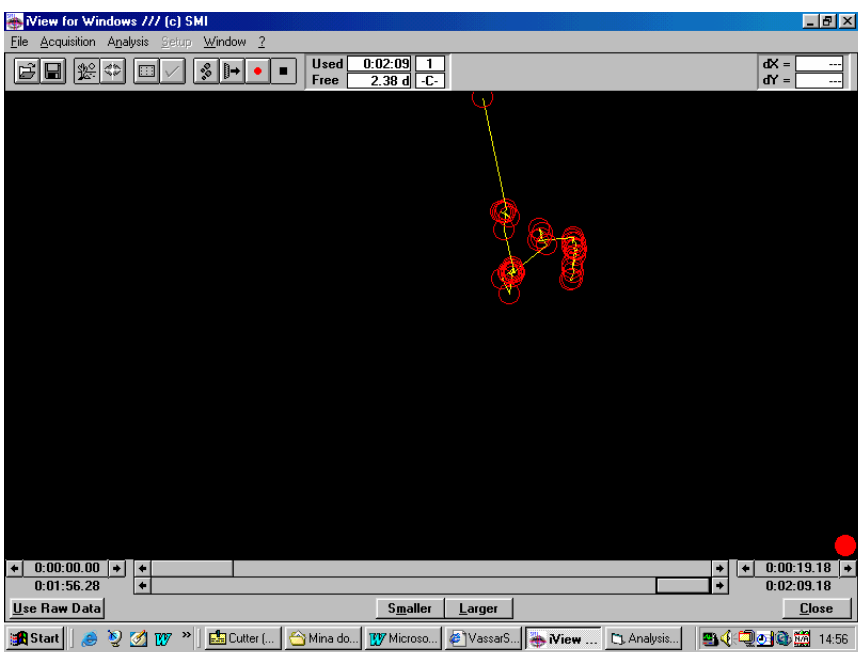
These areas of interest can be seen in picture 3 (the fence and the road were analyzed both as an object below the objects above them and as an object from the left to the right edge).



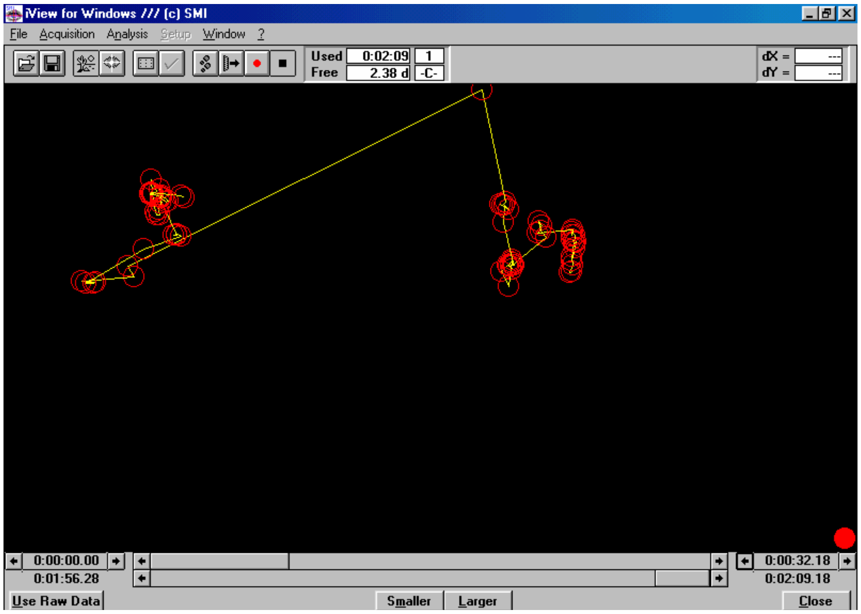
Picture 3 – Areas of interest in the description

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.

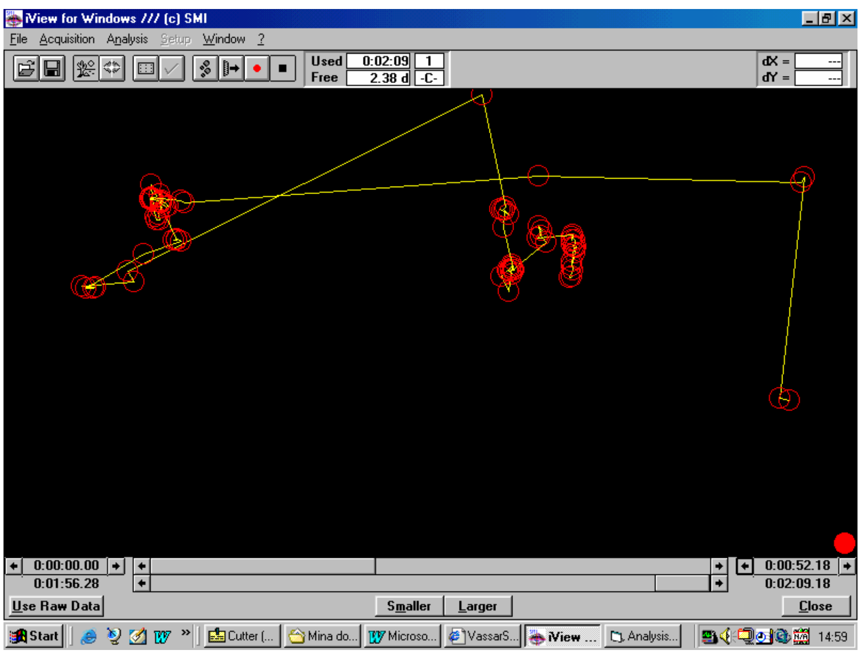
Pictures 4 – 7 are examples of how the eye movements for one subject is represented in iView (circles = fixations, lines = saccades).



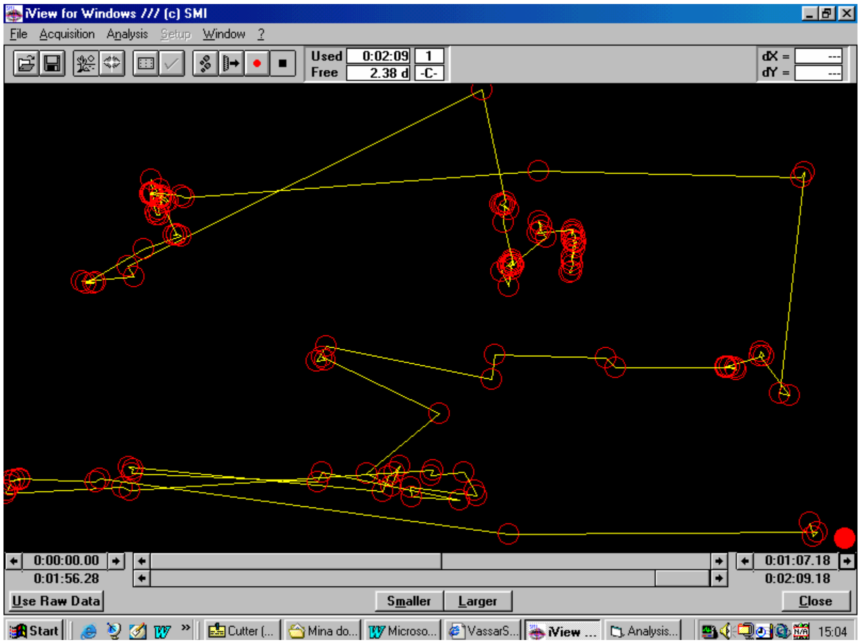
Picture 4 – The fixations and saccades after 19 s in the description, i.e. when the spruce and the bird in top of it had been described.



Picture 5 – The fixations and saccades after 32 s in the description, i.e. when the house with the bird on top of the chimney to the left of the spruce had been described.



Picture 6 – The fixations and saccades after 52 s in the description, i.e. when the tree to the right of the house and spruce had been described.



Picture 7 – The fixations and saccades after 1 min and 7 s in the description, i.e. when the man between the spruce and tree, and the fence below them had been described.

The eye movements of the test subjects were scored as *high correspondence, low correspondence* and *no correspondence*. High correspondence was considered according to the following criteria:

1. In the description phase the eye movement from one position to another must appear within 5 seconds after an object is mentioned in the description.
2. In the question phase the eye movement from one position to another must appear within 5 seconds before or after a subject *starts* to answer a question.
3. The spatial relations that appear in the eye movements must be correct in relation to each other.

The 5 second limit is based on video-based observations of how long it took before an actual eye movement appeared. It was found that this eye movement behavior was somewhat individual between subjects. Some subjects were very fast and the eye movements appeared almost immediately after a new area of interest had been mentioned in the description, or when a question had been asked. Other subjects were slower and it took a couple of seconds before the eye movements appeared.

In order to set an acceptable time limit for how long an eye movement could be delayed one must also consider the time between the mentioning of one object until the next object is mentioned. If the time limit is too long, a later object might interfere with an earlier, and it would be hard to say which object a particular eye movement corresponds to. After careful consideration to the individual differences and to when the objects are mentioned in the description, 5 seconds were considered a sound limit.

For the questions the case was a bit different, because now there were no objects that may interfere with another. The one thing to consider is however how long it took to answer a question when the question was finished. Again the subjects showed individual differences. Some subjects first moved their eyes to a new position and then answered,[4] while others answered fast and then moved their eyes to the position (probably to ensure that they answered correctly). This behavior made it necessary to have a time limit both before and after the question was answered. After careful consideration to the individual differences during the answering of the questions 5 seconds was again considered a sound limit both before and after a subject starts to answer a question.

For example, the situation is that the subject hears the following question:

> *In the described picture there was a house and a spruce. Was the house positioned to the left or to the right of the spruce?*

If the eye movements are considered as high correspondence the eyes have to move to the correct position (to the left) within 5 seconds before or after the subject starts to answer the question. Above this all the positions of the eye movements that appear for every question must be spatially correct in relation to each other. For example if the next question is the following:

> *In the picture there was a raking man wearing a blue overall and a tree with yellow and red leaves. Was the raking man in blue overall positioned to the right or to the left of the tree with the yellow and red leaves?*

If high correspondence is to be considered for these questions the spatial relations for the positions of the eye movements must be correct for all the mentioned objects, i.e. the spruce in the center, the house to the left of the spruce, the tree to the right of the spruce and the man between the spruce and the tree. Example of a finished description with high correspondence for most of the objects can be seen in picture 8.

Low correspondence was considered according to the following criteria:

1. In the description phase the eye movement from one position to another must appear within 5 seconds after an object is mentioned in the description
2. In the description phase the spatial relations that appear in the eye movements must be correct in relation to the previous mentioned object.
3. In the question phase the eye movement from one position to another must appear within 5 seconds before or after a question is answered.
4. In the question phase the spatial relations that appear in the eye movements must be correct for each question separately.

