EYE MOVEMENTS AND SCANPATHS
IN THE PERCEPTION OF REAL-WORLD SCENES

Katherine Anne Humphrey, BSc.

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

Submitted August 2009
Approved Febuary 2010

ABSTRACT

The way we move our eyes when viewing a scene is not random, but is influenced by both bottom-up (low-level), and top-down (cognitive) factors. This Thesis investigates not only what these influences are and how they effect eye movements, but more importantly how they interact with each other to guide visual perception of real-world scenes.

Experiments 1 and 2 show that the sequences of fixations and saccades - 'scanpaths' - generated when encoding a picture are replicated both during imagery and at recognition. Higher scanpath similarities at recognition suggest that low-level visual information plays an important role in guiding eye movements, yet the above-chance similarities at imagery argue against a *purely* bottom-up explanation and imply a link between eye movements and visual memory. This conclusion is supported by increased scanpath similarities when previously seen pictures are described from memory (experiment 3). When visual information is available, areas of high visual saliency attract attention and are fixated sooner than less salient regions. This effect, however, is reliably reduced when viewers possess top-down knowledge about the scene in the form of domain proficiency (experiments 4-6). Enhanced memory, as well as higher scanpath similarity, for domain-specific pictures exists at recognition, and in the absence of visual information when previously seen pictures are described from memory, but not when simply imagined (experiment 6). As well as the cognitive override of bottom-up saliency, domain knowledge also moderates the influence of top-down incongruence during scene perception (experiment 7). Object-intrinsic oddities are less likely to be fixated when participants view pictures containing other domain-relevant semantic information. The finding that viewers fixate the most informative parts of a

scene was extended to investigate the presence of social (people) and emotional information, both of which were found to enhance recognition memory (experiments 8 and 9). However, the lack of relationship between string similarity and accuracy, when viewing 'people' pictures, challenges the idea that the reproduction of eye movements *alone* is enough to create this memory advantage (experiment 8). It is therefore likely that the semantically informative parts of a scene play a large role in guiding eye movements and enhancing memory for a scene. The processing of emotional features occurs at a very early stage of perception (even when they are still in the parafoveal), but once fixated only emotionally negative (not positive) features hold attention (experiment 9). The presence of these emotionally negative features also reliably decreases the influence of saliency on eye movements. Lastly, experiment 10 illustrates that although the fixation sequence *is* important for recognition memory, the influence of visually salient and semantically relevant parafoveal cues in real-world scenes decreases the necessity to fixate in the same order.

These experiments combine to conclude that eye movements are neither influenced by purely top-down nor bottom-up factors, but instead a combination of both, which interact to guide attention to the most relevant parts of the picture.

PUBLICATIONS

Parts of this Thesis have been published as the following articles:

Chapter 2:
Humphrey, K. & Underwood, G., (2008). Fixation sequences in imagery and in recognition during the processing of pictures of real-world scenes. *Journal of Eye Movement Research, 2(2:3)*, 1-15.

Underwood, G., Humphrey, K. & Foulsham, T., (2008). Knowledge-based patterns of remembering: Eye movement scanpaths reflect domain experience. Lecture Notes in Computer Science, 5298, 125-144.

Chapter 3:
Humphrey, K. & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. *British Journal of Psychology, 100(2),* 377-398

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scanpaths in the inspection of real-world scenes: eye movements during encoding and recognition. *Visual Cognition, 17(6),* 812-834.

Underwood, G., Mennie, N., Humphrey, K., & Underwood, J. (2008). Remembering Pictures of Real-World Images Using Eye Fixation Sequences in Imagery and in Recognition. *Lecture Notes in Computer Science, 5329,* 51–64.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Chapter 1: Introduction

When inspecting visual scenes, eye movements are not just a set of random fixations, but occur in a specific pattern, guided by bottom-up visual features and existing top-down knowledge. Studying eye movements is important to understand how the eyes are controlled during the acquisition of information, and how this information is represented and stored (Henderson and Hollingworth, 1998). The focus of this Thesis is to examine eye movements during scene viewing and investigate how these are affected by bottom-up and top-down influences.

The fovea is part of the eye that is located in the centre of the macula region of the retina and corresponds to approximately the central 2º of the viewed scene, and is where visual information of the highest quality is acquired. Due to this, eye movements are made to direct the high resolving power of this part of the eye to different regions of the scene being viewed, to improve discrimination ability of a target location in the periphery (Deubel & Schneider, 1996; Kowler, Anderson, Dosher, & Blaser, 1995; Schneider and Deubel, 2002). Fixations are when gaze is relatively stationary in space and indicate the area where attention is being allocated (Buswell 1935). Saccades are the quick jumps of the eye from area to area, during which vision is essentially suppressed (Hering 1879; Latour, 1962). Together, fixations and saccades produce sequences of eye movements called 'scanpaths', which are a useful, objective, way to measure the allocation of overt attention in real-world scenes. However, it must also be noted that one can

attend to details outside the area of fixation (referred to as covert attention), and it has been found that pre-saccadic perceptual facilitation does not rely on the same processes as those of covert attention (Blangero et al., 2009). This is demonstrated in chapters three (experiment 7) and four (experiment 9) where first fixations are more likely to be on semantically interesting parts of the picture, suggesting that attention was already allocated to these areas before the overt movement of the eyes. Therefore, although scanpaths help identify where an individual is looking (which remains the main focus of this thesis), they may not be a complete representation of the information processing that occurs during scene viewing. Scanpath analyses do, however, allow us to measure how similar eye movements are over multiple viewings and over multiple participants. Similarity in scanpaths (or at least in fixated regions) between participants would suggest common areas of interest, whether it be due a low-level visual saliency (see section 1.2) or shared top-down knowledge about the task or the scene. Investigation of how these and other factors interact to influence eye movements will help to answer the question of *why* certain parts of a scene are fixated over others. Firstly though, this chapter will introduce and discuss some key theories of eye movements that will be referred to throughout this Thesis.

1.1. Scanpath Theory

Scanpath Theory was proposed by Noton and Stark in 1971 and predicts that the fixations made when first looking at a picture are very similar to those made when recognising that picture at a later time. According to this theory, when a person encodes an image, the eye movements they make are stored in memory as a spatial model. When they see the same picture again at a later time, this pattern of fixations and saccades is re-enacted to facilitate recognition of the picture (Noton and Stark, 1971a, b). It was originally suggested that the internal spatial-cognitive model was directly related to the oculomotor movements and neural mechanisms in the brain; however these assumptions are largely unsupported and have attracted criticism (e.g. Henderson, 2003). A problem with Noton and Stark's Scanpath Theory is that it is maintains that scanpaths are generated almost entirely top-down and does not take into account bottom up influences such as low-level visual saliency. It also struggles to account for the variability in scanpaths, both across viewings of the same observer, and between multiple observers (Groner et al., 1984; Mannan, Ruddock, & Wooding, 1997). Lastly, studies have shown that scenes can be identified with a single fixation (Biederman et al., 1982; Schyns and Oliva, 1994; Potter, 1999; Thorpe et al, 1996), challenging Scanpath Theory's assumption that recognition is reliant on replication of the eye movement sequence produced at encoding. Therefore, allusion to scanpaths in this Thesis will refer only to the measurable sequences of fixations and saccades, not the underlying neural mechanisms or the purely top-down stance argued by Scanpath Theory.

A number of studies have found that when participants view a picture for the second time, the scan patterns they produce are very similar to scan patterns produced on first exposure to the picture (Walker-Smith et al., 1977; Stark and Ellis, 1981; Foulsham and Underwood, 2008). For example, in Foulsham and Underwood's (2008) recognition memory study participants first inspected a set of 45 pictures. They were then shown another set of 90 pictures and were asked to decide whether they had seen each picture before. It was found that scan patterns were most similar when compared between two viewings of the same picture (encoding *vs.* old). This similarity was significantly greater than control comparisons (encoding *vs.* new and old *vs.* new). However, there is still some debate over Scanpath Theory, as some studies have failed to find a similarity effect (e.g. Mannan et al., 1997). A main aim of this Thesis is to investigate whether

scanpaths are replicable (at least to some extent) over multiple stimulus viewings and also after varying lengths of time (cf. sections 3.2 and 3.3). Furthermore, if there is a link between memory and eye movements, as suggested by Scanpath Theory, then it is plausible that similar scanpaths would be produced when remembering a picture, even when no visual cues are present. However, presuming that eye movements are influenced by bottom-up as well as top-down factors, this lack of low-level visual information may decrease scanpath similarities, (this possibility is explored in chapter 2 and section 3.4).

Reproducing similar eye movements at test as at encoding has been linked to improvements in recognition memory. For example, outpatients with Schizophrenia were shown to have impaired visual short term working memory, but if scanpaths were replicated at test, this impairment was significantly reduced (*Cocchi, Bosisio, Berchtold, Orita, Debbane, Wood and Schenk, 2009). Memory for visual scenes has been found to be consistently good, even over extended periods of time (e.g.* Hollingworth, 2005), which may be related to the replication of scanpaths. However, eye movements during recognition memory tests have not always been measured, and there is also the question of whether the scanpath similarity is responsible for the memory advantage or a consequence of it. Either way, It h*as been suggested that* scene memory is important in real-world processes such as visual search (Brockmole & Henderson, 2006b; Chun & Jiang, 1998; Henderson, Weeks, & Hollingworth, 1999), as almost all real-world tasks require some degree of visual search, for example looking for a knife when making a sandwich (Hayhoe, 2000; Hayhoe et al., 2003).

As well as *how* we move our eyes, it is important to consider why certain locations are fixated in the first place – a question not really answered by Scanpath Theory. Findlay and Walker (1999) proposed a model of saccade generation with five different levels of processing that incorporates bottom-up and top-down influences on eye movements (see appendix 1 for a diagram). The model has two parallel information and command streams, which run vertically through a hierarchy of processing levels. Information about where a fixation falls is transmitted in spatially mapped pathways, whereas information regarding when a saccade is initiated involves a single nonspatial signal. At level 2 of the model there exists a 'fixate centre' and a 'move centre' which ultimately control when and where the eyes move. Competitive interaction occurs when two centres, or two regions within a centre, are cross-connected with reciprocal inhibitory links. Saccades are produced when activity in the fixate centre decreases, or activity in the move centre increases – this is referred to as 'disengagement'. It has been suggested (Henderson, 1992, Rayner, 1998) that *when* the eyes move is based on the processing of the currently fixated object. The majority of this Thesis will focus on the 'move centre' of the model, i.e. the factors that influence *where* the eyes move. At level 2, the 'move centre' can be influenced by bottom-up low-level visual features of the scene – Findlay and Walker refer to this as the Salience Map. Saliency is an item's quality of being visually distinctive relative to its neighbouring items and has been shown to affect the order and pattern of fixation. Saccades are made to the area with the highest salience (determined by 'peaks' of activation in the map), in a winner-takes-all situation. An inhibition of return mechanism prevents constant refixation of the same area and directs saccades to the next most salient area. This attraction to areas of high visual saliency occurs automatically, without conscious awareness. At level 5, on the other hand, the 'move centre' can be influenced by top-down cognitive factors such as task instructions and domain knowledge (see chapter 3), by either suppressing a saccade or moving the eyes voluntarily. This location of attention to parts of the picture that are semantically relevant addresses the question of not only *where* the eyes move, but also *why* the viewer fixated that part of the picture.

1.2. The effect of saliency on scanpaths

Similar to Findlay and Walker's Salience Map, Koch and Ullman (1985) and Itti and Koch (2000) proposed a 'Saliency Map' whereby attention is drawn to the most salient region in an image first, followed by the second most salient region then the third most salient region, and so on. Attention and eye fixations are attracted to the region identified by the map as being of greatest brightness, colour contrast and orientation change, and once that region is fixated a process of inhibition of return prevents attention from being locked onto any one region, and allows us to saccade to the next most salient region. It is suggested that this winner-takes-all situation is caused by the firing of an integrate-and-fire cell, which generates a sequence of action potentials, causing attention to move to the 'winning' location. Neurological support for this theory comes from spiking patterns of single neurons in response to visual stimuli (Bichot, and Schall, 1999; Bisley, and Goldberg, 2003; Fecteau, et al., 2004; Fecteau, and Munoz, 2003, 2005; Goldberg, et al. 2002; Gottlieb, 2002; Li, 2002; Mazer, and Gallant, 2003; McPeek, and Keller, 2002, 2004; Thompson, and Bichot, 2005).

The influence of saliency on eye movements, attention and memory has been demonstrated in a number of studies. For example, Sheth and Shimojo (2001) briefly displayed a target and then asked participants to point to its previous location. Participants estimated targets to be closer to the centre of gaze, and closer to visually salient markers in the visual display than they actually were. The locations of objects presented earlier were *remembered* falsely as being closer to salient reference frames than they really were. Salient regions attract fixations when viewers are not given an explicit purpose in looking at a picture. Parkhurst, Law, and Niebur (2002), showed viewers a range of images and recorded eye movements. Saliency strongly predicted fixation probability during the first two or three fixations, and the model performed above chance throughout each trial. In contrast to this, Tatler, Baddeley and Gilchrist (2005) found no change in the involvement of image features over time and Tatler (2007) argues that even the *correlation* between features and fixations is minimal. On the other hand, more recent work by Carmi and Itti (2006) on dynamic scenes supports Parkhurst's position.

Further support for a saliency map model of scene inspection comes from Underwood, Foulsham, van Loon, Humphreys, and Bloyce, (2006) and from Underwood and Foulsham (2006), who found that when viewers inspected the scene in preparation for a memory task, objects higher in saliency were potent in attracting early fixations. These studies of the effects of saliency could suggest that scan patterns are similar at encoding and recognition not because of an internally stored sequence of fixations, but because the same bottom-up features are present at both encoding and recognition, and therefore participants just look at the same conspicuous parts of the scene. Lastly, Cerf, Harel, Huth, Einhauser and Koch (2008) investigated whether saliency maps could be used to predict which image observers were viewing given only scanpath data. Using the data from one participant, pictures were correctly identified 82.4% of the time. However, when scanpaths of 9 participants were combined, accuracy fell to 69%, suggesting that although saliency can predict (to an extent) where participants look, there are individual differences in viewing strategies. This thesis, in part, aims to further investigate the effect of saliency on eye movements, and whether is it moderated by other factors such as domain proficiency and emotional stimuli.

1.3. Variation in scanpaths

As well as bottom-up low-level visual features and top-down domain specific knowledge (see chapter 3), scanpaths have been found to be influenced by the nature of the viewing task. For example, Buswell (1935) found that the number of fixations and fixation durations increased when participants were required to search a picture, compared to when they inspected it under free-viewing conditions. Furthermore, when participants were given a written description of the picture before viewing, the number of fixations increased from 61 to 108! A similar study was carried out by Underwood, Jebbett and Roberts (2004), but instead of a simple description of the picture, participants were given a sentence and had to verify whether it was true or false. Compared to seeing the picture first, participants viewing the image after the sentence made fewer fixations overall and these fixations were guided to the objects described in the sentence.

A classic study by Yarbus (1967) found that scanpaths during free-viewing were dramatically different than when instructions such as ''Remember the clothes worn by the people'' were given. It was concluded that fixations fell in regions that were most informative for that particular task. However, Yarbus only tested one participant and the conditions under which eye movements were recorded were not very natural or comfortable, which could have affected the results. For example, the observer's eye was anaesthetized, their eyelids taped open with heated strips of adhesive plaster and their head was constrained using chin and forehead rests. The experiment has recently been replicated by DeAngelus and Plez (in press), who recorded eye movements of 17 naïve observers using a head-free eye tracker. The results were very similar to those reported by Yarbus, with variations in scanpaths depending on task instruction. However, whereas Yarbus's observer refixated on 'informative' regions for the whole 3 minute viewing time, DeAngelus and Plez found that participants began to examine background objects about halfway through viewing. It could be that the observer in Yarbus's study felt obliged to follow the strict instructions due to the uncomfortable equipment set-up, or that having to look at the same painting for over 20 minutes caused them to consciously narrow their gaze based on the instructions. Nonetheless, the effect of task instructions on eye movements continues to be evident in more recent studies. For example, Hayhoe (2003) found that during free-viewing, irrelevant objects were fixated 48% of the time, but when participants were given specific task instructions, this fell to 16%.

For tasks that require visual search, there is some evidence from scanpaths for systematic scanning (Gilchrist and Harvey, 2006). When participants had to confirm whether a target was present in a grid-like display, more horizontal than vertical saccades were produced. This systematic component was not eliminated by disruption of the grid, suggesting that participants employ a cognitive strategy in scanpath generation during visual search. Search has also been found to be directed towards areas where targets are likely to be found and this is reflected in the narrowed focus of fixations (Castelhano, Mack and Henderson, 2009). Castelhano et al also found that when the participants were told to look at the scene in preparation for a memory test, the number of fixations increased and were more distributed than in the search task. These differences in scanpaths have been found to be consistent between participants (Hembrooke, Feusner, and Gay, 2006). Hembrook et al used a multiple sequence alignment algorithm to extract similarities among multiple scan patterns and found that when participants looked at web-pages in preparation for a memory test, saccades were shorter (indicative of reading and close inspection) than in the free-viewing condition.

This top-down effect of task relevance has been shown to override the influences of bottom-up saliency (Underwood, Foulsham, van Loon, Humphreys and Bloyce, 2006; Underwood & Foulsham, 2006; Underwood, Templeman, Lamming and Foulsham, 2008). Underwood,

Foulsham, van Loon, Humphreys and Bloyce (2006) found that saliency played a significant role in a memorisation task, attracting earlier and longer fixations than less salient objects. However, when participants were instructed to search for an object in the scene, saliency was only of secondary importance. Similarly, Einhäuser, Rutishauser, and Koch, (2008) found that during free-viewing, observers' eye-positions were immediately biased toward the high-saliency side of a picture. However, this sensory driven bias disappeared entirely when observers searched for a target embedded with equal probability to either side of the stimulus. When the target always occurred in the low-contrast side, observers' eye-positions were immediately biased towards this low-saliency side, i.e., the sensory-driven bias reversed.

As well as differences due to task, scanpaths have also been found to differ between participants depending on culture. For example, Chua, Boland, and Nisbett (2005) found that American participants fixated focal objects in a scene faster and more often than Chinese participants, who made more saccades to background objects. These differences in scanpaths could be related to differences in perceptual judgment and memory (Masuda, & Nisbett, 2001; Ji, Peng, & Nisbett, 2000; Kitayama, Duffy, Kawamura, & Larsen, 2003). For example, Masuda and Nisbett found that when asked to describe underwater scenes, Americans emphasised focal objects (large, brightly coloured, rapidly moving objects) whereas Japanese participants reported 60% more information about the background (rocks, water colour, small stationary objects). Furthermore, in a recognition memory test of previously viewed animal scenes, changing the background (but keeping the focal animal) decreased the accuracy of Japanese participants compared to American participants (Masuda, & Nisbett, 2001). It is suggested that the Japanese participants used a more holistic processing, binding information about the objects with the backgrounds, so that the unfamiliar new background adversely affected the retrieval of the familiar animal. These cultural differences in attending to objects and context of the scene have also been observed in perceptual judgment tasks (Ji, Peng, & Nisbett, 2000) and change blindness studies (Masuda and Nisbett, 2001). Simons & Rensink, (2005) suggest that the mental representations of scenes may differ with culture, (e.g. Asian participants have more detailed mental representations of scene backgrounds, whereas Westerners have more detailed mental representations of the focal objects). Another explanation is that different cultures have similar mental representations but differ in accuracy for detecting a deviation between the mental representation of the background or focal object and the current stimulus.

Lastly, scanpaths have been found to differ with age and gender. Goldstein, Woods, and Peli (2007) found that when viewing dynamic scenes, male and older subjects were more likely to look in the same place than female and younger subjects, respectively. Despite these differences, for more than half of the time, participants fixated in areas that made up less than 12% of the scene, suggesting that a common component of scanpath generation does exist between participants. The importance and practical applications of scanpath similarity are discussed in the next section.

1.4. Practical Applications of Scanpaths

In this modern, computer-driven age, where camera film and photo albums are being replaced with memory cards and online picture-sharing sites, images are more and more likely to be stored digitally. The volume and diversity of the growing research on picture databases is seen in recent reviews (Eakins and Graham, 1999; Veltcamp, and Tanase, 2002). However, there remain a

number of serious difficulties in both the entry of images into databases and their efficient retrieval that, unresolved, will continue to seriously limit their potential (Enser, 2002).

The two main methods of image retrieval at the moment are the content-based approach (CB) and the description-indexing approach (DI). DI approaches involve textual descriptors and keywords that are used in traditional query languages. An example of a DI image retrieval database is PhotoFinder, which provides a set of visual Boolean query interfaces, coupled with dynamic query and query preview features. However, there are three main difficulties with this approach, i.e. the large amount of manual effort required in developing the annotations, the subjectivity in interpretation of image contents, and inconsistency of the keyword assignments among different indexers (Niblack and Barber, 1994; Flickner, Sawhney, Niblack, Ashley, Huang, Dom, Gorkani, Hafne, Lee, Petkovic, Steele, and Yanker, 1995; Faloutsos, Flickner Niblack, Petkovic, Equitz, and Barber, 1993). As the size of image repositories increases, the keyword annotation approach becomes infeasible.

CB image retrieval (CBIR) on the other hand involves automatic analysis of low-level features such as colours or texture and promise greater automation than DI. Since its advent, CBIR has attracted great research attention, ranging from government (Jain, 1993; Jain, Pentland, and Petkovic, 1995) and industry (Flickner et al., 1995; Dowe, 1993) to universities (Pentland, Picard, & Sclaro, 1996; Huang, Mehrotra, & Ramchandran, 1996; Smith and Chang, 1997; Ma and Manjunath, 1997; Mandal, Aboulnasr and Panchanathan, 1996). Many special issues from leading journals have been dedicated to CBIR (Gudivada and Raghavan, 1995; Pentland and Picard, 1996; Narasimhalu, 1995; Schatz and Chen, 1996) and many CBIR systems, both commercial (Niblack and Barber, 1994; Flickner et al., 1995; Faloutsos, et al., 1993; Dowe, 1993) and academic (Pentland, Picard, & Sclaro, 1996; Huang, Mehrotra, & Ramchandran, 1996; Smith and Chang, 1997; Ma and Manjunath, 1997; Mandal, Aboulnasr and Panchanathan, 1996) have been developed recently.

A problem with CBIR is that there is still a degree of subjectivity involved in defining the visual characteristics of a picture and in many cases domain specific knowledge is required to do so (Amadasun and King, 1988). Even if a computer program was designed for this purpose, it would still require photo interpreters to spend a considerable amount of time in generating the spectral rules used by the expert system. Furthermore, some low-level pictorial qualities are hard to define, for example most people can recognise texture but, it is more difficult to describe (Howarth and Ruger (2004). As a solution, many CBIR systems began to integrate bottom-up visual features and top-down knowledge – the basis of 'computer centric systems'. These system first finds the 'best' representations for the visual features by mapping a high level concept (e.g. a fresh apple) to low-level features (e.g. green and round), then, during the retrieval process, the user selects the visual feature(s) that he or she is interested in. Then the system tries to find similar images to the users query. However, a problem arises when there is a gap between high level concepts and low-level features. For example, it would be difficult to map an ancient vase with sophisticated design to an equivalent representation using low level features (Rui, Huang, Ortega, and Mehrota, 1998). Furthermore, different persons, or the same person under different circumstances, may perceive the same visual content differently. One person may be more interested in an image's colour feature while another may be more interested in the texture feature. Even if both people are interested in texture, the way they perceive the similarity of texture may be quite different. Also, the features identified by the model as the 'best' are intransigent, so cannot effectively model high level concepts or individual subjectivity.

A further development of CBIR is the Human-Computer-Interaction approach. This

approach allows the user to submit an initially broad query and then continuously refine the information based on relevance feedback. Rui, et al (1998) tested over 70,000 images and found that this method greatly reduces the user's effort of composing a query and captures the user's information need more precisely. CBIR systems that have utilised the relevance-feedback approach include 'FourEyes' (Minka and Picard, 1997), which employs a learning algorithm that selects and combines feature groupings based on examples of positive and negative pictures provided by the user. A variation of this is PicHunter (Cox, Miller, Minka, and Yianilos, 1998), which uses an algorithm based on a stochastic-comparison search where the probability of each image in the database being the target is updated thanks to comparisons carried out by the user. This Bayesian relevance feedback process is interesting, since it is not based on binary decisions (relevant or irrelevant). This kind of information is easier to assess, less arbitrary than the binary one, and it takes into account the uncertainty of human judgment. Another interesting system based on interactivity is presented by Schroder, Rehrauer, Seidel, and Datcu (2000). Their program actually allows the user to click on the specific part of a picture that they are looking for (e.g. a lake in a park scene) and the highlighted pixels are associated with the user-defined label for future searches. This way, the user can visually supervise the learning process. The relevance-feedback method has been applied to numerous systems, for beyond the scope of this Thesis, including website deconstruction ("The Collage Machine", Kerne, 2000, 2001), electronic sketchpad drawing ("Epic" Jose, Furner and Harper, 1998), and medical image databases (Howarth, Yavlinsky, Heesch, and Ruger, 2005).

Of particular interest to this Thesis are computer programs that have utilized algorithms that calculate how similar pictures are based on features of a sample picture, for example the weighted dissimilarity function ("RETIN", Fournier, Cord and Philipp-Foliguet, 2001) and sequential pattern recognition ("CAVIAR", Evans, Sikorski, Thomas, Sung-Hyuk, Tappert, Jie, Gattani, and Nagy, 2005). If previous research on the replication of scanpaths holds true and people really do produce highly similar eye movements at encoding and recognition of a picture, then it could be possible to create a CBIR program controlled by eye movements. The participant would simply imagine the picture they want to find in an image library and use their eyes to 'draw' on the blank screen, which is transformed into x-y coordinates and used by the computer program to search the database. This would eradicate a lot of the problems currently associated with CBIR regarding subjective and time consuming labelling of features. Potentially, it would possible to create such an algorithm based on The Hough Transformation. The Hough Transform is primarily for recognising geometric shapes within images, but there is a variant that will allow us to look for partial matches of irregular shapes expressed in the way that the eye tracking software produces them. There is considerable literature in this area that has successfully put these theories into practice. Schau (1992) carried out several experiments that changed the scale and relative rotation of the unknown. In all cases the correct automobile was identified from the set of six prototypes. Eom and Park **(1991) developed algorithms to identify and object based of it's similarity to other objects and also to classify it into the appropriate object category. Out of 1280 unknown objects created by rotating and morphing machine parts and aircraft shapes,** more than 98% of machine parts, and more than 97% of aircraft shapes were correctly classified.

Creating a pictorial library controlled by eye movements could be of great benefit to picture database librarians who have the job of searching through millions of visually similar pictures. These pictures hold similar or identical key words and thus Description Indexing approaches are of limited use. It could also be further developed for use with motor impaired persons, as a way to search pictorial databases without having to use their arms or hands.

However, to create such a program, we first have to determine how similar eye movements are at encoding, imagery and recognition of a picture, and the factors that could potentially affect scanpaths.

