The Role of Effort in Multiple Object Tracking

A Pupillometry Study

Steven Harry Pieter van de Pavert

Master Thesis - September 2011
Master of Psychology - Cognitive Neuroscience
Psykologisk Institutt - Center for the Study of Human Cognition
Universitetet i Oslo

“Everyone knows what attention is. It is the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneous objects or trains of thought.”

William James, 1890

Abstract

We used the multiple object tracking (MOT) paradigm to investigate sustained attention to several objects simultaneously, while directly manipulating task difficulty. By means of pupillometry, we investigated the role of effort in the MOT task and aimed to distinguish between different proposed models for MOT: Pylyshyn’s early-vision model, a purely attentional account of MOT, Yantis’ perceptual grouping model, and a purely serial account of MOT. A phasic increase in pupil size was observed when tracking several objects, while a decrease in pupil size was seen when subjects passively viewed the display. Moreover, the phasic pupil dilation in tracking conditions was proportional to task difficulty. Previously, pupil responses have been demonstrated to have an intimate relation to activation of the Locus Coeruleus, which in turn is thought to have a modulating effect on attention through its norepinephrinergic projections. Importantly, phasic activity of the Locus Coeruleus has been associated with task “exploitation”. The results appear to be in line with Pylyshyn’s early-vision model and a purely attentional account of MOT, whereas other models may have more difficulties explaining the current results. Since the assumed application of effort in MOT differs in these models, suggestions are offered to further distinguish between these models and to clarify which mechanisms make us able to pay attention to several objects simultaneously.

Acknowledgements

The research presented in the current work has been performed as part of a larger project aimed to address attention. More specifically, in the larger project the underlying mechanisms of attention are approached by means of different psychological techniques (e.g. pupillometry and functional magnetic resonance imaging) and different paradigms (e.g. multiple object tracking (MOT) and Lavie's 'perceptual load' task). The project is conducted under the supervision by Professor Bruno Laeng, who, together with Thomas Espeseth (a postdoctoral researcher at the Center for the Study of Human Cognition), supervised me in the research that I conducted as part of this project. I have been involved in multiple parts of the larger project, but decided to focus on attentional effects on pupil dilation in the MOT task. The idea to assess divided attention using the MOT paradigm and to apply pupillometry was posed by Professor Bruno Laeng. Together with Thomas Espeseth and Markus Handal Sneve, a PhD candidate at this institute, we discussed how to approach the research, and during different stages of the project we met regularly to discuss ideas and how the research was developing. The MOT paradigm was implemented and in part developed by Markus Handal Sneve, who made the paradigm directly applicable and ready for use. Together with Silje Jynge, a Master student at the institute, I have approached all participants, tested them and collected all data. We collected data for both the MOT paradigm, which was my paradigm of interest, and the Lavie's 'perceptual load' paradigm, which was Silje Jynge's paradigm of interest. Subsequently, I wrote a computer program which allowed us to preprocess the pupillometry data in order to 'clean' them before statistical analyses. The MatLab-based software was written in a user-friendly and dynamic fashion, so that pupillometry data of multiple studies could be preprocessed by means of this software. Indeed, this software has later been used in future pupillometry studies at the Psychology Institute of the University of Oslo as well. A description of this program can be found in the method section of this Master thesis. Results of the current study have been presented in an oral presentation I gave at the Eye Tracker and Pupillometry Conference at the Uppsala University in May 2010.

I would like to thank Thomas Espeseth and Bruno Laeng for supervising me in the research which I had the honour of participating in. It has been a pleasure collaborating with them and they have taught me a great deal about the fundamentals of academic research. I am convinced that the basis they provided me will prove of great value in future undertakings.

Table of Contents

1. Introduction

At any given moment, one is not able to focus on all presented information simultaneously; certain information is attended to, whereas other information is not. On the other hand, it appears that we are able to focus on several objects at the same time. A central question thus arises: What determines what information is attended to? The underlying mechanisms of attention and their relations to perception have indeed been and still are important themes within psychological research, and date back at least to William James (1890) who famously described attention as follows:

> *"Everyone knows what attention is. It is the taking possession of the mind, in clear and vivid form, of one out of what seem several simultaneous objects or trains of thought."*

Two basic phenomena play a role in attention (Desimone & Duncan, 1995). One has a limited capacity for information processing, so not all presented stimuli can be can be processed simultaneously and given equally much attention, and, consequently, one has to filter out inessential information, so that no (limited) attention is given to irrelevant stimuli. These two phenomena put forward a biased competition model which poses that due to limitations in capacity a competition exists between presented information as to what information will be processed. This competition between information is biased however; behaviourally relevant objects are favoured over behaviourally irrelevant objects. Different cognitive models have been proposed describing the nature of attentional capacity and its relation to perception. Broadbent's (1957) model proposes that irrelevant information after being registered, is not processed further. Deutsch and Deutsch (1963), on the other hand, suggest that the bottleneck of attention lays after perceptual analysis. Their model suggests a bottle neck right before selection of an appropriate response to the stimuli. Alternatively, or perhaps complementary, capacity models of attention suggest that, rather than a bottleneck in the process from sensory registration to response selection, a general limit exists that puts a limit to the amount of information that is attended to.

Kahneman (1973) elaborates on the abovementioned models of attention and stresses the involvement of "effort" in attention. The general capacity limit of attention is thought to be subject to several factors such as effort and arousal. When aroused, one's attentional capacity is larger than when one is in a less aroused state. Also, when more effort is provided, the capacity limit increases, allowing a person to pay more attention. A distinction is made between the *inclusive* and *intensive* aspect of attention. Inclusive attention is considered an automatic form of attention, and it functions in a stimulus-driven manner (e.g. in the detection of salient stimuli). The intensive aspect of attention, on the other hand, is the voluntary allocation of the limited resources of attention and will henceforth be referred to as effort. Any given task requires effort and causes arousal. The amount of effort provided is adjusted according to the demands of the task at hand. An analogy used by Kahneman may aid to clarify: When one puts a slice of bread into a toaster, an additional load is added to the electrical system of the house. The generator has to adjust its power input accordingly to make sure that the demand for energy is met. Similarly, when engaging in a demanding task the capacity limit is increased by an increase in effort to meet the attentional demands of the task.

Note that the current work addresses attention to visual information. However, underlying mechanisms of attention may not be restricted to a single modality, but may involve modality-independent cognitive processes.

1.1 Multiple Object Tracking

Previously it was thought that when several objects are presented at the same time, these objects are attended in a serial manner (e.g. Posner, 1980). However, when Pylyshyn and Storm (1988) developed the multiple object tracking (MOT) paradigm, they demonstrated that we are able to focus on several object simultaneously. The MOT task involves a number of objects displayed on a screen. A subset of these objects are indicated as target objects (by blinking or a change in color), which are to be tracked while the objects move across the display in a random fashion, whereas the objects that are not indicated as targets (distracters) are to be ignored. Importantly, after cueing the target objects are featurally undistinguishable from the distracter objects. Once the objects have stopped moving, the participant is either asked to indicate which objects were previously indicated as targets (full response) or is asked if a certain object is a target or distracter (partial response). Generally one is able to follow four to five objects for a number of seconds, with performance being better with fewer targets. This capacity limit, however, is subject to a number of factors, which will be discussed later on.

Addressing attention by means of the MOT paradigm carries several advantages compared to other paradigms that aid to investigate attention (Cavanagh & Alvarez, 2005; Scholl, 2009). First, tracking several objects in an MOT task, compared to tasks in which one focuses on a single object, has more resemblance to real-world situations, such as following several people in a crowd, keeping track of cars while crossing a street or watching a football match. Second, it measures sustained attention rather than attention to brief events and has a clear active component, as opposed to other paradigms in which one waits for an (unusual) event to happen (e.g. Gabay, Pertzov, & Henik, 2011; Laeng, Orbo, Holmlund, & Miozzo, 2011). Third, the MOT paradigm leaves room for an independent secondary task. Secondary tasks in MOT are generally included in order to identify processes that are involved in the task by investigating whether the secondary task interferes with the ability to track multiple objects or if the secondary task requires entirely different cognitive mechanisms (e.g. Pylyshyn & Storm, 1988; Tombu & Seiffert, 2008). Fourth, difficulty can be manipulated relatively easily. In other attention tasks that aid to investigate attention, difficulty is generally manipulated in an indirect manner through adjustments on the temporary scale. In the MOT the most direct way of manipulating task difficulty is to increase the number of targets to track. Further, increasing the speed of the objects (Liu, et al., 2005), or decreasing the minimum distance between the objects (Pylyshyn, 2003) also leads to a decrease in task performance. However, a study in which the distance between objects, object size, and the speed of the objects were manipulated to investigate these influence of these variables suggested that speed and object size are not directly influencing performance on tracking performance (Franconeri, Jonathan, & Scimeca, 2010). Rather, object spacing seemed to be the determining factor of tracking performance. The reason that speed and object size also seem to interfere with tracking is thought to be due to the fact that larger and faster moving objects get closer to other objects more frequently. Altogether, the abovementioned advantages of the MOT paradigm make it a paradigm which is very suitable to study the underlying mechanisms of attention. It shows a relatively close resemblance to natural situations, it measures sustained attention, it leaves room for a simultaneously presented secondary task, and it allows for direct manipulation of task difficulty.

