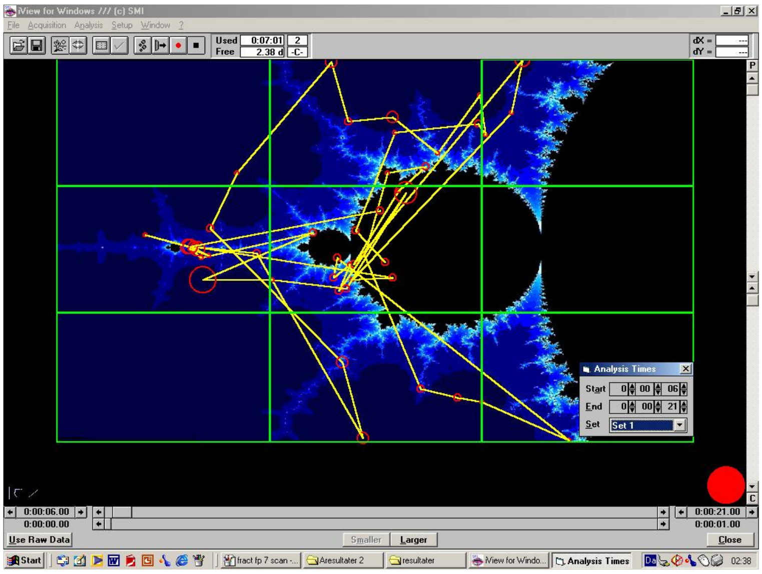
Imagery, scanpaths reenactments and their functionality.
The testing of a hypothesis.



Master's Thesis, Cognitive Science, VT 2004.
Halfdan Fryd Koot
Lund University Cognitive Science
Supervisor
Kenneth Holmqvist

Abstract

This paper centers on visual imagery and the hypothesis put forward by Laeng & Teodorescu regarding imagery or visualization. The hypothesis states that scanpath done during perception of an object or scene is reenacted during imagery of the same object or scene. In addition it claims that the reenactment has a positive functional role in memory retrieval. This paper cannot confirm nor refute the hypothesis put forward regarding scanpath reenactments, due to possible biased data; but on the basis of experimental evidence has to refute the hypothesis regarding scanpaths reenactments functional role in memory retrieval concerning long term memory. These conclusion are then discussed in perspective to the present dominant imagery theories and it empirical implications in relation to other empirical evidence.

The experiment conducted consisted of three phases. First participant viewed four different pictures while eye-movements were measured. Secondly they were instructed to visualize the stimuli and subsequently describe it verbally; eye-movements were measured both during visualization and description. The scanpath generated were then analyzed using to different analysis methods. First an analysis were based on the pictorial material generated using iVeiw 3.01 and subsequently and analysis based on fixations percentage were conducted.

List of Contents:

1. Introduction

This research project centers on a particular elusive subject, mental imagery. Mental imagery, as a research subject, has spurred a continuous stream of publications and articles. One of the main reasons for this is the nature of imagery itself. First of all imagery is a subjective phenomenon. This makes the scientific exploration of imagery a very challenging endeavor. Physical objects, unlike mental images, are directly observable; hence, their properties and functions are more or less directly observable, whereas studies of mental imagery always have to be inferred. Therefore experimental studies of imagery are often open to more than one interpretation or explanation.
Still significant result has been achieved, and the debate is not whether imagery exist or not, but rather of what nature imagery is. Is imagery in its nature quasi-pictorial as proposed by Kosslyn, and therefore in many ways equivalent to visual perception; or is it more like a semantic network propositional in nature and therefore not dependent on the actual sensory modality. This is the central issue of the debate.

Different methods has been deployed and it seems clear, that no method in it itself can yield conclusive evidence on the nature of mental imagery. Consequently the most prudent approach, the approach used by researches today, is to use as many methods as possible to get a result which can lead to new conclusions. These methods are the clinical study of brain damages, questionnaires, experimental research, PET-, fMRI – scans, ERP, EEG and other related methods.
In the recent years another method has seen the light of day in relation to the current imagery debate, even though this method has been around for over thirty years. This method is eye-tracking. It seems paradoxical, because theories regarding eyemovements during perception and subsequent imagery have been present for quite some time. One of the first to propose a correspondence was Hebb in 1968,
"Hebb proposed that eyemovements have an essential organizing function in visual imagery suggesting that if the image is a reinstatement of the perceptual process it should induce eye movement" [1].
The early empirical studies of eyemovements were based on photo techniques. One of the pioneers, Totten, suggested that eye movements correspond to the shape of the object the subjects were asked to visualize[2].
In addition studies conducted by Brown compared ocular activity of vivid and poor visualizers during imagined pursuit and found no consistent differences between groups. There was nothing that suggested that quality or vividness of an imagined picture was correlated to an increase in scanning activity[3].

1.1 The experimental evidence

Brandt and Stark conducted their experiment on perception and subsequent imagery in 1997, by using different checkerboards patterns, measuring eye movement patterns during perception and then subsequently having people visualizing checkerboard patterns while measuring eyemovements.

[1] (Brandt & Stark 1997 p. 27).
[2] (Brandt & Stark 1997)
[3] (Brandt & Stark 1997)

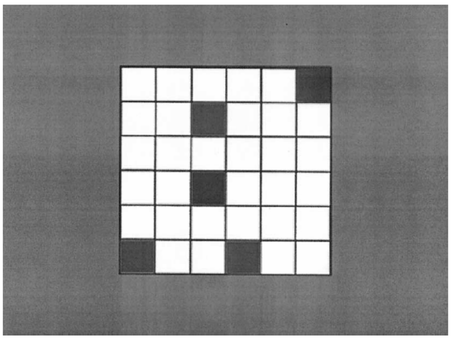


Figure 1: Checkerboard taken from Laeng & Teodorescu 2002

Their experiment showed that there is a correspondence between the eye-movement pattern during perception and imagery. They claim the eye-movements generated during imagery might have a constitutive function. But they did not claim the evidence put forward was conclusive, instead they claimed there was a strong tendency in favor of this interpretation. Furthermore, they could see a tendency that participants made significant fewer fixations during imagery than during perception.

A similar line of experiments conducted by Laeng & Teodorescu in 2001[4] seems to confirm that eyemovements have an essential organizing or constitutive function in relation to the spatial elements of visual imagery.

In their experiment (actually two experiments), the participants viewed a 6 × 6 grid pattern, which contained five black cells. When initially committing the display to memory, one group was instructed to keep their gaze on a central point on the grid; another group was allowed to examine the pattern freely for the same amount of time. Eye movements were recorded while participants studied the displays and then later, when they visualized the memorized displays.

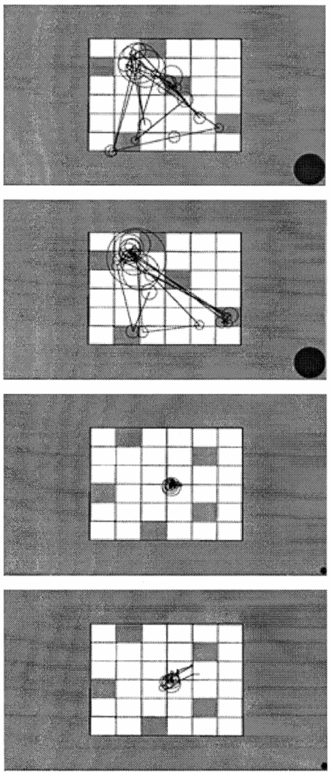


Figure 2: Taken from Laeng & Teodorescu 2002

The eye movements during imagery reflected the conditions in which the participants had studied the stimuli. The participants who maintained fixation while studying the display made almost no eye movements when subsequently visualizing it. Contrarily participants who visually were allowed to explore the pattern while studying it, also made eye movements during imagery. In addition, the sequence of fixations during imagery was similar to the sequence that occurred when the participants studied the pattern during perception i.e. they reenacted. This experiment was directly inspired by Brandt & Starks experiment[5], and had the function of confirming their result. It did deviate in the experimental design. In Brandt & Starks experiment subjects were exposed to the stimuli for twenty seconds, followed by ten seconds imagery and then ten seconds viewing. Because of the small time lapse, the subjects could have stored the stimuli in short term memory[6]. Laeng & Teodorescu[7] therefore ensured that enough time (40 sec.) passed by from stimuli exposure to later visualization, so the stimuli couldn't be stored in short term memory, only in long term memory and to prevent afterimages. Whether or not this amount of time is enough to ensure that the stimuli is only

[4] (Laeng & Teodorescu 2002)
[5] (Brandt & Stark 1997)
[6] (Baddeley 1997; Brandt & Stark 1997; Laeng & Teodorescu 2002)
[7] (Laeng & Teodorescu 2002)

stored in long term memory is questionable, if no intervening noise was put in between the perception and the later visualization. A study by Ishai & Sagi seems to show that some kind of iconic memory exists, in which a memory trace can exist for at least five minutes[8]. In the second experiment, Laeng and Teodorescu altered the display so a fixation cross appeared at the centre of the screen, and the stimuli appeared in one of the four corners of the screen. Additionally they used pictures of natural objects (pictures of fishes) instead of checkerboards. This second experiment included an important clause: free viewing in perception, but fixed gaze during imagery. Based on the previous findings, Laeng & Teodorescu hypothesized that this condition should stop participants using memories of eyemovements to help reconstruct the image. And in fact, the participants were less accurate when asked about visual details while visualizing the previously studied picture. For example, one question required the participants to determine how many stripes were on a fish in the picture they had studied. They as a consequence argued that for the retrieval task in the free-viewing perception but fixed imagery condition, the ocular-motor links established during perception could not be used in the service of building up a mental image, and this limitation impaired recall. Laeng and Teodorescu thus argued against the view that eye movements during imagery are an epiphenomenal by-product of shifting one's attention across an internally generated image. They argue that if this was the case, the subjects should not be impaired when they visualize whilst maintaining fixation. One might argue that this is due to the fact that one has to maintain fixation during visualization, and thereby one's performance is limited, because the subjects' attention is divided[9]. What is very important to notice is that Laeng & Teodorescu do not claim that from scanpath and the re-enactment of those can be said anything about imagery regarding color or anything related to the "what" of visual perception. Their experiment and their conclusions extend only as far as the spatial components of visual perception and imagery.

1.2 The assignment objective

This assignment has four objectives, one is to confirm or refute the hypothesis made by Brandt & Stark and later by Laeng & Teodorescu. This hypothesis claims that eye-movement or more accurately the scanpaths during the perception of one object and the subsequent visualization of the same object corresponds. Thus the null-hypothesis would be to find positive evidence that eye-movement during these two processes in fact do not correspond. Furthermore this study aims at pushing this hypothesis even further, by using pictures of very complex scenes as stimuli, to produce these eyemovements. Thus furthering the ecological validity of this hypothesis. In addition the experiment conducted in conjunction with this study one hopes to shed light on the imagery debate and hopefully contribute to the current debate. Last of all this study aims at discussing from meta-theoretically point of view. Can eye-tracking studies contribute significantly to the imagery debate as a method?

This assignment therefore has four aims:

1. To confirm or refute the hypothesis regarding correspondence of eyemovements during perception and subsequent imagery, and thus contribute to the imagery debate.
2. Secondly confirm or refute the hypothesis regarding the epiphenomenal or functional nature eyemovements in relation to memory retrieval regarding spatial components of a picture, scene or object.

[8] (Ishai & Sagi 1995)
[9] (Laeng & Teodorescu 2002)

3. Explore whether or not reenactments occur during verbal description of a visual scene
4. Can eye-tracking as a method contribute to the current imagery debate

Before going into a discussion regarding ocular motor movements and whether or not these re-enact those of perception during subsequent imagery. It is important to give a short review regarding vision, eyemovements, visual attention, visual representations and imagery theories. The review on vision, eye-movements, visual attention and visual representation is to give the reader an understanding on what happens during perception of a scene. The imagery theories section is to give an understanding of what imagery is and what imagery entails within the different conceptions.
Furthermore a definition of imagery will be given; I have chosen to stick to Finke's definition and it is this definition that will be used.
"Mental imagery" is defined as
"*the mental invention or recreation of an experience that in at least some respect resembles the experience of actually perceiving an object or an event, either in conjunction with, or in absence of, direct sensory stimulation*"[10].

2. Vision

The retina of the human eye consists of ca. 127 millions light sensitive cells. About 120 mill Cells are the so-called rods, which are highly sensitive i.e. they can detect relatively small amount of light and movement, the other cells cones number about 7 million. These cells have less light sensitivity, but can detect colors and have a higher acuity. The cones contrary to the rods are not distributed evenly on the retina, but are concentrated in the very centre of the retina in a depression called fovea. The cones are specialized for detailed vision and have a high acuity, which extends over visual angle approximately 2 degrees. The rest of the retina offers peripheral vision in which motion is easily detected, but color and details are down to 50 % lesser acuity the further removed from fovea.
After the eyes have received a signal the successive processing is done in the brain, both in serial and in parallel. Serially by sending the signal from the retina to thalamus to the visual cortex and in parallel by processing several different type of information at the same time, this happens early in the process. In thalamus there are layers of cells, which begin the processing of different type of information, color, form etc.
After reaching the primary visual cortex, the signal is processed by the secondary visual cortex. The secondary visual cortex is in fact an area covering parts of the occipital and parietal lobes[11]. The signal is processed in the ventral and dorsal streams of the brain. The ventral stream process information regarding object recognition, the dorsal stream processes the spatial information of where a given object is. These two streams shall not been seen as two separate systems, since there is interaction at various levels, but there is a certain amount of specialization. After that the signal is assembled in V5 or MT. In addition in the primary visual cortex, there are different neurons specialized in detecting different features of a visual scene like edges, corners, color called simple, complex and hyper complex cells[12].