1.5. Main methods & equipment used in the experiments

1.5.1. Eye tracking equipment

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker, which uses an ergonomic chinrest and provides very precise data within a gaze position accuracy of 0.2 degrees. The system parses samples into fixations and saccades based on velocity across samples, with a spatial resolution of 0.01°, a processing latency of less than 0.5 milliseconds and a sampling rate of 240 Hz. An eye movement was classified as a saccade when its velocity reached 30 deg/s or when its acceleration reached 8000 deg/s². For experimental analysis throughout this thesis, a fixation is defined as anything above 70ms – micro-fixations below 70ms were discarded.

1.5.2 Scanpath similarity - string editing

There is a certain amount of difficulty in quantifying the similarity between scanpaths elicited at encoding and those made during the recognition test. This difficulty lies in condensing the spatial information of multiple fixations without losing the sequence information inherent in a two dimensional serial scanpath. The most well-known methods are Markov matrices and string-editing. The Markov process is a stochastic model for the probabilities that the viewers' eyes will move from one visual element to another and is based on the assumption that each eye fixation depends only on the previous one. However, an obvious criticism of this is that the sequence-generating process may have a longer history than the immediate past. Furthermore, Abbott and Hrycak (1990) point out that Markov models do not provide a technique for assessing similarity between sequences, categorizing sequences, or identifying typical sequences.

This string editing technique, on the other hand, is described in detail by Brandt and Stark (1997); Choi, Mosley, & Stark, (1995); Hacisalihzade, Allen, and Stark, (1992), Privitera, Stark and Zangemeister (2007) and Foulsham and Underwood (2008) and involves turning a sequence of fixations into a string of characters by segregating the stimulus into labelled regions. The similarity between two strings is then computed by calculating the minimum number of editing steps required to turn one into the other. Three types of operations are permitted: insertions, deletions and substitutions. Similarity is given by one minus the number of edits required, standardised over the length of the string. An algorithm for calculating the minimum editing cost is given in Brandt and Stark (1997) (appendix 1c) and was adapted to analyse string similarity using a Java program (an example of the javacode used can be found in appendix 1d).

For the analyses in this Thesis, a 5 by 5 grid was overlaid onto the stimuli (see Figure 1). The resulting 25 regions were labelled with the characters A to Y from left to right. Fixations were then labelled automatically by the program, according to their spatial coordinates, resulting

in a character string representing all the fixations made in this trial.
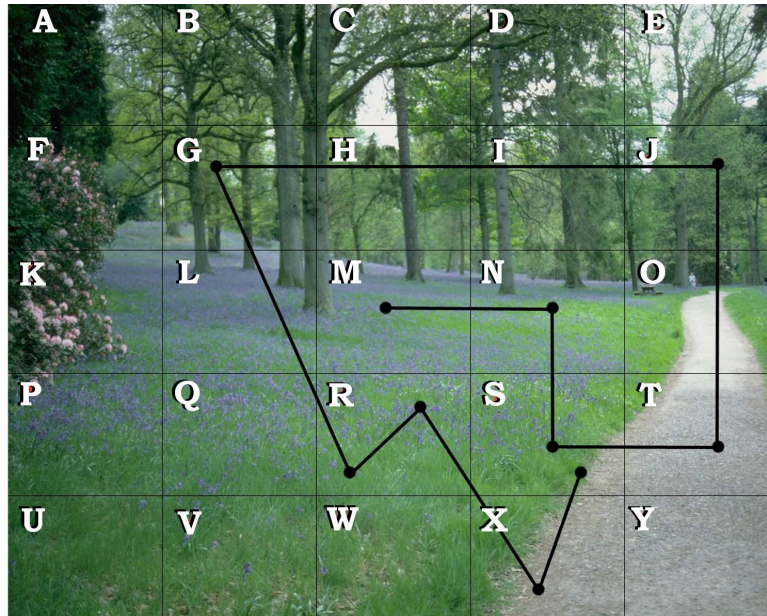
Figure 1: A picture of a natural scene with a 5x5 grid overlaid and an example scanpath superimposed.

For the fixation sequence shown in figure 1, the string would be MNSTJGRRXS. The first fixation, which was always in the centre or region "M", was removed and adjacent fixations on the same regions were condensed into one (making the example NSTJGRXS). Repetitions were condensed because it is the global movements that are of interest here, rather than the small re-adjustments which combine to give one gaze on a region. Strings were cropped depending on the average number of fixations per picture in that experiment (but see sections 2.4 and 6.4)

1.5.3 Measuring Chance

The results were compared against a chance baseline. One way we considered doing this was to compare the experimental data against a random model. For example if more human gazes than randomly generated gazes lie in salient regions then this would suggest the visual system is selecting based on saliency. However, a uniformly distributed random model might lead to a difference purely due to systematic bias in eye movements towards the centre (see Tatler et al, 2005). Therefore, for each picture a participant viewed, the scanpath produced was compared to a scanpath that the participant produced on another a randomly selected picture. This was repeated for every participant and an average similarity score was calculated. This score differed for each experiment as eye movements differed depending on task (however, see chapter 6.3.1 for problems with measuring chance).

Chapter 2 – Scanpaths and Imagery

2.1 Introduction

Imagery plays an important part in human memory (e.g., Paivio, 1969, 1971) and it has been suggested that eye movements or their control systems may play an important intermediate role in imagery (e.g., Hebb, 1968). The strong link between perception and imagery can be seen in the linear relationship between the distance of two imagined points and the time taken to scan between those points, with increased distance resulting in increased scanning time (Kosslyn, 1973; Finke and Pinker, 1982). Mental rotation of 3D objects (Shepard and Metzler, 1971; Cooper, 1975) revealed spatial properties of imagery that are analogous with visual perception. Furthermore, Fink and Schmidt (1977) found that illusions and after-effects normally caused by perception can result from mental images, suggesting that imagery is not just spatial but specifically visual.

It is suggested that this connection between perception and imagery exists because of common neurological underpinnings. For example, various brain scanning techniques have revealed that visual areas in the occipital, temporal, and parietal lobes are also activated during visual mental imagery (see Kosslyn, 1994). The involvement of cortical structures common to imagery and perception in the visual modality is also supported by studies on evoked potentials (Farah, Weisberg, Monheit, and Peronnet, 1989) and regional cerebral blood flow (Charlot, Tzourio, Zilbovicius, Mazoyer, and Denis, 1992; Goldenberg, Podreka, Steiner, Willmes, Suess, and Deecke, 1989; Kosslyn, Alpert, Thompson, Maljkovic, Weise, Chabris, Hamilton, Rauch, and Buonanno, 1993; Mellet, Tzourio, Denis, and Mazoyer, 1995). However, arguments against these shared neurological processes for perception and imagery come from the double dissociation between brain-damaged patients that have impaired vision but intact imagery abilities (Behtmann, Winocur, and Moseovitch, 1992) and patients with visual imagery disorders that exhibit no perceptual impairment (Guaraglia, Padovani, Pantano, and Pizzamaglio, 1993). Goldenberg, Miillbacher and Nowak (1995) report the case of a patient with extensive lesioning of the primary visual cortex. Despite being blind, the patient had intact mental imagery, and more interestingly, believed she could see. This belief could be due to her confusing mental images with real percepts. Due to the extent of her lesioning, however, her mental imagery ability must have relied on higher-level visual areas. Similarly, Servos and Goodale (1995) examined a patient with visual agnosia who nevertheless was able to scan mental images to search for specific features and could generate novel images by combining images of previously encoded stimuli. These studies suggest that while visual perception and mental imagery may be connected, they do not necessarily share the same neurological structures or mechanisms.

To test the extent to which eye movements play a role in mental imagery, Baddeley (Baddeley 1986; recently reported by Postle, Idzikowski, Sala, Logie, Baddeley, 2006) ran three interesting experiments. The first aimed to test whether *involuntary* eye movements induced by spinning in a chair (post-rotational nystagmus) would disrupt imagery. The lack of disruption to imagery or memory tasks indicated that eye movements per se do not have an important *immediate* role in imagery. However, in the second experiment voluntary eye movements, made by watching a moving spot, did disrupt imagery, implying that voluntary eye movements and imagery do share a common processing capacity. The third experiment showed that impairment of performance on the spatial task occurred for voluntary eye movements during both presentation and recall of the memory material. However, because memory was no more impaired when the

eyes were moving all the time compared to only at presentation or only at recall, Baddeley concluded that the link between eye movements and imagery was in the maintenance of the image. This study was followed up by Postle et al (2006) who concluded that that the maintenance component of spatial working memory, independent of encoding and/or response, is sensitive to concurrent saccadic activity. In summary, these four studies show that the control of visual imagery and visual working memory may derive from the same cognitive resources that support eye movement control.

2.2. Experiment 1: Eye movements at Imagery vs. Recognition

2.2.1 Introduction

An argument against Scanpath Theory is that people may not reproduce the same scanpaths over time due to the sequence of eye movements being stored internally or being related to an internal visual image, but they do so by chance because of the bottom-up influences of the visual stimulus. When we view a picture (at least in a free-viewing or in a memory task), our eye fixations are attracted by the visual saliency of the image, with more attention being given to conspicuous regions than elsewhere. When we are shown that same picture again at a later time, perhaps we simply look at the same parts of the picture again, as those parts still hold the same low-level properties as when it was first inspected. By this argument, the re-instatement of a sequence of fixations on separate occasions may be a product of the visual characteristics of the image rather than having any involvement with our memories of the image or of our scanpath on first viewing. Even when saliency is overridden by the task demands, it could still be argued that scanpaths are reproduced because the same semantically interesting parts of the scene are present at encoding and recognition. Repeated scanpaths may be generated by viewers remembering how they inspected a picture when they first looked at it, but it could be that the features of the image - either bottom-up visual features or top-down meaningful features – are what drive the sequence of fixations.

One way to get around these problems is to use an imagery task, so that if scanpaths are reproduced, it cannot be due to external bottom-up influences, as no visual stimulus is present. Brandt and Stark (1997) found substantial similarities between sequences of fixations made whilst viewing a simple checker-board diagram and those made when imagining it later. Since there is no actual diagram or picture to be seen during the imagery period, it is likely that an internalised cognitive perceptual model is in control of these scanpaths. Holsanova, Hedberg and Nilsson (1998) used natural, real life scenes and found results similar to those reported by Brandt and Stark. Pieters, Rosbergen, and Wedel (1999) found that scanpaths remain constant across advertising repetitions, across experimentally induced and naturally occurring conditions, and (like Brandt and Stark), Zangemeister, Oechsner, and Freksa (1995) and Gbadamosi, Oechsner, and Zangemeister (1997), also demonstrated firm evidence for scanpath sequences in the viewing of both real and imagined stimuli.

In a modified version of the imagery experiment, Laeng and Teodorescu (2002) manipulated when participants could move their eyes. Participants that were told to keep their eyes centrally fixated during the initial scene perception did the same, spontaneously, during imagery. Participants that were allowed to move their eyes during initial perception but were told to keep their eyes centrally fixated during imagery exhibited decreased ability to recall the pattern. Laeng and Teodorescu argued that this was because the oculomotor links established during perception could not be used in the process of building up a mental image, and this limitation impaired recall. Eye movements at first viewing

help to encode the picture and reproducing those eye movements at a later stage may help recall the picture. However, it could be argued that when pictures are better recalled, the eye movement patterns during imagery, as a *result*, better match the eye movement patterns during scene viewing. A decrease in recall performance when participants are instructed to keep fixation at imagery could therefore be due to additional cognitive load exhibited by the (additional) task to refrain from naturally moving one's eyes.

One aim of the current experiment is to determine whether scanpaths are reproduced during imagery. This could avoid the criticisms that the reproduction of scanpaths may be due to external bottom-up influences, as this cannot be true if no visual stimulus is present. It would also be interesting to know whether this relationship between imagery and perception persists over time. Ishai and Sagi (1995) have shown, for example, that imagery induced facilitation in a target-detection task decays and is only effective in the first 5 min after the participants saw the stimuli. In Laeng and Teodorescu's (2002) study, the participants performed the imagery task 40 seconds after they studied the stimuli and it was suggested by Mast and Kosslyn (2002) that the sensorimotor trace may be stored only in short-term memory. Therefore, another aim of the current experiment is to determine whether scanpaths at imagery are stable over extended periods of time.

One model that could help explain eye movements during imagery is Kosslyn's (1994) 'visual buffer', which is used to construct an internal image. The visual buffer is located in the working memory, which is topographically organized and has the possibility to represent spatiality. An 'attention window' can be moved to certain parts of the visual buffer, which could be connected to eye movements during imagery. Mental images are generated in the visual buffer, and representations of those images are stored in long term memory. When a scene stored in long term memory is visualized, it is generated (or rather created or re-created) in the working memory and in the visual buffer. Kosslyn's visual buffer model is fundamentally different from Noton and Stark's (1971) Scanpath Theory in the purpose/explanation of the eye movements that occur during imagery. Whereas the visual buffer suggests a 'scanning' of an already existing memory/mental image, Scanpath Theory suggests the reconstruction of an image through scanpath replication (i.e. similar eye movements are essential for accurate picture recognition).

A large amount of criticism against the visual buffer comes from propositional accounts (e.g., Pylyshyn, 2002, 2003), which claim that there are no such things as internal images. Pylyshyn argues that imagined objects and spatial locations are bound to visual features in the external world; these bindings are called 'visual indexes' (Pylyshyn, 2000, 2001, 2002). This theory assumes no pictorial properties whatsoever of the 'projected image', only the binding of imagined objects to real, perceived ones. However, Johansson, Holsanova and Holmqvist (2006) carried out an imagery study in the dark (i.e., without any possible visual features) and still yielded eye movements that reflected objects from both the description and the picture. Therefore, Johansson et al. argued that visual indexes that only assume the binding of propositional objects to real ones cannot explain eye movements during mental imagery.

An additional objective of this experiment is to investigate which account best explains eye movements during imagery, and also whether eye movements at retrieval are affected by different methods of encoding and of retrieval. If Pylyshyn's propositional model holds true, then eye movements should not be affected by such manipulations, as they would not change tacit knowledge (the knowledge of what seeing a specific object would be like). This study also aims to find out if, assuming that a scanpath is reproduced, temporal information is reproduced as well as spatial information. To do this, average fixation duration, average saccadic amplitude, and the number of fixations are calculated at each encoding and retrieval condition. Two procedures were used in the experiment, one in which viewers were required to visualize the picture most recently inspected, and one in which the

imagery task was conducted after the presentation of all of the pictures in the experiment. In both procedures there was an imagery task and a recognition memory task – the order was reversed between procedures. After a two day interval the imagery task was repeated.

2.2.2 Methodology

Participants

Thirty participants took part in the experiments, all of whom were students at Nottingham University. The age range was 18-51 and the mean age was 25.5. The sample comprised 21 females and 9 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and the participants being naïve to eye movements being recorded.

Materials and apparatus

A set of 60 high-resolution digital photographs were prepared as stimuli, sourced from a commercially available CD-ROM collection and taken using a 5MP digital camera. Each picture was distinctly individual, in that given a short sentence describing a picture; it could not be mistaken for any of the others. Examples of these stimuli are shown in Figure 2.

A pilot study was conducted to make sure the stimuli were distinctly individual and could not be confused. Ten participants were given a sheet of 60 pictures and a sheet of 60 descriptive labels, both randomly ordered, and were asked to match the pictures to the labels. All of the participants correctly matched 100% of the stimuli.

Half of each category were designated "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test. New and old pictures were similar in complexity, semantic and emotional content. Pictures were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. The monitor measured 43.5cm by 32.5cm, and a fixed viewing distance of 98cm gave an image that subtended 25.03 by 18.83 degrees of visual angle.

Figure 2: *Examples of two of the distinctively individual pictorial stimuli used in the experiments: 'the penguins', and 'the buttons'.*

Design

The experiment used a between groups design, with 2 groups of participants (15 participants in each group). The independent variable was therefore which group the participant belonged to (The

Imagery First group or The Recognition First group). The dependent variable measures were: accuracy in deciding whether a picture was old or new, average fixation durations, average saccadic amplitude, average number of fixations, and the similarities of scanpaths compared at encoding and imagery, encoding and recognition, encoding and delayed imagery, imagery and recognition, imagery and delayed recognition, and recognition and delayed imagery.

Procedure
Participants were told that their pupil size was being measured in relation to mental workload. They were informed that although their eye movements were not being recorded, it was important to keep their eyes open so pupil size could be reliably measured.

Task1: Imagery Prior to Recognition. Following a 9-point calibration procedure, participants were shown written instructions on the experimental procedure and given a short practice. The first stage involved seeing a picture for 3000 milliseconds then a brightly coloured mask for 1000 milliseconds and then the screen turned blank. The participant then had 5000 milliseconds to visualize the last photograph they had seen. After this time, a fixation cross appeared for 1000 milliseconds to ensure that fixation at picture onset was in the centre of the screen. This experimental procedure is illustrated in figure 3.
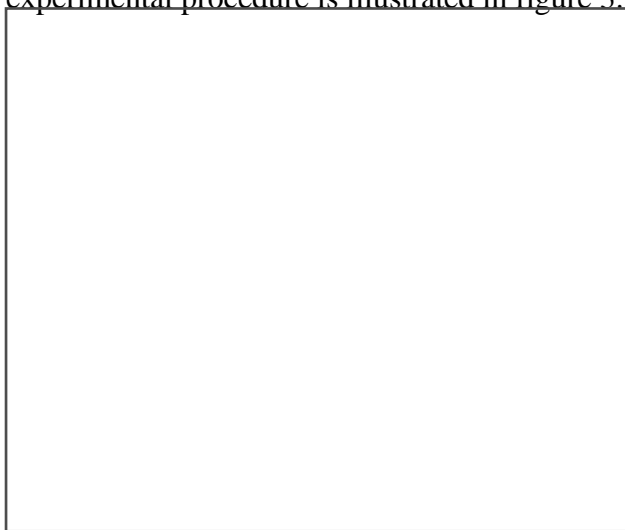
Figure 3: *A diagram of the imagery prior to recognition procedure.*

After participants had seen and visualized 30 stimuli, presented in a random order, they took a short break and were then asked to perform a recognition memory test. Participants saw a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures). They were instructed to press "N" on the keyboard if the picture was new, and "O" on the keyboard of the picture was old. Sixty stimuli were presented in a random order, 30 of which were old and 30 new. In order to facilitate an ideal comparison of scanpaths between encoding and recognition, each picture was shown for 3000 milliseconds and participants could only make a response after this time. This was to encourage scanning of the whole picture. This procedure is illustrated in figure 4.
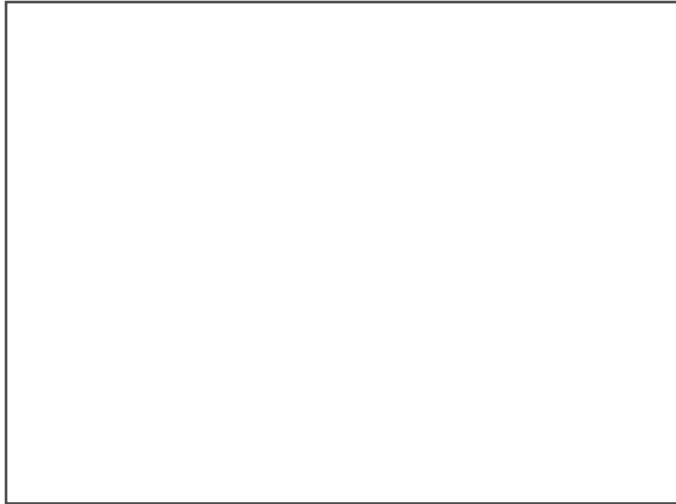
Figure 4: *A diagram of the recognition memory test in the 'imagery prior to recognition' procedure.*

Participants returned approximately 48 hours later to perform another imagery task. This time they saw 30 white screens with a short sentence describing one of the pictures seen 48 hours earlier. All of the pictures described in this task had previously appeared in the first imagery task, and were presented here in a new random sequence. Participants were asked to visualize the picture described and try to remember everything they could about it. Each description appeared for 3000 milliseconds and then the screen went blank for 5000 milliseconds, during which they visualized the stimulus. This procedure is illustrated in figure 5.
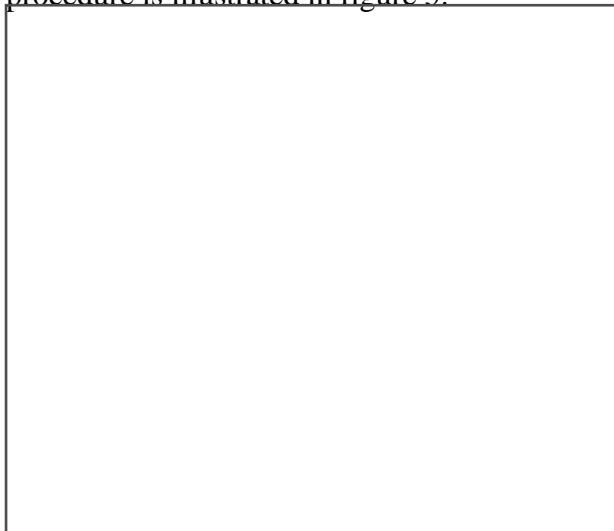
Figure 5: *A diagram of the delayed imagery task in both the 'imagery prior to recognition' and the 'recognition prior to imagery' procedures.*

Task 2: Recognition Prior to Imagery. The difference between The Imagery First group and The Recognition First group was in the order of the imagery and recognition tasks. As before, the first

stage here involved viewing a set of 30 stimuli, presented in a random order, in preparation for a memory test, but no imagery took place at this stage. Each picture was preceded by a fixation cross for 1000 milliseconds, which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 3000 milliseconds, during which time participants moved their eyes freely around the screen.

After all 30 pictures had been presented, participants saw a second set of pictures and had to decide whether each picture was new (not seen before in the experiment) or old (from the previous set of pictures). They were instructed to press "N" on the keyboard if the picture was new, and "O" on the keyboard of the picture was old. Sixty stimuli were presented in a random order, 30 of which were old and 30 new. In order to facilitate an ideal comparison of scanpaths between encoding and recognition, each picture was shown for 3000 milliseconds and participants could only make a response after this time. This was to encourage scanning of the whole picture. See figure 3 for a diagram of this recognition procedure.

After all 60 pictures in the recognition test had been shown, the participants took a break before performing an imagery task. This time they saw 30 white screens with a short sentence describing one of the pictures they had just seen. All the pictures in this imagery task were classified as 'old' but the participants were not informed of this. The pictures appeared in a random order. Participants were asked to visualize the picture described and try to remember everything they could about it. Each stimulus appeared for 3000 milliseconds and then the screen went blank for 5000 milliseconds, in which they visualized the stimulus. See figure 4 for a diagram of this procedure.

Participants returned two days later to perform the last imagery task again (see figure 4). The procedure was identical and all of the descriptions of pictures in this task had previously appeared in the first imagery task, and were presented here in a new random order. Participants were asked to visualize the picture described and try to remember everything they could about it. Each description appeared for 3000 milliseconds and then the screen went blank for 5000 milliseconds, in which time they visualized the stimulus.

2.2.3 Results

In all cases, trials were excluded where the fixation at picture onset was not within the central region (the central square around the fixation cross when the picture was split into a 5x5 grid at analysis), or when calibration was temporarily interrupted (e.g. if the participant sneezed, therefore removing their head from the eye tracker).

There were 2 main types of data, recognition memory data (accuracy), and eye tracking measures – average fixation durations, average saccadic amplitude, average number of fixations, and string analyses.

Although participants in both Tasks performed both the imagery and recognition tests but in different orders, for the sake of clarity Task 1 will be referred to as the 'Imagery First group' and Task 2 will be referred to as the 'Recognition First group'.

At the end of both Tasks, participants filled out a short questionnaire consisting of 9 filler questions (e.g. age, degree course, level of tiredness etc) and one target question asking them about the aim of the experiment. One participant in the Imagery First group guessed the aim of the study and their datum was discarded.

Recognition Memory

Accuracy was measured by the number of pictures participants correctly identified as 'old' (if they were from the previous set) or 'new' (if they had never been seen before). As shown in

Figure 6, both groups performed at a very high accuracy rate (98.10% in the Imagery First group and 97.11% in the Recognition First group).

Datum from one participant in The Imagery First group had to be removed because they pressed the wrong button all the way through the recognition test. A between-groups T-test on the remaining 28 participants showed no reliable difference between the groups: t(26)=0.97, p=0.623 (see appendices 2a and 2b).
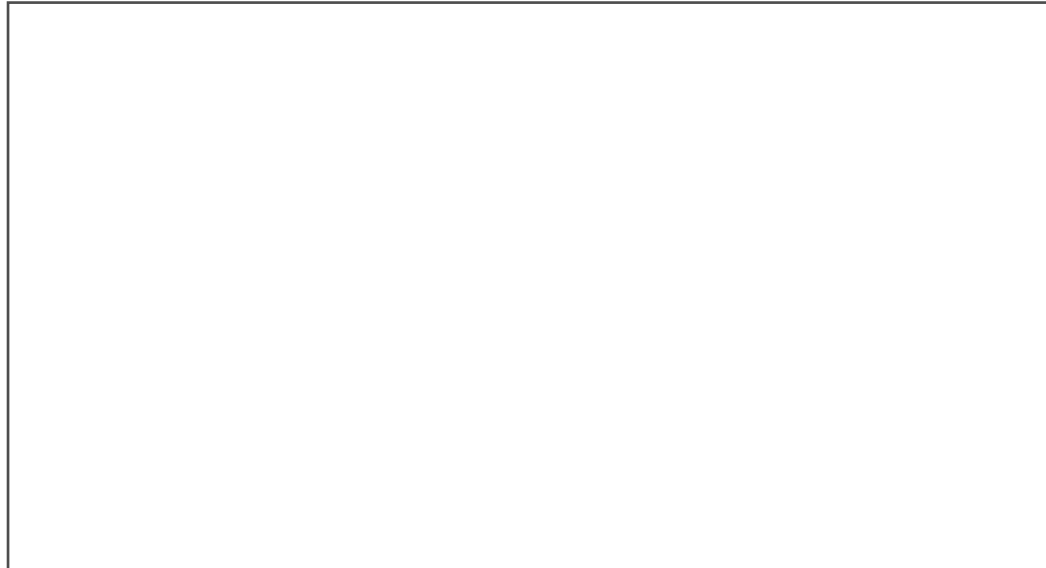
Figure 6: *The mean recognition accuracies in the two Tasks. In The Imagery First group, participants attempted to visualize each picture immediately after viewing it, and in The Recognition First group they performed the imagery task after a recognition test. The error bars here and throughout this Thesis represent Standard Error (SE).*

Eye tracking measures

Average Fixation Duration
Overall, participants in The Imagery First group exhibited shorter fixations than participants in The Recognition First group. These means are shown in Figure 7 (and appendix 3a). In both Tasks, participants made shorter fixations at encoding than at imagery or delayed imagery. Participants also made shorter fixations at recognition (old and new pictures) than at imagery or delayed imagery.

A mixed-design ANOVA showed a reliable main effect of group (Imagery First or Recognition First), $F(1,27) = 17.89$, $MSe = 128692$, $p<0.001$ (appendix 3b), and a reliable main effect of test phase, $F(4,108) = 45.39$, $MSe = 61483.77$, $p<0.001$ (appendix 3c). There was no statistically reliable interaction between group and test phase. From the descriptive statistics in appendix 3d it can be seen that overall the Imagery First group make longer fixations than the Recognition First group.