1.1.1 Proposed Models of Multiple Object Tracking

The underlying mechanisms of attentional tracking have been a topic of debate within attentional research. Debates of MOT have predominantly focused on four topics. First, it is unclear whether MOT is object based or region based. Second, it has been asked whether we track multiple objects by attentionally enhancing the representation of targets, whether we suppress distracters, or if we apply both strategies. Third, it is unclear if working memory plays a role in MOT, or if MOT is a purely attentional process. The role of working memory, however, could not be assessed with the methods applied in the current research, and falls thus outside the scope of the present work. Fourth, a controversy exists whether or not parallel mechanisms are involved in MOT, and if they are involved, what the role of these mechanisms are and how they are applied. Several models have been proposed that explain how we can track several objects simultaneously. Four main theories that have been proposed to explain MOT will be discussed here: Pylyshyn's early vision model (Pylyshyn, 1994; Pylyshyn & Storm, 1988), a purely attentional model of tracking as posed by Scholl (2009), Yantis' model of perceptual grouping (Yantis, 1992), and a serial attention-switching model. The current work will focus on the role of effort in these different models and is aimed to distinguish on these grounds.

In their original paper, Pylyshyn and Storm (1988) propose the so-called "fingers of instantiation" (FINST) model. According to this model, prior to the involvement of attention an automatic cognitively impenetrable mechanism is applied, which provides anchoring points to the target objects during tracking. This primitively parallel mechanism in this way provides 'sticky fingers' which themselves do not provide information but indicate to which objects attention is to be directed. Attention subsequently switches from one object with a 'sticky' index to the next in a serial manner (Pylyshyn, 1994; Pylyshyn & Storm, 1988). The preattentive parallel FINST mechanism is an early visual mechanism that functions in a cognitively impenetrable way and should thus be considered separate from attention or working memory. Limitations on MOT performance are thought to be caused by a limited number (four or five) of indexes that can be applied. Important in this model is the preattentive nature of the FINST mechanism, after which attention serially addresses these objects through these indexes. Especially this preattentive aspect of the FINST model is not generally accepted and several other models have in stead been proposed which do not assume such a mechanism. Mental effort should not be required in the initial assignment of indexes, since it is a preattentive and automatic process. Pylyshyn notes that maintenance of the indexes onto the targets, on the other hand, may be an effortful process. It seems logical to assume that the effort invested in MOT is proportional the number of targets. When tracking more objects, more effort should be required to keep the 'sticky fingers' attached to the targets, than when fewer objects are tracked.

Scholl (2009) questioned whether it is necessary to think that there is more to tracking than attention alone. He argues that we do not have sufficient evidence to think that tracking involves more than purely attention, and that we thus, by reason of parsimony, should adopt the view that no other cognitive processes are implicated in MOT. The role of effort is not explicitly stated in his suggested model, but an increase of effort could be expected that is proportional to the number of targets. Additional attentional capacity should be required when tracking more objects, leading to a greater demand of effort. Note that the application of effort in this model is different from the application of effort in Pylyshyn's early vision model.

Yantis (1992). has suggested that the target objects are the corners of a polygon of which we keep track. According to this theory we are in effect not tracking several objects, but we are automatically grouping these objects together. This would lead us to require only one steady attentional channel to track a single object, instead of several attentional channels or a single channel that switches from object to object. Much like Pylyshyn's FINST model, the grouping of targets is suggested to be an automatic preattentive mechanism. Yantis does not explicitly mention effort in his work, but it seems plausible to suggest that effort is not applied when grouping objects together to a polygon, for this is thought to be an automatic process, but that effort is applied to follow the virtual polygon during tracking. Since only one shape is to be tracked, no additional effort would be required when a polygon should be made up of more objects. On the other hand it may be argued that a virtual polygon that consists of more objects has a more dynamic shape and that more effort should thus be required when more objects are to be tracked.

Finally, the possibility exists of a serially operating attention-switching model. In this model, in contrast to previously mentioned models, no intermediate processes between perception and attention are assumed. Attention would simply 'jump' from one object to the other in a serial fashion. However, due to limitations in switching speed, a purely serial tracking model seems not plausible. Performance on the MOT task is better than could be predicted when applying a purely serial tracking paradigm, which suggests that at least a partial parallel mechanism is involved in MOT (Pylyshyn & Storm, 1988). Nonetheless, in the current work a serial tracking model is included for the sake of completion. Effort in a serial model of tracking is not expected to increase with an increase in number of targets. As mentioned above, if attention is purely serial, no additional demand would exist on attentional function when one has to track more objects, because one object is attended to simultaneously and cognitive processes should not process additional information at any given moment.

Different studies have offered evidence that provide arguments for and against abovementioned models. Evidence against purely unifocal models of attention has been provided by the finding that tracking of one object is not influenced by movement of other objects. Clever designs of paradigm allowed to distinguish between serial and parallel processing techniques (Howe, Cohen, Pinto, & Horowitz, 2009). Objects in the display would only move for a part of the trial; they would either move and stop simultaneously, or objects would move independent in a sequential manner. Because subjects performed better in the condition in which objects moved parallelly, this suggests that multiple object occurs tracking (at least in part) parallelly. In Pylyshyn's (Pylyshyn, 1994; Pylyshyn & Storm, 1988) and Yantis' (1992) models we would require only one focus of attention rather than several foci, for we are tracking only one object at a time. However, a convincing argument has been provided against tracking of single objects in the finding that MOT is hemifield dependent (Alvarez & Cavanagh, 2005). When presenting an MOT task in the left and right hemifields of a display separately, and comparing tracking capacity to another condition in which two MOT tasks were presented in one hemifield of the display, it was found that that the limit of tracking is split between hemifields; "rather than a limit of, say, 4, subjects demonstrated a limit of 2+2". This implies that the limitation on MOT capacity limit is located relatively early in visual processing. Moreover, it argues against the models that propose attention to be unifocal, because at least two foci of attention are needed to explain a split in tracking capacity between hemifields. With regard to the question whether we mainly suppress distracters or whether we enhance targets, a convincing study indicates that we mainly enhance targets, rather than suppress distracters. An event related potential (ERP) study on MOT investigated responses on task irrelevant probes (Drew, McCollough, Horowitz, & Vogel, 2009). Small 'dot-probes' were presented on targets, non targets and static objects in order to determine which objects are attended to which degree. An enhancement of the evoked P1 and N1 signals (two task stimulus dependent ERP signals) was observed when presenting probes on targets compared to presenting probes on non-targets. In addition, subjects that performed better on this task demonstrated larger differences between probes on target than probes on distracters. This study thus indicates that target objects are enhanced early in visual processing.

1.2 Pupillometry

Recently the technique of measuring pupil dimensions within psychology celebrated fifty years. A study by Hess and Polt (1960) made clear that not only luminance, but also emotionally relevant or arousing stimuli lead to a change in pupil size. Subsequently, two landmark *Science* papers within the field of pupillometry found an effect of cognitive processes on pupil dilation. The 1964 paper by Hess and Polt and the 1966 paper by

Kahneman and Beatty demonstrated that performing a task that involves working memory is associated with pupil dilation, and, interestingly, that task difficulty is directly related to the magnitude of dilation. Hess and Polt (1964) demonstrated that pupils were larger when performing a complex multiplication task (e.g. 16 × 23) than when performing a simpler task (e.g. 7 × 8). Interestingly, a study eleborating on this finding (Ahern & Beatty, 1979) showed that more intelligent subjects yielded a smaller increase in pupil diameter than did less intelligent subjecs when performing a demanding task. Kahneman and Beatty (1966) confirmed the suggestion that cognitive effort is an indicator for pupil dilation. To this end, five subjects performed three working memory tasks: They were presented 1) a string of three to seven digits, which were to be recalled immediately, 2) a sequence of four mono-syllabic nouns, also meant for immediate recall, and 3) a series of four digits that require transformation before recall. In figure 1 mean pupil diameter of participants is displayed when they performed the different tasks. A significant increase in pupil dilation was found when performing a working memory demanding task. Furthermore, this increase was dependent on task difficulty; a more demanding task yielded a larger increase in pupil size compared to a less demanding task.

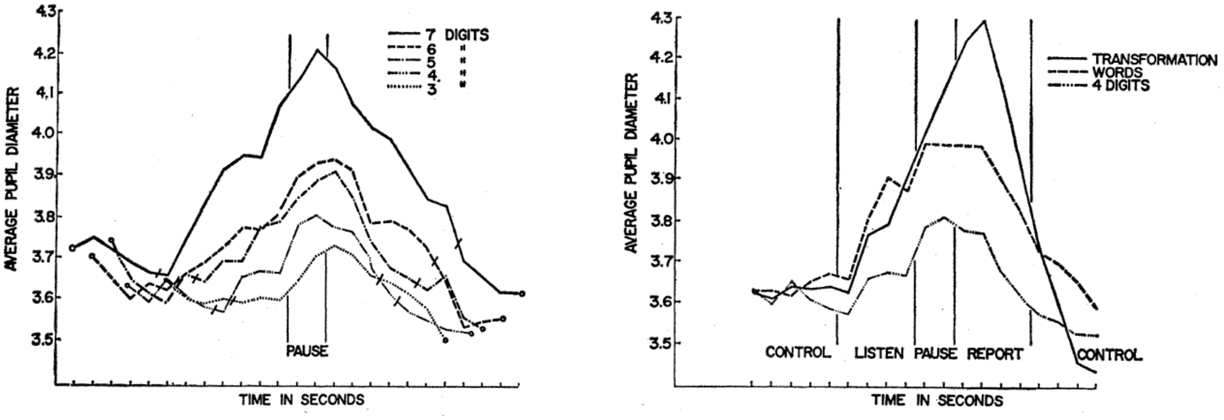

Figure 1. (a) Pupil diameter increases with increase of memory load (i.e. number of digits to be remembered). (b) Involvement in a more complicated tasks yields a larger pupil diameter. From (Kahneman & Beatty, 1966).

Kahneman (1973) remarks that many studies converge on the notion that pupil size increases when performing a more demanding or difficult task, and that increasing evidence suggests that pupil size reflects “the subject’s momentary involvement in the task”.