[10] (Finke 1989 p. 2)

[11] (Duchowski, 2002; Glenstrup & Engnell-Nielsen 1995; Levine 2000)
[12] (Nyberg 2002)

2.1 Eyemovements

Eyemovements have the function of letting the foveal regions of the eyes focus or refocus. Six extra ocular muscles steers the eye and allow the human visual system to function as 3 D gaze control system. The muscles are arranged as 3 antagonistic pairs and enable different type of motions. Two are of particular interest here, saccades and fixations. Saccades function as the principal method of moving the eyes between fixations from one locality to another. It takes about 100 to 300 milliseconds to initiate a saccade and 30 to 120 milliseconds to complete. Saccades can be induced voluntarily or reflectively, but once initiated follows a ballistic pattern of motion. Once set in motion the actual movements cannot be stopped midway, but has to be completed. The speed in which the actual movement is executed renders the viewer effectively blind during its execution.
Fixations actually consist of very small saccades (micro saccades) and drift motions that are involuntarily. These eye-movements function to stabilize the retina over a stationary object of interest, thereby allowing foveal vision to detect minute detail.
The eye-movements are controlled from several different regions in the brain called the ocular movement plant. The signals emanate from occipital cortex (area 17, 18, 19 and 22), the superior colliculus and semicircular canals. In addition eye-movements are connected to the vestibular system and the human primary sensory cortex[13].

2.1 Visual attention

The perception of scenes involves a pattern of fixations, where the eye moves in saccadic movement and then pauses to fixate, and then moves in a saccadic pattern again. Normally this is not a conscious process, mainly because the eyemovements feel like an integrated part of seeing.
But how do we chose were to look, and what to look at, how do we select. This is a matter of attention, and attention can roughly be divided into two different phases; a pre-attentive phase and an attentive phase, two different processes bottom-up and top-down and two different modes of attention overt or covert.
The two processes should be understood as a passive and an active process. Bottom-up is a passive process, which is data-driven. In connection to perception it is guided by the features of the scene and there is no strategic control on the part of the observer. One of the best examples to illustrate this is the pop-out effect. In a scene where one object deviates in color or form from an otherwise uniform background, the object will be detected or selected without conscious control, automatically attracting the beholders attention.
Top-down processes are active processes guided by the observer's expectations or goals; one can say that this process is driven by the need for information or the fulfillment of a certain task. But with this more or less conscious control comes limited capacity, a limitation that entails, that ones attention can only process a limited number of objects at one time.
The pre-attentive stage starts when one initial starts perceiving a scene. In the pre-attentive phase the scene is scanned, but with no limit to capacity, and the scan is working in a parallel pattern over the entire visual field. This process is spatially unlimited, because different locations are simultaneously processed and the basic features of each location in the scene are subjected to local mismatch detection. Thereby the observer gets a notion of where things are, but not what those things are, only basic features like color, size, orientation and presence and direction of motion[14]. The shift form pre-attentive to phase attentive is not entirely clear. One model suggests that features within a limited locality are

[13] (Duchowski 2003; Glenstrup & Engnell-Nielsen 1995; Levine 2000)
[14] (Duchowski 2002; Glenstrup & Engnell-Nielsen 1995)

added together and a difference sum is computed. The object or objects with the highest difference sum are selected first for further inspection. Therefore this is best understood as shifts from a processing in parallel to serial, which have only a limited spatial capacity. This shift in processing has to do with the complexity of the visual scene; a heterogeneous scene consisting of lots of conjoining features (a high difference sum) disables parallel processing. But still if the processing is serial, it is still at this point data-driven or bottom-up steered so the shift from the pre-attentive to the attentive does not necessarily entail a shift from bottom-up to top-down.
Other interpretations suggest that attention functions like glue in binding the different features together creating a detailed percept, the attentive phase and pre-attentive phase are not sharply defined in the theory proposed by Triesman, because the process can short-circuited by the pop-out effect.
Different metaphors have been used to get a better understanding of how visual selective attention might work. One of those is the spotlight metaphor an attention mechanism able to move about the scene in a manner similar to a "spotlight". The spotlight is limited in its spatial extent, but within the “beam” there is the possibility for a higher resolution of details. The spotlight is not the same as foveal vision, but an independent mechanism tied to the human attention system and therefore independent of fixations. Consequently it is possible to direct ones attention covertly towards objects in peripheral vision and then by saccadic movements subsequently fixate the object of interest. Hence it is important to stress the fixation does not necessarily imply attention. It is possible to fixate one locality and having your attention directed at another.
Another is the zoom-lens metaphor, which tries to capture that attention is scalable. One can either have a wide angle of attention, but with a poor resolution in detail or one can zoom in improving resolution, but thereby lose the bigger picture. Kosslyn has suggested a window of attention which is also scalable, but this window is also tied to his theory on the visual buffer, because the scanning procedure in relation to the visual buffer is applicable both to external and internal “stimuli”.
All these theories are based on a bottom-up approach to the human visual system
That is, vision might behave in a cyclical process composed of the following steps:

1. Given a stimulus, such as an image, the entire scene is first "seen" mostly in parallel through peripheral vision and thus mostly at low resolution. At this stage, interesting features may "pop out" in the field of view, in a sense engaging or directing attention to their location for further detailed inspection:

2. Attention is thus turned off or disengaged from the foveal location and the eyes are quickly repositioned to the first region which attracted attention.

3. Once the eyes complete their movement, the fovea is now directed at the region of interest, and attention is now engaged to perceive the feature under inspection at high resolution.

This is a bottom-up model or concept of visual attention[15].

These processes mentioned above is dubbed low-level cognitive process in that regard that it is mostly involuntary and automatic, though most researches acknowledges that this

[15] (Duchowski 2003, p. 12)

division in somewhat artificial and the processes are intertwined, because attention consist of both voluntary (high-level) and involuntary (low-level) procedures. The research on attention steered by intentions, tasks and goals is a bit more unsubstantiated. Studies made on change- and inattentional blindness is related to this subject and as studies has shown, the human capacity to detect changes or detect unanticipated event is hampered. In addition the experiments conducted in conjunction with the anticipatory hypothesis has shown that anticipating a certain event can facilitate detection of that event, it has also shown that anticipation of a certain event can hamper the detection skill if that exact event does not occur[16].

2.2 Visual representation

The nature of the representations formed during visual perception and what and how they are retained, is still not certain. The traditional view was that one had a rich internal visual representation formed by consecutive fixations over the visual field, have largely been abandoned; also called the composite sensory image hypothesis.

Change blindness[17] studies have shown that this view is unlikely because if composite representations were the case changes should be easily detectable. This has lead to claims that no internal representation exists and therefore no representation is retained across saccades. O'Regan has stated that the world functions as an external memory, which is continuously confirmed[18]. This explains why changes are not easily detectable.

The theory advocated by O'Regan and similar theories is called localist-minimalist theories[19]. These states 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 between saccades[20] and only the gist is somewhat retained.

As studies have shown the scene is not completely wiped out, simple features seem to be retained across saccades in an iconic or short term memory store. Hence previewing an object before a saccade is initiated allows for faster identification of object after completion of the saccade.

Therefore most theories agree that a complete internal pictorial representation is not the case, but some theories claim that features are retained and integrated across saccades, stored in short or long term visual memory. These representations are visual, but abstract in the sense that the representation is a structural description.

3. Imagery Theories

There are two major research programs within imagery, not to say that these are the only ones, but these two theories are mentioned continuously within imagery literature.

The two theories are the quasi-pictorial theory and propositional theory. The focal point in the debate is; in which format is the representation in memory is it pictorial in nature or is it more like a semantic network?

Stephen M. Kosslyn has developed the quasi-pictorial theory for over thirty years and is one of its major proponents. The propositional theory has been strongly advocated by Zenon Pylyshyn for the last thirty years. The two theories have competed for the last thirty years and pushed the scientific development to where it is today.

[16] (Finke 1989; Henderson & Hollingworth 2000: Rensink 2002; Simons & Chabris 1999)
[17] (Rensink 2002)
[18] (O'Regan 1992)
[19] (Henderson & Hollingworth 2000)
[20] (Henderson & Hollingworth 2000)

3.1 The quasi-pictorial theory

Most people doing imagery research today acknowledge that the picture metaphor should not be taken too literally. The question of current interest is not whether images function in exactly the same way as pictures, but whether images share *any* properties in common with pictures or if visual imagery shares *any* properties with visual perception.

The quasi – pictorial theory proposes that visual imagery has properties that in some ways are equivalent to pictures as a representational format.

Kosslyn has argued that an image consists of two components. The first component is a "surface" representation, which is a quasi-pictorial entity held within some form of active memory or short term memory. The second component is a "deep" representation, where the information is stored within long-term memory from which the surface representation was derived[21]. The surface representation can be seen as analogous to the subjective experience of imagery, whereas the deep representation is more like a prototype stored within long term memory. In terms of the deep representation, it has similarities with the propositional theory, because in term of format, the deep representation is more a set of abstracted properties, and as such has no pictorial qualities.

Surface representations are constructed on the base of deep representations, the deep representations are retrieved and the surface representation is then constructed in functional unit called the visual buffer. When the construction is complete it is available as an image to consciousness and ready to be "seen" or rather inspected.

The visual buffer is functionally defined neural locus, which Kosslyn and others have advocated is localized in the occipital lobe, to be more precise in primary visual cortex or area 17. The claim is disputed and other researches claim that no significant activation of the primary visual cortex occurs during imagery[22]. This is not crucially important from a theoretically point of view regarding the visual buffer as a component in a procedural model. But in relation to whether visual imagery is analogous to visual perception and therefore uses the same neural substrate, is a different matter.

Kosslyn and others has spurred much empirical evidence in favor of their interpretation. Finke has formulated, on the basis of empirical evidence, five different principles, the first is: The implicit coding principle:

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[23].

This is based on studies showing that visual characteristics are used when we compare object regarding to size, and that we can use visualizations to facilitate memory retrieval about objects by visualizing them.

They found that in relation to visual imagery, there were many correlations to actual perception. In visual perception we have the visual field, which is the field, were we can see objects clearly when they fall within a certain region of space, depending on were our eyes are fixated. Experiments conducted by numerous researchers showed that in imagery we have something similar, images can overflow. Overflowing imply that the different parts of the object cannot be held in sharp focus, this suggest that something analogous to a visual field does exist in mental imagery, because we can only focus on a certain amount of space in mental imagery. Therefore it is only possible to distribute you attention to a certain extent[24]. Additionally Finke & Kosslyn found evidence that that the imagery field is

[21] (Finke 1989; Kosslyn 1980; 1994)
[22] (Cabeza & Nyberg 2000; Finke 1989; Thomas 1999; Thomsen *et al.* 2001)
[23] (Finke 1989 p. 7)
[24] (Finke 1989; Kosslyn 1980)

analogous to perceptual field in shape, both field are elongated horizontally[25]. These studies have made Finke formulate his principle of perceptual equivalences:
Imagery is 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. [26]
Experiments regarding scanning times have shown that when people are asked to scan from one point in an image to another by imagining a tiny speck crossing the map; the time it takes to get from one point in the map is correlated to the actual distance. Although these results seems to be dependable on, that the participant are asked or instructed to imagine crossing the actual distance. If the subjects were asked to go from one point to another on a map, without scanning, the time and distance were not correlated[27]. Still the evidence presented by Finke appears to support that visual images retain spatial characteristics like in normal perception, hence; the principle of spatial equivalence:
The spatial arrangement of the elements of a mental image corresponds to the way objects or their parts are arranged on actual physical surfaces or in an actual physical space[28].
Most studies done, has been associated with static objects. A famous experiment done by Shepard and Metzler showed, that subjects can mentally rotate objects to answer different questions. In their experiment they used 3D geometric objects formed of small cubes. The subject's task was to verify whether or not to figures were the same by mentally rotating one of the figures. The rotations were done in plane and in depth, and the time in took to verify was correlated to the angular difference between the forms. This is one of many studies which have laid the ground for Finke fourth principle: The principle of transformational equivalence:
Imagined transformations and physical transformations exhibit corresponding dynamic characteristics and are governed by the same laws of motion[29].
This holds in most cases although studies have shown that in some cases the transformation is not holistic, but rather the object is transformed in a fragmented way[30].
His last principle the principle of structural equivalence, states that:
The structures 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[31].
This principle implies that images have a structural equivalency to physical objects. Hence imagery can be used to fuse different object while retaining a structural coherence similar to an actual object and thereby allow the emergence of novel structures. This is related to the anticipatory hypothesis were anticipation of a certain sequence of events can help facilitate detection of those, but this has a counterproductive consequence, that in the anticipating a certain outcome one inhibits ones own ability to detect differing outcomes.