A post-hoc paired samples T-tests were carried out to further investigate the main effect of condition

(appendices 3e and 3f). The results showed that fixations were shorter during encoding than during the first imagery phase (*t*(28) = 5.946, *SEM* = 84.16780, *p*<0.001), and the delayed imagery phase (*t*(28) = 7.197, *SEM* = 70.01095, *p*<0.001). Fixations were longer at encoding than during recognition of new pictures (*t*(28) = 3.759, *SEM* = 29.57112, *p*<0.01), but there was no statistically reliable difference between fixation durations at encoding and recognition of old pictures [*t*(28) = 1.767, *p*=0.088]. Fixation durations during imagery, however, were longer than during recognition of both old (*t*(28) = 6.632, *SEM* = 85.96716, *p*<0.001) and new (*t*(28) = 7.363, *SEM* = 83.06748, *p*<0.001) pictures. Similarly, fixation durations during delayed imagery were longer than during recognition of both old (*t*(28) = 8.345, *SEM* = 68.73568, *p*<0.001) and new (*t*(28) = 8.765, *SEM* = 70.17152, *p*<0.001) pictures. Fixation durations were also longer at recognition when viewing old pictures than new ones (*t*(28) = 2.309, *SEM* = 17.94159, *p*<0.05). There was no reliable difference in fixation durations between imagery and delayed imagery [*t*(28) = 0.047, *p*=0.963].

Figure 7: *Differences in average fixation durations between the two Tasks and between phases of the course of the experiment.*

Average Number of Fixations

The mean numbers of fixations made in each phase of the experiment and for each group of participants, are shown in Figure 8 (see also appendix 4a). Participants made more fixations at encoding then at imagery or delayed imagery. Participants also made fewer fixations at imagery and at delayed imagery than at recognition (old and new).

A mixed design ANOVA (appendices 4b and 4c) showed a reliable effect of test phase *F*(4,108) = 20.080, *MSe* = 2601.373, *p*<0.001, but no reliable main effect of group [*F*(1,27) = 3.918, *p*=0.058]. There was no reliable interaction between group and test phase [*F*(4) = 0.503, *p*=0.734].

Post-hoc paired-samples T-tests (appendix 4d and 4e) showed that there were more fixations during

encoding than during imagery ($t(28) = 4.672$, *SEM* = 16.05267, $p<0.001$), and delayed imagery ($t(28) = 5.151$, *SEM* = 15.42290, $p<0.001$). There was no reliable difference in the number of fixations between encoding and recognition of either old [$t(28) = 1.340$, $p=0.191$] or new [$t(28) = 0.657$, $p=0.516$] pictures. There were reliably fewer fixations at imagery than at recognition of both old ($t(28) = 4.437$, *SEM* = 14.93565, $p<0.001$) and new ($t(28) = 6$, *SEM* = 13.91332, $p<0.001$) pictures. Similarly, there were reliably fewer fixations at delayed imagery than at recognition of both old ($t(28) = 4.432$, *SEM* = 15.23354, $p<0.001$) and new ($t(28) = 5.607$, *SEM* = 15.11156, $p<0.001$) pictures. There was no reliable difference in the number of fixations made between imagery and delayed imagery [$t(28) = 0.81$, $p=0.936$].

Figure 8: *Differences in number of fixations between The Imagery First group and The Recognition First group and between phases of the course of the experiment.*

Average Saccadic Amplitude

The average saccadic amplitudes in each phase of the experiment and for each group of participants are shown in Figure 9 (see also appendix 5a). Participants in the Recognition First group produced greater saccadic amplitudes than participants in the Imagery First group. Participants also produced greater saccadic amplitudes at imagery than at encoding and at delayed imagery than at encoding.

Figure 9: *Differences in average saccadic amplitude between The Imagery First group and The Recognition First group and between phases of the course of the experiment.*

A mixed design ANOVA showed a reliable effect of group (Imagery First or Recognition First) $F(1,27) = 13.987$, $MSe = 3795.602$, $p<0.001$ (see appendix 5b). There was also a reliable effect of test phase $F(4,27)= 2.640$, $MSe = 3667.229$, $p<0.05$ (appendix 5c). The descriptive statistics for 'group' (appendix 5d) show that the average saccade amplitude was reliably greater in the Recognition First group then the Imagery First group.

Post-hoc paired-samples T-tests were carried out to further investigate the main effect of test phase (appendices 5e and 5f). The results revealed the saccadic amplitudes were reliably greater at imagery and delayed imagery than at encoding: $t(28) = 2.584$, $SEM = 16.77314$, $p<0.05$; and $t(28) = 3.534$, $SEM = 10.92878$, $p<0.01$, respectively. There was no reliable difference between encoding and recognition of either old or new pictures [$t(28) = 0.728$, $p=0.472$; and $t(28) = 3.534$, $p=0.111$, respectively]. Similarly, there was no reliable difference between imagery and recognition of either old or new pictures [$t(28) = 1.741$, $p=0.093$; and $t(28) = 0.754$, $p=0.457$, respectively]. There was also no difference between imagery and delayed imagery [$t(28) = 0.349$, $p=0.729$], recognition of old and new pictures [$t(28) = 0.913$, $p=0.369$], or between recognition of new pictures and delayed imagery [$t(28) = 0.701$, $p=0.489$]. There was, however, a difference between recognition of old pictures and delayed imagery $t(28) = 2.288$, $SEM = 11.66517$, $p<0.05$.

Scanpaths: String Editing

String editing was used to analyse the similarity between scanpaths produced on encoding and imagery, encoding and recognition, encoding and delayed imagery, imagery and recognition, imagery and delayed recognition, and recognition and delayed imagery. The average number of fixations made by participants in the current experiment was eleven, so strings were cropped to eleven letters for the following scanpath analyses. In those trials where fewer than eleven fixations remained after condensing gazes, the comparison strings were trimmed to the same length.

The results were compared against a chance baseline, calculated by comparing the scanpath produced on each picture a participant viewed to a scanpath that the participant produced on another a randomly selected picture.

This was repeated for all 30 participants and an average chance similarity of 0.1159 was calculated.

Several experiments have shown that subjects rotate, change size, change shape, change colour, and reorganize and reinterpret mental images (e.g. Finke, 1989; Johansson, Holsanova, and Holmqvist, 2006). Although this could be a potential problem for the current experiment, it will also be interesting to see whether scanpaths (and saccadic amplitudes) are highly similar at imagery and delayed imagery, suggesting that the reorganisation occurs mostly between encoding and imagery but then stays relatively stable over multiple imagery tasks.

The results of the comparisons are shown in Figure 10 (see also appendix 6a). In the Imagery First group eye movements were more similar when comparing imagery and delayed imagery than when comparing encoding and imagery or encoding and delayed imagery or Imagery and recognition. Example scanpaths from one participant in The Imagery First group (chosen at random) are also shown in figure 11 and compare encoding, imagery and recognition phases.

Figure 10: *Differences in string similarities between The Imagery First group and The Recognition First group and between string comparison types.*

Figure 11: *Example scanpaths from one participant in The Imagery First group, chosen at random. The blue scanpath represents fixations and saccades at encoding; the red at imagery; and the blue at recognition. Encoding and recognition are slightly more similar than encoding and imagery or recognition and imagery. Note the increased saccadic amplitudes at imagery.*

In The Recognition First group, eye movements were less similar when comparing encoding and imagery then when comparing encoding and recognition, imagery and recognition or imagery and delayed imagery. Eye movements were more similar when comparing encoding and recognition than when comparing encoding and delayed imagery. Eye movements were less similar when comparing encoding and recognition than when comparing imagery and recognition, imagery and delayed imagery or recognition and delayed imagery. Eye movements were less similar when comparing encoding and delayed than when comparing imagery and recognition or imagery and delayed or recognition and delayed.

A mixed design ANOVA (appendices 6c) showed a reliable effect of string comparison type: $F(5,135) = 11.232$, $MSe = 0.010$, $p<0.001$, and a reliable interaction between group (Imagery First and Recognition First) and string comparison type: $F(5) = 3.569$, $p<0.01$. There was no main effect of group [$F(1,27) = 3.268$, $p=0.082$] (appendix 6b).

In the following post-hoc paired samples T-tests, the string similarity scores that are shown in Figure 10 were compared against each other. To identify the source of the interaction, this was done for each of the Task. For the Imagery First group (see appendices 6d and 6e), there were reliable differences between string similarities *encoding vs. imagery* and *imagery vs. delayed imagery* ($t(13) = 4.509$, $SEM = 0.02558$, $p<0.01$), between *encoding vs. delayed imagery* and *imagery vs. delayed imagery* ($t(13)= 2.979$, $SEM = 0.03774$, $p<0.05$), and between *imagery vs.*

recognition and *imagery vs. delayed imagery* (*t*(13) = 3.399, *SEM* = 0.02893, *p*<0.05). In each of these three comparisons, the *imagery vs. delayed imagery* similarity was the greater of the two. There were also a reliable differences between *encoding vs. recognition* and *encoding vs. delayed imagery* (*t*(13)= 2.168, *SEM* = 0.02429, *p*<0.05), and between *imagery vs. recognition* and *recognition vs. delayed imagery* (*t*(13)= 2.346, *SEM* = 0.01444, *p*<0.05).

In The Recognition First group (see appendices 6f and 6g), post-hoc T-tests showed that there were reliable differences between *encoding vs. imagery* and *encoding vs. recognition* (*t*(14) = 2.927, *SEM* = 0.02624, *p*<0.05), with more similarity between scanpaths involving recognition than imagery. There were also differences between *encoding vs. imagery* and *recognition vs. imagery* (*t*(14) = 3.393, *SEM* = 0.05341, *p*<0.01), between *encoding vs. imagery* and *imagery vs. delayed imagery* (*t*(14) = 3.774, *SEM* =0.05049, *p*<0.01), between *encoding vs. imagery* and *recognition vs. delayed imagery* (*t*(14) = 2.708, *SEM* = 0.06392, *p*<0.01). In each of these comparisons the similarity of *encoding vs. imagery* had the smaller magnitude. As with The Imagery First group, the similarity score for *encoding vs. recognition* was greater than that for *encoding vs. delayed imagery* (*t*(14) = 4.241, *SEM* = 0.01972, *p*<0.05), and the score for *imagery vs. delayed imagery* was greater than for *encoding vs. delayed imagery* (*t*(14) = 4.853, *SEM* = 0.04067, *p*<0.001). Other reliable differences were found between *encoding vs. recognition* and *imagery vs. delayed imagery* (*t*(14) = 2.3877, *SEM* = 0.04786, *p*<0.05), between *encoding vs. delayed imagery* and *imagery vs. recognition* (*t*(14) = 4.285, *SEM* = 0.04388, *p*<0.01), and between *encoding vs. delayed imagery* and *recognition vs. delayed imagery* (*t*(14) = 3.491, *SEM* = 0.05153, *p*<0.01).

2.2.4 Discussion

The main aims of this study were to determine whether scanpaths are reproduced when no visual stimulus is present and thus arguing against fixation selection being based on low level factors; to determine whether scanpaths at imagery are stable over time; to determine which account (Visual Buffer/propositional theory) best explains eye movements a imagery; and to determine whether eye movements at retrieval are affected by methods of encoding and of retrieval.

Analyses of recognition memory showed that participants in both Tasks were very good at identifying pictures as old or new. The accuracy was so high because each picture had to be distinctly individual in order for the imagery and delayed imagery tasks to work. This made it easy to decide which pictures had been seen before and which had not.

Average fixation durations were measured and analyses found a main effect of group in that participants in The Imagery First group made shorter fixations than participants in The Recognition First group. Average fixation duration at encoding was almost identical for The Imagery First group and The Recognition First group; suggesting that the groups were well matched and the differences between groups in other conditions were effects of the experimental design. This was also true of number of fixations. Interestingly, there was a difference between the groups at encoding when saccadic amplitude was measured, with participants in the Imagery First group producing smaller saccadic amplitudes than participants in the Recognition First group. One explanation for this could be that because participants in the Imagery First group were visualizing the pictures soon after they had seen them (and thus the pictures would still be in working memory), they may have focused on the main areas of interest. Whereas the Recognition First group had to remember a lot of pictures all at once (which would not be readily available in working memory) so scanned more widely to try

and encode spatial relations between objects.

The lower average fixation duration at imagery for the Imagery First group compared to the Recognition First group suggests that visualizing a scene directly after you have seen it (Imagery First) is less cognitively demanding than visualizing it after the recognition task (Recognition First), where you have to choose from a number of inspected scenes. The lower average fixation duration at delayed imagery for the Imagery First group compared to the Recognition First group suggests that visualizing the scene after the recognition task makes it more cognitively demanding to visualize it again 48 hours later. In accordance with the Visual Buffer model, when you visualize the scene directly after inspection (Imagery First) this process facilitates the long term memory representation of the image, and thus makes it less demanding to visualize it a second time at a later occasion. It is possible that imagining a scene after recognition, where you have to choose from a number of pictures is a process that takes more cognitive processing than the visualizing per se, and therefore this does not facilitate the long term representation, and consequently makes it harder to visualize it a second time.

At recognition, participants in The Imagery First group may have made shorter fixations because they had 'inspected' each picture twice before the recognition test (once during encoding and once during imagery) so recognition may have been easier and less time at each fixation was needed.

Analyses of the number of fixations also varied according to the task being performed. There were more fixations at encoding and at recognition than at imagery or delayed imagery. Considering the above explanations of fixation duration, this makes sense because participants tried to take in as much at encoding as possible, making a greater number of shorter fixations. Research has shown that eye movements at encoding and recognition are similar (e.g. Foulsham & Underwood, 2008; Humphrey & Underwood 2009) and the current results support this, in that the numbers of fixations in these conditions are also similar. In the imagery conditions on the other hand, the longer fixation durations and greater saccadic amplitudes due to the more difficult task of recall with no visual cues may have ultimately lead to a smaller number of fixations in these conditions. This could also be due to the fact that there is less information to fixate on in a "mental image", and also because of reorganizing and resizing shown to occur during imagery. Some previous studies have shown a 'shrinking' of the mental image, (e.g. Finke, 1989; Johansson et al, 2006), though the saccadic amplitude results of this study suggest that a 'stretching' during imagery may also exist.

At recognition, average saccadic amplitudes in the Imagery First group were shorter than those in the Recognition First group. Taking into account the shorter fixation durations and increased number of fixations, this saccadic amplitude data suggest that participants in the Imagery First group focus on a smaller area of the picture. This could be because the participants in this group had, in effect, moved their eyes around the pictures twice before the recognition test – once at encoding and once at imagery, and thus were more familiar with where the areas of interest were situated. They therefore did not have to scan the picture as broadly as participants in the Recognition group, who had only seen the pictures once before.

Overall, average saccadic amplitudes were greater at imagery and delayed imagery than at encoding. This could be explained by the reorganizing and re-shaping shown to occur during imagery. As mentioned above, previous research has indicated a 'shrinking' of the mental image during imagery tasks, whereas the saccadic amplitude data in this experiment suggest enlarging or 'stretching' of the mental image. One possible explanation for this could be a type of boundary extension, which has been shown to occur during imagery as well as perception (e.g. Intraub, Gottesman, and Bills, 1998).

The fact that the results showed no reliable difference between the imagery and delayed imagery conditions suggests that the reorganizing of mental images may take place between encoding and first imagery and then stays relatively stable over multiple imagery tasks.

Scanpaths produced at each condition were compared to every other condition using string analysis to create a similarity score. In The Imagery First group, scanpaths were more similar when comparing imagery and delayed imagery than when comparing encoding and imagery or encoding and delayed imagery or imagery and recognition. This could be explained in terms of mixed and pure process comparisons. When

comparing imagery and delayed imagery, the task was the same in The Recognition First group and very similar in The Imagery First group, in that both conditions involved recalling a memory without any immediate visual cues. This could be referred to as a 'pure process comparison'. Whereas when comparing encoding and imagery or encoding and delayed imagery or imagery and recognition, one of the conditions in each comparison involved visual input from the stimulus and the other involved recalling without any visual input. These could be referred to as 'mixed process comparisons', and produce lower similarity scores.

In The Recognition First group, scanpaths are less similar when comparing encoding and imagery than when comparing encoding and recognition, or imagery and recognition or imagery and delayed imagery. Encoding and imagery is a mixed process comparison and it makes sense that scanpaths in these two conditions would be less similar than when comparing encoding and recognition or imagery and delayed imagery, as these are pure process comparisons. How then can we explain why there is such great similarity between imagery and recognition in The Recognition First group when this is a mixed task comparison, and the same result is not true of this comparison in The Imagery First group? In The Imagery First group, participants visualized the picture shortly after seeing it; therefore the visual image was still in short term memory and imagery involved more reconstruction of the picture rather than retrieval of the memory. It could be said that the spatial information was still in the visual buffer. In The Recognition First group, retrieval was a more competitive process due to the distracter stimuli in the recognition test. Participants had to remember which picture the description was referring to before imagining specific details or features, so this type of imagining is more like the process of recognition. It could be argued that the visual information had to be retrieved from long term memory and re-created in the visual buffer before the picture could be imagined. This also applies to the delayed imagery test and explains the high similarity in between recognition and delayed imagery in both Tasks. In this sense, the comparison between imagery with written cues and recognition is more of a pure process comparison than between encoding and imagery or encoding and recognition or encoding and delayed. The reproduction of eye movements at imagery argues against a purely bottom-up explanation of scanpath similarity, as there is no visual (bottom-up) information at imagery.

The most similar scanpaths came from pure process comparisons where there was similar visual input in each condition (imagery compared to delayed imagery and encoding compared to recognition), and from comparisons that mimicked the same retrieval processes (imagery compared to recognition in The Recognition First group and delayed imagery compared to recognition in both Tasks 1 and 2). Pure process comparisons could also offer an explanation for the similarities between encoding and recognition phases with regards to fixation durations and number of fixations. The lowest scanpath similarity scores came from mixed process comparisons (encoding compared to imagery, encoding compared to delayed imagery, and imagery compared to recognition in The Imagery First group).

Even though the string similarity scores were quite low when comparing encoding and imagery, (Imagery First group = 0.170; Recognition First group = 0.165), the scores were still reliably above chance, suggesting that eye movements are still reproduced even when no visual information is present (during imagery). This argues against a purely bottom-up explanation of scanpath similarity.

The lower scanpath similarity scores when comparing encoding and imagery could be due to reorganizing and re-sizing during mental imagery. However, the greatly increased similarity scores when comparing imagery and delayed imagery (Imagery First group = 0.274; Recognition First group = 0.346) suggest that reorganisation occurs mostly between encoding and imagery but then stays relatively stable over multiple imagery tasks.

Overall, the scanpath analyses have shown that the more similar the retrieval process is to the encoding process, the more similar the scanpaths produced. This suggests that the visual buffer model may be more complicated than simply shifting attention to different parts of an internal image (Kosslyn, 1994). The relationship between the encoding and retrieval process seems to be very important and one

might even suggest the existence of facilitatory and inhibitory pathways within the model. For example, retrieval of a representation from long term memory could be facilitated if exactly the same visual information is present at encoding and recognition, as there are more visual guides and less chance of reorganizing or resizing as the information is transferred from long term memory to the Visual Buffer. The cognitive load on working memory is also lowered.

Propositional accounts such as that of Pylyshyn (2002) argue that there is no such thing as a visual buffer and that when participants are asked to "imagine X" they use their knowledge of what "seeing X" would be like, and they simulate as many of these effects as they can. However, it seems very unlikely that participants are able to mimic behaviour so precisely in their eye movements. In agreement with Johansson et al (2006), the number of points and the precision of the eye movements to each point are too high to be remembered without a support to tie them together in a context, such as an internal image. This is backed up further by the finding that temporal information as well as spatial information is reproduced at retrieval and is consistent over time as long as the same retrieval process is used. Furthermore, if participants did store spatial scene information as a large collection of propositional statements, scanpath similarity should have remained constant across conditions despite changing the retrieval task, but this was not the case.

The finding that scanpaths at imagery were highly similar to those at delayed imagery (48 hours later) suggests that they are stable over time. Furthermore, the similarity between the scanpaths at encoding and imagery challenges the view that eye movements are influenced purely by bottom-up visual features, as none existed during the imagery stage. Instead it suggests that a combination of top-down and bottom up factors interact to guide our eye movements.

To conclude, in accordance with Johansson et al (2006), the results of this experiment lend support for the visual buffer model of imagery (Kosslyn, 1994), and challenge the propositional visual index model (Pylyshyn, 2002). The variations in scanpath similarities caused by manipulation of the retrieval processes suggests that the visual buffer may be more complicated than previously thought, with possible facilitatory and inhibitory pathways. The similarity of scanpaths at encoding and imagery show that eye movement sequences can be, to an extent, replicated at a later time and also argues against the fixation selection being based on low level factors. The lower scanpath similarity scores when comparing encoding and imagery suggests that most of the re-sizing and reorganising of mental images occurs at this stage. The high scanpath similarity scores when comparing imagery and delayed imagery suggests that much less resizing happens once the mental images have been formed and that these scanpaths are relatively stable over time.

2.3. Experiment 2: Controlling for methodological variables – Looking beyond the picture boundaries

2.3.1 Introduction
Although the previous experiment shows novel and promising results, it still remains uncertain why saccadic amplitudes are longer for imagery than other conditions, as previous literature suggests that the resizing and reorganisation during mental imagery result in smaller, not larger scanpaths. One suggested explanation is boundary extension, a memory illusion where participants remember seeing more of the picture than was actually presented. This phenomenon has been shown to occur in both recall and recognition tests (Intraub and Richardson, 1989), after long and short picture presentations (Intraub, Gottesman, Willey and Zuk, 1996), in people of all ages (Seamon, Schlegel, Hiester, Landau and Blumenthal, 2002), and despite prior warning (Intraub and Bodamer, 1993). Boundary extension during imagery could cause the participant to actively look further than the picture boundaries (i.e. off the computer screen). However this is very difficult to measure as any eye movements off the screen are recorded by the eye tracker as a

loss of calibration, thus making it impossible to differentiate between actual calibration issues and fixations out of the picture boundaries.

One solution is to make the picture smaller and central, with a large border. This serves two purposes: firstly to reinforce the picture boundaries by the presence of an obvious border; and more importantly to allow analysis of any fixations that may fall outside the picture boundaries.

This experiment was conducted to determine whether eye movements were affected by the presence of a definite picture border and whether the more extensive scanpaths during imagery could be explained by participants remembering more of the picture than was actually present, causing them to look beyond the picture boundaries.

2.3.2 Methodology

Participants
Fifteen participants took part in the experiment, all of whom were students at Nottingham University. The age range was 18-30 and the mean age was 21. The sample comprised 21 females and 9 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and the participants being naïve to eye movements being recorded.

Materials and apparatus
The same pictures were used as in the previous experiment (60 high resolution, distinctly individual photographs), but they were reduced to half the size (800 by 600 pixels) and were displayed centrally with a black border filling the rest of the screen (see figure 12). Stimuli were presented on a colour computer monitor measuring 43.5cm by 32.5cm, at a fixed viewing distance of 98cm. As in the previous experiment, half of each category were designated "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test. New and old pictures were similar in complexity, semantic and emotional content.
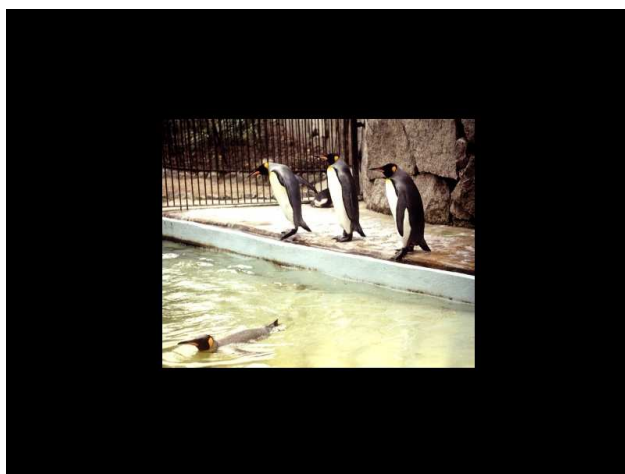

Figure 12: *An example of the stimuli used in the experiment.*

Design

The experiment used a within groups design, with one groups of participants and four main experimental conditions (encoding, imagery, recognition, and delayed imagery). The independent variable was therefore the experimental condition and the dependent variable measures were: accuracy in deciding whether a picture was old or new, average fixation durations, average saccadic amplitude, average number of fixations, the number of fixations outside the picture boundaries and the similarities of scan patterns compared at encoding and imagery, encoding and recognition, encoding and delayed imagery, imagery and recognition, imagery and delayed recognition, and recognition and delayed imagery.

Procedure

Participants were told that their pupil size was being measured in relation to mental workload. They were informed that although their eye movements were not being recorded, it was important to keep their eyes open so pupil size could be reliably measured.

The procedure was identical to the 'Imagery First' procedure of the first experiment. Following a 9-point calibration procedure, participants were shown written instructions on the experimental procedure and given a short practice. The first stage involved seeing a picture for 3000 milliseconds then a brightly coloured mask for 1000 milliseconds and then the screen turned blank. The participant then had 5000 milliseconds to visualize the last photograph they had seen. After this time, a fixation cross appeared for 1000 milliseconds to ensure that fixation at picture onset was in the centre of the screen.

After participants had seen and visualized 30 stimuli, presented in a random order, they took a short break and were then asked to perform a recognition memory test. Participants saw a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures). They were instructed to press "N" on the keyboard if the picture was new, and "O" on the keyboard of the picture was old. Sixty stimuli were presented in a random order, 30 of which were old and 30 new. In order to facilitate an ideal comparison of scan patterns between encoding and recognition, each picture was shown for 3000 milliseconds and participants could only make a response after this time. This was to encourage scanning of the whole picture.

Participants returned approximately 48 hours later to perform another imagery task. This time they saw 30 white screens with a short sentence describing one of the pictures seen 48 hours earlier. The descriptive sentence was in black font on a white background measuring 800 by 600 pixels, with a black border filling the rest of the screen. See figure 13 for an example. All of the pictures described in this task had previously appeared in the first imagery task, and were presented here in a new random sequence. Participants were asked to visualize the picture described and try to remember everything they could about it. Each description appeared for 3000 milliseconds and then the screen went blank for 5000 milliseconds, during which they visualized the stimulus.
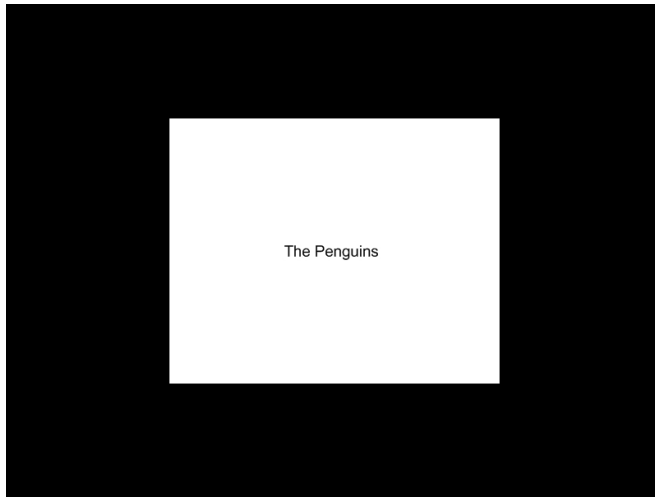
Figure 13: *An example of the stimuli used in the delayed imagery test.*

2.3.3 Results

In all cases, trials were excluded where the fixation at picture onset was not within the central region (the central square around the fixation cross when the picture was split into a 5x5 grid at analysis), or when calibration was temporarily interrupted (e.g. if the participant sneezed, therefore removing their head from the eye tracker).