Kahneman suggests that pupillometry is a useful technique to measure mental effort for two reasons: It allows distinguishing between difficulty of tasks through measuring the effort subjects have to apply to fulfill the task, and it provides information about relatively short-lasting variations in mental effort. Additionally, compared to other techniques used in psychology, pupillometry is a relatively cheap and non-invasive technique.

Importantly, pupil dilation has a unique feature by being nowadays a relatively easily measurable physiological phenomenon that is closely related to a neurotransmitter system. A study by Rajkowski, Kubiak, and Aston-Jones (1993) (revied by Aston-Jones & Cohen, 2005)) showed compellingly that tonic activity of the Locus Coeruleus (LC) has a direct relationship with pupil diameter (figure 2). This is due to the fact that the LC plays a key role in the neural circuitry regulating pupil dilation. Coincidentally, the LC is the sole norepinephrine-releasing nucleus in the human brain. Thus, given the tight link between pupil dilation and the LC, pupillometry provides us with a unique window on the workings of the norepinephrine (NE) system (Laeng, Gredebäck, & Silvois, in press).

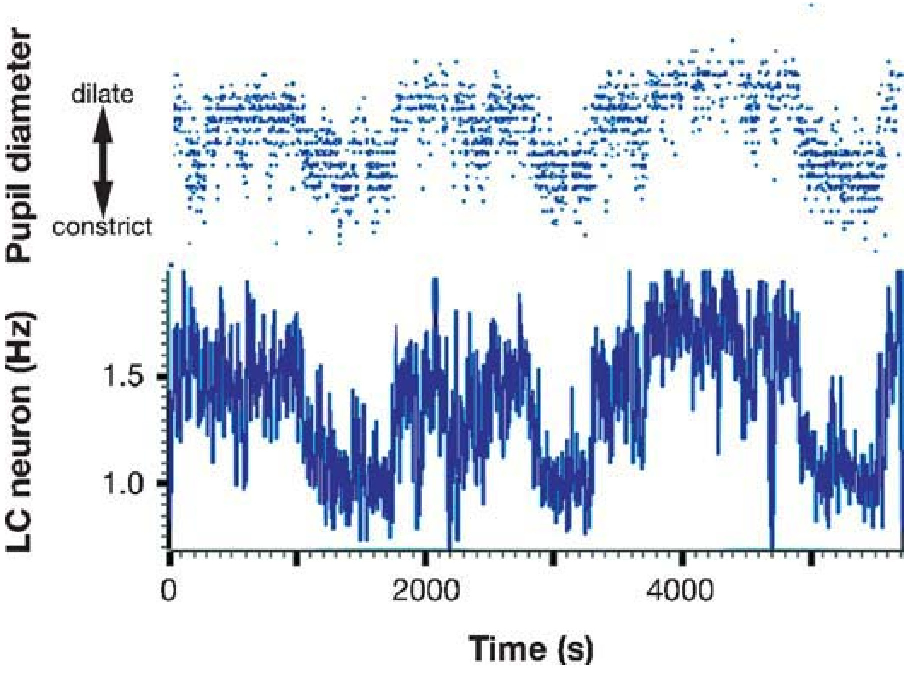

Figure 2. Association between baseline firing rate of an LC neuron in monkey and pupil diameter. From (Aston-Jones & Cohen, 2005)

1.3 The Locus Coeruleus, the Norepinephrine System and its Relation to Attention

The LC is a brain structure which is bilaterally located rostrally of the pons (Aston-Jones & Cohen, 2005). It functions as a hub for the NE system and has projections to virtually the entire brain, with the noteworthy exception of the basal ganglia. The LC-NE system plays an important role in mediating cortical function in concert with other neuromodulators (Sara, 2009). Damage to the LC-NE system is associated with neurological disorders characterized by stress (e.g. anxiety, depression, and panic disorders) and a rat study has shown that pharmacological disruption of the LC-NE system leads to a decrease in cognitive abilities (Mair, Zhang, Bailey, & Toupin, 2005). Through its wide-spread innervations, the LC-NE system plays a role in different cognitive functions. Important for the present article is the theory that the entire attentional system is regulated by the LC-NE system (Corbetta, Patel, & Shulman, 2008). Related, the adaptive gain theory states that phasic activation of the LC optimizes task performance on attentional tasks (Aston-Jones & Cohen, 2005).

The LC-NE system has two modes of activity: tonic activation and phasic activation. Note that generally a dichotomous distinction is made between phasic and tonic activation of the LC, while in reality there may be a continuous range between these states. However, for current purposes we will use the distinction between phasic and tonic LC activity as commonly used in the literature. Both modes of activation are essential for attention, however in different ways. Phasic activity of the LC-NE system is thought to be involved in exploitation of a task at hand (i.e. focusing attentional resources on a single task), whereas tonic activation is thought to be involved in orientation to novel (i.e. paying attention to novel stimuli). Phasic activation is thought to function as an ‘interrupt’ or ‘system reset’ signal to an attentional system that is involved in attentional orienting, the ventral attention system (Dayan & Yu, 2006). Attention is thought to involve two networks with different tasks: a ventral attention system and a dorsal attention system (Corbetta, et al., 2008). The ventral attention network is a network which is responsible for reorienting attention to important novel stimuli, whereas the dorsal network is activated in sustaining attention on a task at hand. The ventral network can, when an important stimulus appears, interrupt the dorsal network in order to focus on the important novel stimulus. Activation of the ventral attention network may be regulated by the dorsal attention system via projections to frontal areas (e.g. the anterior cingulate cortex and frontal operculum) directly, or in an indirect way through the LC-NE system. When performing an attentionally demanding task, phasic signals from the LC are thought to regulate the activity of the ventral reorienting system such that one focuses only on the task at hand. Indeed the temporoparietal junction (TPJ), a key region of the ventral attention network, appears to be deactivated under demanding conditions (Shulman, et al., 2003; Todd, Fougnie, & Marois, 2005). When task utility begins to wane, the LC may enter its tonic state thereby allowing the ventral attention system to reorient attention to other more relevant stimuli in the environment.

Input to the LC is provided by the anterior cingulate cortex (ACC) and the orbitofrontal cortex (OFC) (Aston-Jones & Cohen, 2005). These two frontal areas are involved in evaluation of task utility. The ACC is an area that is concerned with evaluation of the cost of a task (Carter, et al., 1998), while the OFC, on the other hand, is a brain area that plays a role in decision making and reward (Rolls, 2000). The mode of activation of the LC is determined by evaluation of rewards of a task. Both areas receive information from many neocortical and subcortical structures among which the lymbic system and amygdala. Since both areas send axonal projections to the LC, this led to the hypothesis that the LC-NE system regulates attention in a way that is determined by task utility. When performing a task, attention is focused on the task at hand by the dorsal attention system, while activation in the ventral attention system is suppressed. When the OFC and ACC do no longer consider performing the task sufficiently valuable, the LC receives input to enter a tonic state, hereby lowering the threshold for the ventral attention system to respond to external otherwise relevant, but task-irrelevant stimuli. Activation of the ventral attention system may in turn interrupt the dorsal attention system, hereby reorienting attention in order to focus on the more important stimulus. In this manner, at all times a balance is sought between exploring the environment and exploiting a single task within the environment.

A recent study investigated the involvement of the LC-NE system in orienting of attention (Gabay, et al., 2011). Participants performed the Posner cuing task, which allowed distinguishing between volitional (intensive) and reflexive (inclusive) orientation of attention, while activation of the LC-NE system was measured by means of pupillometry. They found a larger phasic dilation of the pupil in a more complicated task than in an easier task. Further, this dilation was found to be time-locked to response, rather than to stimulus onset. This study thus demonstrated the notion that the LC-NE system has two modes of activity, with the phasic mode of activation being involved in more demanding tasks, while the phasic mode is more present in easier tasks.

In contrast to the Posner cuing task, the MOT task requires *sustained* attention, which is the main task of the dorsal attention network. Activation of the ventral attention network should thus be suppressed, so that no reorientation takes place and the dorsal attention network can continue to focus on the MOT task. We thus expect that the LC-NE system will be active in a phasic fashion, so that activation in the ventral attention system may be suppressed during MOT.

1.4 Summary

The underlying mechanisms of attention are a matter of debate within psychological research. Different models are proposed attempting to explain the nature of the limit in attention that causes us to be able to focus on only a limited number of objects. Kahneman (1973) stresses the importance of effort in attention, and states that pupillometry, more than being a measure of working memory load or attentional load, is a measure of cognitive effort. Pupil dilation appears to be closely related to activation of the LC, which is thought to regulate attention mainly through its norepinephrinergic innervations to the ventral attention network. Phasic activity of the LC is expected to silence activation of the ventral attention system through its many innervations to this system. During MOT, the ventral attention system is then not able to interrupt the dorsal attention network in order to abandon the task at hand. More suppression, and thus more LC phasic activity is expected to be required to suppress activation in the ventral attention system, when the task is made more difficult by increasing the number of targets.

The method of pupillometry was chosen to investigate the role of effort in MOT because it provides a unique window on activity of the LC-NE system. Interestingly, previous studies on MOT that included recording of eye movements (Culham, et al., 1998; Fehd & Seiffert, 2008; Pylyshyn & Storm, 1988) did not analyze pupillometry data, even though this additional data is easily obtained. Based upon previous pupillometry studies (Gabay, et al., 2011; Hess & Polt, 1964; Kahneman & Beatty, 1966) an increase in pupil size is expected when performing the MOT task versus passively observing the same display. Furthermore, the magnitude of pupil dilation may depend on the number of targets that are to be tracked; when tracking more objects the pupil may dilate more, because of an increase in cognitive demand. Based upon different predictions of the role of effort in MOT we attempted to distinguish between different models of MOT. The current work focuses on four different proposed models of MOT. First, Pylyshyn’s early-vision model assumes a role for effort in maintaining the preattentive ‘sticky fingers’ onto the objects that are to be tracked. Second, a purely attentional account of MOT (Scholl, 2009) may assume a role for effort in applying additional attentional resources to follow more targets. Thus, in this model an attentional load dependent increase in pupil size is expected. Third, a perceptual grouping model (Yantis, 1992) expects us to only use a single attentional channel. Additional effort should thus not be required with more target objects in this model. Finally, a purely serial account of MOT would expect an equal involvement of effort in tracking few objects as in tracking many objects. Because target objects are attended in a serial manner, there should be no additional demand on attentional processes. However, given that accuracy decreases with increasing targets, one could still argue that pupillary dilations occur as a response to the “stress” of making errors and losing targets along the way while trying to serially track one object at a time and switching attention.