In general when speaking about the quasi-pictorial theory, one can speak of a strong and a weak interpretation. The strong interpretation states that the neural components used in visual imagery are the same as those used in visual perception. The weaker interpretation holds that there are equivalencies between perception and imagery, but not down to the point where the neural components used for both processes are one and the same. Finke for

[25] (Finke 1989; Kosslyn 1980)
[26] (Finke 1989 p. 41)
[27] (Finke 1989)
[28] (Finke 1989 p. 61)
[29] (Finke 1989, p. 93)
[30] (Baddeley 1997; Finke 1989)
[31] (Finke 1989, p. 120)

instance acknowledges that perception and imagery in many ways are equivalent, but contrary to Kosslyn, Finke does not support the idea that these similarities start in the primary visual cortex. These equivalencies are of a higher order and therefore these similarities start in the subsequent processing of visual information, rather than in the primary visual cortex itself[32].
In relation to eye-tracking studies, Kosslyn has taken Laeng & Teodorescu's study as another possible confirmation of the strong interpretation of the quasi-pictorial theory, see Mast & Kosslyn[33].

3.2 The propositional or description theory
This theory, unlike the quasi – pictorial theory claims that the format in which mental representation are stored, has no bearing to which sensory organ the mental representation is related to. That for example visual representations has no unique visual format in which these are stored. All representations are stored in the same format, whether this is a perceptual experience, memory retrieval or mental imagery. The content of a given representation might be perceptual, but there is nothing perceptual about their format i.e. there are not any equivalency between the format and the perceptual experience.
The format should not be understood as a language that describes a given mental representation, but more like a system that holds the abstracted properties of a given object. Therefore one can say that the propositional theory of imagery is related to the functionalist theory of the mind. And that an underlying "language" called mentalese by the functionalist theorists, is used in all mental representations and functions. Furthermore these language-like representations are sufficient to account for all cognitive processes.
In relation to imagery this means that no perceptual mechanisms need to be activated when imagery is experienced.
Recently as acceptance of the quasi-pictorial theory has gradually increased, Pylyshyn has referred to the propositional theory as the natural null-hypothesis in relation to the quasi-pictorial theory. Additionally using a situated cognition approach[34], augmenting that the external world can be used as an external memory; therefore it is unlikely and unnecessary to have an internal "pictorial" representation[35]. These conclusions are related to the study of change-blindness and inattentional blindness[36] which seem to suggest that the memory of a visual scene is not extensive, and only a small part of the scene is retained across saccades. This naturally raises the question; if we have an internal representation of the external world, how come our ability to detect changes seems limited or impaired? In addition Pylyshyn[37] has focused on tacit knowledge as an explanation for the result obtained by picture theorists.
"One of the problems with this research is that nearly all experimental findings cited in support of the picture theory can be more naturally explained by the hypothesis that when asked to imagine something, people ask themselves what it would be like to see it, and they then simulate as many aspects of this staged event as they can and as seem relevant… - it appeals only to the tacit knowledge that people have about how things tend to happen in the world, together with certain basic psychophysical skills"[38]

[32] (Finke 1989)
[33] (Mast & Kosslyn 2002)
[34] (Kosslyn *et al.* 2003)
[35] (O'Regan 1992)
[36] (Henderson & Hollingworth 2000: Rensink 2002; Simons & Chabris 1999)
[37] (Pylyshyn 2000; 2001; 2003a; O'Regan 1992; Rensink 2002)
[38] (Pylyshyn 2003, p. 113)

In relation to eye-movement patterns the question is whether or not people on this psychophysical level are able to control these movements or foresee how their eyes move when seeing a given stimuli. As mentioned by Finke, certain evidence contradicts the quasi - pictorial theory, on issues like mental scanning the evidence brought forward by Lea[39]. Intons-Peterson & Roskos-Ewoldsen has shown that subjects took longer to imagine traversing a familiar route, when they were told they were carrying a cannonball than a balloon, which would involve tacit knowledge. Other studies have shown that mental rotation is not unproblematic[40].

3.3 The perceptual activity theory, the runner up

A new theory is slowly emerging, but a clear cut and finished theory regarding imagery has not yet been accomplished, still it has potential to renew the debate and might become a new dominant theory. The theory was introduced by Neisser in the 1960 but at that time, it did not have enough impact to be a seriously discussed theory.

The perceptual activity theory (PA) is based on a situated approach to cognitive processing and has largely until now been developed within a robotics, in getting them to sense and perceive their surrounding environment.

Before going into its relevance for imagery, it is necessary to sketch out its view on perception in general. The active in PA should be understood in a literal sense and as exceeding the normal top-down processing, by stressing that we absorb information in an active manner by a constant adjusting to the surrounding world, and not by transdussing. The perceptual activity theory aligns itself with the processuralistic view on memory, which stresses that memory is not just storage of old sensory information etc. but a constant updating of procedures or schemata's that specifies how to direct our attention most efficiently[41].

The human sensory apparatus is not understood as five or more separated senses, but as a large array of anatomically overlapping perceptual instruments. These instruments sometimes uses the same resources both regarding receptor cells and muscles, but also the same neural structures. The important thing in relation to perceptual instruments is to notice that PA stresses that the whole body with its muscles and limbs are included in the "box" of perceptual instruments

These perceptual instruments are then used in the exploration of the environment as directed by procedures or schemata. The procedures should not been seen as a static set of instructions but as a set of structures which undergo a constant change as the continuous exploration of the surrounding environment occurs[42]. When we perceive an object questions arises, and at that point we only have a vague pre-attentive preconception of what a things is. Our procedures direct whatever perceptual instrument or instruments seems relevant to the task in determining what object, and then updates the procedures on the basis of new experience, and new questions arise again and so on.

"impulses in the optic nerve fibers at each moment of scanning a scene are the answers, in code, to the "questions" that had been asked at the previous moment. . . . [What] goes on in the retina is not like the recording of a "picture", but the detection of a series of items, which are reported to the brain (Young, 1978[43])."

---

[39] (Finke 1989)
[40] (Baddeley 1997)
[41] (Laeng & Teodorescu 2002)
[42] (Thomas 1999)
[43] (Thomas 1999 p. 218)

Therefore looking at the citation from above that there are no internal images as such. What is imagery then?

"No thing in the brain is the percept or image. Rather, perceptual experience consists in the ongoing activity of schema-guided perceptual exploration of the environment. Imagery is experienced when a schema that is not directly relevant to the exploration of the current environment is allowed at least partial control of the exploratory apparatus. We imagine, say, a cat, by going through (some of) the motions of examining something and finding that it is a cat, even though there is no cat (and perhaps nothing relevant at all) there to be examined. Imagining a cat is seeing nothing-in-particular as a cat".[44]

This implies that during imagery, the procedure or schema is active in much the same way as during the actual perception, a *perceiving as*.

Some of the questions or test directed by exploration of the environment, can according to Thomas[45], be directed at bottom-up neural representations in relation to retinotopically mapped brain regions. So in conjunction to visual imagery, the schema as a byproduct of initiating the relevant procedures for the imagery in question, activates the regions in the visual cortex, and sometimes it does not activate these areas when initiating a schema. This would help explain the contradictory evidence discussed above in the section regarding the quasi-pictorial theory.

4. The experiment.

4.1 Procedure.

The experiment was conducted in three phases: a perception phase, a visualization/description phase and a questioning/debriefing phase.

The participants were instructed to look at four different pictures; the exposure lasted for fifteen seconds for each picture. Subsequently in the second phase of the experiment they were instructed to visualize one picture and afterwards describe the pictures verbally, this was repeated four times, one time for each stimulus.

The participants were exposed to four different pictures in the perception phase, after each picture there were 3 different questions, both to facilitate memory and to ensure that participant hadn't seen the picture before.

The three questions after each picture were:

1. What does this picture represent?
2. Have you seen this particular picture before?
3. Have you seen pictures similar to this before?

In the second phase the participant were asked to visualize each picture for 11 seconds, and afterwards describe the pictures as they remembered them out loud for 25 seconds. Measurements of eye-movement were done for the perception and the visualization/description phases of the experiment. During the debriefing participants were asked some questions regarding pictorial details and were asked to rate the vividness of their visualizations. The reason for the different times during perception and subsequent visualization, was that the pilot experiment had shown, that the questions during the debriefing was to difficult if the participant only were exposed to the pictures for 10 seconds.

---

[44] (Thomas 1999 p. 218)

[45] (Thomas 1999)

The four pictures showed, depicted three natural scenes and geometric pattern: a landscape, a cityscape, a portrait and a fractal. The pictures were presented on a black background.

Figure 3: The pictures used as stimuli taken from the power point presentation.

The four pictures were unlike each other; this was to generate distinct scan paths. The reason to choose pictures of this complexity was to get greater ecological validity. This choice was made to complement the studies already made by Brandt & Stark[46] and Laeng & Teodorescu[47] and thereby putting their hypothesis to the test using even more complex stimuli.
The participant were not at any time instructed to look at certain spots during the perception phase, but they were instructed to look within a white frame and try not to look down during visualization and during their verbal description of the pictures.

In the debriefing, participants were asked to rate the average experience of vividness for all four pictures, and an average of experienced vividness when describing the pictures, and they were asked to give an average vividness rating of mental pictures in their everyday life. In addition they were asked questions about pictorial details. This was done to test what they could remember, compared to the vividness ratings they had given and the scanpath generated.

[46] (Brandt & Stark 1997)
[47] (Laeng & Teodorescu 2002)

4.2. Participants

7 persons participated in the experiment 1 woman and 6 men. The youngest participant was 25 years, the oldest 68 years of age. All had normal or corrected to normal vision, two with glasses. Most of the participants were students and one a pensioners. All volunteered without getting any payment. All had been informed that the experiment was about mental workload, visual perception and mental imagery, and were therefore naive of the underlying hypothesis governing the experiment. They were asked at end of the experimental session about what they thought the purpose of the experiment was and none of the participants had guessed it.

4.3. Apparatus

For measurement of eye-movement a head mounted eye-tracking device from SensoMotoric Instruments (SMI) were chosen. Visual perception is generally binocular, only the right eye was measured and monitored. Infrared corneal reflex video based technology was used to keep track on eyemovements (940 nm), and a Polhemus head-tracker was used, which creates an electromagnetic virtual reality model to compensate for head movements. The system calculates true gaze position as relative to the viewing area (the screen) by using vector analysis, which combines the vectors of eye and head movements. The eyemovements were sampled at 50 Hertz and output consisted of mpegs and eye movements coordinates. The eye-tracking device is mounted on an adjustable bicycle helmet.

The reason for choosing the head mounted tracking device was that this provided the opportunity to get video footage from the participants' perspective, which would generate more research data. This had to be weighed up against that head mounted devices are more intrusive on the participants, but as an advantage, allows more freedom of movements and therefore enables a more relaxed body posture during the experiment compared to remote tracking devices. None of the participant objected or came with any comments that the head mounted device was intrusive or uncomfortable.

The calibration was done using a thirteen point diagram which should give an average error margin of less than 0.5º. The recording of the eye-movement was done using iView version 3.01 from SMI on 400 MHz Pentium computer. The participants were placed in an office chair at an approximately 60 centimeters distance from the screen.

4.4. Stimuli and materials

The participants were exposed to four different pictures downloaded from the internet picture database supplied by Google. The order of appearances had been randomized, but by calculating all the different orders of appearance, and then drawing the different combinations and then make them in Windows Power Point XP pro edition. The presentation was done in Power Point both for the perception phase and visualization phase. The pictures were imported into Power Point in a Jpeg format 800 x 700, the power point presentation was in an 800 x 600 solution because the Power Point doesn't allow an 800 x 700 solution. This provided none distortion of pictures what so ever. The presentation of the stimuli was programmed with exact times for each exposure to give a uniform presentation for each participant. In the visualization phase the pictures where removed and only the frame was left.

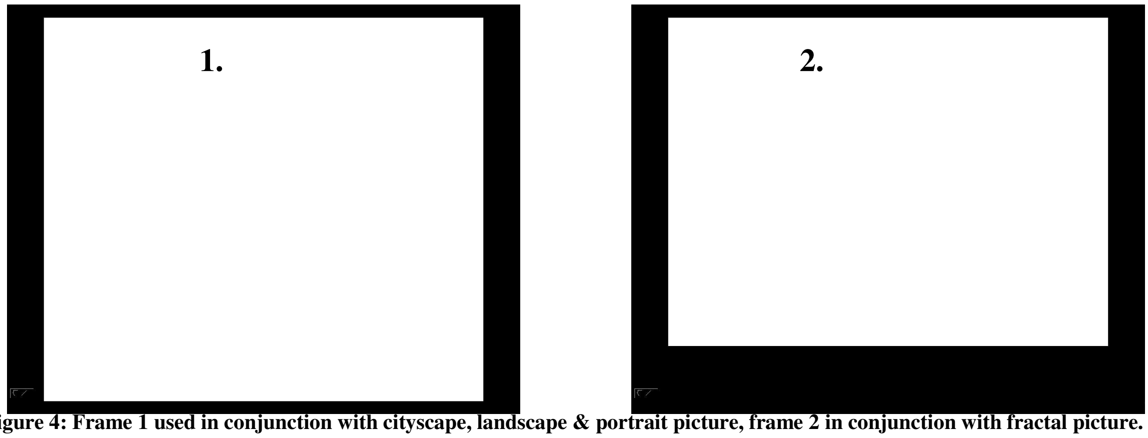


Figure 4: Frame 1 used in conjunction with cityscape, landscape & portrait picture, frame 2 in conjunction with fractal picture.