For low correspondence it is acceptable if the eyes move correctly for each object or question separately. Low correspondence often appears when the subject re-center or changes center, and for subjects that are shrinking the picture to a small area on the white screen. Example of a finished description with mostly low correspondences can be seen in picture 9.

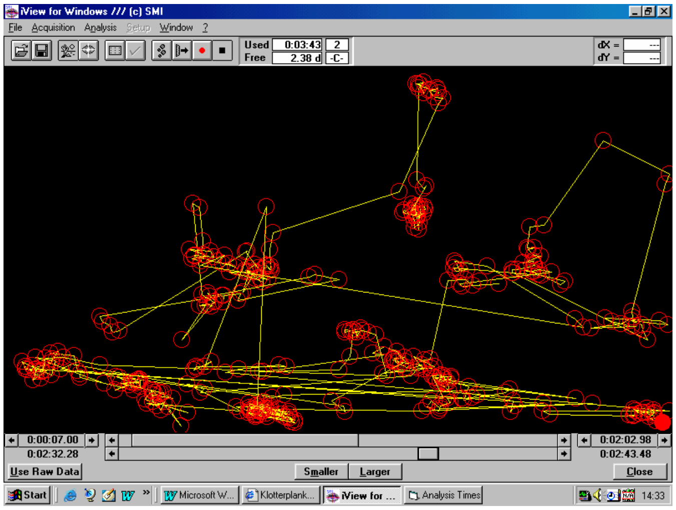
[4] This behavior is consistent with a picture perception study by Holsanova (2001, pp. 104-105), in which it was found that subjects sometimes have a delay for about 2-4 seconds from when a certain area of interest was fixated until it was verbally described.

Sometimes it happens that the subject have occasional saccades that move away from the object in question and then back again (this happens sometimes when the subjects blink there eyes or in some way corrects their gaze). These movements are not considered as a change in location for either the low or the high correspondence, and have been neglected in the analysis.

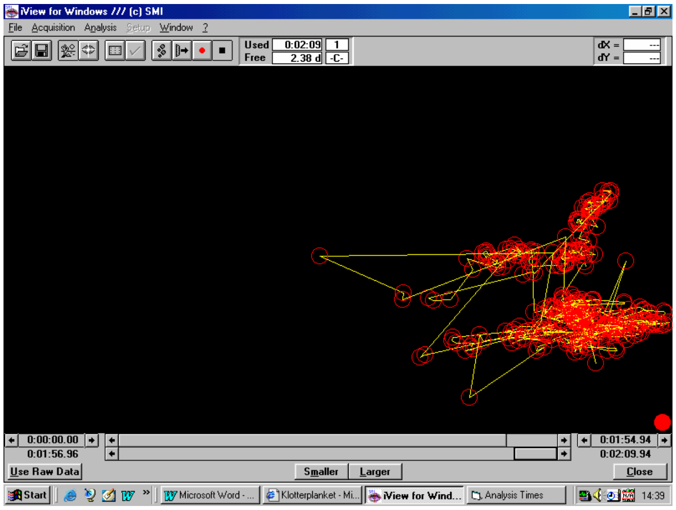
No correspondence was considered according to the following criteria:

1. No spatial relations from the mentally constructed image appear in the eye movements.

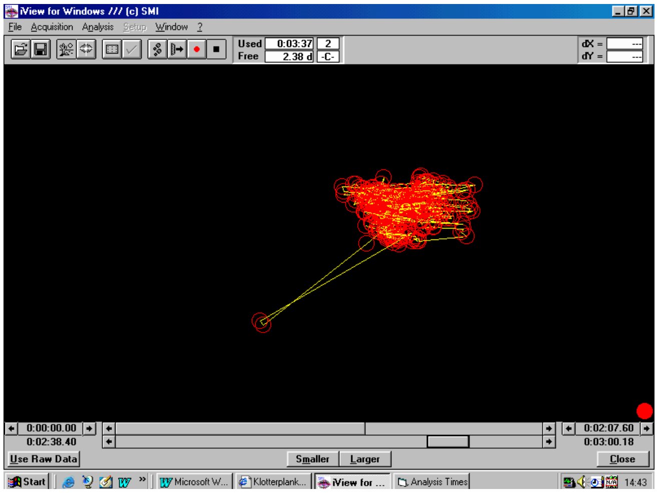
Example of a description with no correspondence can be seen in picture 10.



Picture 9 – Example of mostly high correspondences after the description is finished, i.e. most of the objects from the description have the correct spatial locations towards each other in the eye movements, e.g. in the top center the spruce is positioned, to the left of is the house, to the right is the tree, and below them goes the fence and the road.



Picture 8 – Example of mostly low correspondences after the description is finished, i.e. the eye movements changes the center sometimes and shrinks the picture to a small area on the white screen, e.g. the tree, the girl, the bike, and the boy (they are positioned in the lower blur of fixations) is positioned correctly to one another but because they are in the same vertical plane they are scored as low correspondence. However, it shall be noted that the spruce and the house still have high correspondence.



Picture 10 – Example of no correspondences, i.e. the eyes move randomly in a very small area.

4.3 Results

4.3.1 Correspondence in eye movements

A table was created for the correspondence of the eye movements for each subject and each area of interest in the description. A '**1**' means high correspondence, a '**1***' means low correspondence, a '**0**' means no correspondence, a '**!**' means the subject answered wrong but that the eyes moved in the answered direction, and a '**-**' means that data is missing. Table 1 describes the eye movements during the description and table 2 describes the eye movements during the questions.

Table 1 - Description

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Spruce	**1***	**1**	**1**	**0**	**1**	**1**	**1**	**-**	**1**	**1**	**1**	**1**
Bird in top of spruce	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**-**	**0**	**1**	**1**	**1**
House	**1***	**1**	**1***	**0**	**1**	**1**	**1**	**-**	**1**	**1**	**1**	**1**
Chimney with bird	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**-**	**0**	**1***	**0**	**1**
Tree	**1***	**1**	**1***	**0**	**1**	**1**	**1**	**-**	**1**	**1**	**1**	**1**
Bird above tree	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**-**	**0**	**0**	**1***	**1**
Raking man	**1***	**1**	**1***	**0**	**1***	**1**	**1**	**-**	**1**	**1**	**1**	**1**
Fence (below)	**0**	**1**	**1***	**0**	**1**	**1**	**1**	**-**	**1***	**1**	**1**	**1**
Fence (left to right)	**1**	**1**	**0**	**0**	**0**	**0**	**1***	**-**	**0**	**1**	**0**	**1**
Bike	**1***	**1**	**0**	**0**	**1***	**1**	**-**	**-**	**1***	**1**	**1**	**1***
Mailbox	**1***	**1**	**0**	**0**	**1***	**0**	**-**	**-**	**1***	**1**	**1**	**1**
Cat	**0**	**1**	**0**	**0**	**0**	**0**	**-**	**-**	**0**	**0**	**0**	**0**
Road (below)	**0**	**1**	**1***	**0**	**1***	**1**	**-**	**-**	**1**	**1**	**1**	**-**
Road (left to right)	**1**	**1**	**1**	**0**	**0**	**0**	**-**	**-**	**0**	**1***	**0**	**-**
Girl	**0**	**1**	**1***	**0**	**1***	**1***	**-**	**-**	**0**	**1***	**1**	**1**
Boy	**0**	**1**	**1***	**0**	**1***	**1***	**-**	**-**	**0**	**1***	**1**	**1***
Lady	**1**	**1**	**1***	**0**	**1**	**1***	**-**	**-**	**0**	**1***	**1**	**1**
Bird with worm	**1***	**1**	**0**	**0**	**1**	**0**	**-**	**-**	**0**	**0**	**0**	**0**

Table 2 - Questions[5]

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Question 1	**1***	**1**	**1***	**0**	**1***	**1***	**!**	**-**	**-**	**-**	**0**	**!**
Question 2	**1***	**1**	**1***	**-**	**1**	**1***	**1**	**-**	**-**	**!**	**!**	**1**
Question 3	**1***	**1**	**-**	**0**	**-**	**0**	**!**	**-**	**-**	**1***	**1**	**1***
Question 4	**1***	**1**	**1***	**0**	**1**	**1***	**1**	**-**	**-**	**!**	**1**	**1**
Question 5	**1***	**1**	**1***	**0**	**-**	**1***	**1**	**-**	**-**	**1***	**1**	**1**
Question 6	**1***	**1**	**1**	**0**	**-**	**0**	**1**	**-**	**-**	**0**	**1**	**-**
Question 7	**1***	**1**	**1***	**0**	**1**	**!**	**1**	**-**	**-**	**1***	**1**	**1**
Question 8	**1***	**1**	**0**	**0**	**-**		**1**	**-**	**-**	**0**	**!**	**!**

The data from the tables were then separated into two coding estimations:

<u>Low correspondence coding</u>
High correspondence (1), low correspondence (1*) and wrong answers with correct eye movements (!) are considered as results that support the hypothesis that the spatial locations are reflected in the eye movements.

<u>High correspondence coding</u>
Only high correspondence (1) is considered as a result that supports the hypothesis that the spatial locations are reflected in the eye movements.

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).

A test of significance of differences between proportions that is mathematically equivalent to the χ^2-test under one degree of freedom was employed. This was done by defining the possibility that a test subject moved his or her eyes in the right direction by chance. When the eyes move from one area of interest to another the eyes, in most cases, have 5 possibilities: they can stay in the same position, move up, move down, move to the right, or move to the left. In some cases they even have more possibilities, i.e. when they can move to many positions in one direction. But in some cases, like when the road and the fence are imagined the eyes move from the left to the right, i.e. you could say that the eyes have the possibility to stay, move in the vertical plane or in the horizontal plane. So to be on the safe side you could say that the possibility that the eyes move to a certain position by chance is at least not larger then 1/3 (move vertical, horizontal or stay).

The amounts of correct and false eye movements for low correspondence coding and high correspondence coding for both the description and the questions were counted and tested in direction significance, i.e. with the proportion of 1/3 that the eyes moved to the correct position by chance. These results are presented in table 3 and 4.

How well the eye movements corresponded with the spatial locations for all the areas of interest during both the description and the questions for each subject is presented in diagrams 1-4. This is done both for low correspondence coding and high correspondence coding.

[5] Question 8 was a “trick” question, i.e. it was asked about an object that did not exist in the description. The point was to see if the subjects scanned the entire scene in search of this object.