2. Methods

2.1 Subjects

In the current study 44 subjects were included (28 females). Their age ranged between 20-48 years (Mean = 28.9 years, SD = 7.6 years). All participants had normal or corrected-to-normal-vision.

2.2 Apparatus

Participants were tested in a windowless soundproof room. Participants were instructed to place their head in a headrest so that their head was still and their eyes were at a set distance of 90 cm from the screen. Pupil diameter was recorded from the participant's left eye by means of the Remote Eye Tracking Device (RED), built by SMI-SensoMotoric Instruments® from Teltow (Germany). The RED 2 can operate at a distance of 0.5–1.5 m and the recording eye tracking sample rate is 240 Hz – that is approximately every 4.1667 ms, with resolution better than 0.1°. The eye-tracking device operates based on determining the position of two high contrast elements in the eye: the pupil and the corneal reflection. According to an independent calibration procedure, given the constant distance from the screen, the pupil diameter was recorded with a definition of about 500 pixels/mm. However, given that changes in diameter were used in the interpretation of data, the exact eye diameter was not needed. Illumination of the room does not interfere with the recording capabilities of this apparatus. The coordinates of all boundary points are fed to a computer, which, in turn, determines centroids of the two elements. The vectorial difference between the two centroids is the “raw” computed eye position, which in turn is used to compute the pupil diameter based on the horizontal and vertical projections of the pupil's ellipsoid at the different sampled positions. Stimuli were presented on a 21” EIZO CRT monitor using the Psychophysics toolbox extensions (version 3, for MatLab [MathWorks, Natick, MA]).

2.3 Procedure and Paradigm

The paradigm used was similar to the paradigm described previously (Espeseth, et al., 2010). However, a few parameters were altered for use in the current study. Fewer objects were presented (10 in stead of 12); a passive viewing condition was included in which no objects were to be tracked, whereas the condition in which six object were to be tracked was excluded. Furthermore, the number of trials was reduced due to an increase in trial length (between 19 and 24 seconds, depending on speed of response), because of the slow readjustment of pupil diameter to baseline level.

Each trial began with the appearance for 1.0 seconds of a centrally presented, white 0.2° diameter fixation point, and ten blue 0.7° diameter discs, non-overlapping and randomly spread over the gray 17° × 17° display area (see Figure 3). Importantly, the objects and the display area were isoluminant, so that luminance would not interfere with pupil dilation. Similar to the paradigm of Phylyshyn and Storm (1988), participants were instructed to focus on the fixation point for the entire length of the trial, apart from the response phase of the trial. After 2.5 seconds either none or a subset of two to five discs turned (isoluminant) red for 2.5 seconds before returning to blue; the red colour designated the target discs to be tracked in the current trial. After a brief interval (0.5 seconds) the discs started moving in random directions with a speed of 5.5° per second. To avoid predictable trajectories, each disc made a random change in direction on average once per second. The moving discs bounced off the edges of the display area as well as off each other when they got too close (1°, edge to edge). Additionally, to avoid pulling fixations away from the centre the fixation point also 'repelled' the discs. After 10 seconds the discs stopped moving and the participant, using the mouse cursor, indicated which objects he/she had been tracking (full response). After clicking on the designated number of target discs, the participant received feedback about the number of correctly tracked targets in the trial. Participants completed five practice trials, one per load condition, before commencing on the experimental trials. Each load condition was presented 20 times in the experiment, which consisted of a total of 100 trials. The order of conditions was pseudo-randomized, so that no two subsequent trials were of the same condition. Participants controlled the pace of the experiment by initiating the start of a trial with a mouse click. On average the experiment typically lasted approximately 45 minutes.

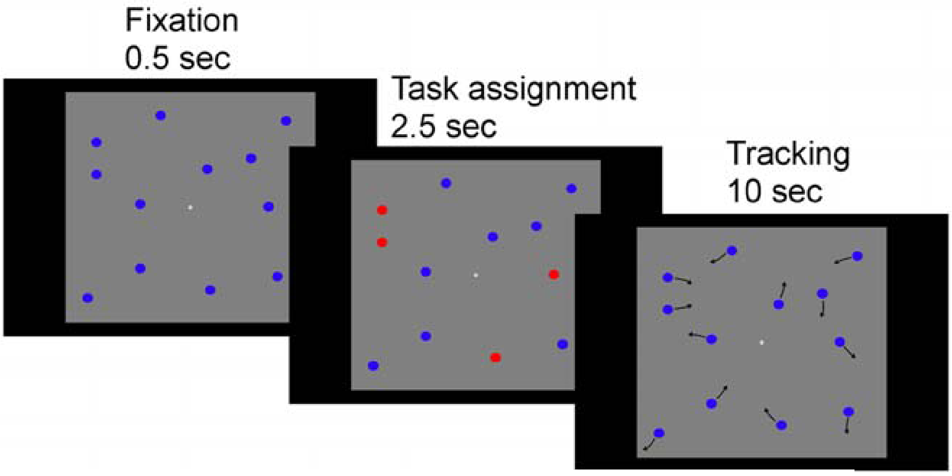

Figure 3. A selection of scenes from a typical trial of the multiple object tracking task. Image from (Espeseth, Sneve, Rootwelt, & Laeng, 2010).

2.4 Data Handling

2.4.1 Accuracy Data

In order to be able to compare accuracy, the proportion of accurately tracked targets was calculated per load condition (number of targets) as follows: The number of correct responses per load condition was divided by the number of possible correct responses for every subject. In the accuracy data one cell remained empty because of insufficient sample points. The missing value was estimated in a similar way as described in step thirteen of the pupillometry data preprocessing.

2.4.2 Pupillometry Data

Prior to statistical analyses, pupillometry data was preprocess in order to 'clean' the data. To this end, in-house software was developed that treated the data as follows.

1. The data had to be down-sampled from 240Hz to 40Hz for processing purposes. This was done by including every sixth sample point in the analysis. The reason for including every sixth data point, instead of including the mean over a bin with a width of six sample points, was to ensure that the data kept a larger resemblance with the raw data. The data were stored in a format containing four columns: Sample point number, set number (This is a variable that is generated by iView X [SensoMotoric Instruments, Teltow, Germany], in order to store in which epoch a sample point is located), the diameter of the pupil in the X-dimension (in pixels), and the diameter of the pupil in the Y-dimension (in pixels).

2. Files that contained data with different time organizations due to technical issues and files that were damaged had to be excluded from analysis. A typical file contained 25 trials. Ideally, every participant is represented by four files, adding up to a hundred trials per participant. Subsequently, the information in the first column was converted from sample point number to milliseconds. This was done by multiplying every value by a factor of 4.1667 (1000 milliseconds / 240 milliseconds). Also, the X and Y diameters per sample point were averaged. The resulting file contained three columns: milliseconds, set number, and the average pupil diameter between X and Y (in pixels).

3. All data were converted from data containing information about pupil diameter into number of pixels and pupil diameter in millimetres. This was done by multiplying all pupil diameter values by 16.72, a conversion factor between pixels and millimetre.

iView X provides a tool to determine this conversion factor. To this end, one places one or multiple black circles in front of the eye tracker's camera. The value that iView X reads in pixels divided by the known actual size of the 'pupil' in millimetres yields the conversion factor.

4. Blinks were excluded, i.e. all sample points that contained the value 0 were replaced with an empty value.

5. Physiologically impossible data were discarded. Based on previous research that indicated that the diameter of human pupils diameter can vary between than 1 millimetre and more than 9 millimetre (Beatty & Lucero-Wagoner, 2000), we decided to exclude all values that were smaller than 1 millimetre or larger than 9 millimetre, and to replace these with a null value.

6. The data were stored so that every trial was represented by one column. The rows indicate the temporal scale and epoch-number. The data are stored so that t = 0 lies at the start of the epoch of the trial previously indicated as the 'target' epoch or 'epoch of interest'.

7. Outliers were deleted within each trial. Of every trial a mean pupil diameter and standard deviation were calculated. Every value that fell outside the range: mean pupil size ± 2.5SD was replaced by an empty value.

8. To remove physiologically impossible increases and decreases in measured pupil diameter, conservative dilation and restriction velocity values were used. Maximum dilation velocity was set at 0.7 millimetres/second and maximum constriction velocity at 3.2 millimetres/second. Isn order to determine a physiologically impossible pupil dilation or restriction an algorithm was created that compared two points of a trial. If the slope between these two points was steeper than could physiologically be expected, the point that was furthest from the mean pupil size of that trial was excluded and replaced by a null value. The distance between these points was chosen to be 250 milliseconds. In this way we were able to remove short physiologically impossible fluctuations in pupil diameter as well as relatively long (up to 500 milliseconds) periods of erroneous tracking. Rapid changes in measured pupil size were often caused by the tracking of objects other than the pupil. In these cases, the

eye-tracker was not measuring the pupil but something else of roughly the same dimensions as a pupil, e.g. a part of the participant's eye-lashes.

9. The trials that contained less than 50% of the data after the cleaning steps were excluded from analysis. In trials that contained 50% or more of the data after previous data cleaning steps, gaps were filled by means of a cubic interpolation method.