4.5. The experimental procedure.

Before the experiment each participant were given an information folder on what the experimental objective was. They had been informed that their pupil size would be measured in relation to the experimental objective on providing new data on mental workload, visual perception and visual imagery.
When the participants arrived, they were asked to fill out a form with their names, age, if they had normal or corrected to normal vision and asked to if they had any visual predicaments.
After filling out the form, they were presented to the equipment that was going to be used during the experiment. When the equipment had been mounted, the calibration was done, of the eye-tracker, by looking at thirteen different dots in succession. After the calibration the participant was shown a demo of both the perception phase and the visualization phase. This was done to show the participants exactly how much time they had available, and to remove any anxiety by letting the participants know exactly what was going to happen and in what order. Furthermore the participants were instructed, that they should answer out loud to the questions presented to them on the screen. In connection to the visualization and the verbal description, the participant were instructed verbally and in written form not to close their eyes for prolonged periods of time, not to look down and try to look on the screen display while visualizing and describing.
When the visualization/description phase was over the equipment was removed, and the participants were asked to sit at a table and answer some questions asked by the experimenter these questions was concerned the vividness of their visualizations and pictorial details. Vividness ratings were given first, because participants might reduce their vividness rating if they couldn’t answer all the questions.

4.6. Experimental problems and discussion.

First of all the normal lab computer used for presentation of stimuli in conjunction with the head mounted eye-tracker could not display Power Point presentations, therefore a Macintosh portable computer had to be used instead. This posed a problem for the calibration and Kenneth Holmqvist the assignment supervisor had to construct a new calibration map. This caused that calibration were made additionally difficult. Furthermore the power point sometimes distorted the picture a little which was notice in one of the pre-runs.

The instructions for the experiment were given in English, because the participants were of different nationalities. This might have caused lower reaction times, due to the fact that

participants had to translate the instructions. Only one participant expressed problems with complying. The problems weren't bigger than the experiment could be completed without any significant problems.
Due to technical problems there was not video footage from one participant and that participant had to be removed from the analysis.
Furthermore people had some difficulties complying with the instructions in the visualization/ description phase. There was a tendency to look down but people in general corrected it themselves, but this still took valuable time. In addition some people complied by visualizing the first time, but not on the subsequent visualizations, they had to be verbally reminded to visualize and this might have affected the experimental data, because they mixed up the order in which to visualize and describe. This flaw was not detected during the pilot experiment, and it would have been more prudent to show the visualization demo just prior to the visualization phase, and it might have helped to have all the visualizations first and then the descriptions instead of visualization - description then visualization – description. Considering the time the participant had to visualize this might have caused the data to be biased. In Brandt & Starks experiment participant were asked to visualize the pattern ten seconds after initial exposure, therefore the percept was still stored in short term memory.
In Laeng & Teodorescu the participant were asked to generate the image and were instructed to say when the image was complete. Therefore the time pressure was different in theirs experiments.
If eyemovements are functionally related to visualization these should be present fairly early, so even if a couple of seconds was distorted there should still be a similar scan-path to the percept, this also apply if these movements are epiphenomenal but re-enacted because of other cognitive activity, internal scanning etc.
Furthermore for reasons unknown, the program for making the analysis iView did not work properly and there were problems importing the stimuli pictures into the program so one could make a percentage analysis.

5. Analysis.

Looking at the scanpath in their pictorial form was quite disturbing and it was relatively clear that something had gone wrong. The hypothesis could be wrong, the problems had effected the data or there was a flaw in the experimental design as discussed in the in the procedure section. Consequently the method developed for an analysis had to take these things into consideration. Therefore it was initially planned to do the analysis only with the pictorial material by analyzing the scanpath in their entirety, meaning the scanpath would be analyzed from start to end as one unit.
After a computer breakdown, and a reinstallation of iView it finally worked as it should and this opened for doing the analysis the way I actually wanted from the start. But with the time factor involved, a full analysis was simply an unrealistic objective to attain, therefore I have chosen to include the methods developed in absence of other possibilities and the analysis made using some of the features available in iView. If one does not agree with the arguments made for the deduction, then one should not consider that part of the analysis valid.

5.1. Spatial orientation and extent analysis.

The first part of the analysis will be based on the following argument, if scanpaths made during a visualization of a visual scene are a re-enactment of those scanpath done during perception of the same visual scene, and then one can expect that the scanpath made during visualization should correlate in overall spatial orientation and in overall spatial extent.

So if there is a positive correlation on spatial orientation and extent between visualization and perception then this should be a confirming instance of the hypothesis regarding the potential re-enactment of the scanpaths. If not, it should add to a possible refutation of the hypothesis.

5.2 Spatial orientation criterion.

The criterion is based on the fact, that the fixations can either cluster around a vertical axis or a horizontal axis, that is, the fixations can have an overall vertical or horizontal orientation.
The method used to determine whether or not a scanpath during perception is oriented horizontally or vertically was to measure if the scanpath extended more in height than in width, arranging itself along a vertical axis, then it was judged vertical in orientation. If the scanpath extended more in width than in height, arranging itself along a horizontal axis, then it was judged horizontal in orientation.
The method used is basically very simple and was done in the following manner: All scanpath relevant to the experiment were made in iView, as iView does not allow one to save single scanpaths separately, screen dumps were taken of all scanpath generated during perception, visualization and description. The pictures were then manipulated in Microsoft Photo Editor and Paint. What is important to be aware is, that this analysis do not look at the orientation of the lines within the scanpath, these lines which represent the general direction of the saccades, these lines can go and mostly goes in many different directions up-down and from side to side across the visual field. What was looked at was the overall orientation using the fixations as marking points.

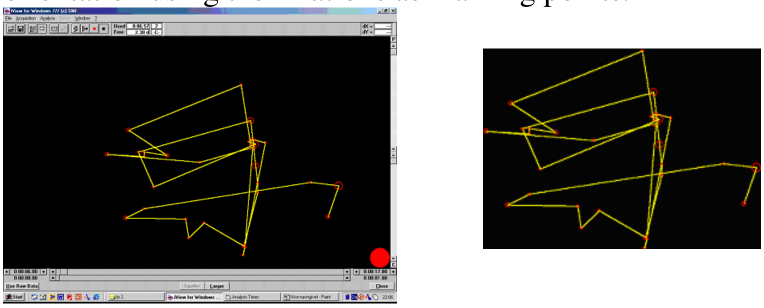


Figure 5: screen dump taken from the pictorial analysis from iView and cut out.

A scanpath generated during exposure to cityscape picture by participant 2 after 11 seconds. Afterwards the same was done for the visualization picture (figure to the right). All pictures underwent the same procedure and where then printed out in pairs (perception – visualization, perception – description) and measurement were made using millimeters as unit. In this case the perception scanpath was deemed horizontal in orientation even though as can be seen there are vertical lines as well. The scanpath of the visualization was judged vertical in orientation.

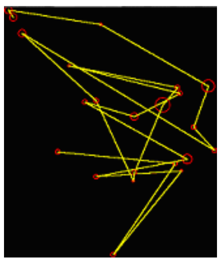
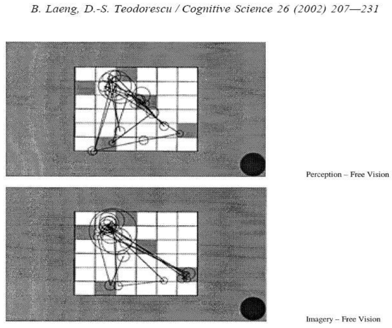


Figure 6: screen dump of visualization cut out.

The result of the analysis was computed using binominal statistics 1 = correspondence, 0 = non correspondence, in this case the example from above would get 0 for non correspondence. The question of course was whether or not this method would give any significant data.

5.3 Spatial extent criterion

This criterion is based on the following deduction; if the visualization is a re-enactment then the spatial extent of the scanpath during visualization should have the same spatial extent as that done during perception. The same procedure as above was deployed. The square area was obtained (in cm.) by multiplying length and width of the rectangles. This was done for perception scanpath, visualization scanpath as well as description scanpath. One can object that this includes a lot of space where there actually are no fixations and saccades. This is correct, but if there is a correspondence and the correspondence is as the one discovered by Brandt & Stark and Laeng & Teodorescu, then the fixations and saccadic movement of the visualization and possibly that of the description should mirror that of perception to a similar spatial extent, and therefore have similar "empty" spaces. Looking at an example from Laeng & Teodorescu one can see the uncanny resemblance between the perception and the visualization. The spatial extent comparison was made by using the following steps: the square value calculated from the perception, was given the value a 100 %.



Then the size of the visualization was computed and the following steps were made. Example: perception equals 30 cm²; this is then considered a 100 percent. The visualization scanpath equals 27 cm² (30/100 = 0.3) then (27/0.3 = 90) the same as 90 percent. Therefore one could say that the visualization is 90 % the size of the original scanpath. In Laeng & Teodorescu's example the visualization is 99% the size of the original scanpath made during perception (see above). The spatial extent of the visualizations and the verbal descriptions will be computed and their size will be given in percent of the percepts size.

In addition the temporal issue has been dealt with in the following manner. The visualization trial lasted for 11 sec. the perception trial 15 sec. a comparison which compared the two pictorial representations without considering the temporal aspect wouldn't be adequate. The same goes for the description trials which lasted for 25 seconds whereas the exposure to the stimuli only lasted for 15 seconds. As a result the material used in comparing the percept and the visualization, the percept has been shortened to match visualization time (11 sec.). In the verbal descriptions case it has been shortened to 15 seconds to match the perception time.

There are of course some problems in the temporal aspect of this analysis. Firstly when the participants were instructed to visualize or describe the picture they had seen, how can the experimenter be sure the participant started complying at exactly the right moment?

In addition this implicit holds, that people when visualizing, they visualize the percept in the same order as when it was perceived. Meaning that, the region they first foveated on during perception is the first region they recall or retrieve during visualization. There is experimental evidence in favor of this interpretation if one look to the results generated by Brandt & Stark and Laeng & Teodorescu[48].

Seen from a statistical point of view, the size of the material presented below does not qualify to say anything significant. Still, possible tendencies can be pointed out, and the material can still be used indicial to confirm or refute the given hypothesis, after all there

[48] (Brandt & Stark 1997; Laeng & Teodorescu 2002)

are 24 trials of visualization and 24 trials of verbal description, where 22 of each obtained data.
All of the analysis is based on pictorial material generated from the analysis program iView version 3.01.12. from SMI and the minimum time threshold for a fixation is set at a 100 ms.

5.4 Spatial orientation results

Spatial orientation table:

Participants:	P 2	P 3	P 4	P 5	P 6	P 7	Sum per trial all P
Landscape picture correspondence of spatial orientation between percept & visualization	1	0	1	1	0	0	3
Cityscape picture correspondence of spatial orientation between percept visualization	0		0	0	1	1	2
Fractal picture correspondence of spatial orientation between percept & visualization	1	1	1	0		1	4
Portrait picture correspondence of spatial correlation between picture & visualization	1	0	1	0	0	0	2
Landscape picture correspondence of spatial orientation between percept & description	0	1	1	0	0	0	2
Cityscape picture correspondence of spatial orientation between percept & description	0	1	0	0	1	0	2
Fractal picture correspondence of spatial orientation between percept & description	1		1	1	1	0	4
Portrait picture correspondence of spatial orientation between percept & description	1	1	1	1		1	5
Sum of correspondence per participants	5	4	6	3	3	3	24
Total number of correspondences in percent of all 22 trials of all participants in relation to percept & visualization	50						
Total number of correspondences in percent of all 22 trials of all participants in relation to percept & description	59,09						
Total number of correspondences of all 44 trials on spatial orientation	54,54						

Key: 1 = spatial orientation correspondence, 0 = non spatial orientation correspondence, = no data

The table on spatial orientation shows that in 54,54% of the trials there is a correspondence between the scanpath during perception and those made during visualization and description. Given that by random chance the probability is 50%, which means that this result is not near significance. In addition the trial where verbal description were the task there is a larger correspondence. This could be caused by describing the pictures the spatial layout of the scene is easily recalled and is reflected in the eye-movements. Still only in 59,09% of cases were there a correspondence which is only slightly above chance. Participant 4 had the highest amount of hits, and seems to have high correspondences, 6 out of 8 trials. The spatial orientation analysis shows little in itself therefore the result has to be seen in relation to the scanpaths spatial extent.

5.5 Spatial extent results

Spatial extent table:

Participants:	P 2	P 3	P 4	P 5	P 6	P 7	Average
Landscape spatial extent: percept & visualization	70,56%	35,73%	77,77%	141,86%	160,13%	10,99%	82,84%
Cityscape spatial extent: percept & visualization	80,11%		8,60%	32,32%	75,06%	34,22%	46,06%
Fractal spatial extent: percept & visualization	56,48%	130,51%	23,95%	46,29%		54,96%	62,44%
Portrait spatial extent: percept & visualization	19,72%	28,33%	85,03%	254,54%	18,70%	61,62%	77,99%
Landscape spatial extent: percept & description	53,04%	11,12%	22,60%	48,19%	74,66%	42,11%	41,95%
Cityscape spatial extent: percept & description	75,75%	120,04%	17,94%	25,79%	17,27%	60,78%	52,93%
Fractal spatial extent: percept & description	45,88%		25,79%	23,62%	86,73%	16%	39,60%
Portrait spatial extent: percept & description	52,10%	13,45%	25,21%	26,95%		16,50%	26,84%
Average total spatial extent of all participants: percept & visualization	68,52%						
Average total spatial extent of all participants: percept & description	40,98%						
Total average of all trials of all participants	54,75%						

Key: Values in percent equals percent size of perception scanpath.