Table 3

Description	# of 1's	# of 0's	Direction significance
Low correspondence coding	121	84	$p \leq .0001$
High correspondence coding	85	120	$p \leq .05$

Table 4

Questions	# of 1's	# of 0's	Direction significance
Low correspondence coding	58	13	$p \leq .0001$
High correspondence coding	26	45	$p = .65$

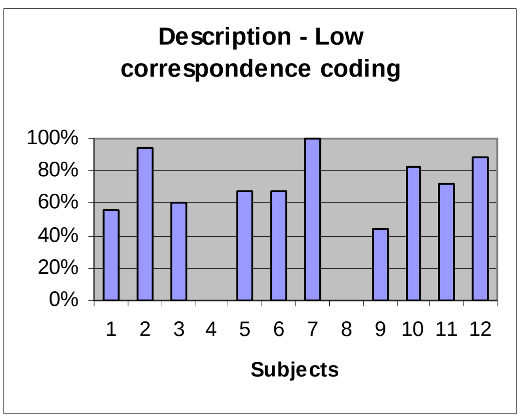


Diagram 1

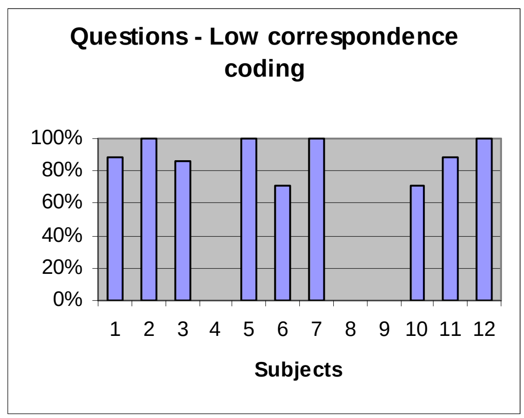


Diagram 2

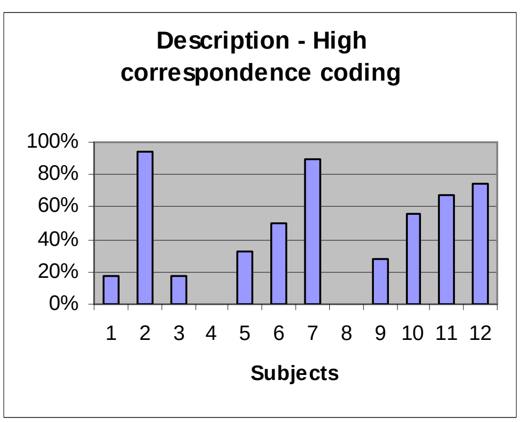


Diagram 3

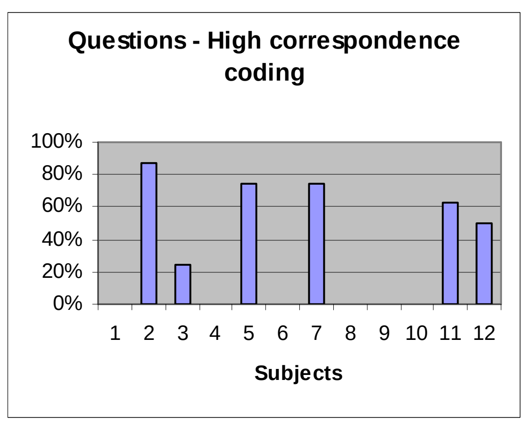


Diagram 4

4.3.2 Subject ratings

To test if there were any relationship between correspondence in eye movements and the subjects vividness ratings a simple correlation test was employed, i.e. to see if high and low correspondence correlate with high and low vividness ratings. However, no correlation was found either for the description phase (correlation coefficient = -0.0054) or the question phase (correlation coefficient = 0.3055).

A similar test was employed between correspondence in eye movements and the subjects decision whether they usually imagine things in pictures or words, i.e. to see if high correspondence correlates with pictures and low correspondence with words. However, no correlation was found either for the description phase (correlation coefficient = -0.0430) or the question phase (correlation coefficient = 0.2782).

4.4 Discussion

The main conclusion from these findings is that eye movements during a visualization of a description have high correspondence with the spatial positions of objects in the description. The eye movement direction was significant both for the low correspondence coding ($p \leq .0001$) and the high correspondence coding ($p \leq .05$). For the high correspondence coding, 6 of the 11 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest. For the low correspondence coding, 10 of the 11 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest. Only one subject showed no correspondence at all.

It was also found that the eye movements during questions about the description corresponded with the spatial positions of the objects in question. But the eye movement direction was only significant for the low correspondence coding ($p \leq .0001$). However, 4 of the 5 subjects that had a 50% or larger correspondence for the high correspondence coding during the description also had a 50% or larger correspondence for the high correspondence coding during the questions. Because of technical problems the data for subject 9 is missing during the question phase.

5. Experiment 2

One group of subjects was instructed to view a picture, which was later to be visually imagined. This was done by letting the subject answer questions about the picture’s objects and spatial relations while he or she looked at a white screen.

Based on the results by Brandt and Stark (1997), and Laeng and Teodorescu (2002) the hypothesis is that the spatial positions of objects in the picture are reflected by the eye movements during the visualization of it (during the questions).

This means that if the subject is to answer if one object in the picture is to the left or to the right of another object, the eye movements should be similar to how the spatial relations of these objects were positioned in the observed picture.

5.1 Method

5.1.1 Participants

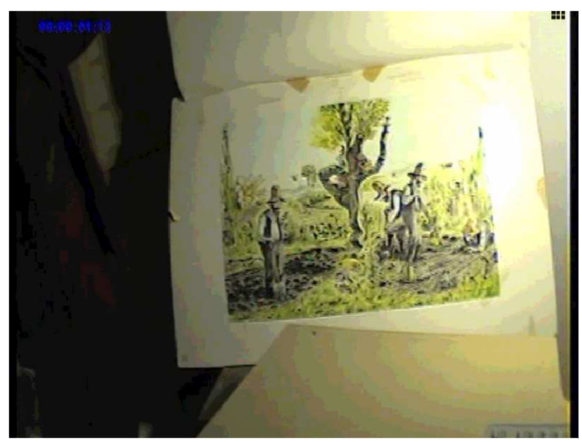
Ten students at the University of Lund, 5 females and 5 males volunteered to participate to an experiment in cognition science. All subjects reported normal vision, or corrected to normal (with contact lenses or glasses). The participants were told that their pupil size was being measured during a visualization task. At the end of each session, participants were questioned about their beliefs about what had been studied in the experiment. It was confirmed that nine were naive about the fact that their eye movements were recorded and that they had no specific knowledge about the experimenters' expectations. One of them did see through the nature of the experiment, and realized that the eye movements were recorded and that there was some sort of mental images versus picture perception that were studied. This participant will be handled separately during the analysis.

5.1.2 Apparatus and stimuli

The output and the eye tracker that was used were the same as in experiment 1.

The visual stimuli used in the experiment consisted of a complex picture (500mm $\times$ 700mm) (picture 11) and a white screen (the same as in experiment 1)[6]. The participants were seated in front of the picture or the white screen at a distance of 150 cm (picture 12).

Picture 11 – The picture that the subjects observed.



Picture 12 - The picture in the environment as it was observed by the subjects.

[6] This picture was chosen because it had been used in several picture perception studies before (Holsanova, 1999; Holsanova, 2001).

5.1.3 Procedure

Before the actual experiment the same calibration procedure as in experiment 1 was done.

The experiment consisted of two main phases, one perception phase and one question phase. Afterwards the subject was asked a few questions about the experiment. Eye movements were recorded both in the perception phase and in the question phase.

At the beginning of the perception phase the subjects received the following instructions:

> *"You will soon se a picture. We want you to study the picture as thoroughly as possible. During your study of the picture we will measure your pupil size"*.

The picture was shown during 1 minute. Then a white screen was placed over the picture and the subjects were told that they now where about to listen to, and answer, a couple of prerecorded questions. 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 screen. They were also informed that their pupil size again would be measured during these questions. These instructions took about 40 sec, in order to prevent afterimages. After the subjects had answered the questions about the picture they received the following instructions:

1) Rate the vividness of your visualization during the imagery phase (a rating scale ranging from 1 to 5).
2) Make an assumption whether you usually imagine things in pictures or words.
3) What do you think we studied in this experiment (to ensure that the subject was naive about the hypothesis of the experiment)?

5.2 Analysis

The analysis of the eye data was done with *iView for windows* as in experiment 1.

The eye movements were also as in experiment 1, and with the same criteria, scored as *high correspondence*, *low correspondence* and *no correspondence*.

5.3 Results

5.3.1 Correspondence in eye movements

As in experiment 1 a table was created (table 5) for the correspondence of the eye movements for each subject and each question.

Table 5 – Questions[7]

	#1	#2	#3	#4	#5	#6	**#7**	#8	#9	#10
Question 1	0	0	0	0	0	0	**0**	0	0	1*
Question 2	1*	1*	1*	1*	0	0	**1***	1*	1*	1*
Question 3	1*	0	1*	0	0	0	**0**	1*	0	1*

[7] Question 8 was as in experiment 1 a "trick" question with the same purpose.

Question 4	1*	!	1*	1*	0	0	**1***	1*	0	0
Question 5	1*	1*	1*	1*	0	0	**0**	1*	1*	0
Question 6	0	0	0	1*	0	0	**0**	0	0	0
Question 7	1*	0	1*	1*	0	0	**0**	1*	0	0
Question 8	1*	0	-	-	-	-	**-**	-	-	-

Test subject 7 was the one who saw through the nature of the experiment.

The data was analyzed with the same coding estimation as experiment 1, and the direction significance was also tested in the same way as in experiment 1 (table 6). How well the eye movements corresponded with the spatial locations during the questions for each subject is presented in diagrams 5 and 6. This is done both for low correspondence coding and high correspondence coding.

Table 6

Questions	# of 1's	# of 0's	Direction significance
Low correspondence coding	30	42	p = .17
High correspondence coding	0	72	p = 1

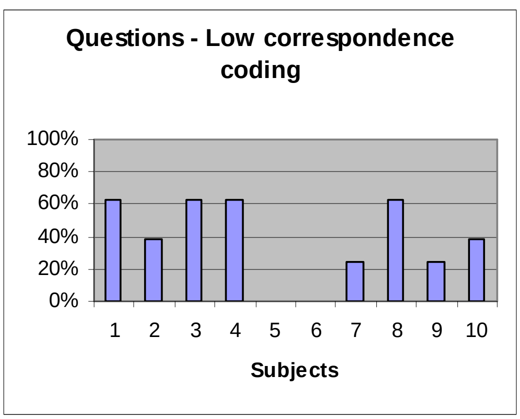


Diagram 5

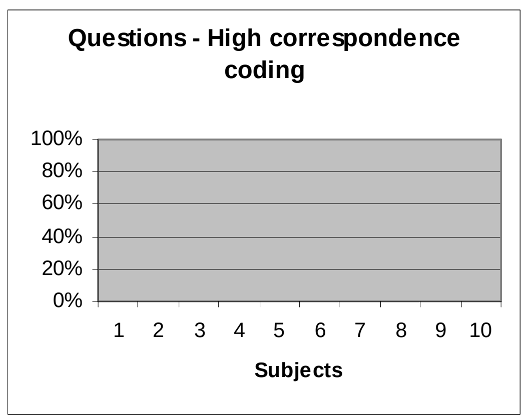


Diagram 6

5.3.2 Subject ratings

To test if there was any relationship between correspondence in eye movements and the subjects vividness ratings a simple correlation test as in experiment 1 was employed. However, no correlation was found (correlation coefficient = -0.2675).