10. Of each participant, trials were deleted if the mean pupil size value over a trial lay 2.5SD from the grand mean pupil size of this participant.

11. We created plots of individual participants' pupil behaviours to be able to visualize the data and check for abnormalities. Moreover, the data were collapsed in different ways to obtain the information required for statistical analyses. Firstly, a file was created in which the tab sheets represent the different participants, the columns represent the within subject factors and the rows the sample points over time. Secondly, from these data we created a file in which all participants were collapsed to obtain information about the pupil diameters over time, separated by condition. Thirdly, over two windows (from 0 to 1000 milliseconds and from 3000 to 10000 milliseconds) within the epoch of interest (i.e. the tracking epoch), a mean pupil diameter was computed, separated by participant and condition.

12. A final outlier deletion step was applied to the data that were collapsed over time. Mean and standard deviations were computed over the values of all participants, separated by condition. Again, a cut-off threshold of 2.5 SD from the mean was chosen.

13. In the mean pupil data six cells remained empty. In order not to exclude subjects on the basis of lacking only one value, empty cells were filled by calculating the average value of the other subjects of the corresponding attentional load (number of targets).

2.5 Statistical Analyses

Statistical analyses were performed using PASW 18 (SPSS Inc., Chicago, IL, USA). Shapiro-Wilk's tests were used to test for normality of distribution. Subsequently, accuracy data was analysed by means of the non-parametric Friedman test for repeated measures. *Post hoc* analyses were performed using Wilcoxon's test for non-parametric comparison. Pupil diameter data were analysed using repeated measures ANOVAs with *post hoc* paired t-tests. Further, to determine whether mean pupil diameter changed during tracking, Bonferroni corrected paired t-tests were applied. The number of targets was the independent within subject factor, with accuracy and mean pupil diameter as dependent variables. Values are presented as mean ± standard error of the mean (SE), unless specified differently. The threshold for significance was set at $p<0.05$.

3. Results

Data of three females had to be excluded from data analysis because of missing or unreliable data due to eye-tracker calibration errors.

3.1 Normality Assumptions

Shapiro-Wilk tests were used to investigate whether the data obtained was normally distributed. Accuracy data on all number of targets appeared not to be normally distributed (all $p<.05$). Shapiro-Wilk normality tests showed that mean pupil diameter data were normally distributed for all conditions (all $p>.05$).

3.2 Accuracy

There was a statistically significant difference in proportion of accurately tracked targets, depending on attentional load, $\chi^2(3) = 83.239$, $p<.001$. *Post hoc* comparisons with Wilcoxon signed-rank tests were conducted with a Bonferroni correction applied, resulting in a significance level set at $p<.0083$. Median accuracy for tracking load of two, three, four and five objects were .98 (inter quartile range ranging from .93 to 1.00), .93 (.88 to .98), .90 (.80 to .96), and .83 (.79 to .92), respectively. *Post hoc*, a significant difference in accuracy between all number of targets conditions could be observed (all $Z< -3.643$, $p<0.001$). Figure 4 shows accuracy as a function of number of targets. Note that chance levels of guessing correctly were .20, .30, .40, and .50, respectively.

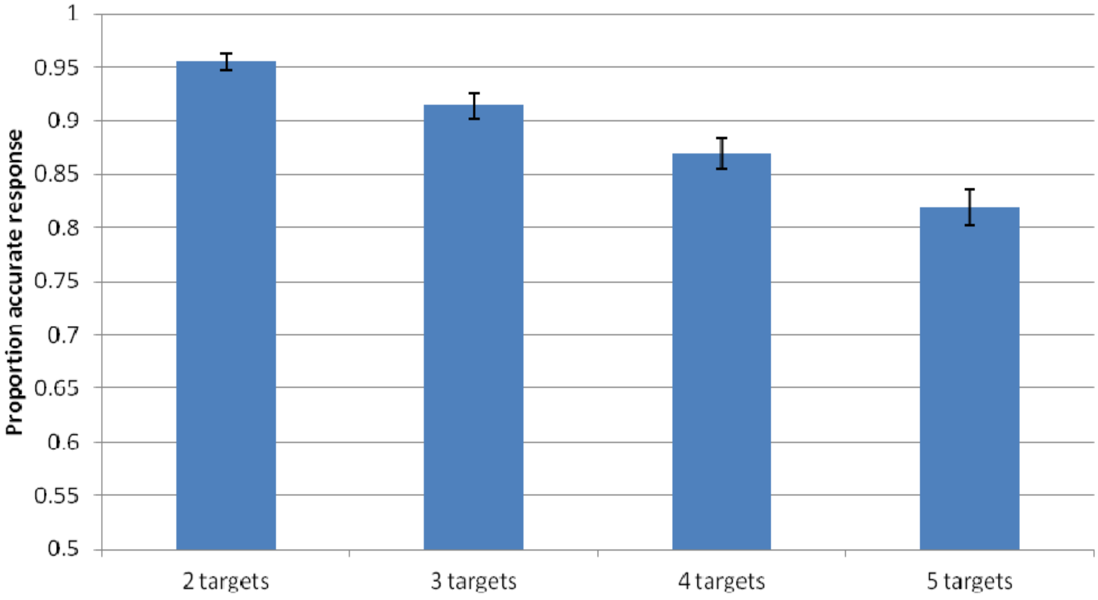

Figure 4. Propotion accurately tracked target objects displayed per total number of targets. Error bars represent standard error of the mean.

3.3 Pupil Diameter

In figure 5 pupil diameter over time is displayed per number of targets. In order to quantify differences in pupil diameter between numbers of targets, mean pupil diameters were calculated over a window within the tracking epoch of a trial, ranging from 3000 milliseconds to 10000 milliseconds, as described in the method section on pupillometry data processing.

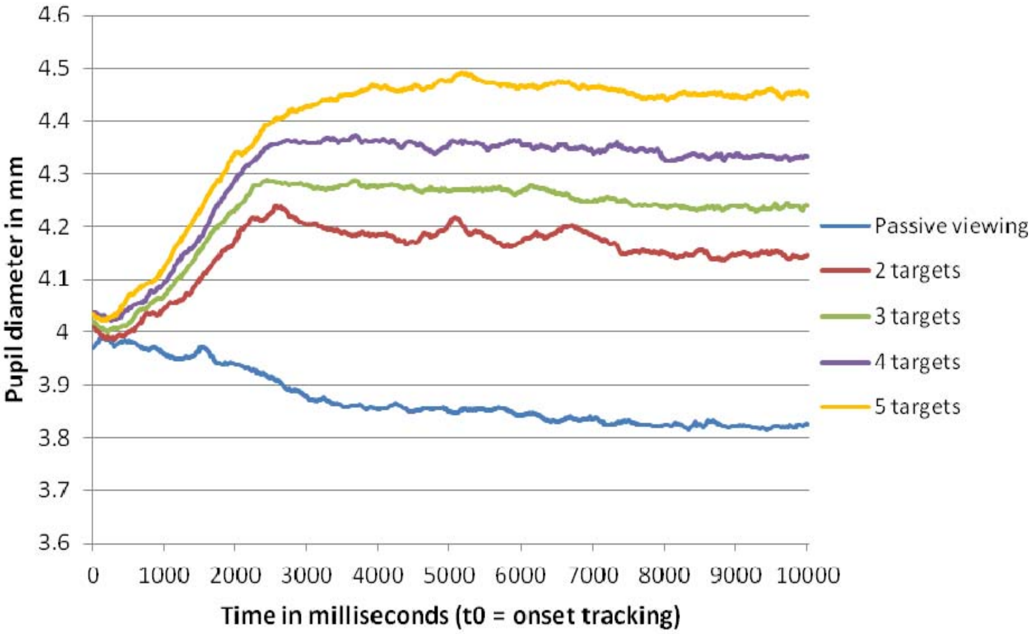

Figure 5. Pupil diameter over time displayed per number of targets

The average pupil size per number of targets is presented in figure 6. Average pupil diameters per number of targets were compared by means of repeated measures ANOVAs. Mauchly's test of sphericity yielded a significant result for mean pupil size (p<.001), indicating that sphericity could not be assumed. Hence, Huynh-Feldt corrections were applied. A Huynh-Feldt-corrected repeated measures ANOVA revealed a significant difference between number of targets in mean pupil diameter F(4, 160) = 84.736, p<.0001. *Post hoc* comparisons were performed to compare pupil diameter between the following pairs: passive viewing and tracking two targets, two targets and three targets, three targets and four targets, and four targets and five targets. Bonferroni corrected paired t-tests indicated a significant difference between pupil diameter within all pairs (all t(40)< -3.740, p<.005), with a significance level set at p=.0125 (α/4 comparisons).

To determine whether pupil diameter changed significantly while tracking, mean pupil diamter of a window (0 to 1000 milliseconds) at the beginning of the trial was compared to a window (3000 to 10000 milliseconds) after the peak (or minimum) pupil diameter had been reached. Bonferroni corrected paired t-tests yielded a significant decrease in pupil size in the passive viewing condition ($t(40) = 6.302$, $p<.0001$. Further, the increase in pupil diameter was significant in the three, four, and five target conditions (all $p>.005$).