The table show that in all trials the spatial extent of the scanpath were 54,75% in average the size of the spatial extent made during perception. In most of the trial the scanpaths generated during visualization and description is smaller than the ones generated during perception. A different pattern can be seen here the visualization scanpaths shows higher near-correspondences in relation to the perception scanpaths whereas the scanpaths generated during description are in general non-correspondent. This might be due to the fact that forcing to verbalize, participants did not intentionally try to generate a reenactment.

Looking at participant 4 again which had the highest correspondence, the results is rather bleak. Comparing the two analyses makes it clear that, only in two trials (landscape and portrait visualization), there is both a correspondence between orientation and a near-correspondence in case of spatial extent. This unfortunately seems to be the case concerning most trials of all participants.

Therefore the next step that was taken was to eliminate all those scanpaths which either deviated in spatial extent or which did not have orientation correspondence. This was done because it is possible that the visualization or description by chance had the same spatial orientation as perception scanpath. The spatial extent a threshold was chosen and trials that deviated more than + or – 30% of the perception scanpaths spatial extent was eliminated. The reason for choosing 30 % percent as threshold was done to allow a total deviancy that was over 50%. One might well argue that this is either too harsh or too lenient a threshold.

Total hit of spatial correspondence and extent:

Participants:	P 2	P 3	P 4	P 5	P 6	P 7	total sum per trial
Landscape visualizations	1	0	1	0	0	0	2
Cityscape visualizations	0		0	0	1	0	1
Fractal visualizations	0	0	0	0		0	0
Portrait visualizations	0	0	1	0	0	0	1
Landscape description	0	0	0	0	0	0	0
Cityscape descriptions	0	1	0	0	1	0	2
Fractal description	0		0	0	1	0	1
Portrait descriptions	0	0	0	0		0	0
Total sum per participants	1	1	2	0	3	0	
Total sum of positive hits	7						

Key: 1 = spatial orientation correspondence and a spatial extent deviancy of less than 30%, 0 = either non spatial orientation correspondence or a spatial extent deviancy of more than 30%, = no data obtained

As can be seen from the table above only seven out of 48 was positive, if the trials were no data was obtained is deducted there is still 44 left. This gives either a positive result for 14,5 % of all trials (48) or 15,90% of the trials (44), when omitting those were no data was obtained. Separating the visualizations and descriptions results: visualizations: 24 trials, 2 omitted trials, 4 positive trials gives 18% percent and the descriptions 24 trials 2 omitted 3 positive trials gives 13,64 %.

Looking at this result, there is no significant correspondence between the scanpath generated during perception and those generated during visualization and description during this experiment using these criteria. Also there is, as can be seen on the table above, a tendency that the visualizations and descriptions generate smaller scanpath than those made during perception. As a method these two analyses most certainly needs refinement, because the gaze behavior is omitted. There is a slight possibility that a scanpath by chance could corresponds in orientation and deviate with less than 30%.

5.6 The grid analysis

This analysis became possible when iVeiw started working properly a cautious approach was selected. A grid was superimposed on the scanpaths generated, this was done using the define objects option in iView, where it is possible to copy the pictures used as stimuli and define different areas of interest by defining different localities on the picture as objects. In stead of defining different object on the pictures a grid was laid out which consisted of 9 equal sized squares covering the whole picture. A special grid was designed in conjunction with the fractal picture, because the picture is of a slightly different size than the others. This was also made of 9 equal sized squares. When using a grid with objects defined as squares as large as these, a correspondence is of course more easily obtainable, when compared to a more finely masked grid.

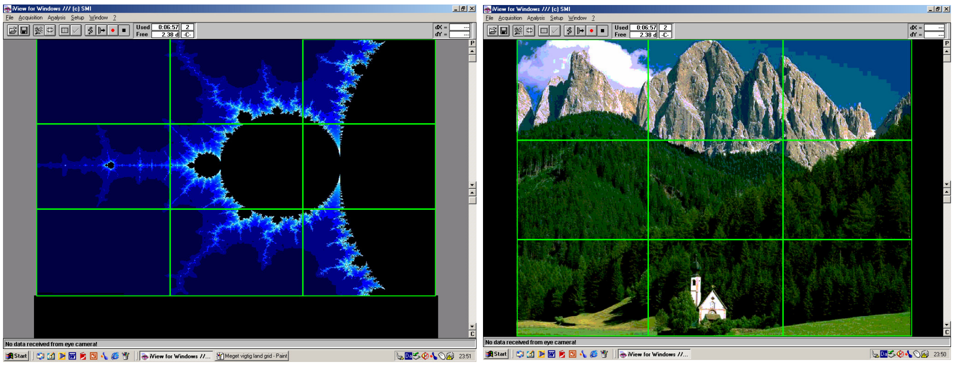


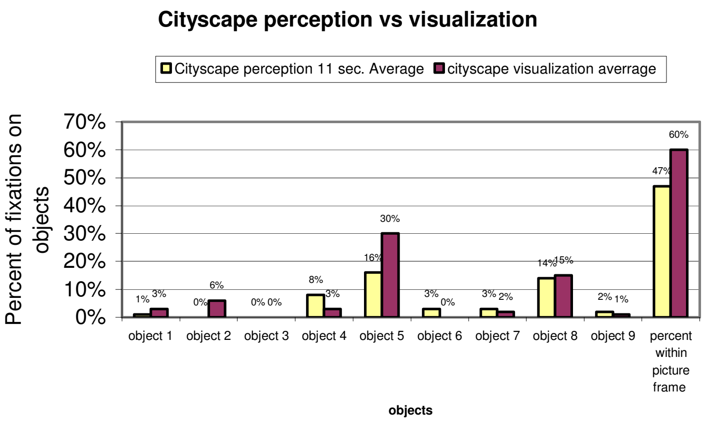
Figure 7: example of the grid laid out on fractal and landscape picture

If the calibration was slightly flawed and features on the picture was designated as objects, for example the tower, the moon or the mountains in the background on the cityscape picture, then these would get the amount of percentages as fixated as what happened during the experiment. Additionally there is a risk, that the participant might not be attracted to the same features as the researcher, and therefore the analysis get biased. One could object that the analysis does not take into account semantics of the respective pictures by choosing this approach. But to obtain whether or not there is a correspondence, one do not need to take these factors into account, as long as one places the same grid on both measurements, the correspondence should still show as similar percentages of fixations on the respective objects, because of similar i.e. reenactments of the scanpath. The percentage distribution should still be correspondent or near correspondent. The reason for choosing a grid consisting of nine squares was taken from Laeng & Teodorescu[49], which used a similar grid in their analysis. The object numbers are placed left to right; Object 1 is top left with object 5 in the centre of the grid.

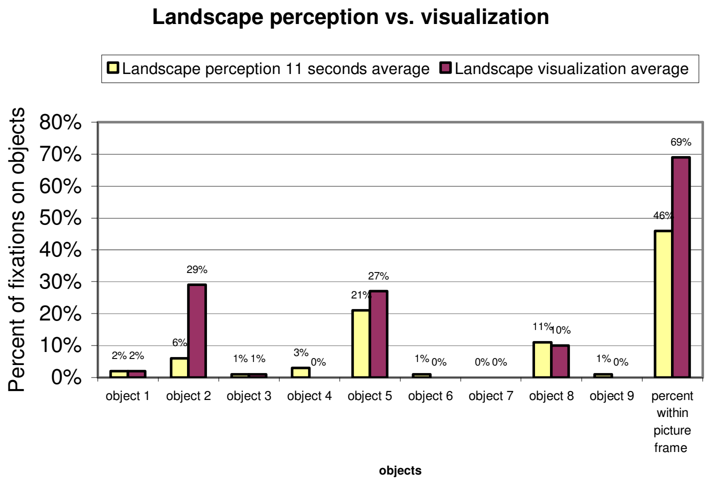
After the objects were defined, this opened the possibility to make percentage analysis of the scanpaths The program (iView) calculates how many percent of the fixations made within a specific amount of time were on the object or objects defined. The fixations were set at 100 Ms. Still it was decided to deal with the temporal issue in the same manner as the analysis concerning spatial orientation and extent. The mean values will be presented below.

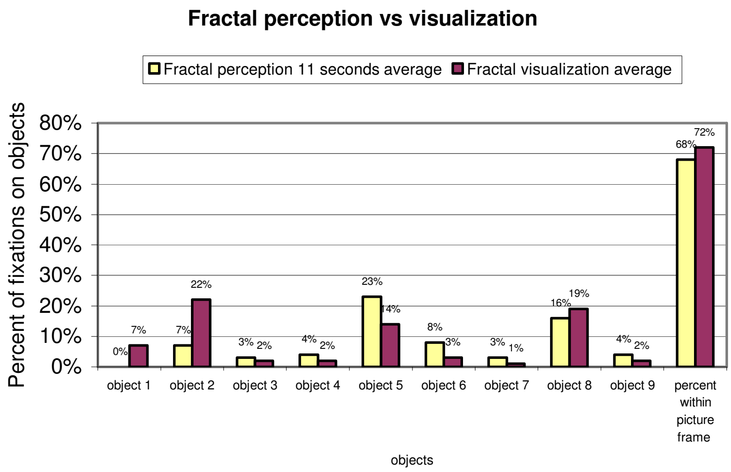
5.7 The grid analysis results

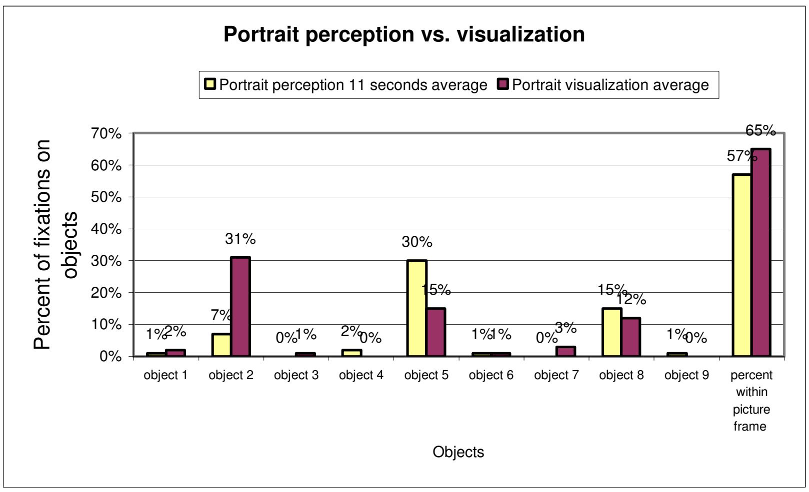
Perception vs. visualization



[49] (Laeng & Teodorescu 2002)







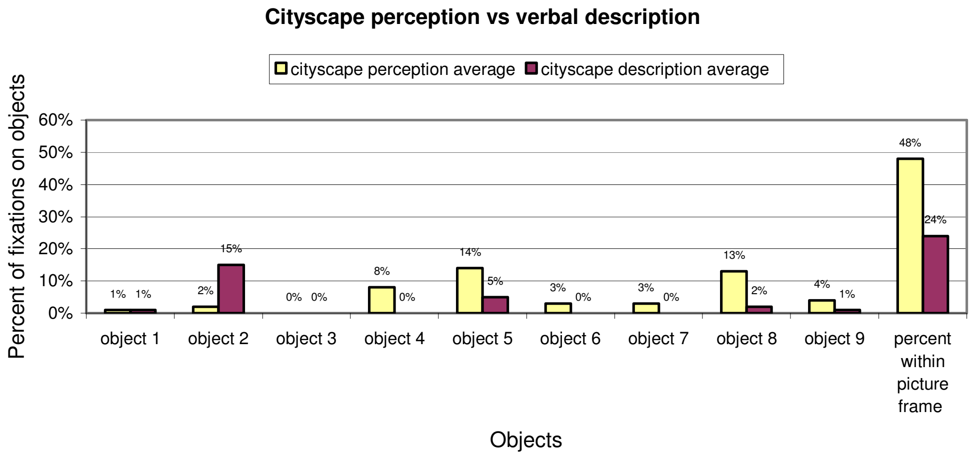
Looking on the graphs above, similarities can be seen. The total percent of fixations within the picture frame is higher for the visualization phase than during the perception phase. The differences are markedly higher for the cityscape picture and the landscape picture whereas the pictures of a simpler complexity have differences of 8% for the portrait picture and 4% for the fractal picture. Therefore the problems getting the participants to comply with the instructions do not seem to have affected the amount of the fixations percentage on the frame itself. Still the scanpath might have been affected. In addition this result is consistent with the results obtained by Brandt & Stark[50], which showed the fixations during visualizations are averagely longer. Glancing at the graph the difference between perception and visualization are substantial, for all pictures. The following table is based on simple deduction. Using the perception data as the prime denominator, the values from the visualization has been deducted. (–) illustrates that the percent of fixations were higher for the visualization than the perception, ( ) illustrates the opposite. 0% means that the fixation percentages are the same for perception and visualization.