There were 2 main types of data, recognition memory data (accuracy), and eye tracking measures – average fixation durations, average saccadic amplitude, average number of fixations, and string analyses.

At the end of the experiment, participants filled out a short questionnaire consisting of 9 filler questions (e.g. age, degree course, level of tiredness etc) and one target question asking them about the aim of the experiment. One participant guessed the aim of the study and their datum was discarded. Out of the 14 remaining participants, only 8 were included in the eye data analysis (fixation duration, saccadic amplitude, number of fixations and string analyses), as the rest did not move their eyes at all during imagery.

Average Accuracy

The average accuracy to identify pictures at recognition test was calculated to be 96.4% (Standard error: 1.2%). Accuracy scores were then calculated to take into account whether the picture was old or new (see figure 14) and a paired samples T-test was carried out.
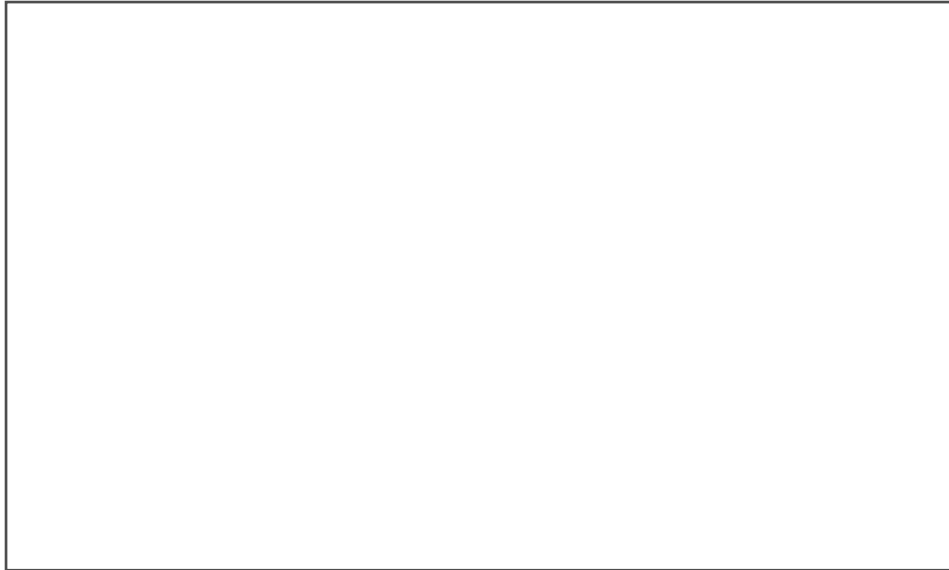
Figure 14: *A bar chart to illustrate the differences in accuracy in recognising old and new pictures at test.*

The T-test revealed a statistically reliable difference between old and new stimuli: $t(13) = 3.017$, *SEM* = 2.20998, $p<0.05$ (appendices 7a and 7b).

Average Fixation Duration
The average fixation duration was calculated for each experimental condition (See figure 15) and a within-group ANOVA was carried out (see appendix 8a). There was a reliable main effect of experimental condition: $F(4,28) = 34.383$, *MSE* = 22170.627, $p<0.001$.

Figure 15: *a bar chart to illustrate the average fixation duration in each of the experimental conditions.*

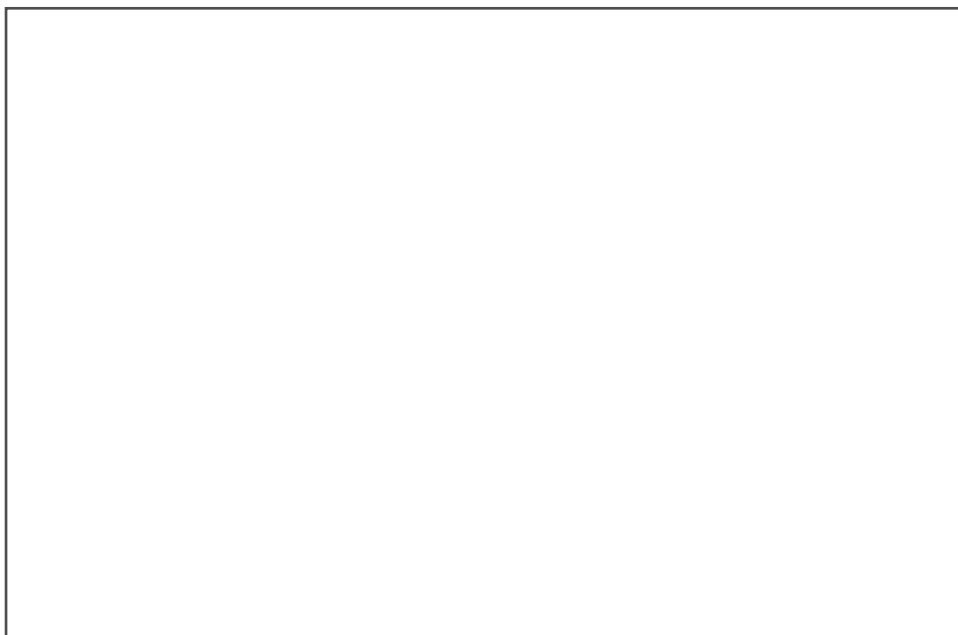Post-hoc paired-samples T-tests were carried out (see appendices 8b and 8c) and found reliable differences between encoding and imagery: $t(7) = 6.597$, *SEM* = 97.15002, $p<0.001$; encoding and delayed imagery: $t(7) = 8.8$, *SEM* = 55.70356, $p<0.001$; imagery and recognition old: $t(7) = 5.116$, *SEM* = 118.02844, $p<0.001$; imagery and recognition new: $t(7) = 5.681$, *SEM* = 113.44837, $p<0.001$; recognition old and delayed imagery: $t(7) = 10.352$, *SEM* = 43.76445, $p<0.001$; and between recognition old and delayed imagery: $t(7) = 12.140$, *SEM* = 40.66865, $p<0.001$. There were no reliable differences between encoding and recognition ($p = 0.451$), imagery and delayed imagery ($p = 0.144$), or recognition old and recognition new ($p=0.075$).

Average Number of Fixations

The average number of fixations per picture was calculated for each experimental condition (See figure 16) and a within-groups ANOVA was carried out (appendix 9a). There was a reliable main effect of experimental condition: $F(4,28) = 18.202$, *MSE* = 0.795, $p<0.001$.

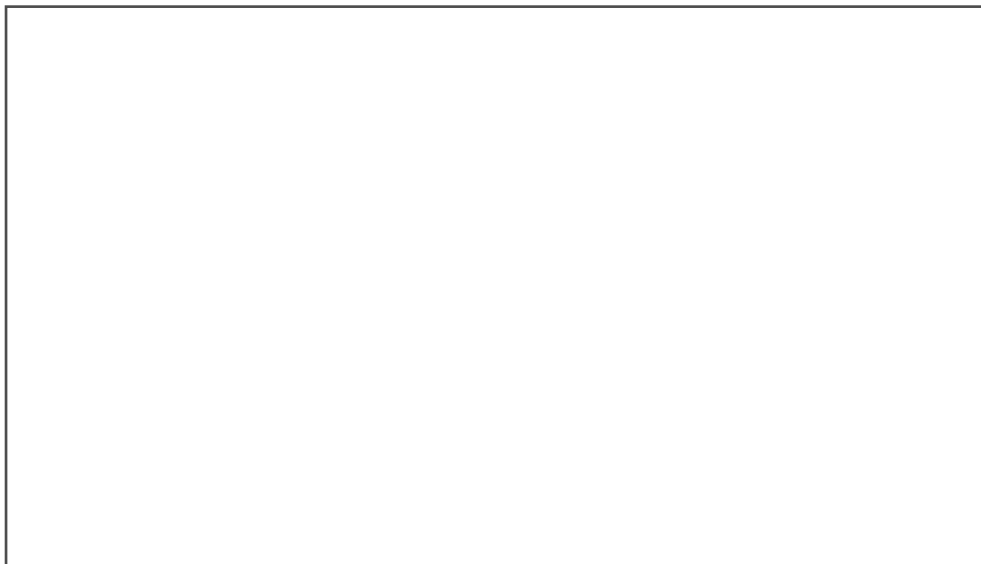***Figure 16:*** *A bar chart to illustrate the average number of fixations per picture, in each of the experimental conditions.*

Post-hoc paired-samples T-tests were carried out (appendices 9b and 9c) and found reliable differences between encoding and imagery: $t(7) = 5.103$, *SEM* = 0.55115, $p<0.01$; encoding and delayed imagery: $t(7) = 4.286$, *SEM* = 55.70356, $p<0.01$; imagery and recognition

old: $t(7) = 3.263$, *SEM* = 0.63203, p<0.05; imagery and recognition new: $t(7) = 6.279$, *SEM* = 0.47979, p<0.001; recognition old and delayed imagery: $t(7) = 3.694$, *SEM* = 0.40158, *p*<0.01; recognition new and delayed imagery: $t(7) = 7.588$, *SEM* = 0.32012, *p*<0.001; and between recognition old and recognition new: $t(7) = 3.801$, *SEM* = 0.24992, *p*<0.01. There were no reliable differences between encoding and recognition old ($p = 0.114$), encoding and recognition new ($p = 0.555$), or imagery and delayed imagery ($p = 0.179$).

Saccadic Amplitude
The average saccadic amplitude per picture was calculated for each experimental condition and a within-groups ANOVA was carried out (appendix 10). There was no main effect of experimental condition: $F(4,28) = 1.085$, *MSE* = 0.480, $p=0.383$.

Fixations outside the picture boundary
An average was calculated for the percentage of stimuli where participants fixated outside the picture boundary. This was calculated for each experimental condition (see figure 17) and a within-groups ANOVA was carried out (appendix 11a). There was a reliable main effect of experimental condition: $F(3,21) = 49.816$, *MSE* = 6.812, $p<0.001$.
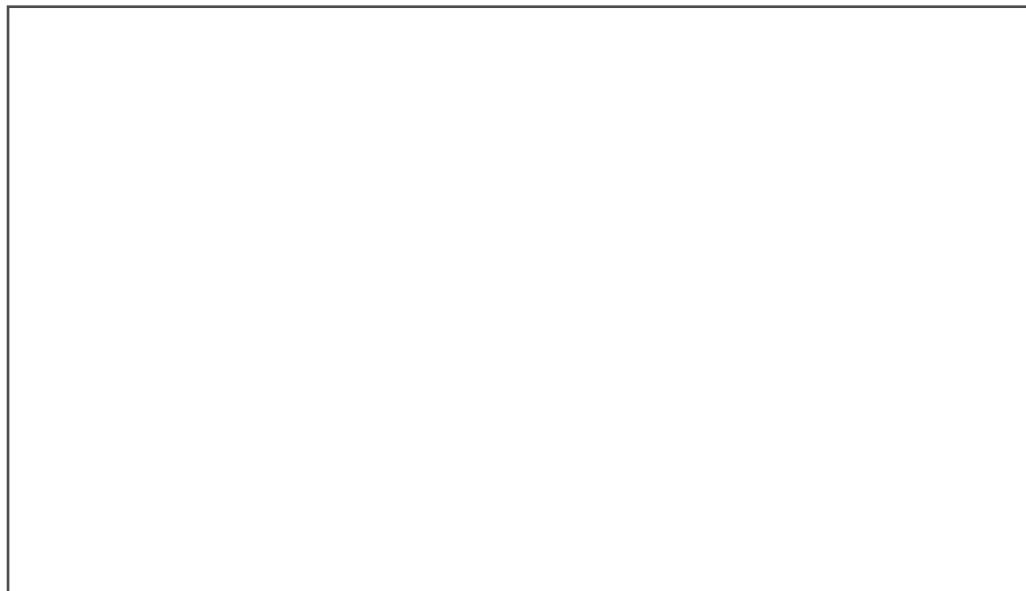

Figure 17: *A bar graph to illustrate the average percentage stimuli where fixations fell outside the picture boundary. Only 'old' stimuli in the recognition condition were analysed, partly because this analysis was carried out to help further explain differences in string similarities (therefore only old stimuli are comparable) and partly because none of the previous eye data*

analyses have found any reliable differences between old and new stimuli at recognition.

Post-hoc paired-samples T-tests were carried out (appendices 11b and 11c) and showed reliable differences between encoding and imagery: $t(7) = 8.037$, *SEM* = 1.39976, $p<0.001$; encoding and delayed imagery: $t(7) = 6.416$, *SEM* = 1.88325, $p<0.001$; recognition and imagery: $t(7) = 8.919$, *SEM* = 1.16794, $p<0.001$; and recognition and delayed imagery: $t(7) = 7.329$, *SEM* = 1.53498, $p<0.001$. There were no reliable differences between encoding and recognition ($p = 0.351$) or between imagery and delayed imagery ($p = 0.170$).

String comparisons

A string editing algorithm was used to analyse the similarity between scan patterns produced at each experimental condition. This was the same method of analysis as the previous experiment and used the same 5x5 grid structure to divide up the picture. The difference this time was that the squares were smaller (160x120 pixels instead of 320x240 pixels) to take into account the smaller image size. Any fixations that fell outside the central image boundaries were assigned numbers, lower-case letters and other symbols, depending on the position of the fixation. A chance baseline was calculated by comparing each scanpath produced by each participant to another randomly selected scanpath from the same participant. This was repeated for all participants and an average similarity score was produced (but see chapter 6 for discussion on calculating chance). Figure 18 illustrates the string similarity scores for each condition comparison.

Figure 18: *A bar chart to illustrate the differences in string similarity scores between the comparison types.*

A between-groups ANOVA (appendix 12a) showed a reliable main effect of comparison type: $F(5,35) = 14.101$, *MSE* = 0.002, *p<0.001*. Post-hoc paired samples T-tests showed were carried out (appendices 12b and 12c) and showed reliable differences between *encoding v imagery* and *encoding v recognition*: $t(7) = 4.339$, *SEM* = 0.0186499, *p<0.01*; *encoding v imagery* and *imagery v recognition*: $t(7) = 2.813$, *SEM* = 0.0157486, *p<0.05*; *encoding v imagery* and *imagery v delayed imagery*: $t(7) = 4.968$, *SEM* = 0.01292128, *p<0.01*; *encoding v recognition* and *encoding v delayed imagery*: $t(7) = 4.578$, *SEM* = 0.0182466, *p<0.01*; *encoding v recognition* and *imagery v delayed imagery*: $t(7) = 2.940$, *SEM* = 0.0218319, *p<0.05*; *encoding v recognition* and *recognition v delayed imagery*; $t(7) = 2.923$, *SEM* = 0.0193295, *p<0.05*; *encoding v delayed imagery* and *imagery v delayed imagery*: $t(7) = 6.935$, *SEM* = 0.0213024, *p<0.001*; *imagery v recognition* and *imagery v delayed imagery*: $t(7) = 3.175$, *SEM* = 0.0317499, *p<0.05*; and between *imagery v delayed imagery* and *recognition v delayed imagery*: $t(7) = 5.167$, *SEM* = 0.0233563, *p<0.01*. There were no reliable differences between: *encoding v imagery* and *encoding v delayed imagery* ($p = 0.901$); or between *recognition v imagery* and *recognition v delayed imagery* ($p = 0.105$). All string similarity scores were reliable greater than chance, apart from *encoding v imagery* and *encoding v delayed imagery* (appendix 12d).

2.3.4 Discussion

This experiment was conducted to determine whether eye movements were affected by the presence of a definite picture border and whether the more extensive scanpaths during imagery could be explained by participants remembering more of the picture than was actually present, causing them to look beyond the picture boundaries.

Analyses show that overall recognition memory is high (consistent with the previous experiment). The reliably higher accuracy for new pictures could indicate a recognition bias, i.e. when participants couldn't remember if they had seen a stimulus before or not, they were more inclined to say 'new'. If the picture was actually new, this increased the average accuracy score for new pictures, however, if the picture was old, the average accuracy for 'old' pictures decreased.

Average fixation durations and the average number of fixations on each picture during each condition were measured and analysed. The pattern of results was almost identical to that of the previous experiment. Participants made reliably more fixations at encoding and recognition than at imagery or delayed imagery, but fixated for longer during imagery and delayed imagery than encoding or recognition. When visual information was available (encoding and recognition), participants made more fixations, trying to take in as much as possible. Whereas when no visual information was available (imagery and delayed imagery), participants looked at each place for longer whilst trying to remember what was there during encoding, meaning they only had time to make a smaller number of fixations. The black border emphasising the picture boundaries therefore did not have an effect on the number of fixations or fixations durations.

Recognition data were split into 'old' and 'new' for two reasons: firstly to remain

consistent with the previous experiment, and secondly because differences were found in the analysis of recognition memory data. There were no reliable differences in the number of fixations or fixation duration between old and new stimuli at recognition.

Saccadic amplitude was measured and analysed. The previous experiment showed that the average saccadic amplitude for the imagery first condition was 3.169 degrees and that saccadic amplitude was reliably greater during imagery and delayed imagery. The current results show greatly reduced average saccadic amplitude (1.687 degrees) and no reliable differences between any of the experimental conditions. This could be because the picture was half the size and thus smaller eye movements were needed to guide attention around the scene (interestingly, the saccadic amplitude is also approximately half the size). This would be true in all conditions, and was reinforced by the surrounding black border, thus differences in saccadic amplitude seen in the previous experiment were reduced to the point of statistical unreliability.

However, participants may have produced smaller saccades but still fixated outside the picture boundaries. The results show that this is the case, on average, for 1.25% of the pictures at encoding, 2.08% of the pictures at recognition, 12.5% of the pictures at imagery, and 13.3% of the pictures at delayed imagery. The reliable difference between the imagery conditions and the conditions where visual information is available suggests that participants are fixating beyond the picture boundaries, despite the presence of the black border. This coincides with previous research into the area (e.g. Intraub and Richardson, 1989). This may also help further explain why string similarities are so low between conditions where visual information is present and those where it is not. As well as the 'pure' and 'mixed' comparison theory, boundary extension at imagery could lead to broader scanpaths, thus reducing their similarity to smaller scanpaths in other conditions. This existence of boundary extension could be confirmed by repeating the experiment and including either a recall memory test or a recognition test where participants are presented with variations of the same picture showing more, less, or the same amount of the scene as was shown at encoding. If the fixations outside the picture boundaries were caused by this memory illusion then it would be predicted that participants would incorrectly choose the pictures showing more of the scene than was present during encoding.

Analysis of the string similarities revealed a pattern almost identical to that of the previous experiment. The decreased scanpath similarity between encoding and imagery could be due to scaling, rather then boundary extention. Pictorial memories have bee found to be distorted in shape and size during imagery, therefore the larger saccadic amplitudes in experiment one, and the fixations outside the picture boundaries could be due to a re-sizing of the scanpath. In effect, participants could be making very similar patterns of eye-movements, but on a larger scale – so the problem is not that they are remembering more of the picture than was present at encoding, but simply that the scanpaths have nbeen scaled up. Further research could test this theory by reanalysing the eye dats, scaling scanpaths and recomputing scanpath similarity.

Further research could also focus on increasing the string similarities between the conditions, especially between imagery and encoding, for possible use in human-computer applications such as content based image retrieval. One way to do this could be to ask participants to give verbal feedback whilst they are imagining the previously-seen picture.

In conclusion, introducing the border to emphasise the picture boundaries did not affect the average number of fixations or average fixation duration, but did help us further understand the increased saccadic amplitudes during imagery and decreased similarity in scanpaths between conditions where visual information is present and those where it is not.

2.4. Experiment 3: Imagery and Verbalisation

2.4.1 Introduction

The link between verbalisation and imagery has been demonstrated on several occasions using the *visual world paradigm,* whereby participants are presented with a visual scene and eye movements are recorded as they either hear an instruction to manipulate objects in the scene or as they listen to a description of what may happen to those objects. This paradigm has been used whilst the picture is still present (e.g. Allopenna, Magnuson & Tanenhause, 1998), and when the visual information has been removed or was never present. For example, Altmann (2004) showed participants a scene, which was then removed and a corresponding sentences was heard. Participants looked towards parts of the blank screne that had previously contained information relevant to the sentence, even though the picture had been absent for over 2 second. Similarly, verbalising the memory of a previously seen picture has been shown to produce eye movements spatially related to both the verbal description and the original picture (Johansson, Holsanova and Holmqvist, 2006). These eye movements were produced when no visual information was present and thus provide evidence for the shared cognitive processes between imagery and eye movements.

This experiment is a simple exploratory study aimed at finding out whether scanpath similarity for a previously seen image can be improved by verbalisation of the memory, compared to simply imagining it.

2.4.2 Methodology

Participants

Sixteen participants took part in the experiment, all of whom were students at Nottingham University (undergraduates and postgraduates). The age range was 18-24 and the mean age was 20.5. The sample comprised 9 females and 7 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration.

Materials and apparatus

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker (see chapter 1.5 for details) A set of 35 high-resolution digital photographs were prepared as stimuli, sourced from a commercially available CD-ROM collection and taken using a 5MP digital camera. The photos were of easily describable scenes (see figure 19) and were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. All scenes were unique but of similar complexity.
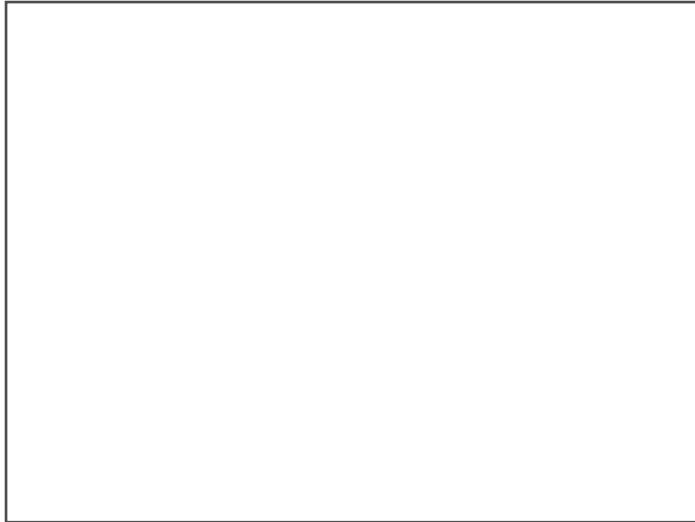
Figure 19: *An example of the stimuli.*

Design

The experiment used a between groups design, with 2 groups of participants (8 participants in each group). The independent variable was therefore which group the participant belonged to (The Imagery First group or The Description First group). The dependent variable measures were the similarities of scan patterns compared at encoding and imagery, and at encoding and verbalization (description).

Procedure

Participants were told that their pupil size was being measured in relation to mental workload. They were informed that although their eye movements were not being recorded, it was important to keep their eyes open so pupil size could be reliably measured. Following a 9-point calibration procedure, participants were shown written instructions on the experimental procedure and given a short practice.

Group 1: Imagery Prior to Description

Each stimulus was presented for 7000 milliseconds, preceded by a 1000 millisecond fixation cross. After each picture, a brightly coloured mask appeared for 1000 milliseconds and then an instruction to "visualise". The screen then appeared blank for 7000 milliseconds whilst the participant visualised the last picture they had seen. An instruction to "describe" was then presented for 1000 milliseconds and the screen appeared blank for another 10,000 milliseconds whilst the participants described the last picture they saw. After this time, the fixation cross reappeared to ensure that fixation at picture onset was in the centre of the screen. This experimental procedure is illustrated in figure 20.

Figure 20: *A diagram of the imagery prior to description procedure.*

Group 2: Description Prior to Imagery

The only difference between the groups was the order that they described and visualised each picture. Each stimulus was presented for 7000 milliseconds, preceded by a 1000 millisecond fixation cross. After each picture, a brightly coloured mask appeared for 1000 milliseconds and then an instruction to "describe". The screen then appeared blank for 10,000 milliseconds whilst the participant described the last picture they had seen. An instruction to "visualise" was then presented for 1000 milliseconds and the screen appeared blank for another 7000 milliseconds whilst the participants visualised the last picture they saw. After this time, the fixation cross reappeared to ensure that fixation at picture onset was in the centre of the screen. This experimental procedure is illustrated in figure 21.
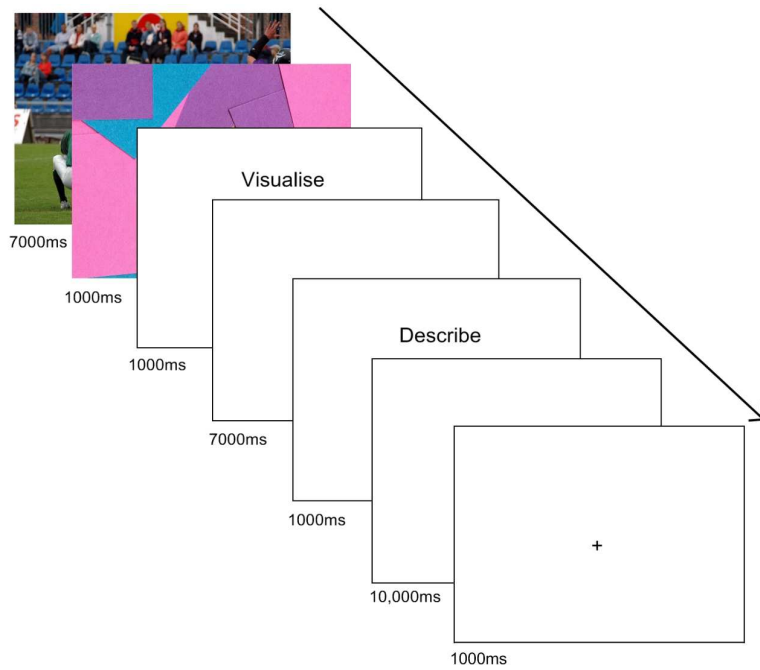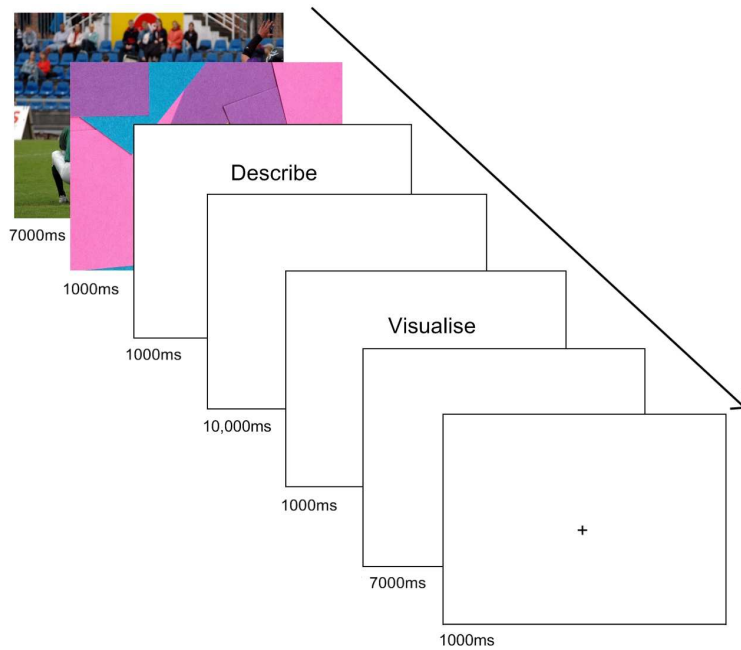
Figure 21: *A diagram of the description prior to imagery procedure.*

In both groups, a head-mounted microphone was worn by the participants. Although the objective of this experiment was purely to find out if the similarity of eye movements differed depending on whether pictures were described or simply visualized, the microphones were used so that participants believed that their descriptions were being recorded and were thus important to the experiment.