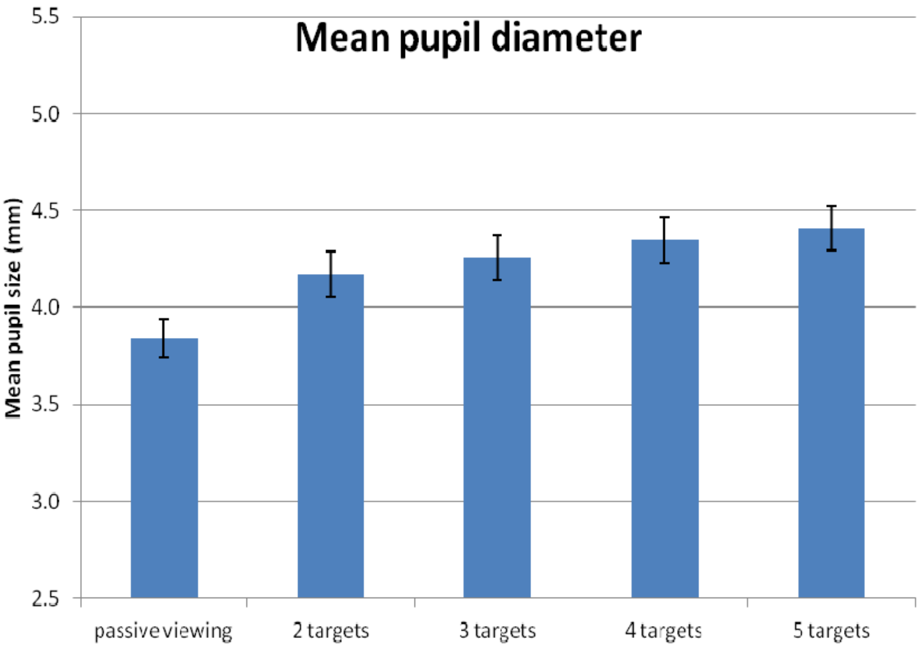

Figure 6. Mean pupil diameter displayed per number of targets, computed over a window from 3000 to 10000 milliseconds from start of tracking. Error bars represent standard error of the mean.

4. Discussion

The current study was aimed to investigate pupil responses depending on attentional effort in MOT. Comparisons of mean pupil diameter, computed over a window starting 3000 milliseconds after onset of tracking and ending after 10000 milliseconds, yielded load-dependent differences, in which tracking more objects was associated with a larger pupil. Differences between all pairs included in the *post hoc* comparisons of different attentional loads on mean pupil diameter were detected, indicating that the effect was present with every increase of attentional load. Importantly, accuracy was significantly lower when tracking more objects than when tracking fewer objects, hereby confirming that more attentional effort is needed when tracking more objects. The information on the screen was the same in all conditions, so the observed effects can be attributed purely to cognitive processes involved in the task at hand and thus reflect effort rather than other possible causes for pupil dilation. It is important to note that the effects found in this study reflect effort and not merely an increase in arousal, for arousal may be a confounding factor in investigating effort (Kahneman, 1973).

As can be observed in figure 5 pupil responses were relatively sluggish. Pupils reached their peak diameter approximately 2500 milliseconds after start of tracking and stayed around their maximum size during the entire length of the tracking period. The peak at 2500 milliseconds is relatively late compared to the latency of the cognitively induced peak in the Stroop task at about 1400 milliseconds after stimulus onset (Laeng, et al., 2011), and very late compared to the peak around 300 milliseconds after response in the posner cuing task (Gabay, et al., 2011). The delay in pupil peak in the current MOT task compared to the Stroop task and the Posner cuing task may be explained by the nature of the task at hand; in these paradigms a stimulus was presented briefly, requiring brief attention and was followed by a response, whereas in the MOT task sustained attention, and thus sustained effort, was required. In the current study, the maximum average increase in pupil size compared to pupil size at beginning of the trial was approximately 0.5 millimetres, compared to a maximum mean increase of approximately 0.16 millimetres in the Stroop task. The delay in maximum pupil diameter in the current MOT task compared to the Stroop task may be caused by a sustained demand of mental effort, leading to prolonged pupil dilation, resulting in later and larger maximum pupil dilation.

In the tracking conditions, we found very robust effects of pupil dilation depending on attentional load. An increase in attentional load was accompanied by a significant increase of pupil diameter on all levels of load. In the passive viewing condition a decrease in pupil size could be observed, suggesting that no additional effort was required in that condition. These results suggest that more mental effort is required when following more objects, while there is evidence for a decrease in demand for effort when passively observing a scene in the decrease in pupil diameter in the passive viewing condition. In line with previous research (Drew, et al., 2009), the current results further indicate that the strategy that is applied to track multiple objects depends predominantly on target enhancement rather than distracter suppression. If the main strategy of following multiple objects would be to suppress distracters, we would have expected an increase in pupil size when more distracters are present (i.e. when fewer targets are to be tracked). In that case the biggest attentional effort would be required in the condition in which two objects were to be tracked and eight objects were to be suppressed, and the least effort would be required in the condition in which five objects were to be tracked and five objects were to be suppressed. However, we found the opposite result; most effort is required in the five-target condition, which indicates that tracking does not depend on distracter suppression. Furthermore, if multiple object tracking would depend purely on preattentive mechanisms, with a maximum capacity of five objects, we would have expected an equal amount of effort needed in the conditions in which two, three and four objects were to be tracked. We would expect this, because no additional effort would be required if the capacity limit is not met. Since we did not observe an equal increase in pupil size in the conditions with fewer targets, we may conclude that tracking is not purely based upon preattentive mechanisms without the involvement of effort.

The current experiment could not determine whether a preattentive mechanism is in place or not. The preattentive mechanism is thought to be automatic and should thus not require effort. The pupillometry data of the epoch in which targets were indicated and the preattentive mechanism should first be active were not analyzed. However, if we would have analyzed these data as well, it would have remained unclear whether a possible effect could be assigned to additional strain on the preattentive mechanism or if a larger pupil size with a larger number of targets merely reflects an increase in arousal. Interestingly, a study aiming to investigate certain proclaimed preattentive mechanisms discovered that any mechanism, however elementary, experiences interference from another secondary task (Joseph, Chun, & Nakayama, 1997). In this particular study, it was found that performance of a supposed preattentive mechanism that is involved in orientation detection is affected by a letter identification task. This shows that an elementary mechanism as orientation does not operate without some form of attention, and can thus not be considered cognitively impenetrable.

Arguably, a more complex mechanism responsible for the parallel application of FINST onto target objects or perceptual grouping could thus also not operate purely preattentively.

4.1 Effort in Models for Multiple Object Tracking

If pupil dilation is indeed a measure of cognitive effort as suggested by Kahneman (1973) we can conclude that increase in the number of targets indeed demands an increase in effort. The results of the current study require different interpretations in the different suggested models suggested to describe MOT, for the proposed models differ in the way they explain the application of mental effort in MOT.

Pylyshyn's FINST model (Pylyshyn & Storm, 1988) suggests that effortin MOT is applied to maintain the 'sticky' indexes onto the target objects. This would lead us to predict an increase in effort depending on the number of targets, since more effort would be needed in maintaining more indexes. The results of the current study are in line with this notion, for indeed more effort is thought to be required with more targets, as indicated by a larger pupil size in conditions with more targets.

A purely attentional model of MOT as proposed by Scholl (2009) would expect a proportional increase in attentional effort depending on the number of targets to track. This is indeed what was found in the current study. More demand on attentional mechanisms would indeed require more effort, for a larger strain is applied to attentional processes. The present results are thus in line with this model of MOT.

Yantis' perceptual grouping model (Yantis, 1992) does not explicitly suggest a role for effort, but effort may come in play during tracking when one is keeping track of the virtual polygon. According to this model, the grouping of the targets to a single virtual object is considered an automatic process. The subsequent tracking of the object should theoretically not require additional effort, for this object is thought only to use a single attentional channel. The size of this attentional channel is larger than it is in other models, for the size of the objects no longer determine the size of the area that is to be attended, but the position of the objects in relation to each other determines this size. The results of the present study may thus be more difficult to explain within this model, for an increase in effort was observed depending on the number of targets. However, possibly the increase in effort was caused by the shape of the polygon which is dependent on the number of targets. A more dynamic polygon could in that way demand more effort to track. The current study was thus not able to unambiguously support or discard this model.

Finally, a purely serial account for MOT would not assume an increase in effort when tracking more objects, for objects are attended in a serial fashion, with cognitive processes handling only a single object at a time. Because no extra information is processed at any given moment, this model does not assume an increase in effort when one is tracking more objects. In the light of this model, the current findings may be difficult to explain.

In summary, the current data are in line with several models for MOT, whereas in other models these data are not expected. The data show an increase in pupil size which depends on the number of targets that are tracked. This indicates an increase in effort in the task with an increase in load. The data are in contrast with Yantis' perceptual grouping account of MOT, for there should be no increase in demand for effort when grouping because this is thought to occur automatically. When tracking, no additional demand for effort should arise, because only one (virtual) object is tracked. But it must be noted that the current data can not provide evidence against this model conclusively, for the role of effort in this model of MOT is undefined and unclear. Further, the data are in contrast with a purely serial account for MOT, because no additional strain on effort should be expected in this model, since information is processed simultaneously and no additional strain is expected at any given moment. Pylyshyn's early vision model expects the effect found, because more 'sticky' indexes are to be maintained onto the targets. A purely attentional account of MOT, finally, also assumes an increase in effort in MOT with an increase in load, because more attentional processes would experience additional strain. Note that even though the models that expect the observed effect, assign a different underlying application for effort. In Pylyshyn's early vision model effort is thought to maintain 'sticky' indexes, whereas in a purely attentional account of MOT effort is required for attentional processes.

4.2 Neural Correlates of Multiple Object Tracking

The first paper of a study that was aimed to identify brain areas that are involved in tracking multiple objects was published in 1998 (Culham, et al., 1998). Culham and colleagues used functional magnetic resonance imaging (fMRI) to distinguish between areas that are involved in attentional tracking and areas involved in viewing the objects passively. To this end twenty-two participants were included to perform the MOT task. The study showed that during MOT, asides from several mostly occipital areas involved in visual perception, certain attentive-tracking related areas were activated. Bilaterally the frontal cortex, parietal cortex and the MT complex appeared to be activated when comparing attentive tracking to passive viewing. Specifically, of the frontal cortex the FEF and precentral sulcus appeared to be activated. Of the parietal cortex, the intraparietal sulcus (IPS), postcentral sulcus, superior parietal lobule (SPL), and precuneus demonstrated more activity during tracking than viewing. The MT complex (an area directly involved in motion perception) showed only a mild increase in activity between attentive tracking and passive viewing, and in some subjects this increase was lacking altogether. Together these results imply that a frontoparietal network is activated when one is engaged in multiple object tracking.