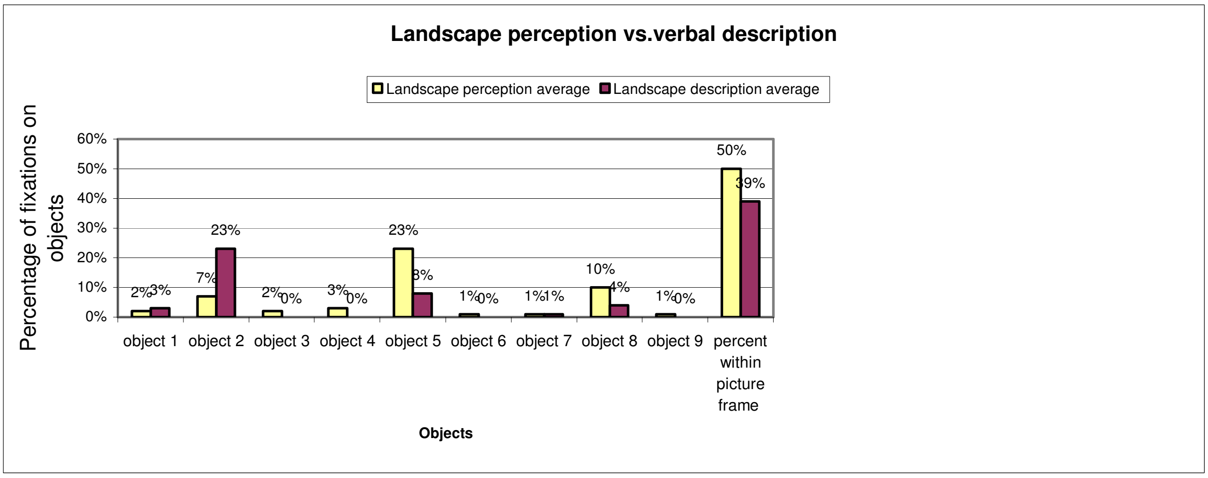
	Cityscape differences in percent	Landscape differences in percent	Fractal difference in percent	Portrait differences in percent
object 1	-2%	0%	-7%	-1%
object 2	-6%	-23%	-15%	-24%
object 3	0%	0%	1%	-1%
object 4	5%	3%	2%	2%
object 5	-14%	-6%	9%	15%
object 6	3%	1%	5%	0%
object 7	1%	0%	2%	-3%
object 8	-1%	1%	-3%	3%
object 9	1%	1%	2%	1%
percent within picture frame	-13%	-23%	-4%	-8%

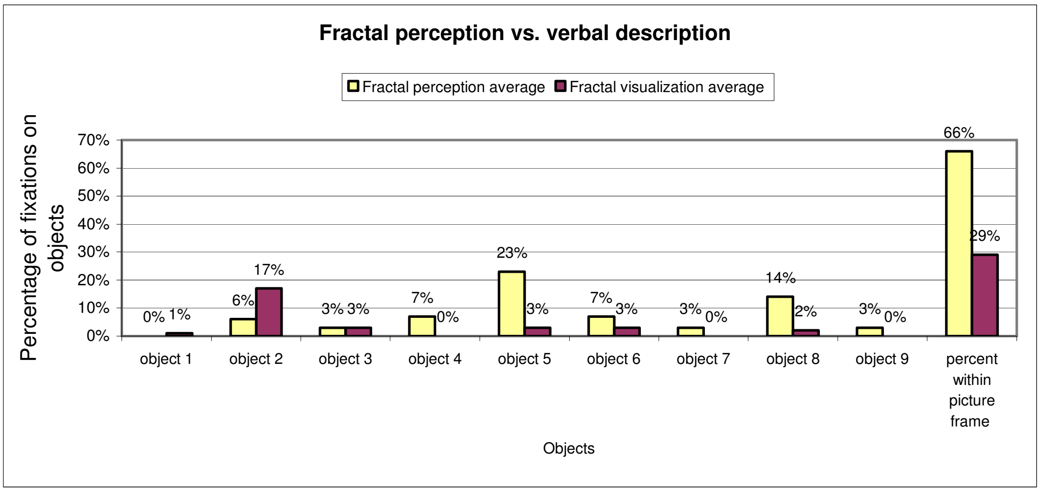
The differences are highest for object 1, 2 & 5 respectively. The differences in some cases can be partly explained by longer fixation times in general, although this not the case in two instances regarding object 5, and the differences are rather substantial. It has to be taken as indicial that the participant had problems generating a reenactment and therefore chose in general to fixate more on the top, the top left and at the center of the picture frame. But as also should be noticed in many instances there is a correspondence or a near correspondence. So as noticed in the analysis of spatial extent, in nearly all instances the scanpath generated during visualization are markedly smaller, and hence the increased percentage distribution in conjunction to object 1, 2 and 5. The possibility of scanpath generated during visualization on average are smaller is of course there, but this seems inconsistent with the empirical data gathered by Finke[51], where the imagery field is analogous to the perceptual field and also because there in some cases during visualization the scanpath actually have a larger spatial extent.

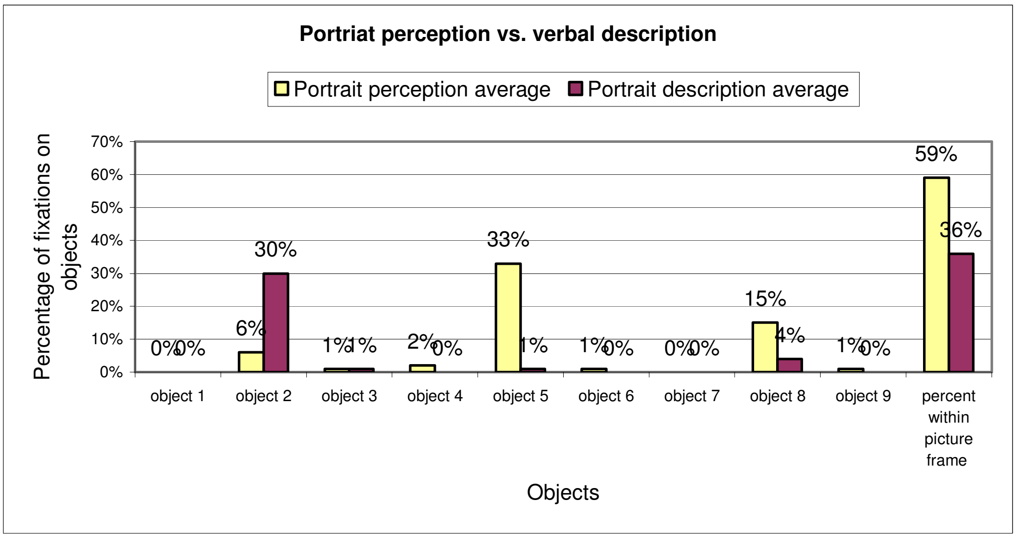
[50] (Brandt & Stark 1997)
[51] (Finke 1989)

The perception versus verbal description









At a glance we see similarities between the percentage distribution compared to the comparison on perception and visualization. Although there is a substantial difference in fixation percentage on objects whereas during the visualization task the total percentage was higher, it is clear that during verbal description it is lower. This could be caused that when performing the verbal description, the participants unconsciously tried to focus at the experimenter looking for a reaction, while at the same time holding their head still. None of the participants did turn their heads during the experiment phase two, but this does not imply that did not have an urge to turn their heads and that this did not affect eyemovements.

Using the same procedure as above to display the differences, using the perception data as the prime denominator, the values from the description has been deducted. (–) illustrates that the percent of fixations were higher for the description than the perception, ( ) illustrates the opposite. 0% means that the fixation percentages are the same for perception and visualization.

	Cityscape description	Landscape description	Fractal description	Portrait description
object 1	0%	-1%	-1%	0%
object 2	-13%	-16%	-11%	-24%
object 3	0%	2%	0%	0%
object 4	8%	3%	7%	2%
object 5	9%	15%	20%	32%
object 6	3%	1%	4%	1%
object 7	3%	0%	3%	0%
object 8	11%	6%	12%	11%
object 9	3%	1%	3%	0%
Total percent of fixations within grid	24%	11%	37%	23%

Here a tendency to gaze upon the top-center object (2) is clear, and the differences in percentage distribution are more explicit. Also contrarily to the visualizations there is a tendency that only on object 1 and 2 are the fixation percentage higher. This means also looking at the other distributions, that reenactment did not occur in the case of verbal descriptions.

5.8 Summary result of grid analysis

The grid analysis shows that scanpaths made during perception and visualization and perception and description does not corresponds, the percentage differences are too substantial, to claim that there is a correspondence. Still, despite the differences there are similarities. On the basis of this analysis one has to reject the hypothesis regarding reenactments. What is important to notice even though the percentage distributions are unaligned, the correspondences were higher in case of the visualizations compared to the verbal descriptions. This should be expected considering that the hypothesis put forward by Laeng & Teodorescu[52] only concern visualizations. This analysis cannot confirm their hypothesis, but can positively state that visualization scanpath has a higher degree of similarity with the scanpath generated during the actual perception, than the one generated during verbal description. As seen the biggest difference is concerning object 2 (top-center). This variance in the data is first of all due to participant 4:

P 4 average percent of visualizations and description	0%	73%	0%	0%	16%	1%	0%	2%	0%

The participant displayed a gaze behavior where during visualizations and descriptions he almost fixated object 2 all the time. One could have chosen to strike him from the analysis, and then similarities in percent would properly have been larger, there might have been correspondence, but this would have been a step which meant manipulating the data to match the hypothesis, and this would have been a very unhealthy scientific approach. If my sample had been larger this participant would not have affected the data as much, if his gaze behavior was just an irregularity. Looking at the percentage distribution in Laeng & Teodorescu table 1[53]

Table 1

Experiment 1. Mean percentages of time spent in each of the 9 regions of interest for each experimental group (free vision and central fixation) during perception and imagery. Region 5 corresponds to the center of the grid and region 1 to the leftmost and topmost.

Region	Free vision		Central fixation	
	Perception mean %	Imagery mean %	Perception mean %	Imagery mean %
1	5.9	5.8	0.2	0.0
2	6.3	5.8	0.2	0.4
3	3.3	3.9	0.0	0.0
4	8.4	11.1	3.8	1.4
5	35.2	39.4	87.3	88.1
6	5.6	6.4	1.7	3.9
7	8.7	6.5	0.8	0.7
8	13.2	10.5	1.1	2.1
9	4.4	3.7	1.0	0.1

It is evident that the difference is not as high as in the experiment generated here; the highest difference is in conjunction to object 5 which is 4.2%. Whereas the highest difference in distribution concerning visualization for this experiment is 24%, substantially higher. In addition the participants in Laeng and Teodorescu's experiment 1 did not have a higher percentage of fixations on object 1 or 2. This indicates that the participants in the experiment, conducted in conjunction to this paper, could not make a proper reenactment. At the same time it has to be acknowledged that the distributions in many cases show a near correspondence.

[52] (Laeng & Teodorescu 2002)
[53] (Laeng & Teodorescu 2002)

5.9 Questionnaire regarding pictorial details

Questioning was done during the debriefing of the experiment and without measuring eyemovements. The questions were given 1 for right answer and 0 for wrong answer and no value at all, if the question wasn't answered. The, "if yes where", questions were only asked if the participant has answered for example that the moon was present on the cityscape picture. If they placed the moon in the wrong side of the picture they would get 0 for incorrect answer.

All questions regarding vividness and pictorial details were asked after the actual experiment. From exposure to the last stimulus to the question were asked went at least 5 minutes and at most 10 minutes. This means that details concerning pictorial information had to be stored in long term memory.

Participants:	P 1	P 2	P 3	P 4	P 5	P 6	P 7
1. Cityscape: Was the moon present on the picture?	1	1	1	1	1	1	1
2. Cityscape: If yes where?	1	1	0	1	1	1	1
*3. Cityscape: Was there an airplane present on the picture?	1	1	1	1	1	1	1
4. Cityscape: If yes where?							
5. Cityscape: Were there any mountains present on the picture?	1	1	1	1	1	1	1
6. Cityscape: If yes where?	1	1	1	1	1	1	1
*7. Fractal: Were the color yellow present in the picture?	0	1	1	1	0	0	0
8. Fractal: If yes where?	0				0	0	0
9. Fractal: Which shape did the geometric figures have?	1	1	1	1	1	1	1
10. Fractal: What was the dominant color?	1	1	1	1	1	1	1
11. Landscape: Were there any clouds present on the picture?	1	1	1	1	0	0	1
12. Landscape: If yes where?	1	1	0	1	0	0	1
13. Landscape: Can you recall how many windows the building had?	0	0	0	1	0	0	1
*14. Landscape: Were there any people present on the picture?	1	1	1	1	1	1	1
15. Landscape: If yes where?							
16. Portrait: What color were his eyes?	1	1	1	1	1	1	1
17. Portrait: Did the man have a necktie on?	1	0	1	0	1	0	1
18. Portrait: If yes which color?	1	0	1	0	1	0	1
19. Portrait: Did the man have any hair?	1	1	1	1	1	1	1
20. Portrait: If yes which color?	1	1	1	1	1	1	1
Number of correct answers	15	14	14	15	13	11	16
Mean of correct answer from all participants	14						
Correct answers mean in percent	70%						

Key: 1 = correct answer, 0 = incorrect answer, = no answer. * = trick question,

<u>Cityscape:</u>

None of the participants had problems answering the questions; this might indicate that the questions were too easy. Only one answer was wrong and this was only in relation to the spatial placement of the moon in the picture.