As in experiment 1 a similar test was employed between correspondence in eye movements and the subjects' decision if they usually imagine things in pictures or words. However, no correlation was found (correlation coefficient = 0.0945).

5.4 Discussion

The main conclusion from these findings is that eye movements during the questions of the observed picture did not show significant correspondence. The eye movement direction was significant neither for the low correspondence coding ($p = .17$) nor for the high correspondence coding ($p = 1$). For the high correspondence coding none of the subjects' eye movements showed correspondence with the spatial relations of the questions. For the low correspondence coding, 4 of the 10 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest.

Because test subject 7 saw through the nature of the experiment this subject was intended to be treated separately, but because the results did not reach significance even when this subject was included there was no reason to exclude the subject.

If the results from experiment 1 and experiment 2 are compared, it is found that for low correspondence the mean correspondence for the low correspondence coding is 82% in the question phase in experiment 1 and 40.5% in experiment 2. For the high correspondence coding the mean correspondence is 36.6% in experiment 1 and 0% in experiment 2.

To test if there were any significant difference for the eye movement correspondences in experiment 1 and 2 a test of significance of differences between proportions that is mathematically equivalent to the χ^2-test under one degree of freedom was employed (table 7).

A diagram that compares the mean value of subjects' eye movement correspondence during the retelling of the description and the retelling of the picture for both low correspondence and high correspondence coding is presented in diagram 7 and 8.

Table 7

Questions	Picture	Verbal	Significance
Low correspondence coding	40,5%	81,7%	$p \leq .0001$
High correspondence coding	0%	36,6%	$p \leq .0001$

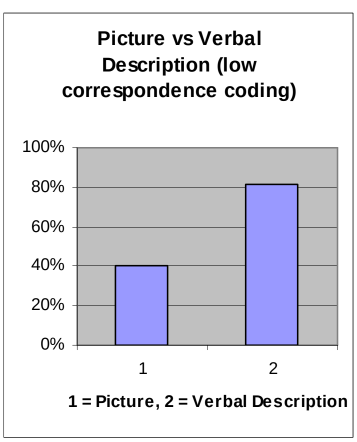


Diagram 7

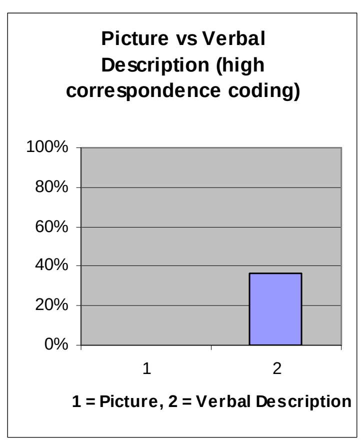


Diagram 8

These results, with the significance $p \leq .0001$ both for the low correspondence coding and the high correspondence coding, show that there is a significant difference in eye movement correspondences for experiment 1 compared to experiment 2.

So why are the results not as good in experiment 2 as in experiment 1? Could it be that the eyes reflect a mental image during a verbal description but not when imagining an earlier observed picture? This explanation was not found very likely when previous studies had indicated the opposite. The answer was considered to be found in *how* the experiment was designed. One design problem could be that the questions about the picture were too easy to answer after observing the picture for a whole minute, i.e. the subjects did not have to employ eye movements when recalling a mental image of the picture to answer the questions. It is, for example, easy to remember the amount of men and cats from the picture. Another explanation can be found from a study by Demarais and Cohen (1998). They mean that many experimental imagery tasks do not involve the inspection of the extremities of a linearly extended image, but, rather, internal structural details of an image that may be relatively compact, i.e. large eye movements are suppressed because they tend to move the image and disrupt the inspection. This means that most eye movements are more likely to be low amplitude “fixation” movements (microsaccades) rather than the large, saccades evoked by spatially linear transitive inference tasks. It is possible that the design of the questions were of a nature that suppressed these larger eye movements.

To eliminate these possible problems another experiment was created in which the subjects only were to observe the picture in 30 seconds and not answer questions about the picture, but instead with their own words retell what they had seen in the picture.

6. Experiment 3

One group of subjects was instructed to view a picture, which was later to be visually imagined when they freely with their own words were asked to describe the picture while watching a white screen. The picture was the same as in experiment 2.

The hypothesis is that the spatial positions of objects in the picture are reflected by the eye movements during the visualization of it (during the description).This means that when objects from the picture are mentioned, during the retelling, the eye movements should be similar to how the spatial relations of these objects were positioned in the picture.

6.1 Method

6.1.1 Participants

Twelve students at the University of Lund, 6 females and 6 males volunteered to participate to an experiment in cognition 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 there eye movements were recorded and that they had no specific knowledge about the experimenters’ expectations. Two of them discussed the possibility that what you looked at during and after the picture was measured, but it was not considered that they had seen through the nature of the experiment.

6.1.2 Apparatus and stimuli

The output and the eye tracker that was used were the same as in experiment 1 and 2.

The visual stimuli used in the experiment were the same as in experiment 2.

6.1.3 Procedure

Before the actual experiment the same calibration procedure as in experiment 1 and 2 was done.

The experiment consisted of two main phases, one perception phase and one retelling phase. Afterwards the participant was asked a few questions about the experiment. Eye movements were recorded both in the perception phase and in the retelling phase.

At the beginning of the perception phase the subjects received the following instructions:

> *"You will soon se a picture. We want you to study the picture as thoroughly as possible. During your study of the picture we will measure your pupil size"*.

The picture was shown in 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 screen. They were also informed that their pupil size again would be measured during these questions. These instructions took about 40 seconds, in order to prevent afterimages. After the subjects had answered the questions about the picture they received the following instructions:[8]

1) What do you think the meaning of this experiment was (to ensure that the subject was naive about the hypothesis of the experiment)?
2) Rate the vividness of their visualization during the imagery phase (a rating scale ranging from 1 to 5).
3) Make an assumption whether you usually imagine things in pictures or words.

6.2 Analysis

To analyze the data the test subjects' descriptions were first transcribed so it was possible to analyze *when* certain objects are mentioned. Example:

00:42 – And there is a tree in the middle
00:54 – In which it lives small animals and stuff like that

[8] The question to test if the subjects were naive of the experiments hypothesis was slightly different from experiment 2 and was asked before the ratings, because it was considered that the rating questions could give the subjects clues about the experiments nature. To ask about purpose instead of what had been studied was considered less leading.

The first numbers are the time when the sentence starts, and then the test subject's description follows.

The eye movements of all test subjects were as in the earlier experiments scored as *high correspondence*, *low correspondence* and *no correspondence*. But because the test subjects themselves freely are describing the picture the eye movements are analyzed according to certain *areas of interest* from their descriptions. But these areas can be mentioned in a number of ways and can be divided into smaller units that together form a *superfocus* (Holsanova, 1999, pp. 16-17). Example of a superfocus:

> 01:20 – And ehhh to the **left** in the picture
> 01:23 – there are large **daffodils**
> 01:26 – it looks like there also sat some **animals there** perhaps

In this example the flowers to the left in the picture is described by the direction left, they being daffodils and that there were animals on them.

Another difference in this experiment is that when you describe something with your own words the eye movements can move to a certain area before you start to describe that area. Therefore high correspondence was in this experiment slightly different than in experiment 2. High correspondence was considered according to the following criteria:

1. The eye movement from one position to another must appear within 5 seconds before or after an area of interest is mentioned.
2. The spatial relations that appear in the eye movements must be correct in relation to each other.

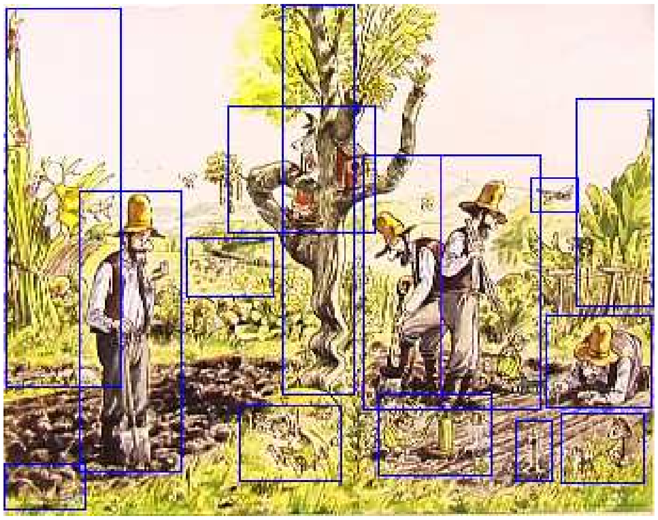
Low correspondence was considered according to the following criteria:

1. The eye movement from one position to another must appear within 5 seconds before or after an area of interest is mentioned.
2. The spatial relations that appear in the eye movements must be correct in relation to the previous mentioned area of interest.

The time limit 5 seconds was used again. The reason to this is that in a retelling case there is again the possibility that one area of interest may interfere with another if you have too generous a time limit, e.g. if a new area of interest is being retold before the time limit for the previous one has ended. Above this the subjects did also show individual differences in eye movements and the retelling of an area of interest. Some subjects first moved their eyes to a new position and then started the retelling of that area of interest,[9] while others started the retelling of an area of interest and then moved their eyes to the new location. This behavior made it necessary to have a time limit both before and after an area of interest was mentioned. After careful consideration to the individual differences and the possibility that one area of interest interfere with another the 5 seconds were again considered a sound limit both before and after a subject starts the retelling of an area of interest.

[9] This behavior is again consistent with the study by Holsanova (2001, pp. 104-105), in which it was found that, in picture perception, subjects sometimes have a delay for about 2-4 seconds from when a certain area of interest was fixated until it was verbally described.