Two independent human fMRI studies on MOT replicated the above mentioned findings (Culham, Cavanagh, & Kanwisher, 2001; Jovicich, et al., 2001). Moreover, a distinction was made between areas that are involved in perception of the task-elements, areas that are involved in tracking, and task-supporting areas. Slightly different approaches were used in the two studies. Participants in the Jovovich et al (2001) study could for instance move their eyes freely, whereas participants in the Culham et al. (2001) study were instructed to focus their eyes on a fixation point. Also, slightly different parametric fMRI analysis methods were applied. However the slight differences between the two studies, they yielded comparable results. In short, areas that demonstrated activation that was proportional to the amount of objects tracked were considered areas that are directly involved in MOT, whereas areas that are activated in a more or less dichotomous manner (i.e. either showing activation or not showing activation, independent of attentional load) are thought to be involved in task supporting processes that may be independent of attentional processes (e.g. suppression of eye movements). Regions that are associated with task-supporting functions, because activation in these region reflected a task-only function, more so than a load-dependent function, included the FEF, SPL, and the larger part of the MT+ complex. Activation in the FEF is suggested to be a reflection of (suppression of) eye movements. Two hypotheses are put forward by Culham et al (2001) to explain activation in the FEF. Eye movement may be constantly planned to move the subject’s gaze to the optimal place to observe all objects. The second hypothesis states that, in addition to planning eye movements, the FEF may also play a role in attention. The SPL can also play a role in eye movement planning or in attentional mechanisms. However their real functions, it is unlikely that these areas are involved tracking of multiple objects directly. Areas that both studies consider areas in which the activation associated with load was larger than activation for task-supporting functions most notably included the precentral sulcus, the IPS, and a region at the junction of the transverse occipital sulcus. Several other areas were found in both studies, but due to different applied methods, different regions are identified to be involved. The regions that showed load-dependent activation are thought be involved in MOT directly. Several of these areas have indeed been suggested to be involved in (visual) working memory and attention. It is important to note that the SPL and FEF, which show a task-only function in MOT, have been suggested to belong to the dorsal attention network which is thought to be involved in top-down processes controlling attention (Corbetta, et al., 2008).

Combined, these neuroimaging studies support the notion that two independent attentional networks are involved in MOT. The two attentional networks are responsible for attentional processes and can be distinguished in a dorsal and ventral attentional system. Important is that the LC-NE system is thought to play an active role in mediating cognitive performance through projection to these networks. The pupillometry data of the current study are thus in line with neuroimaging data on MOT, and provide information as to how activation of the LC-NE system is related to the two attentional networks.

4.3 Attentional Networks and Activation of the Locus Coeruleus-Norepinephrine System

Pupillometry allows us to investigate the role of the LC-NE system and its relation to different attentional systems. According to Posner and Fan (2008) a distinction can be made between three systems regulating attention, viz. the alerting, orienting and executive network. These systems can be distinguished on basis of their anatomy and pharmacology (Marrocco & Davidson, 1998). Manipulating the influence of the respective neurotransmitters reveals a double dissociation between the systems (i.e. manipulating one neurotransmitter system has an effect on one system, while it has little or no effect on the functionality of another). Executive control over the orienting and alerting systems is carried out by a network that is controlled by dopamine, consisting notably of the ACC (which is thought to regulate LC activity), the lateral ventral cortex, the prefrontal cortex and areas of the basal ganglia. The alerting network is innervated by the NE system, and is comprised of the LC, right frontal cortex and regions of the parietal cortex. The orienting network is regulated through acetylcholine (ACh) and is thought to involve the superior parietal cortex,TPJ, FEF and the superior colliculus. The predominantly frontal executive network is involved in attention in case of conflict. When for instance subjects are required to name the ink colour in the Stroop task, which carries a conflict between colour and colour-name, frontal areas appear to be activated, whereas less activation is observed when they are asked to merely read the text of the presented words (Bush, Luu, & Posner, 2000). Specifically, medial frontal areas, such as the anterior cingulate cortex, play a role in conflict detection, while working memory is carried out by more lateral prefrontal regions.

Also in functionality a distinction can be made between the three attentional systems (Fan, McCandliss, Sommer, Raz, & Posner, 2002). In this study a cuing approach embedded in the attention network test (ANT) is used to investigate the role of the three proposed attention systems. The ANT is best described as a combination between the cued reaction time (RT) task (Posner, 1980) and the flanker task (Eriksen & Eriksen, 1974). By means of the different components within the task a distinction can be made between three different attentional functions, corresponding to the three proposed networks (i.e. alerting, orienting, and executive control). Importantly, no correlation between performances on the three domains of attention could be observed. In other words, performance in one domain does not influence performance in another.

In a study addressing uncertainty the neuromodulators ACh and NE have been suggested to have antagonistic functions (Yu & Dayan, 2005). Uncertainty in decisions is regulated by these two neurotransmitter systems and they play complementary roles. A distinction is made between expected and unexpected uncertainty. Expected uncertainty is proposed to be the domain of ACh, whereas unexpected uncertainty is controlled by NE. An extended version of the Posner task (Posner, 1980) was used to make this distinction between expected and unexpected uncertainty. The paradigm consisted of a sequence of trials in which a number of cues (coloured arrows) was presented. Participants were to establish which of the cues corresponded to the subsequently presented target in order to anticipate the location of this target. A certain error was included in the predictability of the predictive cue. This error was known to the subjects and was thus considered an expected uncertainty. In addition, a second, unexpected, uncertainty was included in the paradigm by altering the predictive cue after a number of trials (instead of, say, the red arrow that used to be the predictor, the blue arrow would now be the predictor). Mathematical analysis of the accuracy data on this task indeed showed a distinction between two separate processes that, importantly, correspond well with existing theories on NE and ACh and their role in attention. This lead to the suggestion that ACh and NE interact closely and that “expected uncertainty, signalled by ACh, gates the effectiveness of NE in controlling representational learning”. In line with the theory proposed by Posner and Fan (Posner & Fan, 2008) this model indicates that a central executive system, mediated by ACh, is involved in attention and uncertainty, whereas the orienting system, mediated by NE, is responsible for alerting in the case of novel unexpected information.

Elaborating on this notion, Corbetta, Patel and Shulman (2008) describe how two frontoparietal networks cooperate to guide our attention. They pose that orienting to stimuli is the main task of a dorsal frontoparietal network through top-down control of sensory input, whereas a ventral attention network is responsible for reorienting of attention in the case of important novel information. The dorsal attention network most notably consists of the dorsal parietal cortex, IPS, SPL, and dorsal frontal cortex along the precentral sulcus, near or at te FEF (Desimone & Duncan, 1995). Note that tracking-related cognitive processes are associated with activation of many of these regions (Culham, et al., 2001; Jovicich, et al., 2001). The main goal of the dorsal network is to regulate sensory input in a top-down fashion to match pre-existing information in order to achieve present goals. An important observation for this notion is that the dorsal network appears to be activated in anticipation of stimuli at a certain location or with certain traits, when a specific response is prepared, or by short-term memory of a visual scene. The ventral frontoparietal network, on the other hand, is an attentional network responsible for recognizing behaviorally relevant stimuli in a bottom-up manner, and is markedly comprised of the superior temporal sulcus, superior temporal gyrus , the ventral part of the supramarginal gyrus and ventral frontal cortex, including parts of middle frontal gyrus (MFG), inferior frontal gyrus (IFG), frontal operculum, and anterior insula. If a stimulus is considered relevant enough, then the ventral attention system 'interrupts' the dorsal attention system in order to reorient attention to the relevant simulus. Interestingly, salient unimportant stimuli do not activate the ventral attention network, while less salient but more important stimuli do indeed activate the ventral network (Corbetta, et al., 2008). This shows that the ventral network does not simply respond to prominent stimuli, but has a preference to behaviorally relevant stimuli. Importantly, two studies (Shulman, et al., 2003; Todd, et al., 2005) reviewed by Corbetta et al. (2008) show that when subjects had to perform tasks in which they had to suppress task-irrelevant stimuli a decrease in activation could be observed of the TPJ and other parts of the ventral attention system (viz. MFG and IFG). Moreover, the study by Todd et al found that the right TPJ showed a deactivation that was related to working memory load. The TPJ, in other words, appears to be suppressed in a manner that is related to the load of the task at hand. Together, these studies point to the suggestion that the dorsal attention system mediates sustained attention, and is suppresses activation of he ventral attention system, which, when behaviourally relevant stimuli are presented ‘interrupts’ the dorsal attention system in order to reorient to the newly presented information.

Altogether, the above mentioned studies indicate that three anatomically separate systems control alerting, orienting and executive control. Moreover, they are functionally independent mechanisms and should thus be considered independently. When involved in MOT the dorsal attention network appears to be activated, while activation in areas, e.g. the TPJ, of the ventral attention network is suppressed.