<u>Landscape:</u>

Several participants had difficulties answering the questions, but the one question they had no trouble answering was the trick question, there were no people on the picture, which indicates that the participants had a good memory of the picture in general, but not minute details.

<u>Fractal:</u>

There is nothing to be noticed other than people had difficulties answering the trick question, was the color yellow present on the picture, this is properly due to the fact that a light green was present on the picture, but still none of the four answering incorrectly named the same location as containing the yellow color.

Portrait:
These questions seemed fairly easy and the experimenter expected 100 % correct answers, but this wasn't the case, 3 participants failed to notice that the man was wearing a necktie, which was surprising. This can be interpreted as a slip or can be seen that viewing behavior when looking at faces are steered towards obtaining emotional information as seen in the experiment conducted by Yarbus[54]. To this picture there were no trick questions.

The correct answer rate was 14 or 70% which is quite high. If one strikes the three "if yes where" questions in conjunction to the trick questions which could not be correct is eliminated, then it is 82%. This means" that overall the questions might have been too easy, or that the participants in general had a good memory of pictorial details. Furthermore the questions answered and the scanpaths generated do not seem to bear any relationship. Participant four had a tendency to fixate at the same locality during visualization and description, but did not have any additional difficulties answering the questions. This is interesting considering the condition set for the participants in Laeng & Teodorescu in the free perception & fixed imagery.

5.10 Subjective vividness rating questionnaire

Participants	P 1	P 2	P 3	P 4	P 5	P 6	P 7	Average	Total average value
How vivid or clear was you mental images of the perceived pictures in general visualization:	5	6	4	5	4	6	6	5,14	5
How vivid or clear was you mental images of the perceived pictures, when you were describing what you saw in general:	6	6	4	5	3	5	5	4,86	
How vivid or clear do you think that your mental images are in general in your everyday life:	7	4	3	7	5	4	5	5	

The value are taken from the questionnaire, where there were used letters A signified the highest value (highest vividness), while G signified the lowest. This gives 7 different values where A has been given the value 7 as the highest and G has been given 1 as the lowest.
As can be seen from the average values generated they are all fairly high, and all the participants had a subjective experience of mental images which in some way or the other has pictorial qualities. The verbal descriptions have a slightly lower value than the visualization, but this can be due to the fact that the participants were not told explicitly to visualize while describing.

5.11 Subjective workload questionnaire

Participants	P 1	P 2	P 3	P 4	P 5	P 6	P 7	Average	Total average value
How easy do you think it was to visualize:	4	4	3	3	2	4	4	3,42	3,21
How easy do you think it is to visualize in your everyday life	4	2	2	4	3	3	3	3	

The values are based on letters from A to E, A signifies the highest value (very easy), E signifies the lowest (very difficult). As can be seen from looking at the table the participants thought of the task, as neither easy nor difficult.

[54] (Yarbus 1967)

5.12 Vividness score of pictures.

Pictures	Cityscape	Landscape	Fractal	Portrait
Participant 1:were there any of the pictures, that you felt was more vivid than others, if any which one	0	1	0	0
Participant 2: were there any of the pictures, that you felt was more vivid than others, if any which one	0	0	0	0
Participant 3: were there any of the pictures, that you felt was more vivid than others, if any which one	0	0	0	1
Participant 4: were there any of the pictures, that you felt was more vivid than others, if any which one	0	1	0	0
Participant 5: were there any of the pictures, that you felt was more vivid than others, if any which one	0	1	0	0
Participant 6: were there any of the pictures, that you felt was more vivid than others, if any which one	0	0	1	0
Participant 7: were there any of the pictures, that you felt was more vivid than others, if any which one	0	0	0	1
Total score pictures	0	3	1	2

As we can see from the table the high score was obtained by the landscape picture, but in case of the cityscape picture which had the lowest score, the guess would be that this picture was more complicated than the others. First of all this picture contained the most depth, second it contained the most differentiated objects. Still the landscape scored the lowest in the questionnaire regarding pictorial details, but the highest rating here. This highlights that the subjective experience of high vividness does not necessarily entail a superior recollection of pictorial details.

5.13 Ease or difficulties of visualization score.

Pictures	City-scape	Land-scape	Frac-tal	Por-trait
Participant 1: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	1	0	0
Participant 2: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	0	1	0
Participant 3: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	0	0	1
Participant 4: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	1	0	0
Participant 5: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	1	0	0
Participant 6: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	0	1	0
Participant 7: Did you feel that, there was one picture that was easier to visualize, if any which one?	0	1	0	0
Total score pictures	0	4	2	1

The landscape picture tops the list again as the most easy to visualize and this seems logical in the light of the above table, again the cityscape picture does not get any score at all which is odd looking at the table below.

Pictures	Cityscape	Landscape	Fractal	Portrait
Participant 1: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	1	0
Participant 2: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	0	0
Participant 3: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	1	0
Participant 4: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	1	0
Participant 5: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	1	0
Participant 6: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	1	0	0	0
Participant 7: Did you feel that, there was one picture that was more difficult to visualize, if any which one?	0	0	1	0
Total score pictures	1	0	5	0

One would expect that the cityscape picture would get the highest score, simply because it is the only picture which did not gain any point on the two last tables. This is not the case, the fractal picture was the most difficult to visualize for the participants.

5.14 Analysis summery

Looking at both analyses results (spatial orientation and extent and the grid analysis), the results are consistent. They show that the visualizations demonstrated a higher correspondence with those of perception, than description. What also can be seen is, if one only had the spatial orientation/extent analysis one would have to state that reenactments do not occur. This method needs refinement, because it does not captivate that the differences although substantial, are not as poor as one would believe. Taken together these two methods give a nuanced picture of the material generated. The questionnaire indicates that scanpaths reenactments and general memory of pictorial detail are not directly connected. The subjective experience of vividness and mental workload does not give any bearing on memory retrieval (correctly answered questions). 4 out of 7 stated that the landscape picture were the easiest to visualize, still looking at the mean answer rate this shows the minute pictorial details does not seem to be encoded. The fractal picture was the hardest to visualize, still the correct answer rate was not substantially worse than the answer rate concerning the landscape picture.

5.15 Reflections concerning the experimental result.

Now that the analysis is done and the experiment is finished one has to ask whether or not the experimental design was flawed. As noted in the procedure section of this paper nearly all participants had difficulties complying with the instructions during the experiment. So

the question is, if this experiment was to be repeated what should be changed to get a more consistent result.
The number of pictures used as stimuli was four, which should be adequate, but if a second experiment was an option, it would have been more prudent to use maybe 2 pictures as stimuli, and prolonging the exposure time instead to 25 seconds, considering the pictorial complexity. By using the Power Points program option of automated slideshow, the whole experiment was automated and this put a time pressure on the participant, which is good if one want to generate an experiment where this factor matters or it is a part of accentuating the result. It did not matter in this experiment and put the participants under an unnecessary time pressure.
Secondly all instructions and questions was written and that caused for example the participant to be put in a situation were they got the following instruction appearing on the screen “please start visualization the landscape picture” this would be on the screen for 5 seconds then the screen would go black for one second and then the frame appeared which the participants should gaze upon during visualization. This, I believe, stressed the participants and could have slowed reaction times i.e. delayed the point in time where the visualization actually started; instead in a future experiment another method would be used. There are in fact two options one could expose the participant to the stimulus for 25 seconds, let the screen go blank, wait for 30 seconds to prevent afterimages and then by verbally instructing the participant to look at the empty frame on the screen which would appear. Then verbally instruct the participant to visualize the picture just seen for 25 seconds, then again wait 30 seconds and instruct the participant to describe the picture while looking at the frame. This would prevent any confusion regarding which stimulus one was to visualize or describe and would be flexible enough to prevent time pressure. This would ensure that the participant acted accordingly to the instructions. In addition instead of using a white background for the empty frame a color would be used blue or grey which are relatively neutral.
The other option would be to use the approach used in the actual experiment, in which the participants was first exposed to all stimuli and then the visualization and the description would begin. But the experiment would be in the following steps 1.Perception 2.Visualization and 3.Verbal description. But to ensure less confusion and stress, the picture that the participant was first exposed to during perception, would be the first picture which should be visualized and in the third step of the experiment describe. By randomizing the visualizations and descriptions, and having different times for exposure, visualization and description during the actual experiment, the task became unnecessarily difficult for the participants, because the intervening exposures, visualizations and descriptions might accidentally have acted as noise[55]. To be sure that intervening pictures, visualizations and description would not act as noise, the first regarding option experimental design would properly be the most prudent, as a result, if the experiment was to be repeated the following design would be chosen:

Perception Stimulus 1. →	Visualization Stimulus 1.→	Verbal description Stimulus 1.

This procedure would of course be repeated for the next stimulus and so on, all instructions would be verbally given.
This would allow the use of Power Point by using it as manually operated slideshow instead of an automated one, which would be fine as a medium for presentation.

[55] (Ishai & Sagi 1995)

These changes would be adequate to get a successful experiment in relation to participants delivering an optimal result. The question is however if these changes enhance the ecological validity of experiment?

6. General Discussion

6.1. Noise and ecological validity

There are of course other possibilities why the experimental data produced this result. In an article by Ishai & Sagi[56] proved that low-level memory traces of sensory information are withheld in a form of iconic memory for up to five minutes. When they repeated the experiment after ten minutes the effect disappeared. If the scanpaths generated during perception reacts as a low-level memory trace there is a possibility that the visualizations and verbal description in the experiment, in some instances exceeded the five minute limit. Some of the scanpath generated during visualizations and verbal descriptions were made 5 minutes after initial exposure and this opens the possibility that the sensory low-level memory traces vanished and therefore the ocular-motor movement that were displayed were not a reenactment.

As mentioned before, another possibility is accidental noise, when, in their experiment placing noise after each trial, no imagery effect was seen. It is therefore likely that the question posed visually to each picture afterwards and the subsequent exposure to stimuli acted as noise and therefore disrupted the scanpath stored in memory. If this is the case then there is one conclusion to be drawn.

Scanpath reenactment is an experimental entity, a thing of the laboratory, because in normal everyday life people are exposed to infinite numbers of visual stimuli, that can act as noise and from perception of a scene to a point in time where they had to recall it, intervening events happen. Eyemovements can therefore not be functionally related to memory retrieval of spatial components of a given scene. Not in a way which inhibits memory retrieval of a given scene or object in a vital degree. If eyemovements during retrieval of spatial information is functional, and are related to the initial perception of a given object or scene, it should endure the passage of time. If not, then it is very possible that these traces are stored in short term visual memory, but not in long term memory.

As mentioned under the section imagery theories one of the big issues is whether or not the primary visual cortex is used in conjunction with imagery. If low level memory traces decay such as ocular motor movement traces, then eye-tracking as a methodological application cannot help clarify this issue, which is acknowledged by Mast & Kosslyn[57]. Another possibility raised by active perception theory is that the procedure or schema generated in the act of *seeing as*[58] does not activate the ocular motor plant that activates these movements. If this is the case one has to reject that eyemovements has a significant functional role in the remembrance of spatial components of a given picture. The questions posed to the participants seem to confirm this rejection. The participants did not have any difficulties answering most questions regarding the gist of the pictures and most did not have any problems regarding placing the objects at their spatial locality. Therefore it seems unlikely that scanpath reenactments have a constitutive role in memory retrieval.

In the case of participant four he maintained, for reasons unknown, his gaze fixed during visualizations and descriptions and had one of the highest sums of correctly answered

[56] (Ishai & Sagi 1995)
[57] (Mast & Kosslyn 2002)
[58] (Thomas 1999)

questions. This does not correspond well with the result achieved by Laeng & Teodorescu. According to their results for the free and fixed group he should have scored the lowest[59]. The likelihood of their result of lesser accuracy for this group seems to be more a question of, that setting this condition produced anxiety or divided the participant's attention, inducing a lesser result. Or it means that there does not have to be a correspondence between percept and visualization, but only the presence of individual unconstrained eyemovements, if subjects are not allowed to move their eyes as they wish, it impedes the appropriate activation of the ocular motor plant and the retrieval of a scenes spatial component.

6.2. Checkerboards versus natural scenes

The experiments regarding eyemovements, that I have referred to in this paper both used checkerboards as stimuli, and one might argue that these stimuli are to simplistic to have any ecological relevance. Furthermore in one of the experiments using checkerboards the grid remained but without the black fillings during visualization[60], therefore it cannot be rejected with absolute certainty that these results are not due to cuing. The grid could have induced or facilitated the right scanpath behavior and helped memorizing the position of the filled squares and reduce the experiments ecological validity. Laeng & Teodorescu did try to enhance ecological validity by introducing real world objects. This consisted of using pictures of different fish. Only one fish were used as stimuli at a time.