2.4.3 Results

The similarity of eye movements at encoding and visualisation and at encoding and description were analysed using a string editing technique. The similarity scores from these analyses were then compared.

Trial exclusions comprise of 27.3% of trials from the encoding and visualisation comparison and 7% of trials from the encoding and description comparison, due to participants not moving their eyes at all.

As in previous string editing analyses, strings were cropped to five letters to provide standardised and manageable data sets that were still long enough to display any emerging similarity (Foulsham and Underwood, 2008).

Paired samples T-tests found no reliable difference between the comparisons [$t(15) = 0.408$, SEM $= 0.01656$, $p=0.689$] (appendix 13a), but both string similarities for *encoding vs. visualisation* and for *encoding vs. describe* were reliably greater than chance (appendix 13b): $t(15) = 2.718$, SEM $= 0.01436$, $p<0.05$ and $t(15) = 3.674$, SEM $= 0.00879$, $p<0.01$, respectively. See figure 22.

Figure 22: A bar chart to show the similarity of eye movements at encoding and visualisation and at encoding and description, compared to an average chance similarity score of 0.1037. Stings in this analysis were cropped to 5 letters.

The average number of fixations were calculated for each of the 3 tasks (encode, visualise and describe). The least number of fixations were made in the visualisation task (average 13). The string editing analysis was repeated using this average as the string length instead of cropping to 5 letters. Paired samples T-tests found no reliable difference between the comparisons [$t(15)$ = 0.471, SEM = 0.01610, p=0.645] (appendix 13c), but both string similarities for *encoding vs. visualisation* and for *encoding vs. describe* were reliably greater than chance (appendix 13d): $t(15)$ = 2.595, SEM = 0.01744, p<0.05 and $t(15)$ = 4.574, SEM = 0.00824, p<0.05, respectively. See figure 23.
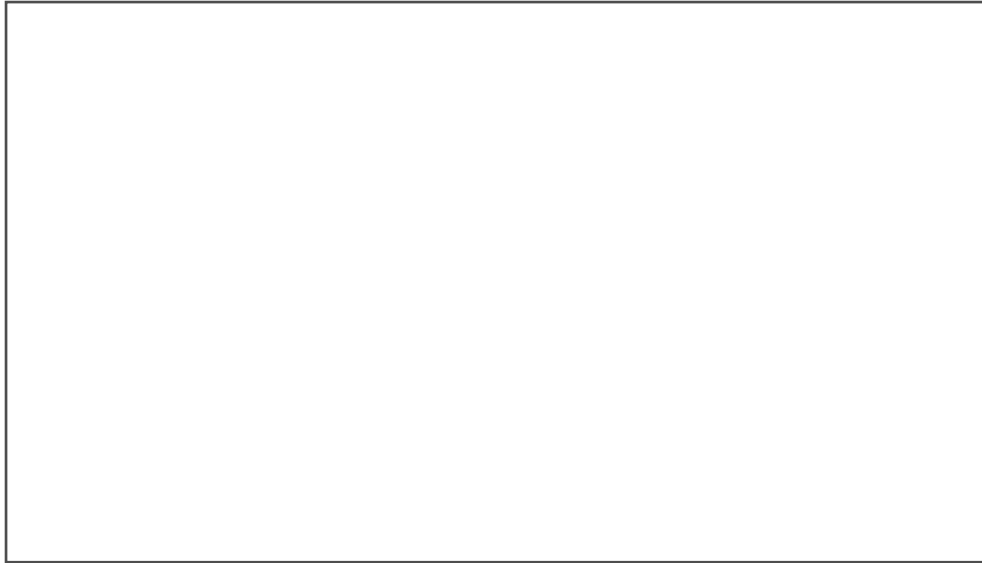

Figure 23: A bar chart to show the similarity of eye movements at encoding and visualisation and at encoding and description, compared to an average chance similarity score of 0.1. Stings in this analysis were cropped to 13 letters (based on the average number of fixations in the visualisation

task).

The above string analysis was repeated using unrestricted string lengths. A reliable difference was found between the string comparisons, $t(15) = 5.069$, SEM $= 0.00623$, $p<0.001$ (appendix 13e). Eye movements were reliably more similar to encoding when pictures were described from memory than when they were simply visualised. The encoding-describe comparison was reliably greater than the average chance similarity of 0.1463, $t(15) = 5.425$, SEM $= 0.00599$, $p<0.05$, but the encoding-visualisation comparison was not, $t(15) = 0.116$, SEM $= 0.03131$, $p = 0.909$. See figure 24 (and appendix 13f).
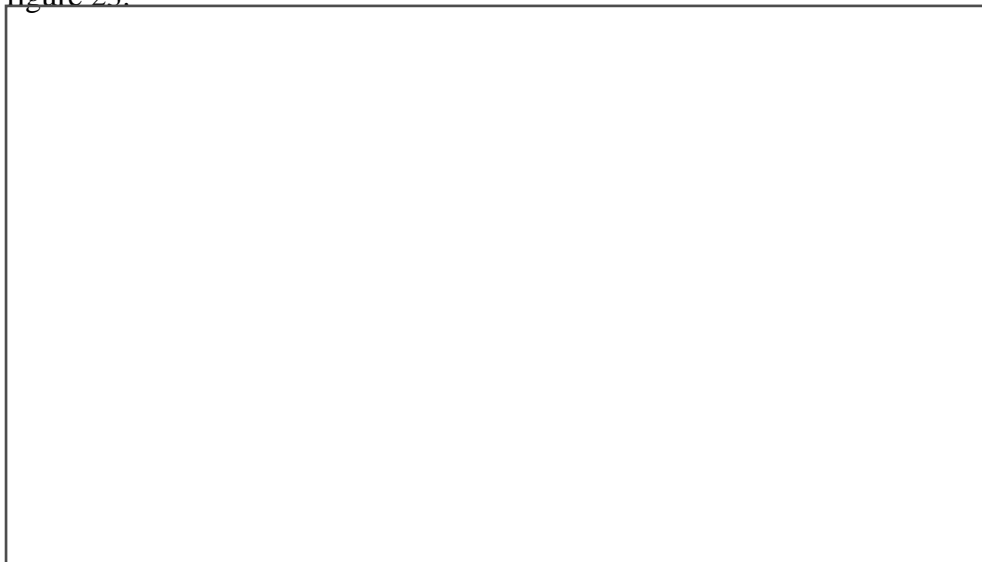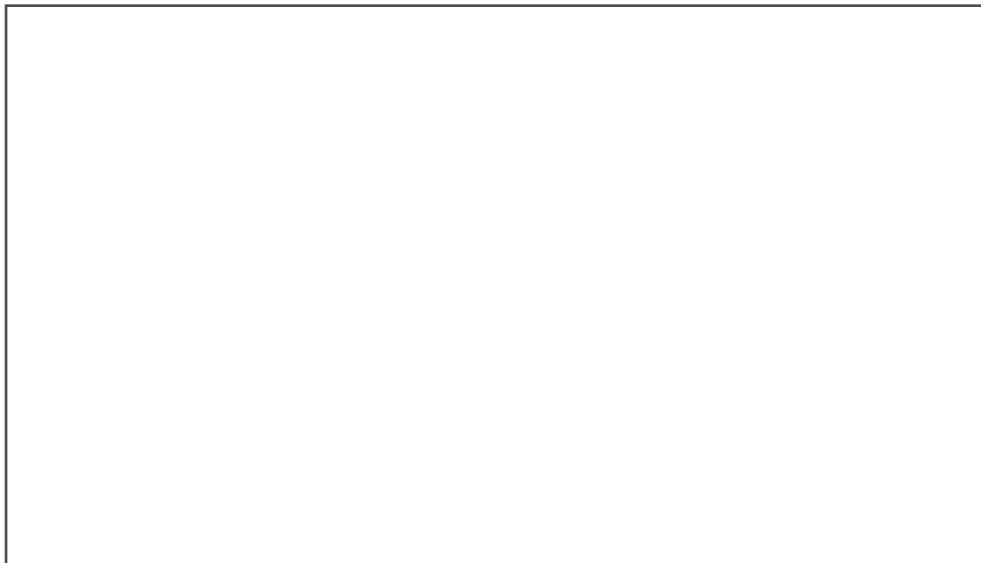

Figure 24: A bar chart to show the similarity of eye movements at encoding and visualisation and at encoding and description, compared to an average chance similarity score of 0.1. Stings in this analysis were unrestricted in length, although the computer program cropped to the length of the shortest string in the comparison.

2.4.4 Discussion

This experiment was a simple exploratory study aimed at finding out whether scanpath similarity for a previously seen image can be improved by verbalisation of the memory, compared to simply imagining it. Scanpath analyses cropped to strings of 5 and 13 letters showed no reliable differences between the groups or between the comparisons. However, when the strings were left unrestricted (and automatically cropped to the length of the shortest string in the comparison by the computer algorithm), eye movements were found to be reliably more similar to encoding when pictures were described from memory than when they were simply visualised. It could be that in the describe task, participants spent the first few fixations thinking what to say and the later fixations were the important ones related to description. As a result, participants made reliably more fixations during description than during visualisation but when the strings were cropped, these later fixations were lost. This could explain why string similarity was reliably increased when string lengths were unrestricted. A consequence of having more letters in a string is that each incorrect match between the two comparable strings counts for less and the overall score is

statistically more likely to be higher (more similar). For example, if strings are cropped to 5 letters and 4 or those letters are incorrect matches, the similarity score falls to 0.2. However, if the strings are unrestricted in length and contain, for example, 30 letters, then 4 incorrect matches, would give 0.93. Of course, the longer the string, the more chance of getting incorrect matches, but with 25 squares in the analysis grid, there is more chance of replicating a scanpath over the whole grid in 30 fixations than in 5, thus increasing the chance of similarity. This explains why the chance level is so high. Refer to chapter 6 for further discussion on string analysis.

Overall, there is some evidence for increased scanpath similarity by describing previously viewed scenes compared to simply imagining them (p<0.05). This finding can now be elaborated on by investigating which bottom-up and top-down factors affect these scanpaths during imagery and verbalisation. More exhaustive analyses can also be employed such as recognition accuracy, number of fixations, fixation durations, locations of fixations and content of verbalization. Further research should also take into account methodological issues such as giving the participants enough time to prepare before verbalisation.

2.5. Chapter 2 conclusions

In support of previous eye movement theories (e.g. Hebb, 1968; Postle, Idzikowski, Sala, Logie, Baddeley, 2006) this chapter has provided evidence for a link between eye movements and imagery. This can be concluded due to the fact that participants moved their eyes during imagery without being told to do so and secondly that scanpaths at encoding and imagery were reliably more similar than would be expected by chance. Furthermore, the finding that eye movements are still reproduced even when no visual information is present argues against a purely bottom-up explanation of scan pattern similarity (Brandt and Stark, 1997; Laeng and Teodorescu, 2002). Evidence was presented for the stability of scanpaths at imagery over time, suggesting that movement sequences are (to an extent) replicable over multiple viewings and also that they do not solely rely on bottom-up visual features.

This chapter has also shown that the reorganization of a mental image during the imagery stage can lead to increased saccadic amplitudes, implying a 'stretching' of the metal image. This is contradictory to previous findings of Finke (1989) and Johansson et al (2006) that suggested the opposite affect: a 'shrinking' of the mental image. Further experiments concluded that these greater saccadic amplitudes were due to participants extending their eye movements outside the boundaries of the picture ('boundary extension'), which coincides with previous research into the area (e.g. Intraub and Richardson, 1989). The fact that the results showed no reliable difference in saccadic amplitudes between the imagery and delayed imagery conditions suggests that the reorganizing of mental images may take place between encoding and first imagery and then stays relatively stable over multiple imagery tasks.

Furthermore, the most similar scanpaths came from 'pure process' comparisons where there was similar visual input in each condition (imagery compared to delayed imagery and encoding compared to recognition), and from comparisons that mimicked the same retrieval processes. This suggests that the visual buffer model may be more complicated than simply shifting attention to different parts of an internal image (Kosslyn, 1994). The relationship between the encoding and retrieval process seems to be very important and one might even suggest the existence of facilitatory and inhibitory pathways within the model.

Chapter 3: Top-Down expertise vs. Bottom-Up saliency

3.1 Introduction

Expertise can be defined as skilfulness by virtue of possessing special knowledge and has been shown to affect memory, eye movements and, as a top-down cognitive advantage, moderate the influences of bottom up factors such as saliency. Brain imaging studies have suggested underlying neurological mechanisms involved in expertise. This so-called 'expert recognition pathway' (Draper, Baek, and Boody, 2004) begins in the early visual system (retina, LGN/SC, striate cortex) and is defined by subsequent diffuse activation in the lateral occipital complex (LOC) and sharp foci of activation in the fusiform gyrus and right inferior frontal gyrus. This pathway recognizes familiar objects from familiar viewpoints under familiar illumination. More interestingly, Tarr and Gauthier (2000) demonstrated, through fMRI, that this expert recognition pathway could be trained. They found that training participants to recognise and classify 'greebles' resulted in increased activation of the fusiform gyrus.

Chase and Simon (1973a, 1973b) propose that the enhanced memory experts demonstrate is due to encoding and storing information in 'chunks', which is then stored in "retrieval structures" (Chase and Ericsson,1982; Staszewski, 1990) in long term memory, allowing for easy retrieval of associated memories. Chess masters are a classic example of an expert group who exhibit this memory advantage and can recall several chess boards that have been presented successively (Cooke, Atlas, Lane & Berger, 1993; Gobet & Simon, 1996). Ericsson and Kintsch (1995) suggested that the "retrieval structures" in chess memory are hierarchical and mapped onto the 64 squares of a chess board. However if this was true, individual pieces could be associated with individual squares, thus accurately encoding countless random configurations into memory, which has not been found to be the case.

This 'chunking' technique has been linked to faster recognition of notes (Bean, 1938; Salis, 1980; Sloboda, 1978) and enhanced encoding of musical information (Clifton, 1986; Halpern & Bower, 1982; Sloboda, 1976; Thompson, 1987) in skilled music readers, and increased recall of domain-specific knowledge in baseball experts (Voss, Vesonder, & Spilich, 1980) and athletics enthusiasts (Chase & Ericsson, 1982). This enhanced recall and recognition performance increases linearly with level of expertise, regardless of aptitude level (Walker, 1987), and can be applied to solve domain-related problems (Chi et al., 1982; Siegler & Richards, 1982). Myles-Worsley, Johnston and Simons (1988) found that radiological experience was linked to increased recognition memory for abnormal x-ray films but decreased memory for normal x-ray films. They suggested that these experts learn to recognise and diagnose abnormalities faster (Christensen et al., 1981) and more efficiently, but as a consequence do not encode information irrelevant to their jobs (i.e. 'normal' x-rays). Haider and Frensch (1999) called this that 'information-reduction' hypothesis where with increasing expertise people learn to distinguish between relevant and irrelevant information and therefore concentrate on processing mostly relevant information, Jarodzka, Scheiter, Gerjets and Van Gog (2009, *in press*) found that experts attend more to relevant aspects of the stimulus, use more heterogeneous task approaches, and use knowledge-based shortcuts.

Memory for experiences and events can effect eye movements for previously viewed stimuli. Smith and Squire (2008) found that participants explore old scenes less than new ones, and when shown an old scene that had been modified, participants preferentially look at the changed region, but only if they are aware of the change. Smith and Squire concluded that eye

movements were linked to hippocampus-dependant memory, a fundamental characteristic of which is awareness of what is learned. The pattern of eye movements also varies with the level of experience, for example Manning, Ethell, Donovana and Crawford (2006) found that experts covered the visual scene in longer eye movements without fixating at all on large amounts of the image, and non-experts were more likely to show shorter saccades thereby fixating on a greater amount of the image. The chunking method of encoding information used by chess experts results in fewer fixations, and fixations between rather than on individual pieces (Chase and Simon, 1973a,b). Such a visual-span advantage would also mean that while examining structured, but not random, chess configurations, experts may make greater use of parafoveal processing to extract information from a larger portion of a chessboard during an eye fixation. Evidence for increased visual span in chess experts comes from Reingold, Charness, Pomplun, and Stamper (2001) who tested chess masters using a combination of the gaze-contingent window paradigm, the change blindness flicker paradigm and a check detection task. Experts made fewer fixations per trial than less-skilled players, and had a greater proportion of fixations between individual pieces, rather than on pieces.

This experience-dependant variation in eye movements has also been documented in the domains of music, driving, industry and leisure. For example, Goolsby (1994a, 1994b) found that skilled music readers used more, but shorter, fixations and also looked father ahead in the notation than less skilled readers. When comparing music scores, more experienced musicians do so faster with fewer, and shorter, glances between the patterns (Walters, Underwood and Findlay, 1997). Experienced musicians also show a greater degree of accuracy in sight-reading and comparison tasks (Gilman & Underwood, 2003; Walters & Underwood, 1998). In driving research, novices have been found to hold their attention on hazards longer than more experienced drivers (Chapman & Underwood, 1998) and in both actual and simulated driving situations, fixation durations were longer for the experienced drivers than for the novice drivers (Carter and Laya, 1998). This, however, has been found to be dependant on the complexity of driving situations as well as driving experience (Crundall and Underwood, 1998). For example, on dual-carriageways, experienced drivers made shorter and more spatially varied fixations than novices, but on rural roads novices made shorter, more spatially varied fixations than experienced drivers. One particular group of driving experts is the police. Police drivers have been found to have a wider search strategy and spend more time inspecting peripheral hazards such as parked vehicles and side roads (Crundall, Chapman, France, Underwood & Phelps, 2005). This could be due to the nature of their jobs, i.e. the constant visual search for potential dangers.

Similarly, pilots and aviation specialists make different eye movements to novices in the field. Experts were found to spend less time finding and fixating on navigational landmarks, while novices exhibited greater difficulty finding landmarks and extracting useful data from them, resulting in increased response times (Ottati, Hickox, and Richter, 1999). Furthermore, professional radar operators have been found to be better at tracking multiple targets and are less susceptible to interference than novices (Allen, McGeorge, Pearson and Milne, 2004). In an airport, the importance of expertise in visual search extends beyond the people who fly the planes. Especially in this day and age, security is more important than ever before and much research has been invested into identification of potentially dangerous items in luggage screening. For example, training in the recognition of knives in x-ray images resulted in faster fixation and enhanced recognition of the target (McCarley, Kramer, Wickens, Vidoni and Boot, 2004).

In a different type of x-ray scanning, radiologists show an enhanced ability to identify abnormalities than novices, and do so within the first 0.5 seconds of viewing (De Valk and

Eijkman, 1984; Nodine, Mello-Thoms, Kundel, and Weinstein, 2002). Such increased effectiveness of visual scanning could reflect strategic expertise in planning scan paths (Kundel & La Follette, 1972) or perceptual expertise in noticing and guiding the eyes toward peripherally viewed targets (Kundel, Nodine, & Toto, 1991).

The effect of expertise on eye movements has also been seen in leisure activities such as sports and computer gaming. For example, expert gymnastic coaches make longer and fewer fixations than novices when inspecting videos of gymnastic techniques. With relation to Haider and Frensch's (1999) information-reduction hypothesis, expert coaches attended to more informative (i.e., relevant) areas and ignored uninformative (i.e., irrelevant) ones (Moreno, Reina, Luis, and Sabido, 2002). Furthermore, evidence from both cricket and football show that eye movement patterns differ between beginner and expert players (Land & McLeod, 2000; Williams & Davids, 1998; Williams, Davids, & Williams, 1999). When a ball is bowled in cricket, the batsman makes an anticipatory saccade to the location where the ball will bounce before it reaches him/her (Land & McLeod, 2000). More skilled players fixate the 'bounce point' sooner, suggesting that they are better at determining the trajectory of the ball when it is pitched. In football, when determining the direction of an oncoming kick, more skilled players tend to fixate the midsection of the opponent player rather than fixating the legs or ball, as unskilled players do (Williams & Davids, 1998). This difference suggests that the skilled players have developed an eye movement strategy that focuses centrally, relying on their peripheral attention to monitor the movements of the opponents' limbs. When watching more complicated situations that involved everyone on the field, the experts tended to make many more saccades with quicker intermittent fixations throughout the field of play while the novices showed a slower and less comprehensive search pattern (Williams et al., 1999). These examples show that eye movement strategies develop with experience as knowledge about the information most critical for selection is gained.

Experienced video game players demonstrate enhanced visual acuity (Green & Bavelier, 2007), maintain a wider field of attention (Green & Bavelier, 2003; 2006a), have a higher resolution of temporal attention (Green & Bavelier, 2003), can better track multiple moving objects (Green & Bavelier, 2006b), and they generally respond faster (e.g., Castel et al., 2005; Green & Bavelier, 2007). Importantly, these benefits represent skills enhanced via videogame experience - when novices are trained in videogame play, they demonstrate specialist-like advantages (e.g., De Lisi & Cammarano, 1996; Gopher, Weil, & Bareket, 1994; Green & Bavelier, 2003, 2006a, 2007). Experienced video-game players are less susceptible to crowding, whereby visual processing is hindered with decreased distance between distracters and the target (Green and Bavelier, 2007). This suggests enhanced spatial processing in experienced players. Computer expertise has also been linked to faster and more efficient visual search performance (Aula, Majaranta, and Räihä, 2005).

Face recognition is an area where we are all considered experts (excluding patients with brain damage that affects their ability to recognise faces, e.g. prosospagnosia). We have a natural ability to recognise and discriminate between faces, even when we are very young. There are, however, still variations in scanpaths to faces that are *more* familiar than others. Althoff and Cohn (1999) found that when asked to judge how famous a person was, participants viewing famous faces (the 'domain' we have more experience in) made fewer fixations, sampled less regions, and fixated facial features more symmetrically then when viewing non famous faces. When asked to judge the emotion of a famous face, participants produced less constrained, more symmetric scanpaths than when viewing non famous faces.

However, despite this wealth of evidence for experience moderated eye movements, some

studies have failed to find an effect. For example, Abernethy and Russell (1987a,b) found no difference in eye movements between expert and novice badminton players. However, they did find attentional differences, in that experts were able to pick up anticipatory information earlier than novices. Furthermore, Croft, Pittman and Scialfa (2005) measured eye movements of five experienced and 5 novice 'spotters' whilst carrying out an air-to-ground search task. Inter-fixation amplitude was significantly related to task success, which was independent of fixation rate, fixation duration, and inter-fixation duration. They found that experience did not predict task success.

Evidence has been found for a moderating effect of domain knowledge upon the influence of visual saliency in scene recognition. For example, Parkhurst, Law & Niebur (2002) found that people looked in salient regions more for pictures where top-down knowledge was limited, e.g. pictures of computer-generated fractals. Henderson, Brockmole, Castelhano, and Mack (2007) found that during an active search task, neither region-to-region saccades nor saccade sequences were predicted any better by visual saliency than by a random model. They conducted additional analyses on the fixated regions in their scenes and found that these were more likely to provide meaningful information about the scene i.e. be semantically informative. This cognitive override of visual saliency is consistent with the Findlay and Walker's (1999) model of saccade generation, which acknowledges top-down influences in eye guidance. The process of "intrinsic saliency" uses the viewers own knowledge of the scene to guide fixations to probable target locations. Underwood, Foulsham and Humphrey, (2009) researched the bottom-up effect of visual saliency, in conjunction with the top-down effect of cognitive knowledge in the form of domain proficiency. The effect of saliency was supported in determining scanpaths but this effect was found to be reduced when specialists viewed images from their domain. Domain proficiency therefore caused an overriding effect of visual saliency in determining scanpaths. However, participants do not necessarily have to possess a specific domain expertise to exhibit this overriding effect. Henderson, Malcolm, and Schandl (2009, *in press*) found that simple knowledge of where things are likely to be in scene is enough to moderate the influence of saliency. They found that participants fixated objects that were in semantically relevant places rather than objects in the scene that were more visually salient.

Lansdale Underwood and Davies (2009) argue that low-level analysis of visual saliency has been found to occur in both untrained and expert viewers; however the guidance of eye movements, (i.e. attending to visually salient regions) is regulated by the viewer. In domain-specific tasks, experts selectively fixate regions on the basis of semantically salient features as opposed to the more visually salient, whereas untrained viewers' fixations fall upon visually salient items since their lack of domain-proficiency means that semantic saliency information is not available to them. In certain tasks, such as the spatial memory task in Lansdale et al.'s study, fixating on visually salient regions can benefit performance. This explains why a strong influence of visual saliency was found, even in experts, in this task, as they attended to such regions selectively.

In summary, domain expertise has been shown to enhance both recall and recognition memory for domain-specific material, affect the way in which we move our eyes, and have an overriding effect of low-level visual saliency. This chapter will further explore the effects of expertise and saliency on scanpaths, extending methods used in chapter 2 such as imagery and verbalisation, as well as introducing new concepts (incongruence) and methods (comparative visual search tasks) and also investigating the stability of these effects over extended time periods of time.

3.2. Experiment 4: The effects of expertise and saliency on scanpaths at encoding and recognition

3.2.1. Introduction

The Saliency Map model of eye guidance proposes that attention is drawn to the most salient region in an image first, followed by the second most salient region then the third most salient region, and so on (Koch and Ullman, 1985; Itti and Koch, 2000). Support for the effect of saliency on fixation locations comes from Parkhurst, Law, and Niebur (2002), who showed viewers a range of images and recorded eye movements. Saliency strongly predicted fixation probability during the first two or three fixations, and the model performed above chance throughout each trial.

An interesting question is whether the effects of saliency are stable over time and over multiple viewings. Parkhurst et al. concluded that saccades are guided by a low-level saliency and more so when top-down factors can play less of a role, as, for instance, in natural scenes with many objects. They argue that saliency has a greater impact in the early fixations and decreases over the time of the trial. However, Tatler, Baddeley and Gilchrist (2005) argued that the finding that saliency is more involved in targeting early fixations than later ones is due to methodological limitations with the technique applied by Parkhurst et al (2002). Tatler et al. reported that they also observed such an interaction when they failed to correct for central fixation bias but this disappeared when the appropriate correction was used. On the other hand, more recent work by Carmi and Itti (2006) on dynamic scenes supports Parkhurst's position.

Scanpath Theory states that eye movements made when first viewing a picture, help encode the picture into memory and that recreating those eye movements facilitates recognition. A bottom-up explanation for similarities in scanpaths at encoding and recognition could be that fixation locations are at least partly determined by salience and this remains constant over viewings. However, it could be argued that this bottom-up effect of saliency could be reduced by increasing top-down knowledge of the scene. Furthermore, if an effect of domain knowledge on saliency affects eye movements, it would be interesting to see if it is consistent over multiple viewings (i.e. if scanpaths remain less affected by saliency when viewing the same picture a second or third time). This has not been specifically investigated before in non-search tasks, although there have been studies that have found a cognitive override of saliency in search tasks (e.g. Underwood et al., 2000). This said, no studies to date have yet considered whether these findings are constant over multiple viewings.