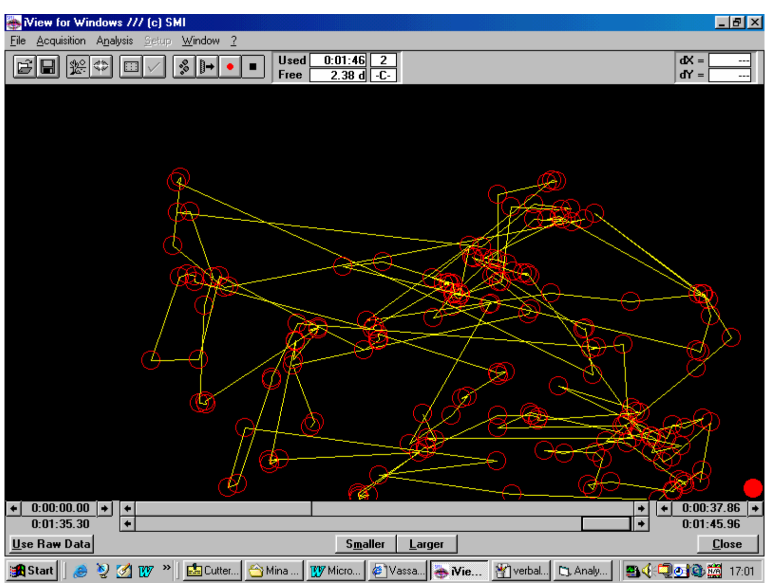
The areas of interest that were mentioned by the subjects varied depending on how much they remembered and what they considered of interest. But after transcribing all the subjects' retellings the areas of interest in picture 13 were defined.



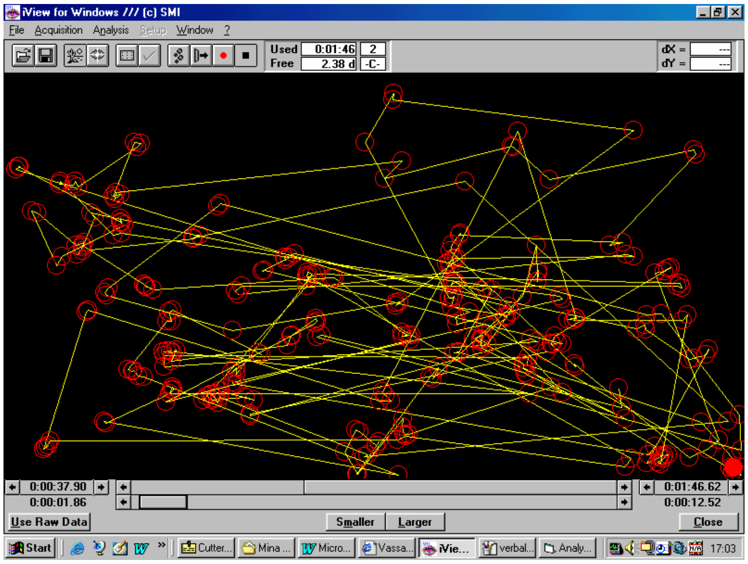
Picture 13 – Areas of interest in the picture

Because it was common to describe the versions of the man like: *"there were four versions of a man"*, and the cat like: *"there were four versions of the cat"*, those as a unit were also considered as an area of interest. For example, if a subject said there were four men the eyes had to go from the left edge to the right edge (or vice versa) to get correspondence.

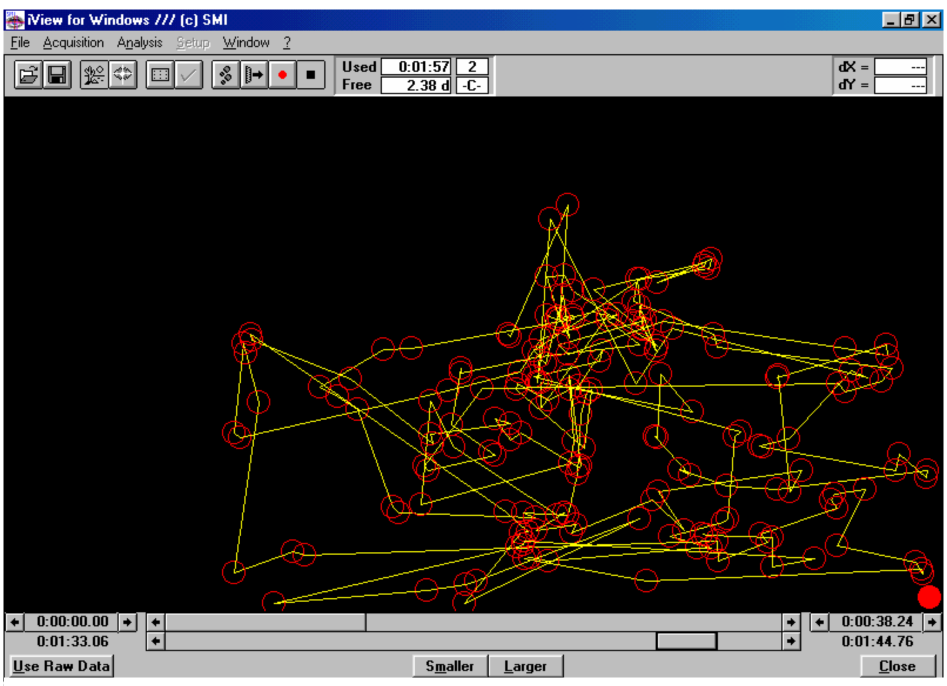
When the picture was described it was common that the subjects shrunk the dimensions, i.e. that the saccades were smaller like Brandt and Stark observed (1991, pp. 32-33), or that they smeared it out a bit (probably because the white screen was slightly larger then the picture). This can be seen in pictures 14 – 17. As previously it was also common that the subjects re-centered and changed the center from time to time (low correspondence).



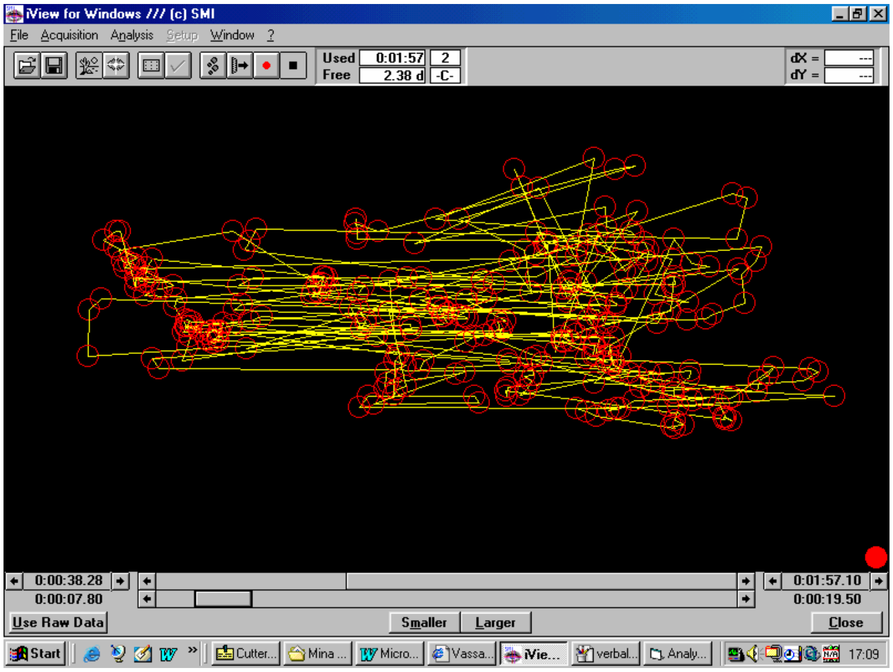
Picture 14 – One test subject's fixations and saccades after the observation of the picture.



Picture 15 – The fixations and saccades for the same test subject after the retelling of the scene. This subject smears the picture on a larger area (probably because the white screen where slightly larger than the picture).



Picture 16 – One test subject's fixations and saccades after the observation of the picture.



Picture 17 – The fixations and saccades for the same subject after the retelling. This subject shrinks the picture in the vertical plane.

6.3 Results

6.3.1 Correspondence in eye movements

As in experiment 1 and 2 a table, with the same criterion, was created for the correspondence of the eye movements and the described areas of interests (table 8).

Table 8 – Retelling of the picture

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Tree	**1**	**1**	**-**	**1***	**1**	**1**	**1**	**1**	**1**	**1**	**1***	**0**
Daffodils	**1**	**1**	**-**	**1***	**-**	**1**	**1**	**-**	**1**	**1**	**-**	**0**
Cows	**1**	**1***	**-**	**1***	**-**	**-**	**-**	**-**	**-**	**0**	**1***	**0**
Versions of the man	**1***	**1**	**1***	**1***	**1**	**1**	**1**	**1**	**1***	**1**	**1**	**0**
Nesting box, birds in the tree	**1**	**-**	**1***	**1***	**1**	**1**	**1**	**1**	**1**	**1**	**1***	**0**
The man raking	**1**	**1**	**-**	**0**	**0**	**1***	**1***	**-**	**1**	**1***	**-**	**0**
Dragonfly	**1**	**-**	**-**	**-**	**-**	**1**	**-**	**-**	**-**	**-**	**-**	**0**
Versions of the cat	**0**	**1**	**1***	**1***	**1**	**1**	**1**	**-**	**0**	**1**	**1***	**0**
Sign	**1**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0**
The man sawing	**1**	**1**	**1***	**0**	**-**	**1***	**-**	**1**	**1**	**1***	**1**	**0**
Up to the right	**1**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**1**	**-**	**0**
Cart	**1**	**-**	**-**	**1***	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0**
The man digging	**-**	**1**	**-**	**0**	**1**	**1**	**1***	**-**	**-**	**1**	**1**	**0**
Sky above	**-**	**-**	**-**	**-**	**1**	**-**	**-**	**-**	**1**	**-**	**-**	**0**
The cat with waterpump	**-**	**-**	**-**	**-**	**-**	**1***	**-**	**1**	**-**	**-**	**-**	**0**

The direction significance was also tested in the same way as in experiment 1 and 2 (table 9), and how well, i.e. for each area of interest, the eye movements corresponded with the spatial locations during the retelling for each subject is presented in diagrams 9 and 10. This is done both for low correspondence coding and high correspondence coding.

Table 9

Retelling	# of 1's	# of 0's	Direction significance
Low correspondence coding	77	26	$p \leq .0001$
High correspondence coding	56	46	$p \leq .0001$

6.3.2 Subject ratings

To test if there were any relationship between correspondence in eye movements and the subjects vividness ratings a simple correlation test as in experiment 1 and 2 was employed. However, no correlation was found (correlation coefficient = -0.2202).

As in experiment 1 and 2 a similar test was also employed between correspondence in eye movements and the subjects decision whether they usually imagine things in pictures or words. However, no correlation was found (correlation coefficient = 0.3163).