An objection that states that the stimuli complexity used, pushed the hypothesis to hard and caused it to fail is, at a glance a reasonable objection. Further scrutinizing will reveal that it is not, if one states that small steps has to taken to slowly further ecological validity and the reach of the hypothesis, then one has to take into consideration the claims stated by the hypothesis itself. Laeng & Teodorescu does not limit their hypothesis to only explain an experimental result. They claim that reenactments do happen and that these have a function in relation memory retrieval. This claim transcends the confines of the laboratory and checkerboard stimuli and extent into the world. The claim made by Brandt and Stark are not as extensive therefore it is not as easily refuted, but at the same time does not contain as high a potential explanatory power. Furthermore small steps forward is a viable approach, one just has to keep in mind, that this approach illustrates the uncertainty of how far a given hypothesis extends.

One can state that the experiment conducted failed in confirming the given hypothesis and by its size of sample lacked statistical validity, but an open question remains. Did it in reality fail because of flawed design or did it fail because it lacked the conditions to induce an experimental phenomenon, which have no ecological validity?

6.3 Reenactments

The experiment conducted here and other experiment conducted with a similar agenda raises the issue of what is a reenactment or what is similar? This issue has to be discussed, because it affects what one defines as correspondence. No one mentioned in this paper assume that a reenactment has to be 100% identical to the scanpath generated during perception. Reenactment in itself means to repeat actions of past events. The assumption which lays implicit in the analysis is reenactment entails similar geometric form on the scanpaths generated both regarding orientation and extent. Similar or very near the same distributions fixation percentages on objects. In addition the sequence of fixations should be the same. These criteria are all fulfilled in connection to the experimental result

[59] (Laeng & Teodorescu 2002)
[60] (Laeng & Teodorescu 2002)

generated by Laeng and Teodorescu, but not in relation to the experimental results generated here.

6.4 Visualizations versus descriptions

As discussed in the analysis above, visualizations were closer to be reenactments than descriptions. This might not be surprising as the hypothesis discussed in this paper only extends as far as visualizations. One possibility was that the higher level of correspondence was task induced. Before the visualization/description phase it was emphasized that during visualizations, the participant should try to see the pictures with their "minds eye". This emphasis was not done in conjunction to verbal description. The raise the issue regarding tacit knowledge, but augmenting that tacit knowledge is at play here entails the following. People has exact notions on how they look at pictures, in addition they have exact notion on how they look a different pictures. This means in relation to eyemovements, that people can differentiate between pictures and the related eyemovements, and then simulate how they for example look at the portrait picture and induce the correlated eyemovements. If tacit knowledge were the case the level of correspondence ought to have been higher between visualization and perception. Furthermore the subjective experience of perceiving belies tacit knowledge in relation to eyemovements. The subjective experience is that we perceive one coherent picture, which is not the case, regardless of which theory one cling to regarding visual representation. This entails that tacit knowledge somehow extends down to the psychophysics level, which a majority of studies seems to deny[61]. A more plausible explanation would be that certain schemata or procedures are instantiated and therefore emulate the motions of seeing the given picture. Still this entails that the results are somewhat task induced which is consistent with the result made in conjunction to internal scanning[62]. Additionally the information needed to complete the two different tasks could be stored differently. In conjunction to visualizing in some visual memory while the knowledge needed during verbal description in some kind of semantic or episodic memory[63].

6.5 The theories of imagery

Turning to another aspect of the discussion, one of the aims of this assignment was to contribute to the theoretical discussion in connection with imagery.
If we look at the experimental results for answers taking it at face value, the answer has to be that the quasi-pictorial cannot be right in what I described as the strong interpretation if imagery is analog to perception. But using this line of argument on the basis of the experiment would be unjust to the theory involved. One can dismiss eyemovements as being evidence in favor of the strong interpretation, but that does not dismiss the strong or the weak interpretation. Imagery can still be analog to perception, and therefore in the case of visual imagery activate the primary visual cortex, when we look at other lines of evidence. So in rating the different theories we have to use another approach. I have chosen to start with the theory I find least likely to be true on the basis of empirical evidence, the propositional theory. The propositional theory in general, does not succeed in given a satisfactory account of imagery and perception. The fact is that activations of brain regions connected to visual imagery have been found. The debate concerning primary visual cortex rages on, but that does not diminish the fact that activations occur[64]. The propositional theory has not accounted for these findings, and a satisfactory account within a

[61] (Finke 1989)
[62] (Baddeley 1997; Finke 1989)
[63] (Baddeley 1997)
[64] (Cabeza & Nyberg 2000; Finke 1989; Thomas 1999; Thomsen *et al.* 2001)

propositional framework is very difficult if one insists that only one format exists which is disconnected from perception. The question and main argument against the pictorial theory remains tacit knowledge. Still evidence for and against tacit knowledge has been provided. A situated approach has been used to revive the propositional theory, but this does not seem to be enough if one does not choose to compromise and acknowledge that more than one format exists, pictorial or not, to account for regional activation within the brain. And doing that means abolishing the main tenet of the theory.
This takes us to the quasi-pictorial theory; still the problem regarding imagery and tacit knowledge persists. This is mainly due to the agenda of reductionism the theory has. The aim of understanding imagery in its most reduced form so it is generally applicable. But insisting that imagery functions on nearly the same principles as the laws of physics[65] prevents taking into account, cultural differences, personal knowledge and experience[66]. Here the PA theory can adjust and argument that personal knowledge affects imagery as well as perception, because the schemata and procedures vary from individual to individual based on past experience[67].
The PA theory has the advantage of being a flexible theory, but lacks the empirical foundations to contest the quasi-pictorial theory. But regarding explanatory power, it can explain why eyemovements or brain regions sometimes are activated and some times not, by using the procedures or schemata as an explanation, but not actually how it happens[68] In addition this theory can explain why imagery seems to be based on the same neural substrate as perception, the *seeing as* approach. This is also possible from a quasi-pictorial standpoint, if a sharper definition between imagery short term and long term memory is accomplished. It is plausible that the primary visual cortex and reenactment of scanpath only are activated in relation to short term or iconic memory. Still if we have an internal “pictorial” representation that is retained across saccades the problem explaining change blindness and inattentional blindness persists[69]. When the percept is processed from the visual buffer and becomes a deep representation the low-level cognitive processes might not be stored with it and during subsequent retrieval not present. This could explain some of the contradictory data from PET and fMRI studies and give a plausible explanation of only higher level activation in some cases. All the theories have inconsistencies some greater than others. The PA theory has the potential to wrest the propositional theory from the “throne” as the main competitor to the quasi-pictorial theory by taken a situated approach, and still have an plausible scientific explanation for the physical phenomena concerning imagery and accounting for tacit knowledge. Still there is a long stretch forward; if the PA theory is to be taken seriously, empirical evidence has to be produced.

7. Conclusion & future research

Now reflecting on the result achieved, did these fulfill the objectives of this assignment?
No, the results are not adequate enough to either confirm or refute the hypothesis concerning scanpath reenactments. Many factors could have played a significant role in affecting the results achieved. One can state that the visualizations had a higher level of correspondence than the verbal descriptions hence reenactments do not seem to occur in relation to verbal description. Two plausible explanations resides, one is that different procedures are initiated; the other is that the information used to complete the two tasks are stored differently, causing retrieval of different information.

[65] (Finke 1989)
[66] (Pylyshyn 2003a)
[67] (Thomas 1999)
[68] (Thomas 1999)
[69] (Henderson & Hollingworth 2000, Rensink 2002; Simons & Chabris 1999)

Regarding the second hypothesis, concerning the functionality of the scanpath reenactment as a facilitating mechanism in memory retrieval of spatial components, the result achieved by this assignment is ambiguous; reenactment of scanpath if these exist are epiphenomenal and does not appear to have any influence on memory retrieval itself. It is possible that eyemovements in general facilitate memory retrieval, but when one is inhibited from making these, it inhibits memory retrieval. Another interpretation would be that the actual scanpaths is stored as a trace in iconic or short term memory, and plays no functional role in long term memory retrieval.
So in furthering the ecological validity of the hypothesis the experiment failed, and in one instance led to a weak refutation of the hypothesis concerning the functionality of the scanpaths.
Whether or not eye-tracking as a methodological approach can contribute to the imagery debate the answer has to be yes with reservations. An effort into clarifying some issues has to be done first, to get a notion of what impact this method can have on the field of inquiry. More research effort has to be put into experiments that clarify if reenactments happen. If these happen, is the ocular motor movement's activation of the relevant motor areas stored in a short or long term memory. In addition the reenactments vulnerability to noise and intervening visual stimuli has to be clarified.
If these uncertainties are clarified eye-tracking as a method has the possibility of having a lasting impact on imagery research.

The discussion above brings one to reflect over the enterprise of science, validity and experiments in general. Experiments are in a sense a controlled environment were the experimenter know at precisely at what time a stimulus was introduced and what stimulus it was, and what factors could affect the result. This is the ideal of a laboratory enterprise if not, it is not possible to argument that this effect, condition or event was due to this cause or this effect and when one cannot establish these things, the experiment in reality tells you nothing. But it is possible to create an experimental environment that is so disconnected from the world, that it set out to explore, that it on the other hand again tells you nothing. The aim of science is the exploration of ourselves as human beings and the world we live in, and therefore some of the conditions in an experiment have to coincide with conditions of the external world and its environment to be credible or have any value other than that of a mere curiosity.

8. References

Baddeley, A. 1997. HUMAN MEMORY Theory and Practice. Psychology Press Ltd. U.K

Brandt, S.A. &. Stark, L.W. 1997. Spontaneous Eye movements During Visual Imagery Reflect the Content of the Visual Scene. Journal of Cognitive Neuroscience 9.1 pp. 27 – 38. 1997.

Cabeza, R. & Nyberg, L. Imaging cognition II: An empirical review of 275 PET and fMRI studies. Journal of Cognitive Neuroscience, 12(1).

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

Duchowski, A.T. 2003.Eye-tracking Methodology theory and practice. Springer. U.K. 2003
Finke, R.A. 1989. PRINCIPLES OF MENTAL IMAGERY. A Bradford Book. The MIT Press. U.S.A.

Glenstrup , A.J. & Engnell-Nielsen , T. 1995. Eye Controlled Media: Present and Future State

Henderson, J.M. & Hollingworth, A. 2000. Eye Movements, Visual Memory, and Scene Representation. Michigan State University Eye Movement laboratory Technical Report, 2000, 5, 1-17.

Holšánová, J., Hedberg, B. & Nilson, N. 1997. VISUAL AND VERBAL FOCUS PATTERNS WHEN DESCRIBING PICTURES. Poster at the 9th European Conference on Eye Movements (ECEM 9) in Ulm, Germany, September 1997.

Josephson, S. & Holmes, M. E. 2002, Visual Attention to repeated Internet images: Testing the scanpath Theory on the World Wide Web. ETRA'02 New Orleans Louisiana USA.

Kosslyn, S.M. Ganis, G. & Thomson, W.L. 2003. (Letters) Mental imagery: against the nihilistic hypothesis. Trends in Cognitive sciences Vol. 7 No. 3 March 2003.

Kosslyn, S.M. 1994. Image and Brain The resolution of the imagery debate. A Bradford Book. MIT press. U.S.A. 1994.

Kosslyn, S. M. 1980. Image and Mind. Harvard University Press. U.S.A.

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

Levine, M.W. 2000. Levine & Shefner's Fundamentals of Sensation and Perception. Third edition. Oxford University Press Inc. New York.

Mast, F.W. & Kosslyn, S.M. 2002. Eye movements during visual mental imagery. Trends in Cognitive Sciences Vol 6 No. 7 July 2002.

Nyberg, L. 2002. Kognitiv neurovetenskap studier av sambandet mellan hjärnaktivitet och mentala processor. Studentlitteratur, Lund. Sweden.

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

Pylyshyn, Z.W. 2003a. Return of the mental image: are there really pictures in the brain? Trends in Cognitive Sciences Vol. 7 No. 3 March 2003.

Pylyshyn, Z.W. 2003b. (Letters response) Explaining mental imagery: now you see it, now you don't. Trends in Cognitive sciences Vol. 7 No. 3 March 2003.
Pylyshyn, Z.W. 2001. Visual indexes, preconceptual objects, and situated vision. Cognition 80 2001 p. 127-158.

Pylyshyn, Z.W. 2000. Situating vision in the world (Pylyshyn – situated vision, opinion). Trends in Cognitive Sciences – Vol. 4. No. 5, May 2000.

Rensink, R.A. 2002. Change Detection. Annual. Review. Psychol 2002. 53.

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

Simons, D, J. 2000. Attentional capture and inattentional blindness. Trends in cognition science vol. 4. No 4, April 2000

Simons, D.J. & Chabris, C.F. 1999. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 1999 vol. 28 pages 1059-1074.

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) 1999 pp. 207 – 245.

Thomson, W. L. Kosslyn, S. M. Sukel, K. E. & Alpert, N. M. 2001. Mental Imagery of High- and Low-Resolution Gratings Activates Area 17. NeuroImage 14, 454 – 464. 2001

Wexler, M. Kosslyn, S. M. & Berthoz, A. 1998, Motor processes in mental rotation. Cognition 68 77 – 94. 1998.