The observed continuous dilation of the pupil may reflect sustained phasic activity of the LC-NE system, which regulates performance of the dorsal and ventral attentional networks. Phasic activity of the LC is thought to keep the ventral attention network from interrupting the activation of the dorsal network. Indeed, the adaptive gain theory suggests that phasic activation of the LC is involved in task optimization (Aston-Jones & Cohen, 2005). As described in the introduction, a previous study (Gabay, et al., 2011) that aimed to address activation of the LC-NE system through pupillometry supported the idea of Corbetta et al. (2008) that this system has two modes of activation that have different roles in attention. In that study a larger phasic pupil dilation was observed in when one performed a more demanding task than when performing an easier task which generated a tonic pupillary response. This indicates that phasic activation of the LC-NE system, which underlies the pupillary response, is larger in more demanding tasks. This idea is in line with the data of the current study, that also show an increase in phasic activation of the LC-NE system in more demanding conditions. Altogether this suggests that phasic activation of the LC-NE system optimizes task performance, through the direct link of the LC with the two frontoparietal attention networks. Furthermore, in contrast to the suggestion that the pupillary response is caused by mechanisms that initiate a response (Gabay, et al., 2011; Simpson, 1969), the constant dilation of the pupil in the current MOT study indicates that the pupil response is based upon cognitive mechanisms that are related to task-related processes and not on mechanisms that are involved in preparing a response.

4.4 Methodology

The current paradigm and apparatus allowed us to investigate relatively small changes in effort when performing a task requiring sustained attention. Additionally, it allowed us to distinguish between conditions that required more effort by directly manipulating task difficulty. As mentioned above, the MOT paradigm is unique in this way, by requiring sustained attention and by allowing for direct manipulation of the demands of the task. The current paradigm could aid to distinguish between several proposed models of MOT, by investigating the role of effort in this task. The current paradigm could, however, not determine the application of effort, whether it is involved in the maintenance of ‘sticky fingers’ onto targets, or if it applies purely to attentional processes. Hence, a suggestion is made for a modification of the current paradigm, in order to further distinguish between the different models and the application of effort in the MOT task.

4.5 Future Perspectives

To further investigate the different proposed models for MOT, modifications could be made to the current paradigm. As explained above, the applications of effort in the suggested models differ. Pylyshyn’s early vision model assumes a role for effort in maintaining the indexes onto the targets while a purely attentional account of MOT assumes effort for attentional processes. Yantis’ perceptual grouping model does not explicitly state a role for effort, but effort may be assumed to be involved in keeping track of the virtual polygon. Finally, a purely serial account of MOT may assume a role of effort in attending every object in turn. The current paradigm aided to distinguish between paradigms that assume an increase in effort with an increase in number of targets and paradigms that do not assume such an increase. The models that assume such an increase (i.e. Pylyshyn’s early-vision model, a purely attentional account of MOT, and potentially, however to a lesser degree, Yantis’ perceptual grouping model) can be further distinguished through manipulation of the size of the objects. Manipulating the size of the objects should have an effect on required effort in certain models, while it should not have an effect in other models. In Pylyshyn’s early vision model, decreasing the size of the objects should have an effect on effort, for it may be assumed that maintaining ‘sticky’ indexes onto smaller objects is more demanding than maintaining them on larger objects. Effort Yantis’ perceptual grouping model, on the other hand, should not increase when decreasing the size of objects. Targets are though to be grouped together automatically, after which we merely keep track of the virtual polygon. Small objects should serve as corner object of this virtual polygon equally well as large objects. A future study could thus use object of different sizes and compare whether more effort is required to track smaller targets.

Further, to confirm that mechanisms involved in tracking are indeed the causal factor of pupil dilation and not the preparation of a response, it may be interesting to investigate whether aborting tracking results in a decrease in pupil size. In a future study participants could be instructed to abandon tracking at a certain point in the task. Alternatively, participants could be asked to indicate when they no longer succeed to follow the target objects by pressing a button, although this latter method is prone to more variability and possible errors. Both ideas may indicate that, after being dilated during tracking, pupils decrease in size, for after abandoning the task, participants are merely passively observing the display, without engaging in an effort-demanding task. This may indicate that no effort is invested in the task after the person stops tracking, which would confirm that effort is indeed the factor that is related to pupil dilation.

To further investigate the relation of the LC-NE system and the two frontoparietal attention networks, fMRI studies may be combined while simultaneously obtaining pupillometry data. Activity of the LC is difficult to assess by means of fMRI, but efforts are underway to simplify *in vivo* measurement of LC activity (Keren, Lozar, Harris, Morgan, & Eckert, 2009). Combining pupillometry and fMRI may aid us to investigate whether activation of the LC in humans also is related to pupil dilation. This in turn can then aid us to understand how the LC is involved in attention.

5. References

Ahern, S., & Beatty, J. (1979). Pupillary responses during information processing vary with Scholastic Aptitude Test scores. *Science, 205*(4412), 1289-1292.

Alvarez, G. A., & Cavanagh, P. (2005). Independent resources for attentional tracking in the left and right visual hemifields. *Psychol Sci, 16*(8), 637-643.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci, 28*, 403-450.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary & B. G.G. (Eds.), *Handbook of Psychophysiology* (2 ed., pp. 142 - 162). Cambridge Cambridge University Press.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychol Rev, 64*(3), 205-215.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn Sci, 4*(6), 215-222.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280*(5364), 747-749.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends Cogn Sci, 9*(7), 349-354.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. H. (1998). Cortical fMRI Activation Produced by Attentive Tracking of Moving Targets. *Journal of Neurophysiology, 80*(5), 2657-2670.

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron, 32*(4), 737-745.

Dayan, P., & Yu, A. J. (2006). Phasic norepinephrine: a neural interrupt signal for unexpected events. *Network, 17*(4), 335-350.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18*, 193-222.

Deutsch, J. A., & Deutsch, D. (1963). Some theoretical considerations. *Psychol Rev, 70*, 80-90.

Drew, T., McCollough, A. W., Horowitz, T. S., & Vogel, E. K. (2009). Attentional enhancement during multiple-object tracking. *Psychon Bull Rev, 16*(2), 411-417.

Eriksen, B., & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics, 16*(1), 143-149.

Espeseth, T., Sneve, M. H., Rootwelt, H., & Laeng, B. (2010). Nicotinic receptor gene CHRNA4 interacts with processing load in attention. *PLoS One, 5*(12), e14407.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *J Cogn Neurosci, 14*(3), 340-347.

Fehd, H. M., & Seiffert, A. E. (2008). Eye movements during multiple object tracking: where do participants look? *Cognition, 108*(1), 201-209.

Franconeri, S. L., Jonathan, S. V., & Scimeca, J. M. (2010). Tracking multiple objects is limited only by object spacing, not by speed, time, or capacity. *Psychol Sci, 21*(7), 920-925.

Gabay, S., Pertzov, Y., & Henik, A. (2011). Orienting of attention, pupil size, and the norepinephrine system. *Atten Percept Psychophys, 73*(1), 123-129.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science, 143*(3611), 1190-1192.

Howe, P., Cohen, M., Pinto, Y., & Horowitz, T. (2009). Distinguishing between parallel and serial accounts of multiple object tracking. *Journal of Vision, 9*(8), 239.

James, W. (1890). *The Principles of Psychology* (Vol. 1). New York: Holt.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997). Attentional requirements in a 'preattentive' feature search task. *Nature, 387*(6635), 805-807.

Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., & Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *J Cogn Neurosci, 13*(8), 1048-1058.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(756), 1583-1585.

Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., & Eckert, M. A. (2009). In vivo mapping of the human locus coeruleus. *Neuroimage, 47*(4), 1261-1267.

Laeng, B., Gredebäck, G., & Silvois, S. (in press). Pupillometry: A window to the pre-conscious? *Perspectives on Psychological Science*.

Laeng, B., Orbo, M., Holmlund, T., & Miozzo, M. (2011). Pupillary Stroop effects. *Cogn Process, 12*(1), 13-21.

Liu, G., Austen, E. L., Booth, K. S., Fisher, B. D., Argue, R., Rempel, M. I., et al. (2005). Multiple-object tracking is based on scene, not retinal, coordinates. *J Exp Psychol Hum Percept Perform, 31*(2), 235-247.

Mair, R., Zhang, Y., Bailey, K., & Toupin, M. (2005). Effects of clonidine in the locus coeruleus on prefrontal- and hippocampal-dependent measures of attention and memory in the rat. *Psychopharmacology, 181*(2), 280-288.

Marrocco, R. T., & Davidson, M. C. (1998). Neurochemistry of Attention. In R. Parasuraman (Ed.), *The attentive brain* (Vol. xii, pp. 35 - 50). Cambridge, MA: The MIT Press.

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol, 32*(1), 3-25.

Posner, M. I., & Fan, J. (2008). Attention as an organ system. In J. R. Pomerantz (Ed.), *Topics in integrative neuroscience: from cells to cognition* (1 ed., pp. 31-61). Cambridge, UK: Cambridge University Press.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*(1-3), 363-384.

Pylyshyn, Z. W. (2003). *Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities*.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat Vis, 3*(3), 179-197.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). *Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention*. Paper presented at the Society for Neuroscience.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cereb Cortex, 10*(3), 284-294.

Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nat Rev Neurosci, 10*(3), 211-223.

Scholl, B. J. (2009). What have we learned about attention from multiple-object tracking (and vice versa)? In D. Dedrick & L. Trick (Eds.), *Computation, Cognition, and Pylyshyn* (pp. 49 - 77): Massachusetts Insitute of Technology press

Shulman, G. L., McAvoy, M. P., Cowan, M. C., Astafiev, S. V., Tansy, A. P., d'Avossa, G., et al. (2003). Quantitative analysis of attention and detection signals during visual search. *J Neurophysiol, 90*(5), 3384-3397.

Simpson, H. M. (1969). Effects of a task-relevant response on pupil size. *Psychophysiology, 6*(2), 115-121.

Todd, J. J., Fougnie, D., & Marois, R. (2005). Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. *Psychol Sci, 16*(12), 965-972.

Tombu, M., & Seiffert, A. E. (2008). Attentional costs in multiple-object tracking. *Cognition, 108*(1), 1-25.

Yantis, S. (1992). Multielement visual tracking: attention and perceptual organization. *Cogn Psychol, 24*(3), 295-340.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*(4), 681-692.