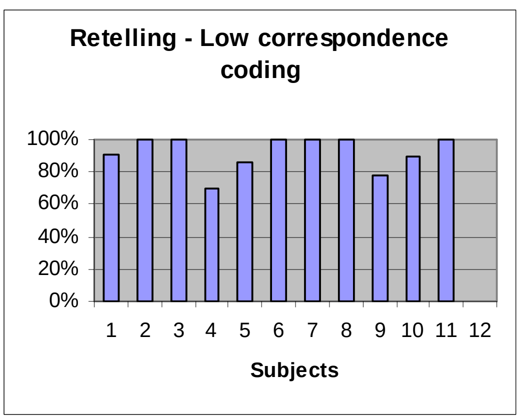


Diagram 9

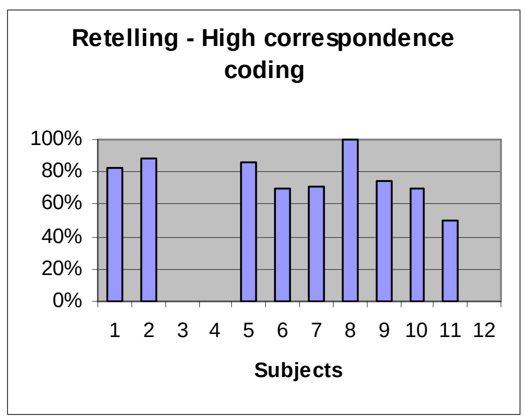


Diagram 10

6.4 Discussion

The main conclusion from these findings is that eye movements during the retelling of the observed picture did show significant correspondence. The eye movement direction was significant both for the low correspondence coding ($p \leq .0001$) and for the high correspondence coding ($p \leq .0001$). For the high correspondence coding 9 of the 12 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest, and for the low correspondence coding 11 of the 12 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest. This experiment shows strong indication that the eye movements are reflected when a picture is described during a mental visualization of it. The results are even better than the results from the questions of the verbal description in experiment 1. This observation suggests that it is possible that experiment 1 in some way also had similar problems as experiment 2. With this possibility in mind, and in order to make it possible to compare the results from experiment 3 with a verbal experiment, a new experiment was created. This experiment was the same as experiment 1 with the exception that there were no questions. The subjects were now to retell the description in the same manner as in experiment 3.

7. Experiment 4

One group of subjects was instructed to listen to a prerecorded verbal description, which was later to be visually imagined when they freely with own words, while watching a white screen, were told to describe the description they had listened to. The description was the same as in experiment 1.

The hypothesis is that the positions of objects in the description are reflected by the eye movements during the description and during the retelling of it. This means that when the subject listens to the verbal description the eye movements should follow the spatial relations of the objects in the description, and when the objects from the description are mentioned, during the retelling, the eye movements should be similar to how the spatial relations of these objects were positioned in the description.

7.1 Method

7.1.1 Participants

Twelve students at the University of Lund, 6 females and 6 males, volunteered to participate to an experiment in cognition 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 there eye movements were recorded and that they had no specific knowledge about the experimenters' expectations. Although two of them discussed the possibility that what you looked at during and after the description/picture was measured, but it was not considered that they had seen through the nature of the experiment.

7.1.2 Apparatus and stimuli

The output and the eye tracker that was used were the same as in experiment 1, 2 and 3.

The stimuli were the same as in experiment 1.

7.1.3 Procedure

Before the actual experiment the same calibration procedure as in experiment 1, 2 and 3 was done.

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. Afterwards the participants were asked a few questions about the experiment. Eye movements were recorded both while subjects listened to the verbal description and while they retold it.

At the beginning of the description phase the subjects received the following instructions:

> *"You will soon hear a prerecorded verbal description. The description will describe a two dimensional picture. We want you to listen to the description as carefulyl as possible and to visualize 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 freely look wherever you want on the white screen".*

When the description finished the subject was told with own words freely to describe the description they just had listened to. 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 screen. They were also informed that their pupil size again would be measured during these questions. These instructions took about 40 sec, in order to prevent afterimages. After the subjects had answered the questions about the picture they received the following instructions:

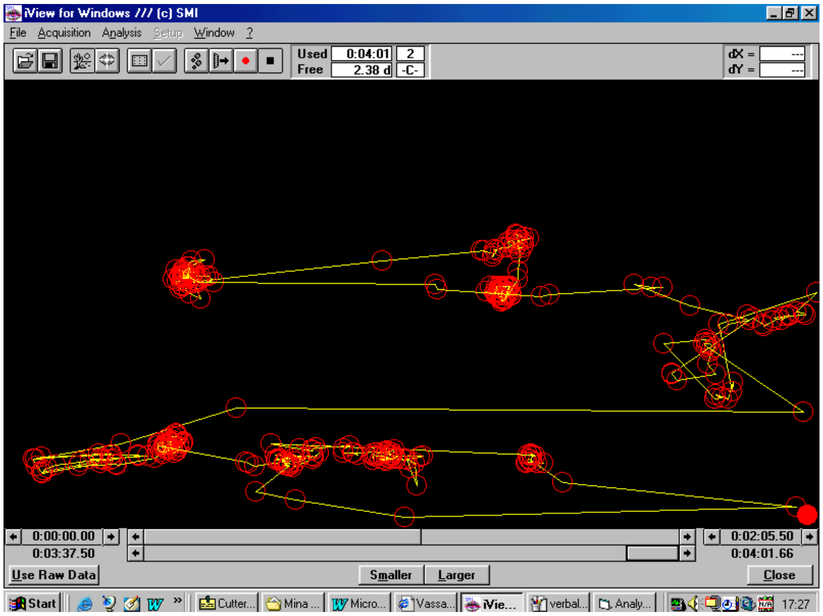
1) What do you think the meaning of this experiment was (to ensure that the subject was naive about the hypothesis of the experiment)?
2) Rate the vividness of your visualization during the description phase (a rating scale ranging from 1 to 5).
3) Rate the vividness of your visualization during the retelling phase (a rating scale ranging from 1 to 5).
4) Make an assumption whether you usually imagine things in pictures or words.

7.2 Analysis

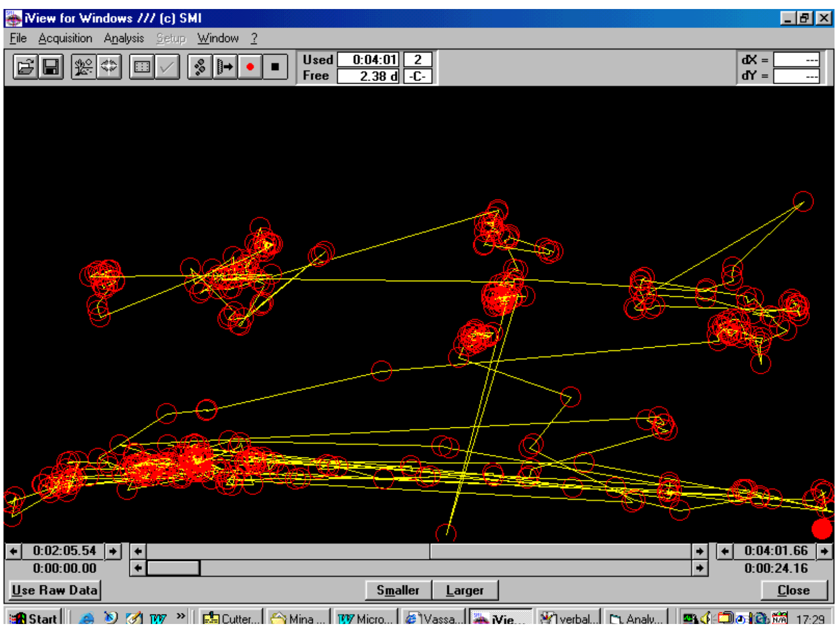
The analysis of the eye data was done with *iView for windows* as in experiment 1, 2 and 3.

The retellings were transcribed in the same way as in experiment 3 and the areas of interest were the same as in experiment 1. The eye movements were also coded as in experiment 1 and 3 (during the description as in experiment 1 and during the retelling as in experiment 3), and with the same criteria scored as *high correspondence*, *low correspondenc*e and *no correspondence*.

The eye movements were like in the previous experiments sometimes re-centered and shrunk into a smaller area. But the smearing effect that was found in experiment 3 was not found in this experiment (probably because the white screen was the same in both the description phase and the retelling phase). The eye movements almost always had the same proportions during the description phase and the retelling phase. An example of this can be seen in pictures 18 and 19.



Picture 18 – The fixations and saccades for one subject after the description phase.



Picture 19 – The fixations and saccades for the same subject after the retelling phase.

7.3 Results

7.3.1 Correspondence in eye movements

As in experiments 1, 2 and 3 tables were created, with the same criterion, for the correspondence of the eye movements for each subject and each area of interest during the description (table 10) and the retelling (table 11).

Table 10 – Description

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Spruce	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Bird in top of spruce	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**1***	**0**	**0**
House	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**
Chimney with bird	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**
Tree	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Bird above tree	**1**	**1**	**0**	**0**	**-**	**-**	**1**	**0**	**0**	**0**	**0**	**0**
Raking man	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Fence (below)	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Fence (left to right)	**-**	**1***	**0**	**1***	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**0**
Bike	**-**	**1**	**1***	**1***	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**0**
Mailbox	**-**	**1**	**0**	**1***	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**0**
Cat	**-**	**1***	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**
Road (below)	**-**	**1**	**0**	**1***	**1**	**0**	**1**	**1**	**1**	**1**	**0**	**0**
Road (left to right)	**-**	**1***	**0**	**1***	**1**	**1**	**1***	**1***	**0**	**0**	**0**	**0**
Girl	**-**	**-**	**1**	**1***	**1**	**1**	**1**	**1**	**1***	**1**	**0**	**0**
Boy	**1***	**-**	**1**	**1***	**1**	**0**	**1**	**0**	**1***	**1**	**0**	**0**
Lady	**-**	**-**	**1**	**1***	**-**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Bird with worm	**-**	**-**	**0**	**1***	**-**	**1**	**0**	**1**	**1**	**0**	**0**	**0**

Table 11 – Retelling of the description

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12
Spruce	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1***	**1***
Bird in top of spruce	**1**	**1**	**0**	**1***	**1**	**1**	**-**	**0**	**1**	**-**	**0**	**0**
House	**1**	**1**	**1**	**1**	**1**	**1**	**-**	**1**	**1**	**1**	**1***	**1***
Chimney with bird	**0**	**1***	**0**	**1***	**1**	**1**	**-**	**-**	**1**	**-**	**0**	**0**
Tree	**1***	**1**	**1***	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1***	**0**
Bird above tree	**-**	**-**	**1***	**0**	**1**	**1***	**1***	**0**	**1**	**1**	**1***	**0**
Raking man	**1***	**1**	**1***	**1***	**1**	**-**	**1**	**1**	**1***	**1**	**1***	**0**
Fence (below)	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1***
Fence (left to right)	**0**	**1**	**1***	**-**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1***
Bike	**1**	**1**	**1***	**1***	**1**	**1**	**1**	**-**	**-**	**1**	**0**	**0**
Mailbox	**-**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Cat	**-**	**1***	**0**	**0**	**1***	**0**	**-**	**1***	**-**	**0**	**0**	**0**
Road (below)	**-**	**1**	**-**	**0**	**1**	**1**	**1**	**-**	**1**	**1***	**0**	**0**
Road (left to right)	**-**	**1**	**-**	**-**	**1**	**1**	**1**	**-**	**1**	**1***	**0**	**0**
Girl	**0**	**1**	**1***	**0**	**1**	**1**	**-**	**1**	**1!**	**1**	**0**	**0**
Boy	**-**	**1**	**1***	**1***	**1**	**1**	**-**	**1**	**1**	**1***	**0**	**0**
Lady	**-**	**1**	**1**	**1***	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**
Bird with worm	**-**	**-**	**0**	**0**	**-**	**1***	**-**	**-**	**1**	**-**	**0**	**0**

The data was analyzed with the same coding estimation as experiment 3, and the direction significance was also tested in the same way as in experiment 1, 2 and 3 (table 12 and 13), and how well, i.e. for each area of interest, the eye movements corresponded with the spatial locations during both the description and the retelling for each subject is presented in diagrams 11-14. This is done both for low correspondence coding and high correspondence coding.