In the current experiment, participants consisted of two groups of domain specialists (Engineering Undergraduate students, and American Studies Undergraduate students) and a control group. Although the specialists cannot be strictly classified as 'experts', they did have a high level of domain-specific knowledge, an imperative component of expertise, into which there has been much research.

Expertise has been shown to enhance memory and performance on cognitive tasks, for example Walker (1987) found that on both recall and recognition tests, performance was a function of level of expertise in the domain. This enhanced performance has also been shown for visual search tasks, for example, McCarley, Kramer, Wickens, Vidoni, & Boot, (2004) examined visual performance in a simulated luggage-screening task. Sensitivity and response times to find knives hidden in x-ray images improved reliably with expertise.

Recognition accuracy could be used as a baseline to justify whether participants have sufficient domain-specific knowledge to be classed as 'specialists'. If, for example, participants

were just as accurate (or inaccurate) at recognising pictures from their own domain as they are with control pictures, then their position as 'specialists' might be questioned. This is important because in order to investigate the true effect of domain knowledge on eye movements, one has to ensure the participants are really specialists in that domain.

As mentioned above, the scanpaths of non-specialists are influenced by saliency, but if domain specialists produce different eye-movements to non-specialists on the same picture, it would provide support for the overriding effects of domain knowledge. Research has shown that experts' eye movements do differ from non-experts, for example, experienced football players have been found to have a higher search rate, involving more fixations of shorter duration than novice players, (Williams & Davids, 1998).

Non-specialists have also been shown to focus on small detail rather than the bigger picture. For example, Lowe (2004) found that when subjects were asked to make predictions on weather maps, novices adopted more 'local' strategies, examining the maps bit-by-bit at the expense of the more global details. Similar results have been found in the detection of pulmonary nodules in radiology. Expertise, experience and training were associated with larger saccade amplitudes. Experts were more likely to cover the visual scene in longer, sweeping eye movements leaving larger areas of the image un-fixated whereas the non-experts used short-interval, point-by-point examination of the visual scene (Manning, Ethell, Donovan & Crawford, 2006).

However, there is little evidence of how saliency has differing effects on the eye movements of domain-specialists and non specialists in a recognition task. Furthermore, if eye movements are related to memory, then the overriding effect of domain knowledge should be constant over time, producing similar scanpaths on multiple viewings of the same stimulus. Although research has shown that specialists are consistently more accurate at recognition of domain specific targets (McCarley et al, 2004), and that they consistently produce scanpaths reliably different from non-specialists (Manning et al 2006), there have been no scanpath comparisons carried out. Therefore, the links between scanpath replication, saliency and expertise cannot yet be identified from previous research.

The current experiment offers the chance to further investigate the similarity of scanpaths on encoding and second viewing of a naturalistic picture and how this is influenced by saliency. It also offers the opportunity to explore how domain knowledge influences the relationship between saliency and scanpaths, and whether a combination of top-down and bottom-up factors determine scanpaths during visual inspection.

3.2.2 Methodology

Participants
All participants were students at Nottingham University (undergraduates and postgraduates), and consisted of 15 Engineers, 15 American Studies students and 15 non-specialists (control group). A requirement of the American Studies group was that they had to have taken a core module on The American Civil War.

The age range was 18-30 and the mean age was 22. The sample comprised 24 females and 21 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and three participants had to be replaced due to technical difficulties.

Materials and apparatus

A set of 90 high-resolution digital photographs were prepared as stimuli, sourced from a commercially available CD-ROM collection and taken using a 5MP digital camera. Of this set of 90, 30 were engineering-specific, 30 were Civil War specific, and 30 were of natural scenes such as gardens, parks and landscapes (control stimuli) – see figures 25a, 25b, and 25c for examples of the stimuli.

Figure 25a: *An example of the 'civil war' category of stimuli.*

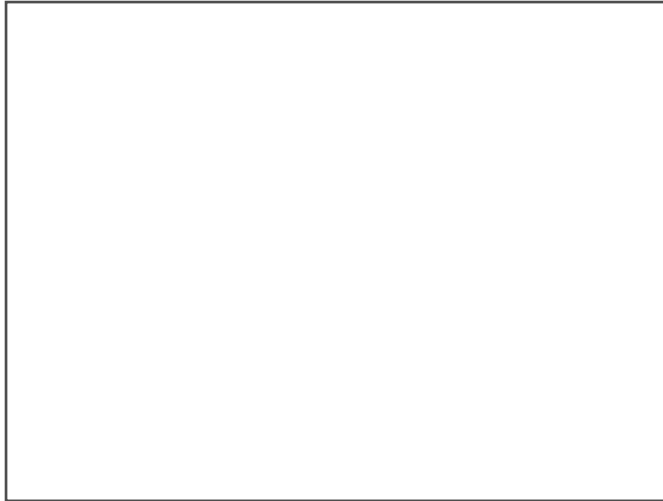***Figure 25b:*** *An example of the 'engineering' category of stimuli.*

Figure 25c: An example of the 'neutral' category of stimuli.

Half of each category were designated "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test. Pictures were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. The monitor measured 43.5cm by 32.5cm, and a fixed viewing distance of 98cm gave an image that subtended 25.03 by 18.83 degrees of visual angle.

Saliency maps were generated, using a version of Itti and Koch's (2000) model (for an overview, see appendix 1b) compiled from source code available at http://ilab.usc and downloaded in May 2004. Standard parameters were used, as far as possible. For example the default setting for the FOA is a size equivalent to 1/16th of the image, which it is argued is a realistic estimate of the resolution of human visual attention. These maps were produced for the first four simulated shifts and thus indicate the first five most salient regions for each picture (see Figure 26 for an example). The only further criterion for stimuli was that all 5 salient regions were non-contiguous. Those pictures where the same or overlapping regions were re-selected within the first 5 shifts were replaced.
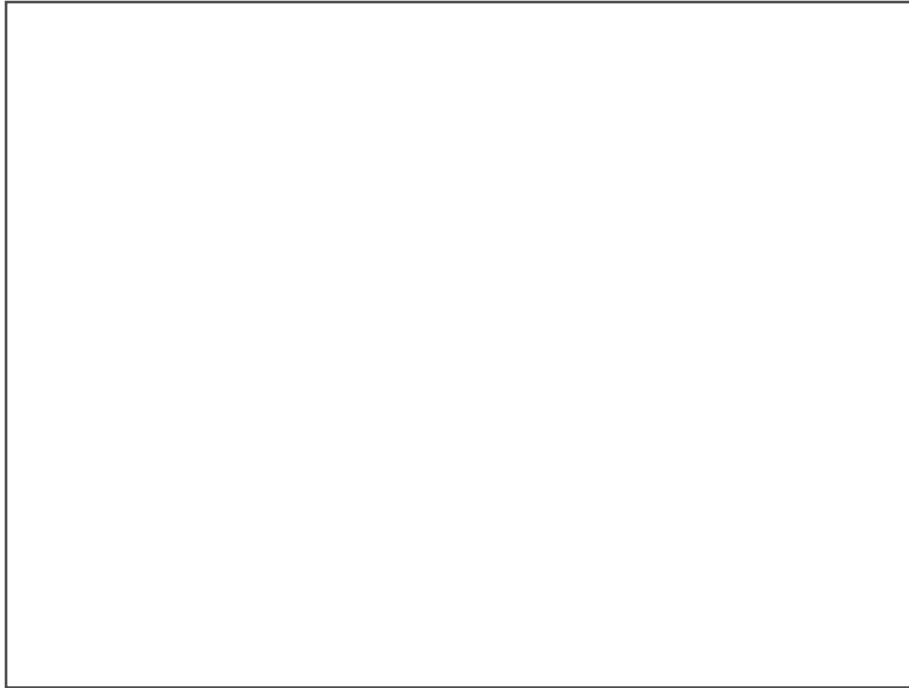
Figure 26: *An example of a scanpath predicted by the saliency map model. The five most salient areas are marked by yellow circles and the red arrows specify the order of the sequence of eye movements. The stimulus is an example from the 'Civil War' category.*

Design

The experiment used a three-by-three mixed design; with three specialist groups of participants and three specific types of stimuli. The specialist groups were Engineers and American Studies students, who were both domain specialists in their field. The third group consisted of non-specialists (also students at Nottingham University), who acted as controls. The independent variables were therefore which group the participant belonged to and the type of stimulus being shown. The dependant variable measures were: accuracy in deciding whether a picture was old or new; location of fixations relative to salient regions; total number of fixations; saccadic amplitude; average fixation duration; the similarity of the scanpath compared to that predicted by the saliency map; the similarity of the scanpath when comparing recognition and saliency; and the similarity of the scanpath when comparing encoding and second viewing of a picture.

Procedure

Following a 9-point calibration procedure, participants were shown written instructions asking them to inspect the following pictures in preparation for a memory test.

In a practice phase, participants were shown a set of five photographs that were similar to the ones in the experimental set, but did not fall into any of the three distinct experimental categories. The practice aimed to familiarise participants with the equipment, the displays and the task. Participants were not told to look for anything in particular in any of the pictures but were asked to look at them in preparation for a memory test. Following the practice phase, the first stage of the experiment began. There were 45 stimuli (15 engineering pictures, 15 Civil War pictures and 15 natural scenes) presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen. Each picture was

presented for 3000 milliseconds, during which time participants moved their eyes freely around the screen.

After all 45 stimuli had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures). Participants were instructed to press "N" on the keyboard if the picture was new, and "O" on the keyboard of the picture was old.

During this phase, 90 stimuli were presented in a random order; 45 of these were old and 45 new (though the participants were not informed of this fact). In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 3000 milliseconds and participants could only make a response after this time. This was to encourage scanning of the whole picture so that scanpaths from the first and second phases of the experiment could be compared. At the start of the second phase, participants were given a practice of the task, using 10 photographs that were similar to the ones in the experimental set, but did not fall into any of the three distinct experimental categories. Five of these were the practice photographs from the first part of the experiment. Feedback was given in the practice phase as to whether or not the participant gave the correct response of "old" or "new". No feedback was given in the experimental phase.

3.2.3. Results

Trials were excluded where the fixation at picture onset was not within the central region (the central square around the fixation cross when the picture was split into a 5x5 grid at analysis), when participants looked away from the screen (e.g. to the keyboard), or when calibration was temporarily interrupted (e.g. if the participant sneezed, therefore removing their head from the eye tracker).

There were 2 main types of data, recognition memory data (accuracy); and eye tracking measures – mean number of salient fixations, mean total number of fixations per stimulus, average saccadic amplitude and string analyses (encoding compared to second viewing, and encoding compared to saliency).

3.2.3.1 Recognition Memory

Accuracy
Engineers were more accurate with engineering stimuli and American Studies students seem to be more accurate with Civil War stimuli. Accuracy was measured by the number of pictures participants correctly identified as 'old' (if they were from the previous set) or 'new' (if they had never been seen before). This was done for each category of stimuli, and was out of 30 (as there were 30 pictures in each category in the second phase of the experiment). A mixed-design ANOVA was conducted and found these differences to be statistically reliable (see appendix 14a for descriptive statistics).
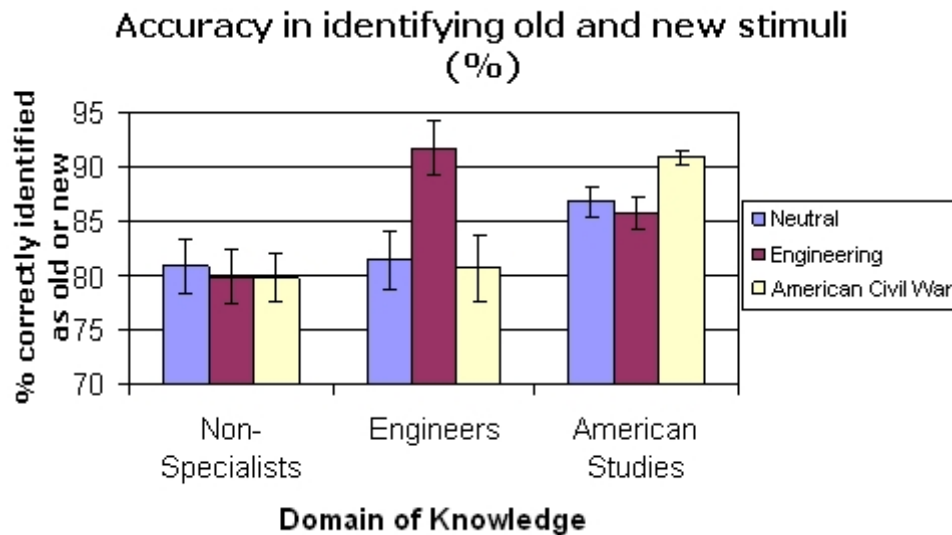
Figure 27: *A bar chart showing the percentage of pictures correctly identified as old or new on the recognition memory test for the three groups of participants on the three categories of stimuli. The error bars represent standard error.*

There was a main effect of group, $F(2,42) = 3.9$, $MSe = 15.527$, $p<0.05$ (appendix 14b). There was no statistically reliable main effect of stimuli, [$F (2,84) = 2.984$, $MSe = 2.729$, $p = 0.056$], but there was an interaction between group and stimuli, $F(4) = 10.004$, $MS = 27.296$, $p<0.05$ (appendix 14c).

Paired samples T-tests were carried out to further investigate the interaction between group and stimuli (appendices 14d and 14e). American Studies students were reliably more accurate at recognising Civil War stimuli than Engineering stimuli or neutral stimuli: $t(14) = 4.380$, $MSe = 0.35006$, $p<0.01$; and $t(14) = 3.674$, $MSe = 0.32660$, $p<0.01$, respectively. American Studies students were also reliably more accurate at recognising civil war stimuli than Engineers and control participants: $t(14) = 3.419$, $MSe = 0.89691$, $p<0.01$; and $t(14) = 4.799$, $MSe = 0.69465$, $p<0.001$, respectively. Engineers were reliably more accurate at recognising Engineering stimuli than Civil War stimuli or neutral stimuli: $t(14) = 3.927$, $MSe = 0.84890$, $p<0.01$; and $t(14) = 3.570$, $MSe = 0.87759$, $p<0.01$, respectively. Engineers were also more accurate than control participants at recognising Engineering stimuli: $t(14) = 3.603$, $MSe = 0.99905$, $p<0.01$.

3.2.3.2 Eye-tracking measures

Locations of fixations
It was found that domain-specialists made fewer fixations in salient areas when looking at domain-specific stimuli. Figure 28 illustrates this interaction.
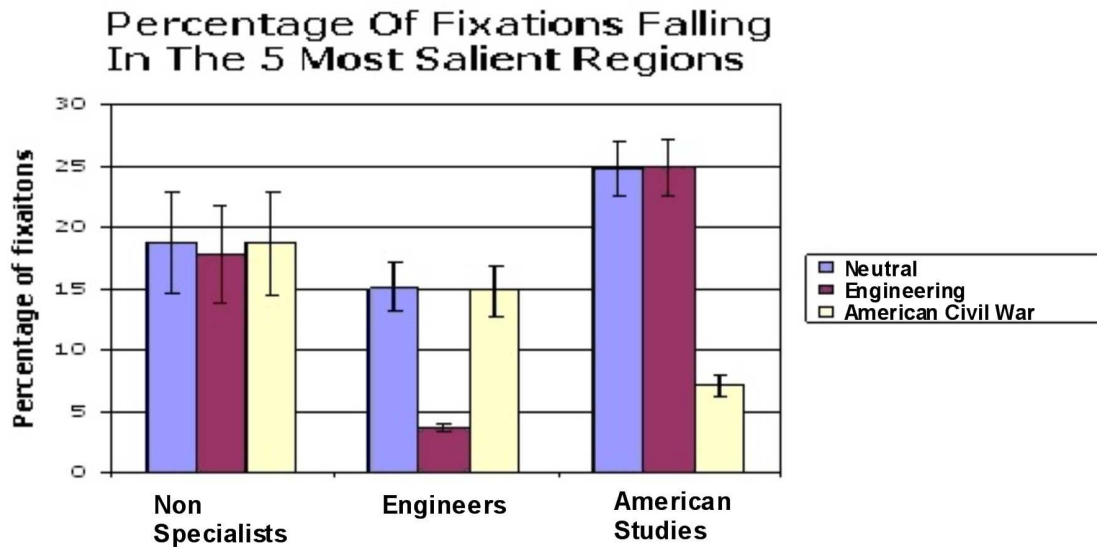
Percentage Of Fixations Falling In The 5 Most Salient Regions

Figure 28: *A bar chart to show the percentage of all fixations made by each group of participants for each type of stimulus that fell into salient regions. Salient regions were as predicted by a saliency model (Itti & Koch; 2000). The error bars represent standard error.*

Out of all the fixations each participant made on each stimulus, the proportion of these that fell in five most salient areas was calculated. A salient region was defined by an area centred on the peak identified by the saliency algorithm, and with a radius of 2 degrees of visual angle. A mixed design ANOVA was carried out (appendices 15a, 15b and 15c) and found no statistically reliable main effect of group, $[F(2,42) = 3.066$, $SEM = 329.325$, $p=0.057]$, but there was a statistically reliable effect of stimulus, $F(2,42) = 25.563$, $SEM = 15.884$, $p<0.001$. There was also an interaction between group and stimulus, $F(4) = 58.065$, $p<0.001$.

Post-hoc paired samples T-tests were conducted and found American Studies students made reliable fewer fixations in salient regions of Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 8.302$, $SEM = 2.14370$, $p<0.001$; and $t(14) = 7.617$, $SEM = 2.48954$, $p<0.001$, respectively. American Studies students also made reliably fewer fixations in salient regions of Civil War stimuli compared to Engineers and control participants: $t(14) = 3.343$, $SEM = 2.32034$, $p<0.01$; and $t(14) = 2.539$, $SEM = 4.57165$, $p<0.05$, respectively. There was no reliable difference in the number of fixations American Studies students made to salient areas of Engineering and neutral stimuli $[t(14) = 1.328$, $p = 0.205]$. Engineers made reliable fewer fixations in salient regions of Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 5.733$, $SEM = 1.93713$, $p<0.001$; and $t(14) = 5.826$, $SEM = 1.67888$, $p<0.001$, respectively. Engineers also made reliably fewer fixations in salient regions of Engineering stimuli compared to American Studies students and control participants: $t(14) = 9.247$, $SEM = 2.28687$, $p<0.001$; and $t(14) = 3.581$, $SEM = 3.91208$, $p<0.01$, respectively.

As many of the scanpath analyses are restricted to the first five fixations in this paper, it is useful to repeat this analysis of fixation locations using only the first five fixations from each stimulus. A mixed design ANOVA (appendices 16a, 16b and 16c) found a main effect of stimulus, $F(2,84) = 26.058$, $MSe = 15.100$, $p<0.001$. There was no statistically reliable main effect of group, $F(2,42) = 2.301$, $MSe = 246.794$, $p = 0.113$, but there was an interaction between group and stimulus, $F(4) = 70.405$, $p<0.001$.

Post-hoc paired samples T-tests (appendices 16d and 16e) showed that American Studies students made reliable fewer fixations in salient regions of Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 9.940$, *SEM* = 1.84817, $p<0.001$; and $t(14) = 9.754$, *SEM* = 1.88346, $p<0.001$, respectively. American Studies students also made reliably fewer fixations in salient regions of Civil War stimuli compared to Engineers and control participants: $t(14) = 6.011$, *SEM* = 1.77460, $p<0.001$; and $t(14) = 3.133$, *SEM* = 4.03542, $p<0.01$, respectively. There was no reliable difference in the number of fixations American Studies students made to salient areas of Engineering and neutral stimuli [$t(14)=0$, $p=1$]. Engineers made reliable fewer fixations in salient regions of Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 7.343$, *SEM* = 1.79958, $p<0.001$; and $t(14) = 6.688$, *SEM* = 1.86952, $p<0.001$, respectively. Engineers also made reliably fewer fixations in salient regions of Engineering stimuli compared to American Studies students and control participants: $t(14) = 8.565$, *SEM* = 2.44222, $p<0.001$; and $t(14) = 3.789$, *SEM* = 3.75335, $p<0.01$, respectively. There was no reliable difference in the number of fixations Engineers made to salient areas of Civil War and neutral stimuli [$t(14)=0.821$, $p=0.425$]. Figure 29 illustrates these results.
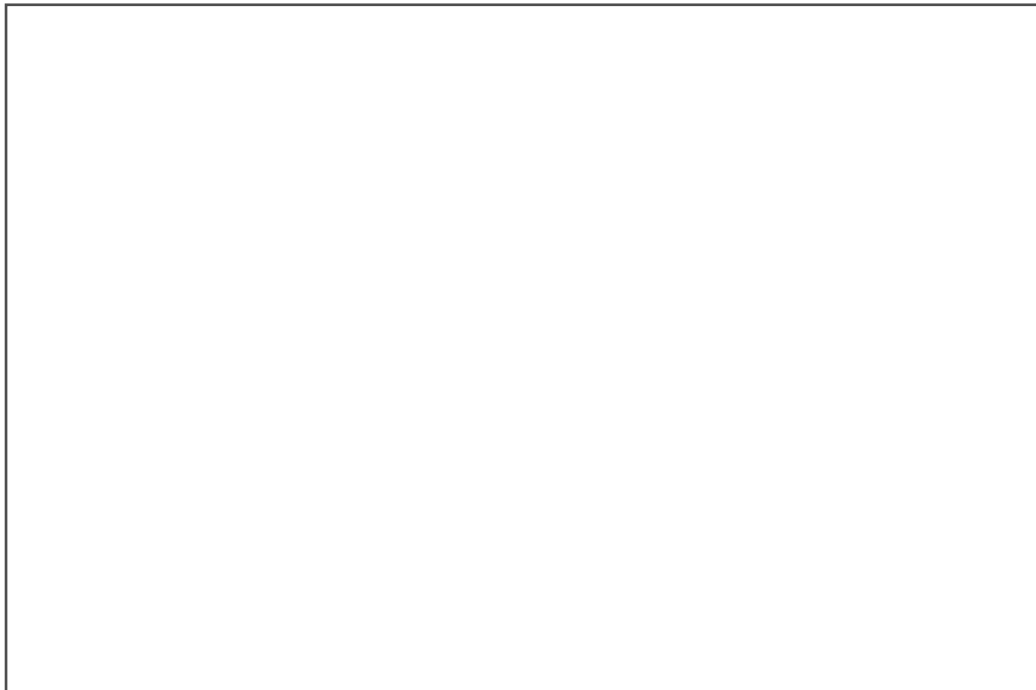

Figure 29: *A bar chart to show the percentage of the first 5 fixations made by each group of participants for each type of stimulus that fell into salient regions. Salient regions were as predicted by a saliency model (Itti & Koch; 2000). The error bars represent standard error.*

Total Number of Fixations

Specialists made fewer overall fixations when looking at stimuli from their own domain. Figure 30 illustrates these results.

Figure 30: *A bar chart to show the average number of fixations made by each group of participants on each type of stimulus. The error bars represent standard error.*

The total number of fixations on each stimulus for each participant was calculated. A mixed design ANOVA was carried out (appendices 17a, 17b and 17c) and found a main effect of stimulus, $F(2,84) = 9.340$, $MSe = 0.250$, $p<0.001$. There was no main effect of group, [$F(2,42) = 0.521$, $MSe = 4.904$, $p = 0.521$], but there was an interaction between group and stimulus, $F(4) = 45.830$, $p<0.001$.

Post-hoc paired samples T-tests (appendices 17d and 17e) showed that American Studies students made reliable fewer fixations when viewing Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 9.510$, $SEM = 0.18459$, $p<0.001$; and $t(14) = 9.238$, $SEM = 0.20110$, $p<0.001$, respectively. American Studies students also made reliably fewer fixations when viewing Civil War stimuli compared to Engineers and control participants: $t(14) = 2.490$, $SEM = 0.54860$, $p<0.05$; and $t(14) = 3.126$, $SEM = 0.38480$, $p<0.01$, respectively. There was no reliable difference in the number of fixations American Studies students made when viewing Engineering and neutral stimuli [$t(14)=1.246$, $p=0.233$]. Engineers made reliable fewer fixations when viewing Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 7.730$, $SEM = 0.19375$, $p<0.001$; and $t(14) = 3.963$, $SEM = 0.25648$, $p<0.001$, respectively. Engineers also made reliably fewer fixations when viewing Engineering stimuli compared to American Studies students and control participants: $t(14) = 3.713$, $SEM = 0.50834$, $p<0.01$; and $t(14) = 3.760$, $SEM = 0.32975$, $p<0.01$, respectively.

Average Saccadic Amplitude
Specialists produced greater saccadic amplitudes when looking at stimuli from their own domain. Figure 31 illustrates these results.
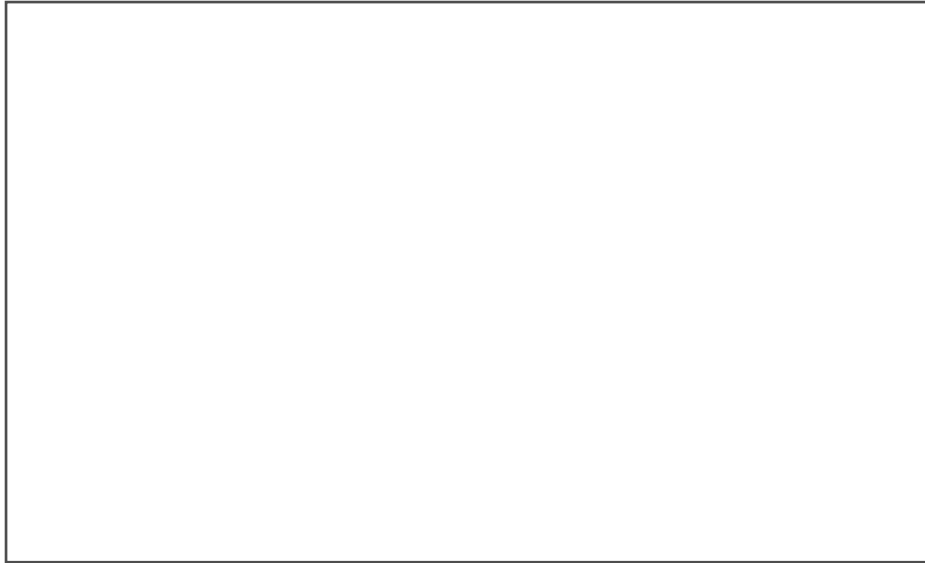
Figure 31: A bar chart to show the average saccadic amplitudes for each stimulus type for each participant group. The error bars represent standard error.

The average saccadic amplitude on each stimulus for each participant was calculated. A mixed design ANOVA (appendices 18a, 18b and 18c) revealed a main effect of stimulus $F(2,84) = 7.938$, $MSe = 0.110$, $p<0.05$. There was no main effect of group, $[F(2,42) = 0.450$, $MSe = 0.492$, $p = 0.641]$, but there was an interaction between group and stimulus, $F(4) = 22.926$, $p<0.001$.