Table 12

Description	# of 1's	# of 0's	Direction significance
Low correspondence coding	128	71	$p \leq .0001$
High correspondence coding	109	90	$p \leq .0001$

Table 13

Retelling	# of 1's	# of 0's	Direction significance
Low correspondence coding	137	46	$p \leq .0001$
High correspondence coding	101	82	$p \leq .0001$

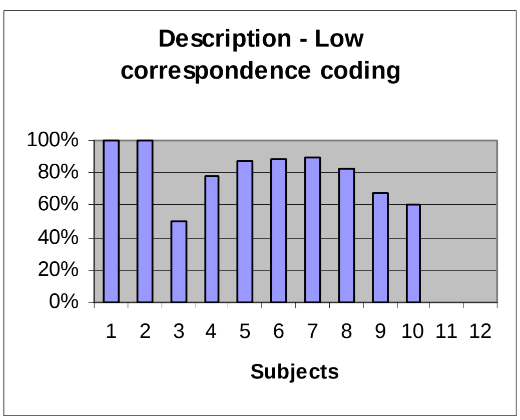


Diagram 11

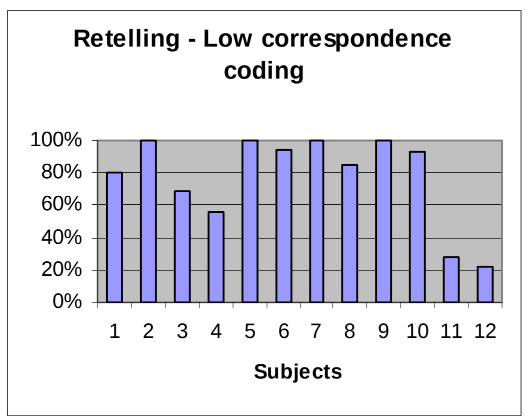


Diagram 12

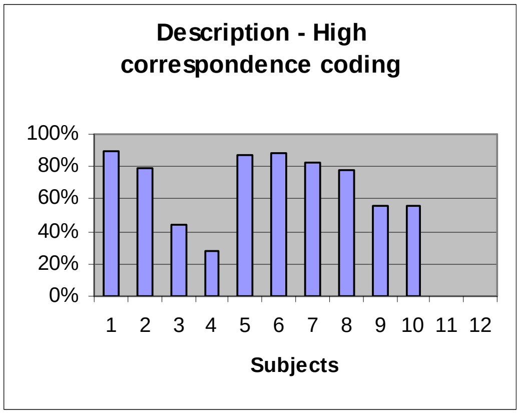


Diagram 12

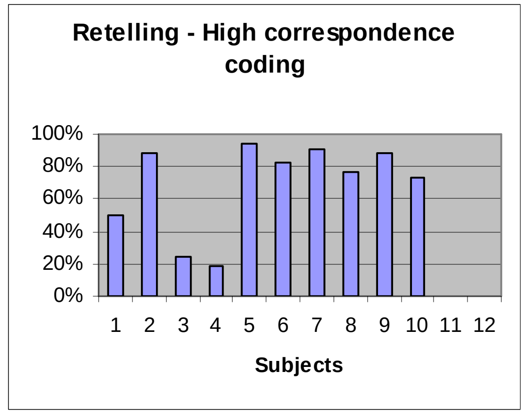


Diagram 14

7.3.2 Subject ratings

To test if there was any relationship between correspondence in eye movements and the subjects' vividness ratings a simple correlation test as in experiment 1, 2 and 3 was employed. However, no correlation was found either for the description phase (correlation coefficient = -0.1615) or the retelling phase (correlation coefficient = -0.1393).

Also as in experiment 1, 2 and 3 a similar test was employed between correspondence in eye movements and the subjects decision whether they usually imagine things in pictures or words. However, no correlation was found either for the description phase (correlation coefficient = -0.0330) or the retelling phase (correlation coefficient = -0.0119).

7.4 Discussion

The main conclusions from these findings are that eye movements during the retelling of the description did show significant correspondence. The eye movement direction was significant both for the low correspondence coding ($p \leq .0001$) and for the high correspondence coding ($p \leq .0001$). For the high correspondence coding, 8 of the 12 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest. For the low correspondence coding, 11 of the 12 subjects had eye movements that had a 50% or larger correspondence with the spatial relations of the analyzed areas of interest. This experiment shows strong indication that the eye movements are reflected when a description is retold from a visualization of it.

If the results from experiment 3 and experiment 4 are compared, it is found that for low correspondence the mean correspondence for the high correspondence coding is 54.4% in experiment 3 and 55.2% in experiment 4. For the low correspondence coding the mean correspondence is 74.8% in experiment 3 and 74.9% in experiment 4.

An interesting observation is that subjects that mostly had high correspondences during the description also mostly had high correspondences during the retelling, and subjects that mostly had low correspondences during the description mostly had low correspondence during the retelling. Although a strange effect can be seen for subjects 11 and 12, they had no correspondence during the description but had some correspondence in the low correspondence coding basis during the retelling. A possible explanation could be found in Johnson-Lairds theory (1981) that language can be mentally stored both as linguistic representations and as a form that are akin to perception and internal images. It is possible that these subjects had stored the description as linguistic representation in the brain, i.e. not as a mental model of the scene that had been described, but when they were to retell the description this turned out to be a harder task then just listening to the description and they had to construct certain parts of the description into a mental model that they used when they recalled the description.

To test if there were any significant difference for the eye movement correspondences in experiment 3 and 4 a test of significance of differences between proportions that is mathematically equivalent to the χ^2-test under one degree of freedom was employed (table 14).

A diagram that compares the average value of subjects' eye movement correspondence during the retelling of the description and the retelling of the picture for both low correspondence and high correspondence coding is presented in diagram 15 and 16.

Table 14

Retelling	Picture	Verbal	Significance
Low correspondence coding	54,4%	55,2%	p = 1
High correspondence coding	74,8%	74,9%	p = 1

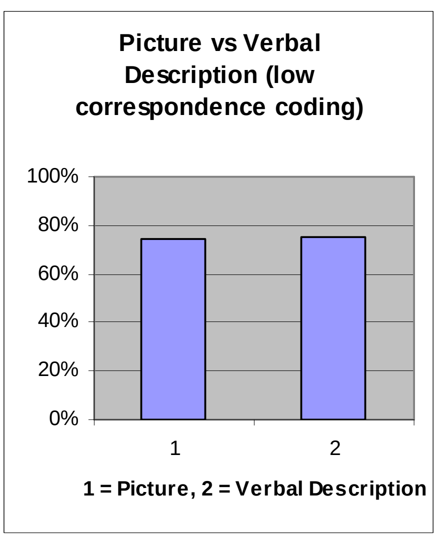


Diagram 15

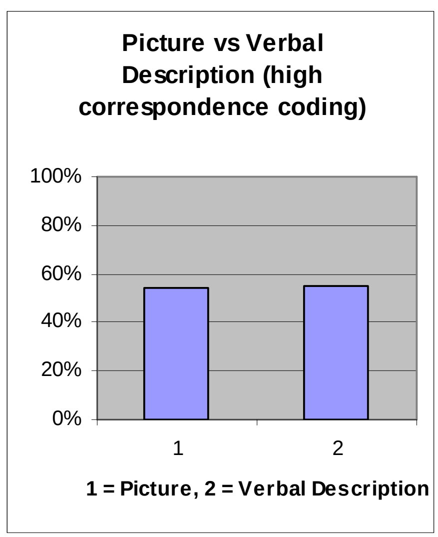


Diagram 16

These results, with the significance p = 1 both for the low correspondence coding and the high correspondence coding, show that there is no significant difference in eye movement correspondences during a visualization of a verbal description or a complex picture. For the low correspondence coding basis the correspondence is almost 75% and for the high correspondence coding basis the correspondence is over 50%.

8. General discussion

The experiments that have been presented above were designed to test the hypotheses that the spatial positions of objects in a picture are reflected by the eye movements during a visualization of that picture, and that the spatial positions of objects in a verbal description are reflected by the eye movements during a visualization of that description. An open question was if the reflection in eye movements is similar or different in these two situations.

In experiment 1 test subjects listened to a verbal description while watching a white screen and then answered questions about it (still watching the white screen). In experiment 2 test subjects viewed a picture and then answered questions about it while watching a white screen. While these experiments involved some design problems two new experiments were designed. In experiment 3 test subjects viewed a picture and then retold what they had seen while watching a white screen. In experiment 4 test subjects listened to a verbal description while watching a white screen and then retold it (still watching the white screen). The eye movements were recorded in all the experiments.

How well the eye movements corresponded with the spatial locations during imagery was considered for two coding bases. Low correspondence coding means that the eye movements are considered to move correctly if they moved in the directions left, right, up, and down when they were supposed to. High correspondence coding means that the eye movements are considered to move correctly not only if they moved like they should in the directions left, right, up, and down, they also had to be located in the correct position with each other, i.e. if there are 3 objects positioned to the left it is not enough that the eyes move to the left when the subject imagines one of these objects, the positions of the eye movements must be correct with each other for all of these objects. Mean values of how well the eye movements corresponded, i.e. either they did or they did not, in each experiment are presented in table 15.

The direction significance was also tested for the eye movements during imagery in all the experiments. These results are presented in table 16.