Post-hoc paired samples T-tests (appendices 18d and 18e) showed that American Studies students made reliable longer saccadic amplitudes when viewing Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 4.570$, $SEM = 0.17165$, $p<0.001$; and $t(14) = 4.766$, $SEM = 0.17555$, $p<0.001$, respectively. American Studies students also made reliably longer saccadic amplitudes when viewing Civil War stimuli compared to Engineers and control participants: $t(14) = 4.412$, $SEM = 0.21221$, $p<0.01$; and $t(14) = 3.037$, $SEM = 0.19262$, $p<0.01$, respectively. There was no reliable difference in the average saccadic amplitude for American Studies students when viewing Engineering and neutral stimuli $[t(14)=1.611$, $p=0.130]$. Engineers made reliably longer saccadic amplitudes when viewing Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 4.706$, $SEM = 0.15528$, $p<0.001$; and $t(14) = 4.115$, $SEM = 0.16867$, $p<0.05$, respectively. Engineers also made reliably longer saccadic amplitudes when viewing Engineering stimuli compared to American Studies students and control participants: $t(14) = 3.874$, $SEM = 0.14942$, $p<0.01$; and $t(14) = 2.327$, $SEM = 0.20692$, $p<0.05$, respectively. There was no reliable difference in the number of fixations Engineers made to salient areas of Civil War and neutral stimuli $[t(14)=0.808$, $p=0.432]$.

3.2.3.3 Scanpath Analyses
Upon inspection of the sequences of fixations made when viewing a picture on the first (encoding) and second (recognition) occasion, some repetitions are evident. Often similar regions were inspected soon after picture onset and in some cases scanpaths were identical for the first few fixations.

Five letters were used because the mean number of gazes in different regions (i.e. excluding adjacent fixations on the same regions) on each stimulus was five, with a range of four to seven gazes. This gave a more standardised and manageable data set, and was long enough to display any emerging similarity. In those trials where fewer than five gazes remained after condensing fixations, any comparison strings were trimmed to the same length. Once the strings had been produced for all trials, they were compared using the editing algorithm and an average string similarity was produced across trials.

Encoding Vs. Model-predicted saliency scanpaths
Scanpaths at encoding were similar to those predicted by the saliency model, apart from when specialists viewed pictures from their own domain, in which case the similarity decreased. Figure 32 illustrates this interaction.
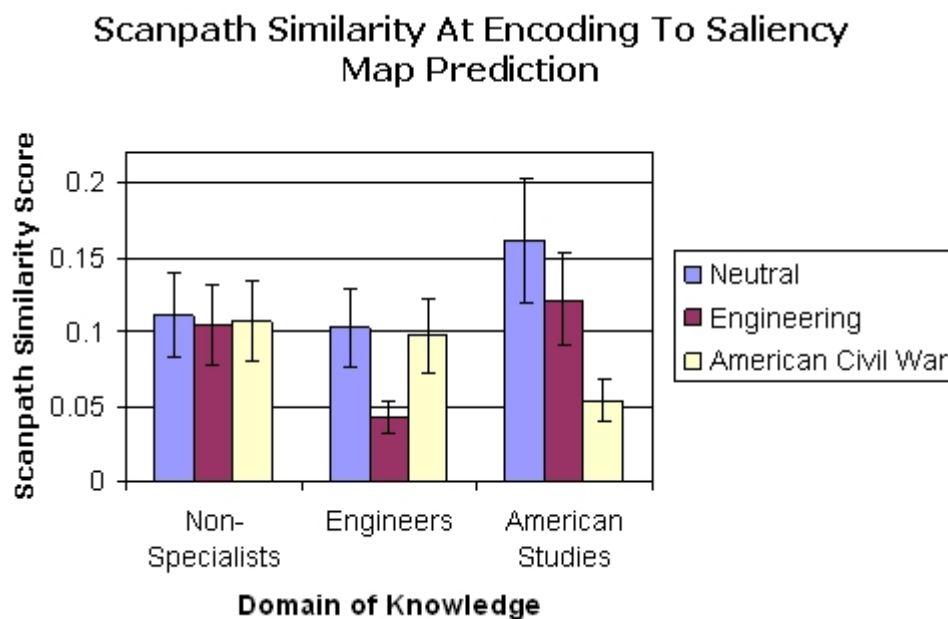

Figure 32: The mean similarity scores for encoding of a stimulus and predicted scanpaths based on the saliency model. A score of 1 would be 100% identical. The error bars represent standard error.

The scanpaths generated from encoding of a picture were compared to respective scanpaths predicted by the saliency model (Itti & Koch, 2001). Each scanpath was given a score depending on how similar eye movements were at encoding compared to scanpaths predicted by saliency.

A mixed design ANOVA (appendices 19a, 19b and 19c) revealed a main effect of stimulus, $F(2,84) = 8.785$, $MSe = 0.002$, $p<0.001$. There was no statistically reliable main effect of group, $[F(2,42) = 2.374$, $MSe = 0.005$, $p = 0.105]$, but there was an interaction between group and stimuli, $F(4) = 8.572$, $p<0.001$.

Post hoc paired sample T-tests (appendices 19d and 19e) found that scanpaths for American Studies students at encoding were reliably less similar to model-predicted saliency scanpaths when viewing Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 5.313$, $SEM = 0.01277$, $p<0.001$; and $t(14) = 5.101$, $SEM = 0.02108$, $p<0.001$, respectively.

Scanpaths for American Studies students at encoding were also reliably less similar to model-predicted saliency scanpaths when viewing Civil War stimuli compared to Engineers and control participants: $t(14) = 2.814$, *SEM* = 0.01567, $p<0.05$; and $t(14) = 3.123$, *SEM* = 0.1710, $p<0.01$, respectively. Scanpaths by Engineers at encoding were reliably less similar to model-predicted saliency scanpaths when viewing Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 3.550$, *SEM* = 0.01556, $p<0.01$; and $t(14) = 4.983$, *SEM* = 0.01201, $p<0.001$, respectively. Scanpaths by Engineers at encoding were also reliably less similar to model-predicted saliency scanpaths when viewing Engineering stimuli compared to American Studies students and control participants: $t(14) = 5.380$, *SEM* = 0.01468, $p<0.001$; and $t(14) = 5.276$, *SEM* = 0.01176, $p<0.001$, respectively.

An example of the differences in scanpaths at encoding between a non-specialist looking at a picture from their domain and a specialist looking at the same picture can been seen by comparing Figures 33a and 33b with the saliency-predicted scanpath in Figure 26.

Figure 33a: *A scanpath made by an engineer, looking at a Civil War stimulus. Comparing this with Figure 1, it can be seen that a large proportion of the fixations fall into salient regions and that the scanpath sequence is very similar to the predicted scanpath. The diagram was produced using the eye-tracking computer software 'Begaze'.*
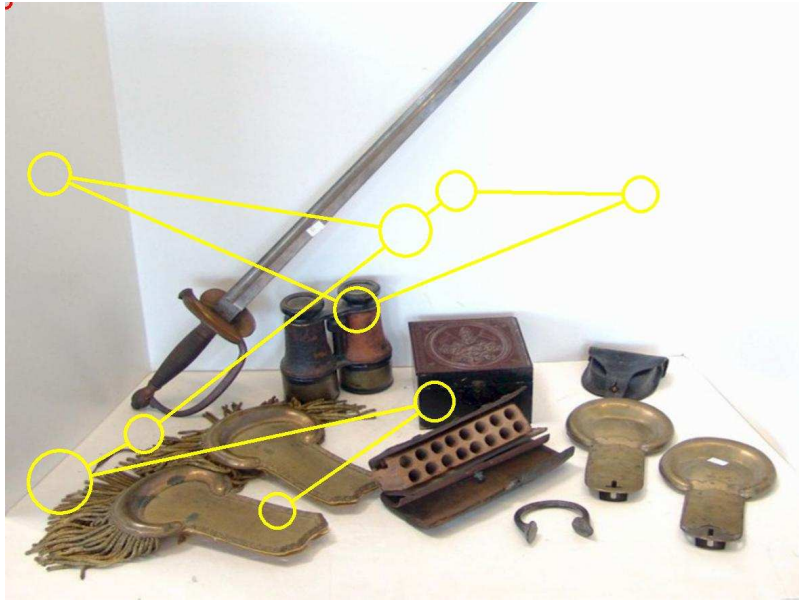
Figure 33b: *A scanpath made by and American Studies student, looking at a Civil War picture. Comparing this with Figure 1 it can be seen that very few fixations fall within salient areas and the scanpath sequence is different from that predicted by saliency.*

Second viewing (recognition) vs. Model-predicted saliency scanpath

The second viewing (i.e. when participants were given the recognition test) was compared to the model predicted saliency sequence to see if the effect of saliency changed over time and after multiple exposures. The results showed that with recognition as well as encoding, saliency had less of an effect on experts' scanpaths when viewing domain-specific stimuli. Figure 34 illustrates this data.
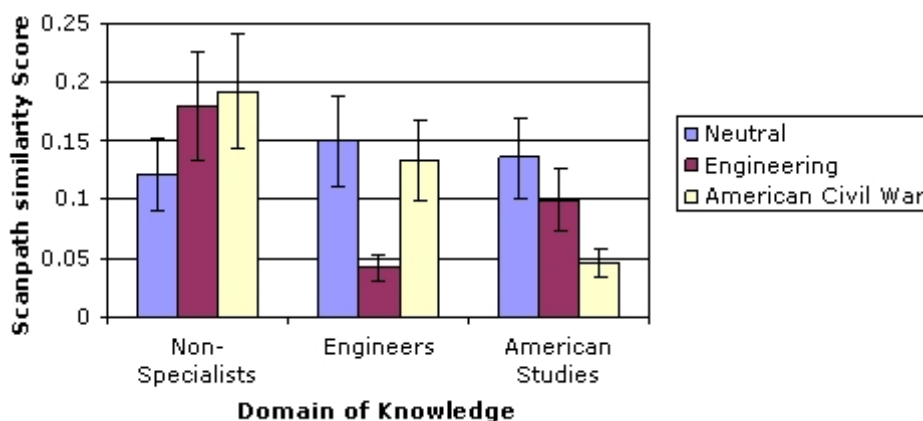

Figure 34: *The mean similarity scores for scanpaths produced at the recognition test and those predicted based on the saliency model. The error bars represent standard error.*

A mixed design ANOVA (appendices 20a, 20b and 20c) revealed no statistically reliable main effects of group, [$F(2,42) = 1.754$, $MSe = 0.036$, $p=0.185$], or stimulus, [$F(2,84) = 0.867$, $MSe =$

0.011, $p=0.424$]. However, there was a statistically reliable interaction between group and stimulus, $F(4) = 4.377$, $p<0.01$.

Post-hoc paired samples T-tests (appendices 20d and 20e) showed that scanpaths for American Studies students at recognition were reliably less similar to model-predicted saliency scanpaths when viewing Civil War stimuli compared to Engineering and neutral stimuli: $t(14) = 3.743$, $SEM = 0.01423$, $p<0.01$; and $t(14) = 6.414$, $SEM = 0.01390$, $p<0.001$, respectively. Scanpaths for American Studies students at recognition were reliably less similar to model-predicted saliency scanpaths when viewing Civil War stimuli compared to Engineers and control participants: $t(14) = 3.379$, $SEM = 0.02578$, $p<0.01$; and $t(14) = 2.167$, $SEM = 0.06754$, $p<0.05$, respectively. Scanpaths by Engineers at recognition were reliably less similar to model-predicted saliency scanpaths when viewing Engineering stimuli compared to Civil War and neutral stimuli: $t(14) = 4.029$, $SEM = 0.02676$, $p<0.01$; and $t(14) = 5.494$, $SEM = 0.01667$, $p<0.001$, respectively. Scanpaths by Engineers at recognition were also reliably less similar to model-predicted saliency scanpaths when viewing Engineering stimuli compared to American Studies students: $t(14) = 4.129$, $SEM = 0.01398$, $p<0.01$.

Encoding vs. Second Viewing (recognition)
Scanpaths at encoding and test were reliably more similar for all participant groups than would be expected by chance. This data is shown in Figure 35.
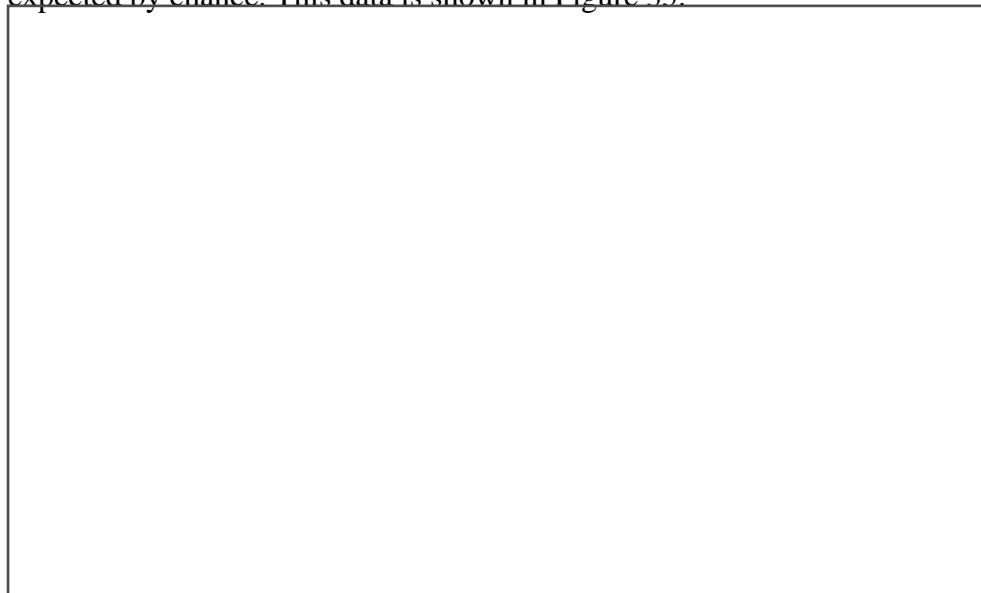

Figure 35: A bar chart to show the mean string similarity scores of encoding compared to test. Chance was calculated as 0.1148 and represented on the graph by the bar labelled 'chance'. The error bars represent standard error.

The scanpaths generated from encoding of a picture were compared to those on second viewing during the recognition test. A repeated measures ANOVA (appendices 21a, and 21b) revealed no statistically reliable main effect of group, [$F(2,28) = 1.693$, $MSe = 0.005$, $p=0.202$].

Overall, there were string similarities of 0.221303 for control participants (non-specialists), 0.257670 for Engineers and 0.267778 for American Studies students. All of these string similarities were reliably greater than the calculated chance value of 0.1148 ($p<0.01$). See

appendix 21c.

This comparison was broken down further to investigate whether scanpaths from encoding and recognition differed according to stimulus type for each participant group (see appendices 21d and 21e). There were no statistically reliable main effect of stimulus type [$F(2,84) = 2.350$, $MSe = 0.005$, $p = 0.102$] and no reliable interaction between stimulus type and participant group [$F(4) = 1.069$, $p = 0.337$]. This suggests that scanpaths at encoding and second viewing were reliably similar for all participant groups across all stimulus types.

3.2.4 Discussion

Does knowledge of a domain affect the relationship between saliency and scanpaths when viewers look at images from within their domain?

The analyses of recognition memory show that the specialist groups were more accurate in identifying pictures from their own domain. American Studies students were more accurate in identifying Civil War pictures, and engineers were more accurate in identifying engineering pictures. There was no significant difference in accuracy across the groups when identifying neutral pictures and non-specialists were equally accurate over all the stimuli. This result suggests that the participants in each group showed true domain-specialised knowledge for their area, which provides a valid basis for scanpath comparison between these groups. This is consistent with Walker (1987) who found that on recognition tests, performance was a function of level of expertise in the domain.

All the fixations made on a particular stimulus were compared to the five most salient areas of that stimulus. In previous research, saliency effects have been found when memory tasks were performed (e.g. Underwood, et al, 2006). In this experiment, it was found that overall the specialist groups made fewer fixations in salient areas when the pictures were from their own domain, i.e. Engineers made fewer fixations in salient areas of Engineering pictures, and American Studies students made fewer fixations in salient areas of Civil War pictures. This suggests that when viewing their own area of expertise they were less constrained to looking at salient regions. There was no significant difference between groups when looking at neutral pictures and non-specialists showed no significant difference across stimuli types. This analysis was repeated using only the first five fixations in order to remain consistent with the scanpath analyses and to rule out the possibility that the saliency effect might be biased by later fixations. The same results were found. This lends support for the saliency map theory (Koch & Ullman, 1985) that suggests that saliency influences eye-movements. It is also consistent with Parkhurst et al (2002), who concluded that saccades are guided by low-level saliency and more so when top-down factors can play less of a role. The present study shows a correlation between saliency and eye-movements but this becomes limited when domain-specific knowledge comes into play. It can be seen in Figures 4 and 5 that Engineers make reliably fewer fixations in highly salient regions when viewing engineering pictures. These results could also be interpreted in terms of Haider and Frensch's (1999) information-reduction hypothesis, which states that with experience, people learn to ignore task-redundant information and limit their processing to task-relevant information. Further evidence in the literature for the information-reduction hypothesis comes from Charness, Reingold, Pomplun, and Stampe, (2001) who found that chess experts fixated proportionally more on relevant pieces than non-expert players.

The apparent cognitive override of saliency in the current study may seem intensified because the interesting parts of the stimuli to the Engineers may have been, by chance, of

particularly low saliency, and thus it is almost like they were actively seeking out low-salient regions, which would not have been of interest to non-specialists. It could therefore be argued that salience is not necessarily informativeness, in that it plays a part in allocating attention *unless* there are more semantically interesting parts of the scene, in which case those more informative areas are fixated.

The cognitive override effect that has been found is consistent with previous investigations of saliency influences (e.g., Underwood et al., 2006; 2007) in a search task, but when an encoding task was used, as here, the saliency map did predict fixation locations. In the present experiment, saliency was a good predictor of fixation locations, apart from when specialists viewed pictures from their own domain. However, others have proposed alternative causal factors that could result in fixations within salient locations or similarities between sequences of locations fixated. For example, Torralba, Oliva, Castelhano and Henderson's (2006) Contextual Guidance model proposes that image saliency and global-context features are computed in parallel, in a feed-forward manner and are integrated at an early stage of visual processing (i.e. before initiating image exploration). For example, Engineers might have known where the more interesting parts of a car engine would have be located and thus this knowledge resulted in similar fixation locations on first and second viewing.

The overall number of fixations and the saccadic amplitudes for stimulus type for each participant group were calculated. It was found that specialists make fewer fixations and produce greater saccadic amplitudes when viewing stimuli from their own domain. This suggests that specialists have a wider spatial distribution of eye movements and coincides with previous research such as that by Reingold, Pomplun and Stampe (2001) who also found that experts made fewer fixations per trial and greater amplitude saccades than non-experts.

The main analyses in this experiment were scanpath comparisons. Overall, scanpaths produced on encoding of a picture compared to those produced on second viewing were more similar than would be expected by chance. This was consistent across all participants, regardless of group or stimulus type, providing evidence in favour of the replication if scanpaths over time and multiple viewings. One has to keep in mind that although the similarity seen here is significant, there is still a large amount of variance unaccounted for. Previous demonstrations of scanpath similarity have largely used simple patterns or line drawings, with fewer and larger regions of interest. It is likely that the much more complex photographs used here led to less scanpath repetition, possibly due to a greater appreciation of the scene semantics by knowledgeable viewers.

Scanpaths from the encoding of a picture were compared to computer-generated scanpaths predicted by saliency. It was found that scanpaths of American Studies students were least similar to the saliency scanpaths when looking at Civil War pictures; and scanpaths of Engineers were least similar to the predicted scanpaths when looking at Engineering pictures. Very similar results were found when scanpaths from the second viewing of the picture were compared to computer-generated saliency scanpaths. In other words, scanpaths were similar to those predicted by the saliency model, apart from when the stimulus was associated with the participant's specialist domain. This result reinforces the mean salient fixation analysis findings (that saliency is less of an influence when the participant is a specialist in that domain) but also incorporates the important concept of sequence and the order of fixations. This is consistent with the notion that bottom-up saliency guides eye movements, but can be overridden by top-down cognitive domain-specific knowledge, and that this effect is constant over time. Furthermore, it supports previous findings that scanpaths from encoding and second viewing of a picture are more similar than

would be expected by chance. However, the effects found when comparing second viewing with saliency, although statistically reliable, were weaker than when comparing encoding with saliency. The main reason for this is that in the first comparison, American Studies students were reliably different to both Engineers *and non-specialists* on Civil War stimuli, and Engineers were significantly different to both American Studies students *and non-specialists*, thus making the overall effect very strong. However, on the second half of the experiment, non-specialists showed large variance amongst the group. Therefore, on the second comparison, Engineers were only reliably different from American Studies students and vice versa, therefore the overall effect was decreased. Due to the large variance in the non-specialist group, it makes it hard to interpret this result, although it does not look to be related to the main hypothesis.

In conclusion, there is a relationship between saliency and eye movements, shown by the similarity of actual scanpaths to those predicted by the saliency model (Itti & Koch, 2000). However, domain-specific knowledge can act as an overriding factor, weakening this relationship between saliency and eye movements.

3.3. Experiment 5: The stability of scanpaths over time

3.3.1 Introduction

It has been found that participants have good memory for the visual detail of fixated regions of a previously viewed scene and are capable of retaining these visual details across delays of at least 24 hours (Castelhano and Henderson, 2005; Hollingworth & Henderson, 2002; Hollingworth, 2005). Such findings date back to classic experiments by Shepard (1967) where participants were able to correctly identify 'old' (previously seen) stimuli in a 2-aternative forced choice task, 98% of the time. Similarly, Standing et al (1970) demonstrated that participants were able to correctly identify 90% of previously seen pictures even after extended periods of time (3 days), short stimuli exposure (1 second per picture) and when pictures were reversed in orientation. Scanpath Theory suggests that recognition memory for visual stimuli is facilitated by reproducing similar eye movements to when the picture was first encoded. Therefore, if recognition accuracy is high after extended periods of time, it follows that scanpaths should also be highly similar. This experiment partly aims to determine whether scanpaths are stable over multiple viewings, experimental sessions and extended time periods.

Furthermore, experts have been found to have better long-term memory for domain-specific material (Hayes-Roth, 1983; Postal, 2004; Drai-Zerbib and Baccin, 2005) than irrelevant material and better long term memory than control participants, who have been shown to rely more on working memory (Sohn and Doane, 2004). This experiment also aims to test whether long-term recognition memory for domain-specific and domain-irrelevant pictures differs with expertise, and whether this affects scanpath similarity.

3.3.2 Methodology

Participants
30 participants took part in the experiment, 15 Engineers and 15 American Studies students. A condition of the American Studies group was that they had to have taken a core module on The American Civil War. Three other people took part in the pilot study, although the pilot data were not used in the analysis and these participants were not included in the sample. Opportunity sampling was employed to recruit participants, who were all students at Nottingham University

(undergraduates and postgraduates). The age range was 18-27 years and the mean age was 21.7 years. The sample comprised of 24 females and 6 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and not having taken part in the previous experiment.

Materials and apparatus

Eye movements were recorded using an SMI iVIEW X Hi-Speed eye tracker. The same stimuli as in section 3.3 were used (see section 3.3.2 for examples). Saliency maps were generated, using Itti and Koch's (2000) model with standard parameters. These maps were produced for the first five simulated shifts and thus indicate the first five most salient regions for each picture (see section 3.3.2 for an example). The only further criterion for stimuli was that all 5 salient regions were non-contiguous; those pictures where the same or overlapping regions were re-selected within the first 5 shifts were replaced.

Design

The experiment used a two-by-three mixed design; with two specialist groups of participants and three specific types of stimuli. The specialist groups were Engineers and American Studies students, who were both domain specialists in their field. The independent variables were therefore which group the participant belonged to and the type of stimulus being shown. The dependant variable measures were: accuracy in deciding whether a picture was old or new on immediate test compared to accuracy on delayed test; average fixation durations, mean number of salient fixations and scanpath similarity between encoding and model-predicted saliency, encoding and immediate recognition test, and encoding and delayed recognition test.

Procedure

The initial procedure was identical to the previous experiment in section 3.3 (for a systematic representation of the experimental procedure, see section 3.3.2). The only difference was that, this time, participants were called back one week later and were given a second recognition memory test. The procedure and conditions at the 1-week follow-up were the same except there was no practice trial. Forty-five of the pictures were the "old" stimuli from the first half of the experiment (1 week prior) and 45 were completely novel to the participants. As in the first half, the participants had to identify whether the picture was old or new.

3.3.4 Results

There were many different possible measures to extract from the raw data showing fixation locations and durations for each subject on each picture. In all cases, trials were excluded where the fixation at picture onset was not within the central region, when participants looked away from the screen (e.g. to the keyboard), or when calibration was temporarily interrupted.

There were 2 main types of data, recognition memory data (accuracy and response time) and eye tracking measures (average fixation durations, mean number of salient fixations and scanpath comparisons). Scanpath similarity scores were calculated for encoding compared to saliency, encoding compared to immediate recognition test, and encoding compared to delayed recognition test.

3.3.4.1. Recognition Memory

Accuracy

Engineers were more accurate with engineering stimuli and American Studies students were more accurate with Civil War stimuli. Accuracy was measured by the number of pictures participants correctly identified as 'old' (if they were from the previous set) or 'new' (if they had never been seen before). This was done for each category of stimuli, and was out of 30 (as there were 30 pictures in each category in the second phase of the experiment). Accuracy was measured at immediate and delayed test.
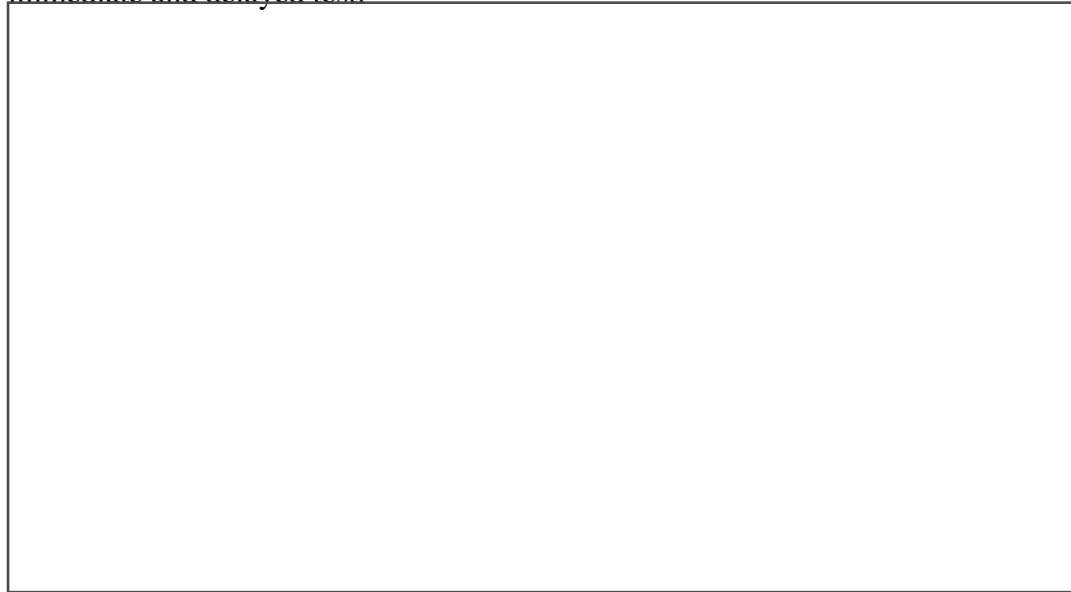

Figure 36: *Accuracy of the two groups of participants at immediate and delayed test for the three sets of stimuli.*

A 3x2x2 mixed design ANOVA (appendices 22a, 22b and 22c) revealed a reliable main effect of picture type: $t(2,56) = 7.147$, $MSe = 6.956$, $p<0.01$; and a reliable interaction between picture type and group: $t(2) = 16.245$, $p<0.001$. There was no reliable main effect of group [$t(1,28) = 0.563$, $MSe = 26.685$, $p = 0.459$], and no reliable main effect of recognition test (immediate or delayed) [$t(1,28) = 0.311$, $MSe = 8.647$, $p = 0.582$] and no other interactions were statistically reliable.