Table 15

	Exp1 - Description	**Exp1 - Questions**	**Exp2 - Questions**	**Exp3 - Retelling**	**Exp4 – Description**	**Exp4 - Retelling**
Low correspondence coding	59,0%	81,7%	40,5%	74,8%	64,3%	74,9%
High correspondence coding	41,5%	36,6%	0%	54,4%	54,8%	55,2%

Table 16

	Exp1 - Description	**Exp1 - Questions**	**Exp2 - Questions**	**Exp3 - Retelling**	**Exp4 – Description**	**Exp4 - Retelling**
Low correspondence coding	$p \leq .0001$	$p \leq .0001$	$p = .17$	$p \leq .0001$	$p \leq .0001$	$p \leq .0001$
High correspondence coding	$p \leq .05$	$p = .65$	$p = 1$	$p \leq .0001$	$p \leq .0001$	$p \leq .0001$

As can be seen in table 16 the eye movements directions for experiment 1 and 2 were not significant during the questions (during imagery) for the high correspondence coding either for the verbal description (experiment 1) or the picture (experiment 2), and there were no correspondences in the eye movements for the picture (experiment 1) during the questions. The reason to these bad results was probably the design of the questions. This problem was eliminated in experiments 3 and 4 where the subjects instead of questions were to retell the picture and the verbal description with own words. As can be seen in table 15 and 16 the correspondence in eye movements were much better, and the eye movement directions were significant. However, a funny thing was that for the low correspondence coding the 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 all eye movements that appeared were not related to image scanning. It has been found that eye movements also appear during verbal tasks such as general arousal, orienting reactions, and/or in cognitive change (Demerais & Cohen, pp. 230-231). 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 15 the results during the verbal description and the retelling of it are almost identical (54.8% and 55.2%). Another interesting observation from table 15 is that the results of the retelling of what could be seen in the picture (experiment 3) and the retelling of the verbal description (experiment 4) 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 can also be seen in diagram 15 and 16.

So what do these results say about my hypotheses and the open question about if the reflections in eye movements are similar or different in these two situations?

While the results for the low correspondence coding can be criticized (because of the possible eye movements that were not related to image scanning) the results for the high correspondence coding are very solid. The mean results for the subjects in experiment 4 with high correspondence coding show that the eye movements correspond with the spatial locations during imagery to a degree over 50%. This is very strong evidence that the spatial positions of a verbal description are reflected by the eye movements during a visualization of that description.

While the results for the low correspondence coding can be criticized also for picture imagery (the effects with eye movements that are unrelated to image scanning might also appear in the retelling of a picture) the results for the high correspondence coding again are very solid. The mean results for the subjects in experiment 3 with high correspondence coding show that the eye movements correspondence with the spatial locations during imagery to a degree over

50%. This is very strong evidence that the spatial positions of objects in a picture are reflected by the eye movements during a visualization of that picture.

Diagram 15 and 16 show how well the eye movements corresponded during the retelling of the verbal description (experiment 4) and the picture (experiment 3). The results are almost identical. In table 14 it can also be seen that there was no significant difference for eye movement correspondences during the visualization of the verbal description or the complex picture. This is very strong evidence that the reflection in eye movements is the same during the visualization of a picture and during a verbal description. These results are consistent with Johnson-Laird's (1981) theory that language and discourse sometimes are represented in a form akin to that of perception and internal images.

In all the experiments the subjects were also told to rate the vividness of their visualizations during imagery and to make an assumption whether they normally think in pictures or words. However, it was found that there were no correlation between the subjects' ratings and their eye movement results, neither for visualization ratings nor for the picture/word assumption. These results strongly suggest that people in general are not aware of *how* they are thinking, i.e. these results show that when studying mental images, e.g. if at all we have any, introspection is not a plausible way to find out (as in all kind of research).[10]

So what do these results say in a wider perspective, e.g. how do they correspond with previous research? And how do they stand in relation to the different theories about eye movements and imagery?

Compared to the studies by Brandt and Stark (1997), and by Laeng and Teodorescu (2001) it can be concluded that it is much harder to study the exact scanpaths that appear during the mental reconstruction of a picture of the complexity that was used in experiment 2 and 3. However, even if the scanpaths are harder to study for complex pictures, the results that spatial locations of the picture's objects to a high degree were preserved during the visualization are consistent with the studies by Brandt and Stark, and by Laeng and Teodorescu.

Compared to the studies by Demerais and Cohen (1998), Spivey, Tyler, Richardson, and Young (2000), and by Spivey and Geng (2001) it can be concluded that their findings of eye movements moving to the directions left, right, up, or down when a visualization of a scene that is generated by a description that moves in one of this directions also appeared in experiment 1 and especially in experiment 4. These results are even extended by showing that the subjects' eye movements to a high degree build up an entire scene with the correct spatial locations for each of the objects and for their relations to each other. Another extension is that the subjects also retold the description and it was found that the spatial relations to a high degree were preserved in the eye movements.

But although it has been shown that the eye movements reflect the spatial relations from the mental visualization of both a complex picture and a complex verbal description, does this mean that the eyes play a functional role in the image generation process? Or are there other explanations?

[10] These suggestions are, however, based on the assumption that we *have* mental images and that they are reflected in our eye movements during imagery, i.e. a person who is good at visualizing pictures should have eye movements that correspond with the spatial relations from the mental image.

From a quasi-pictorial theory point of view the 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. Although this is a healthy interpretation of my results it is not totally unproblematic. It gets difficult to explain why there were no correspondences in the eye movements either in the low correspondence coding or the high correspondence coding for some of the subjects. One explanation could be that during imagery we tend to have low amplitude "fixation" movements (microsaccades) (Demerais & Cohen, 1996), or like Brandt and Stark (1997) state that in imagery there is no need for eye movements with the full range of amplitudes as in perception. One could then argue that the amplitude of the saccades for these subjects could be so small that they were impossible to analyze in this way. It is also possible that the effect was absent because these subjects shrink the picture so much that they could "scan" most of their mental image covertly, i.e. that their inner attention was shifted without eye movements.

If instead the results are interpreted from a description theory point of view the explanation to the results would possibly be that the correspondence occurred because of tacit knowledge, i.e. that the subjects were simulating their past perceptual behavior. Although this critique could be used towards the eye movements generated by the picture it is more difficult to apply it to the eye movements that were generated by the verbal description, i.e. what tacit knowledge makes the eyes move to the same spatial locations as are told in a description? I do not find it very likely that we mimic a perceptual behavior when listening to a verbal description or when we are retelling this description. If this was the case we would mimic a perceptual process that never took place. I assert that tacit knowledge can not be used towards my results from the verbal description experiment (experiment 4), and while the results are almost identical for the eye movements generated by the picture (diagram 15 and 16) I assert that tacit knowledge can not be the case for eye movements generated by the picture either. With respect to the results for the high correspondence coding it does also seem unlikely that we are able to mimic a behavior in eye movements that is so good that entire scenes of objects with correct spatial locations are built up.

Another description theory explanation could be task-induced demand characteristics, e.g. that the environment of the experiment, with the head mounted eye tracker and the calibration routine, made the participants realize what were expected of them, and therefore intuitively made the expected eye movements. But this explanation is also rather weak, because all the participants were asked after each experiment what they thought the meaning of the experiment was. Almost none of the participants thought that their eye movements had been recorded. However, there were some subjects that discussed the possibility that it was studied *what* they observed in the picture, and some subjects that pretty vaguely discussed a comparison of mental images and perception, but it was not considered that they had seen through the hypotheses of the experiments. There was also one participant in experiment 2 that was spot on the hypothesis. Although this experiment suffered from design problems an interesting observation is that this subject did not show better results (rather the opposite) than the other subjects.

It is also possible, from a description theory point of view, to argue that the eye movements only are a by-product. But this argument can not be used towards my results that have direction significance both for the eye movements generated by the verbal description and by the picture (table 16).

If the third theory, the perceptual activity (PA) theory, is used to interpret the results in its proceduralist approach we have another healthy interpretation. The PA theory would say that the results of the experiments supports the theory that eye movements during mental imagery happen because of the procedures that take control of the exploratory apparatus during both the experiences of mental images generated by pictures and those generated by verbal descriptions. The PA theory is also able to explain why reflections in eye movements are absent for some of the subjects. This could be because the procedures that are activated have individual differences, which for example is the reason why the reflections for some subjects were totally absent and why for others they shrunk or re-centered the experienced image. This explanation gets even more plausible when the results of a subject during the verbal description are compared to the results during the retelling of it, i.e. the same individual differences that happened during the description also happened during the retelling.

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 reflections were also as good for eye movements generated by the complex picture as for eye movements generated by the complex verbal description, which suggests that discourse can be represented in a form similar to that of perception and internal images.

I have argued that the description theory do not have good arguments towards the results of my experiments. The results can not be explained in terms of tacit knowledge, task-induced demand, or that the eye movements only are a by-product. Therefore the results do suggest that we indeed have mental images (if interpreted quasi-pictorial), or that the eye movements that occur during imagery at least are involved in the procedures that make us experience mental images (if interpreted in line with the PA theory). But it is hard from this study to say whether the quasi-pictorial theory or the PA theory is the most plausible explanation for the results. But either way, the results do suggest that “lower-level” motor processes (e.g. the eyes) are not separated from “higher-level” cognitive processes (e.g. imagery). Instead they are connected and perhaps dependent of each other when mental operations are executed.

9. References

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.

Duchowski, A. T. (2003). Eye Tracking Methodology: Theory and Practice, London, Springer-Verlag.

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

Glenstrup, A. J., & Engell-Nielsen, T. (1995). Eye Controlled Media: Present and Future State, URL : http://www.diku.dk/~panic/eyegaze

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

Henderson, J. M., & Hollingworth, A. (2003). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics, 65, 58-71*.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 113-136.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review, 8,* 761-768.

Holsanova, J. (1999). Speaking of pictures: Focusing attention and structuring spoken descriptive discourse in monological and dialogical settings. Licentiate dissertation. *Lund University Cognitive Science 78*.

Holsanova, J. (2001). Picture Viewing and Picture Description: Two Windows on the Mind. Doctoral dissertation. *Lund University Cognitive Studies 83*.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye Movements and scene perception: Memory for observed things, *Perception & Psychophysics*, 2002, 64 (6), 882-895.

Ishai, A., & Sagi, D. (1995). Common Mechanisms of Visual Imagery and Perception, *Science,* New Series, Volume 268, Issue 5218, 1772-1774.

Johnson-Laird, P. N. (1983). Comprehension as the Construction of Mental Models, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences,* Vol. 295, No. 1077, 353-374.

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.

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

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

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

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

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

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. (2000). Situating vision in the world, *TRENDS in Cognitive Science,* 4 (5), 197–207.

Pylyshyn, Z. W. (2003). Return of the mental image: are there really pictures in the brain?, *TRENDS in Cognitive Science, Vol. 7 No.3.*

Rensink, R. A. (2002). Change Detection. *Annual Review of Psychology*, 53:245-277.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes, *Psychological Science,* 8 (5): 368-373.

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

Spivey, M., Tyler, M., Richardson, D., & Young, E. (2000). Eye movements during comprehension of spoken scene descriptions. *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, 487-492.

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).*

Wolfe, J. M. (1998). Visual Search, *Attention,* H. Pashler (Ed.), London, UK: University College London Press.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.