Post-hoc Independent samples T-tests (appendices 22d and 22e) showed that when viewing Civil War pictures at immediate recognition test, American Studies students were reliably more accurate than Engineers: $t(28) = 2.618$, $SED = 1.09487$, $p<0.05$, although this was not statistically reliable at delayed recognition test [$t(28) = 1.388$, $SED = 1.24849$, $p = 0.176$]. At both immediate and delayed recognition tests, Engineers were reliably more accurate at identifying Engineering pictures than American Studies students: $t(28) = 2.879$, $SED = 1.20423$, $p<0.01$; and $t(28) = 2.408$, $SED = 1.19070$, $p<0.05$, respectively. There was no difference between the groups for neutral stimuli for immediate or delayed recognition test [$t(28) = 1.212$, $SED = 1.10007$, $p = 0.236$; and $t(28) = 0.339$, $SED = 1.17972$, $p = 0.737$], respectively.

Post-hoc paired samples T-tests were carried out for American studies students (appendices 22f and 22g) and Engineers (appendices 22h and 22i). American Studies students were found to be reliably more accurate at identifying Civil War stimuli at immediate recognition test than Engineering stimuli or neutral stimuli: $t(14) = 6.485$, $SEM = 0.83267$, $p<0.001$; and $t(14) = 4.962$, $SEM = 0.83305$, $p<0.001$, respectively. American Studies students were also reliably

more accurate at identifying Civil War stimuli at delayed recognition test than Engineering stimuli or neutral stimuli: $t(14) = 3.063$, $SEM = 1.02291$, $p<0.01$; and $t(14) = 2.656$, $SEM = 0.85338$, $p<0.05$, respectively. There was no reliable difference in accuracy for American Studies students in identifying Civil War stimuli at immediate and delayed recognition tests [American Studies students were reliably more accurate at identifying Civil War stimuli at immediate recognition test than Engineering stimuli or neutral stimuli: $t(14) = 1.258$, $SEM = 0.84778$, $p = 0.229$]. No other comparisons for American Studies students were statistically reliable. There were no reliable differences in accuracy between the stimuli at immediate and delayed recognition test for Engineers.

3.3.4.2. Eye-tracking measures

Locations of fixations
Out of all the fixations each participant made on each stimulus, the proportion of these that fell in five most salient areas was calculated. A salient region was defined by an area centred on the peak identified by the saliency algorithm, and with a radius of 2 degrees of visual angle.

It was found that domain-specialists made fewer fixations in salient areas when looking at domain-specific stimuli. Figure 37 illustrates this interaction.
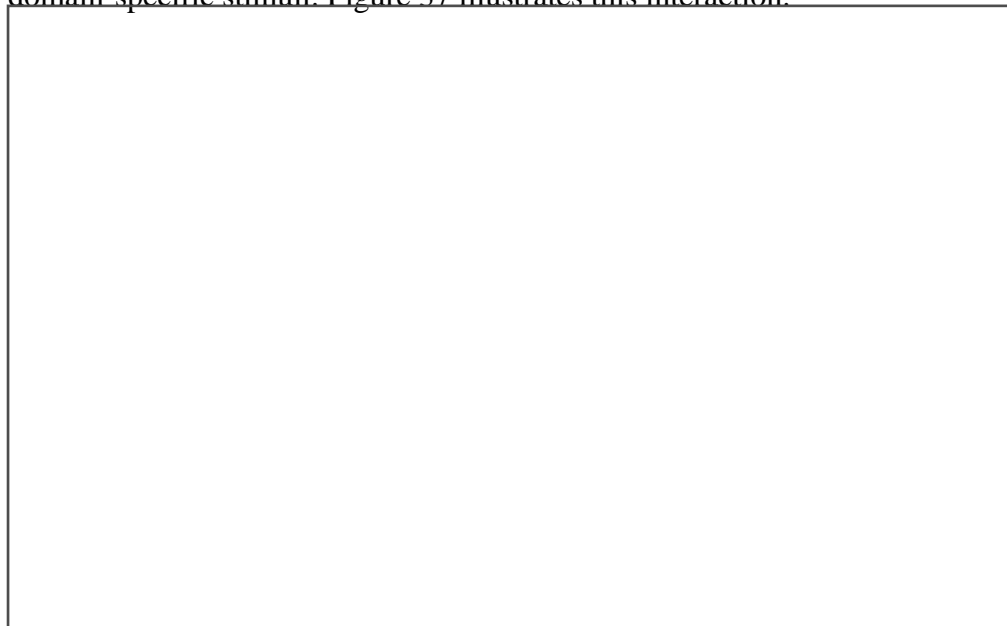
Figure 37: A bar chart to show the percentage of all fixations made by each group of participants for each type of stimulus that fell into salient regions. Salient regions were as predicted by a saliency model (Itti & Koch, 2000).

A mixed design ANOVA (appendices 23a, 23b and 23c) revealed a reliable main effect of picture type: $F(2,56) = 56.659$, $MSe = 0.003$, $p<0.001$; and a reliable interaction between picture type and group: $F(2) = 127.497$, $p<0.001$. There was no reliable main effect of group [$F(1,28) = 1.362$, $MSe = 0.010$, $p = 0.253$].

Post-hoc independent samples T-tests (appendices 23d and 23e) showed that American

Studies students made reliably fewer fixations in salient regions on Civil War pictures than Engineers did: $t(28) = 8.385$, $SED = 0.02746$, $p<0.001$; and that Engineers made reliably fewer fixations in salient regions on Engineering pictures than American Studies students did: $t(28) = 13.528$, $SED = 0.01564$, $p<0.001$. There was no reliable difference between the groups for neutral stimuli [$t(28) = 1.656$, $SED = 0.03397$, $p = 0.109$].

Post-hoc paired samples T-tests (appendices 23f and 23g) showed that American Studies students made reliably fewer fixations to salient regions on Civil War pictures compared to Engineering and neutral pictures: $t(14) = 12.072$, $SEM = 0.01623$, $p<0.001$; and $t(14) = 11.380$, $SEM = 0.01768$, $p<0.001$, respectively. Similarly, Engineers made reliably fewer fixations to salient regions on Engineering pictures compared to Civil War and neutral pictures: $t(14) = 8.899$, $SEM = 0.02764$, $p<0.001$; and $t(14) = 3.732$, $SEM = 0.00727$, $p<0.01$, respectively.

3.3.4.3 Scanpath Analyses

Strings were cropped to five letters, and were computed for each subject viewing each stimulus in the experiment. Five letters were used because the mean number of fixations made on each stimulus was five, with a range of four to seven fixations. This gave a more standardised and manageable data set, and was thought long enough to display any emerging similarity. In those trials where fewer than five fixations remained after condensing gazes, any comparison strings were trimmed to the same length. Once the strings had been produced for all trials, they were compared using the editing algorithm and an average string similarity was produced across trials.

Encoding vs. model-predicted saliency scanpath

The scanpaths generated from first viewing of a picture were compared to respective scanpaths predicted by the saliency model (Itti & Koch, 2002). Observed scanpaths were more similar to those predicted by the model when stimuli were *not* domain-specific. When stimuli *were* domain specific, similarity dropped to (or below) chance.
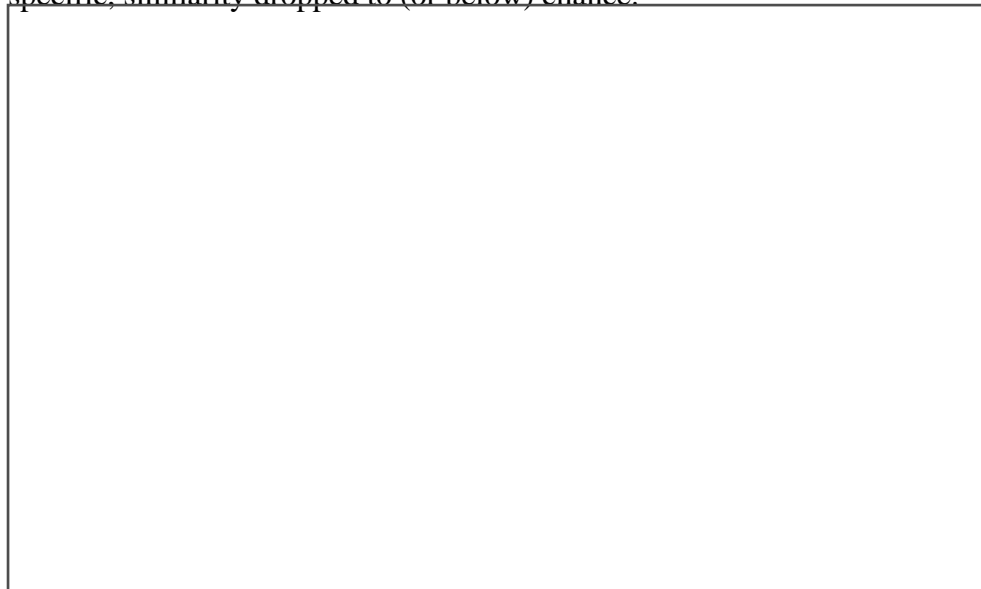

Figure 38: *Average string similarities between encoding and model-predicted saliency for each groups on each type of stimulus. A score of 1 would be a perfect match.*

A mixed-design ANOVA (appendices 24a, 24b and 24c) revealed a reliable main effect of picture type: $F(2,56) = 28.801$, $MSe = 0.002$, $p<0.001$; and a reliable interaction between picture type and group: $F(2) = 49.848$, $p<0.001$. There was no reliable main effect of group [$F(1,28) = 1.567$, $MSe = 0.003$, $p = 0.221$].

Post-hoc independent samples T-tests (appendices 24d and 24e) showed that when viewing Civil War stimuli, scanpaths produced by American Studies students were reliably less similar to model-predicted saliency scanpaths than Engineers' scanaths compared to model-predicted saliency: $t(28) = 7.246$, $SED = 0.01586$, $p<0.001$. Similarly, when viewing Engineering stimuli, scanpaths produced by Engineers were reliably less similar to model-predicted saliency scanpaths than American Studies students' scanaths compared to model-predicted saliency: $t(28) = 6.982$, $SED = 0.01358$, $p<0.001$. There was no reliable difference between the groups for neutral stimuli [$t(28) = 1.183$, $SED = 0.02082$, $p = 0.247$].

Post-hoc paired samples T-tests (appendices 24f and 24g) were also carried out to further investigate the main effect of stimulus type. Overall, the average string similarities for Civil War and Engineering stimuli were reliably different to neutral stimuli: $t(29) = 5.088$, $SEM = 0.01327$, $p<0.001$, and $t(29) = 4.676$, $SEM = 0.01516$, $p<0.001$.

Paired samples T-tests were carried out to compare the string similarities for each stimulus type for each group to chance (0.0417). For American Studies students, when viewing Civil War stimuli, the similarity of actual scanpaths to those predicted by the saliency model dropped reliably lower than chance: $t(14) = 3.088$, $SEM = 0.00487$, $p<0.01$; and when viewing Engineering stimuli, the similarity of actual scanpaths to those predicted by the saliency model increased reliably higher than chance: $t(14) = 6.727$, $SEM = 0.01285$, $p<0.001$. For Engineers, when Engineering stimuli, the similarity of actual scanpaths to those predicted by the saliency model decreased and was not reliably different from chance [$t(14) = 1.906$, $SEM = 0.00439$, $p = 0.077$]. When viewing Civil War stimuli, the similarity of actual scanpaths made by Engineers compared to those predicted by the saliency model increased reliably higher than chance: $t(14) = 6.617$, $SEM = 0.01509$, $p<0.001$. These results are illustrated in Figure 39.

[pic]

Figure 39: *An example of a scanpath predicted by the saliency model (yellow circles) on a Civil War Stimulus. The blue scanpath was created by an Engineer and the green scanpath was created by an American Studies student. Notice that the blue scanpath is much more similar to that predicted by the saliency model than the green scanpath.*

Encoding v. Immediate Test

The scanpaths generated at encoding were compared to those during the immediate test. Scanpaths at encoding and immediate test were more similar when pictures were domain-specific. This data is shown in Figure 40.
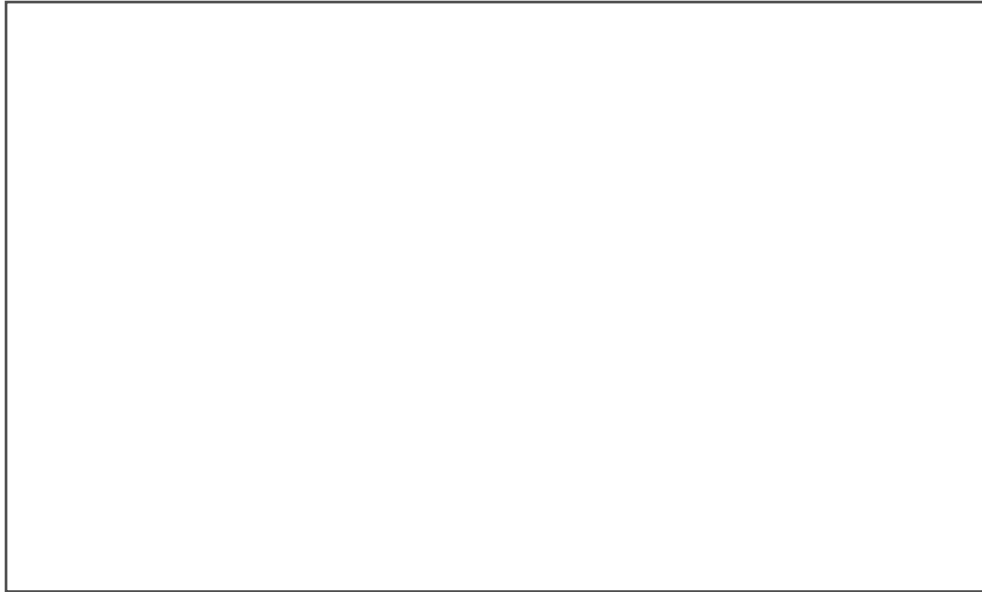
Figure 40: *Average string similarities between encoding and immediate test, for each groups on each type of stimulus. A score of 1 would be a perfect match.*

A mixed-design ANOVA was carried out (appendices 25a, 25b and 25d) and found a reliable main effect of group $F(1,28) = 5.831$, *MSE* $= 0.011$, $p<0.05$. From the graph and the descriptive statistics (appendix 25c) it can be concluded that overall, string similarities between encoding and immediate test were higher for Engineers than American Studies students. There was no main effect of stimulus type: $[F(2,56) = 1.385$, *MSE* $= 0.003$, $p = 0.259]$; but there was a reliable interaction between stimulus type and group: $F(2) = 4.409$, $p<0.05$.

Post-hoc independent samples T-tests (appendices 25e and 25f) showed a reliable difference between the groups for Engineering stimuli: $t(28) = 4.343$, *SED* $= 0.022306$, $p<0.001$. Scanpaths for Engineering pictures at immediate test were reliably more similar to scanpaths at encoding for Engineers than for American Studies students. There was no reliable difference between the groups for Civil War stimuli or neutral stimuli $[t(28) = 1.457$, *SED* $= 0.02776$, $p = 0.155$; and $t(28) = 0.695$, *SED* $= 0.03106$, $p = 0.493$, respectively]. The mean string similarity scores were then compared to chance, using paired-samples T-tests (see appendices 25g and 25h). Randomly generated strings would give a value of approximately 0.0417, and all string similarities were shown to be reliably greater than chance ($p<0.001$).

Encoding v. Delayed Test
The scanpaths generated from encoding of a picture were compared to those during the delayed test, to create string similarity scores for each group on each type of stimuli. Scanpaths were reliably more similar between encoding and delayed test if stimuli were domain-specific. Figure 41 shows this interaction.
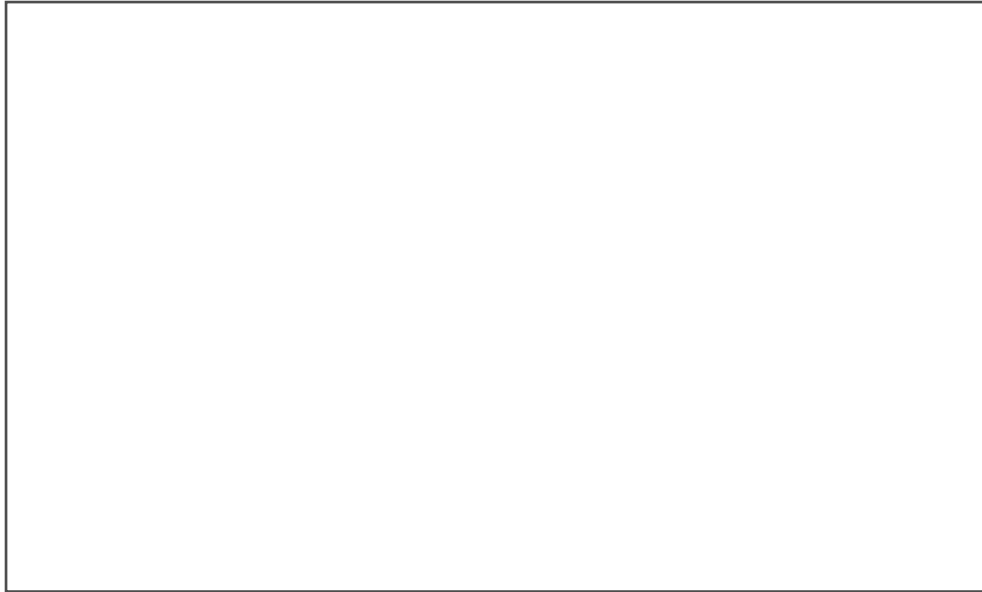
Figure 41: *Average string similarities between encoding and delayed test, for each groups on each type of stimulus. A score of 1 would be a perfect match.*

A mixed-design ANOVA was carried out (appendices 26a, 26b and 26d) and found a reliable main effect of stimulus type: $F(2,56) = 18.357$, $MSE = 0.003$, $p<0.001$; and a reliable main effect of group: $F(1,28) = 7.286$, $MSE = 0.004$, $p<0.05$. These results and the descriptive statistics (appendix 26c) show that scanpaths for Engineers were overall more similar at encoding and test than American Studies students' scanpaths were. Post-hoc paired-samples T-tests were conducted (appendices 26g and 26h) to further explore the main effect of stimulus type and found that overall, similarity scores were reliably higher for Civil War and Engineering stimuli than for neutral stimuli: $t(29) = 3.401$, $SEM = 0.1428$, $p<0.01$; and $t(29) = 4.127$, $SEM = 0.02136$, $p<0.001$, respectively.

The ANOVA also found a reliable interaction between stimulus type and group: $F(2) = 46.351$, $p<0.001$. Post-hoc independent samples T-tests (appendices 26e and 36f) found a reliable difference between the groups for Civil War stimuli: $t(28) = 3.335$, $SED = 0.02467$, $p<0.01$; and for Engineering stimuli: $t(28) = 7.482$, $SED = 0.02468$, $p<0.001$. Scanpaths for pictures at delayed test were reliably more similar to scanpaths at encoding test if they were domain-specific to the participants' area of expertise. There was no reliable difference between the groups for Civil War stimuli [$t(28) = 0.357$, $SED = 0.01240$, $p = 0.724$]. The mean string similarity scores were then compared to chance, using paired-samples T-tests (appendices 26i and 26j). Randomly generated strings would give a value of approximately 0.0417, and all string similarities were shown to be reliably greater than chance ($p<0.05$).

Immediate Test v. Delayed test
The scanpaths generated during immediate test were compared to those generated during the delayed test, to create string similarity scores for each group on each type of stimuli. Scanpaths were reliably more similar between immediate and delayed tests if stimuli were domain-specific. Figure 42 shows this interaction.
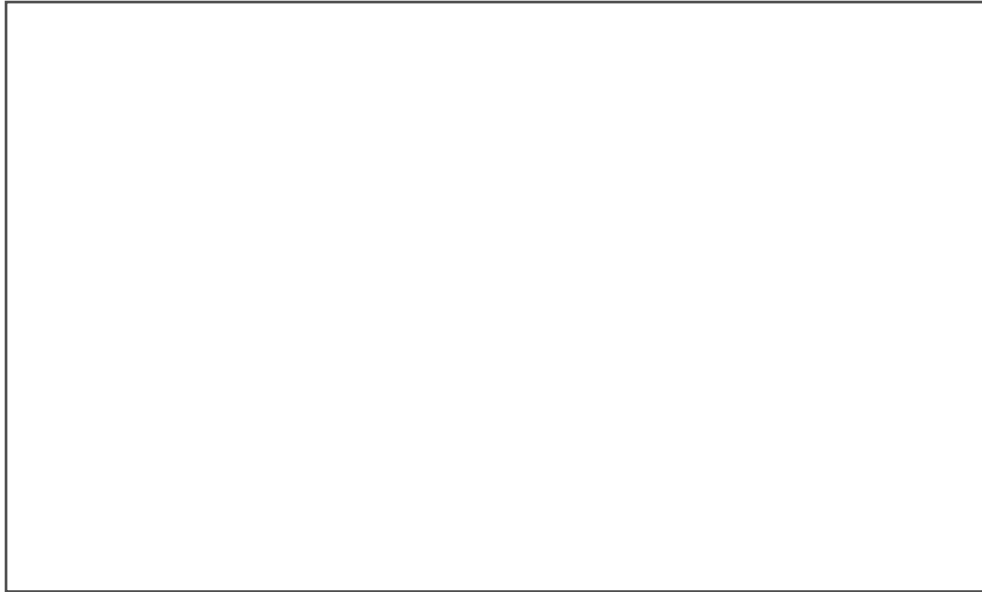
Figure 42: Average string similarities between immediate recognition test and delayed recognition test, for each groups on each type of stimulus. A score of 1 would be a perfect match.

A mixed-design ANOVA was carried out (appendices 27a, 27b and 27c) and found a reliable main effect of stimulus type: $F(2,56) = 31.840$, $MSE = 0.002$, $p<0.001$. There was no reliable main effect of group [$F(1,28) = 0.876$, $MSE = 0.003$, $p = 0.357$]. Post-hoc paired-samples T-tests were conducted (appendices 27f and 27g) to further explore the main effect of stimulus type and found that overall, similarity scores were reliably higher for Civil War and Engineering stimuli than for neutral stimuli: $t(29) = 4.410$, $SEM = 0.1398$, $p<0.001$; and $t(29) = 4.821$, $SEM = 0.02080$, $p<0.001$, respectively.

The ANOVA also found a reliable interaction between stimulus type and group: $F(2) = 68.248$, $p<0.001$. Post-hoc independent samples T-tests (appendices 27d and 27e) found a reliable difference between the groups for Civil War stimuli: $t(28) = 8.347$, $SED = 0.1532$, $p<0.001$; and for Engineering stimuli: $t(28) = 6.515$, $SED = 0.02556$, $p<0.001$. Scanpaths for pictures at delayed test were reliably more similar to scanpaths at immediate test if they were domain-specific to the participants' area of expertise. There was no reliable difference between the groups for Civil War stimuli [$t(28) = 0.858$, $SED = 0.00999$, $p = 0.398$]. The mean string similarity scores were then compared to chance, using paired-samples T-tests (appendices 27h and 27i). Randomly generated strings would give a value of approximately 0.0417, and all string similarities were shown to be reliably greater than chance ($p<0.05$).

Figure 43 is an example of an engineer viewing an engineering stimulus at encoding (pink scanpath), immediate test (yellow scanpath) and delayed test (green scanpath).
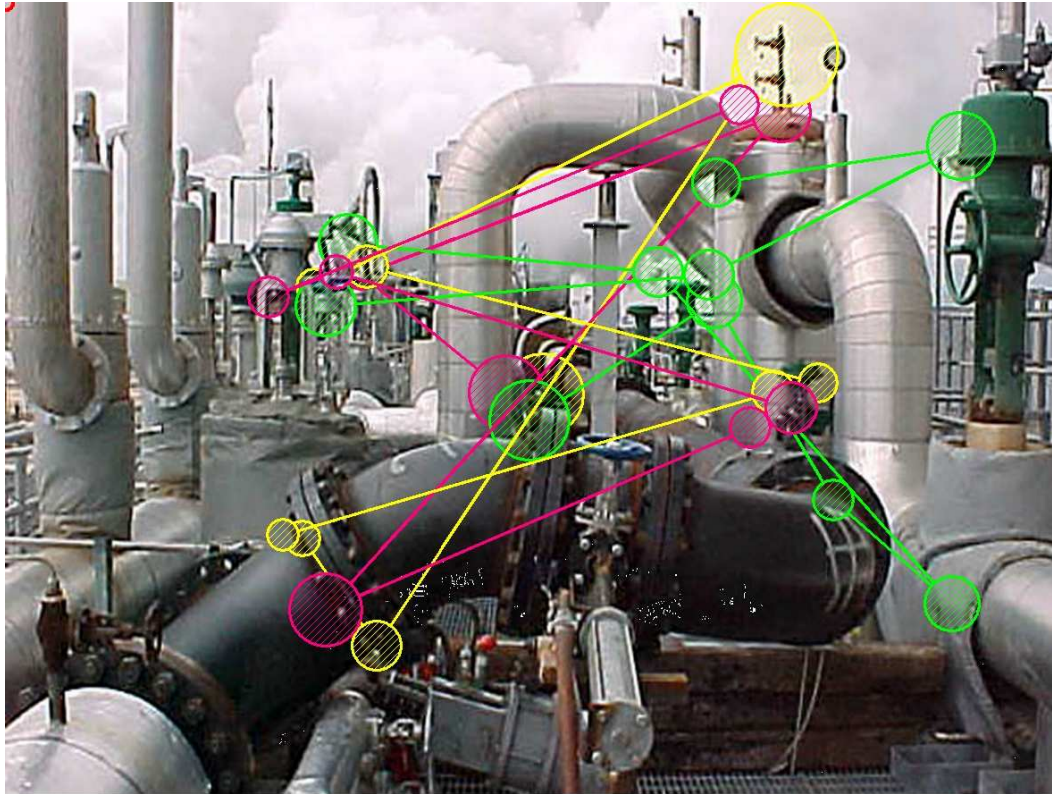
Figure 43: An example of an engineering stimulus. The pink scanpath was generated at encoding, the yellow scanpath was generated during immediate test and the green scanpath was generated during delayed test. Notice the similarities between the scanpaths, even at delayed test one week later.

3.3.4 Discussion

This experiment aimed to find out whether scanpaths are stable over time and whether accuracy and eye movements differed with expertise. Accuracy was measured as the number of pictures at test correctly identified as 'old' or 'new' and was measured immediately after the encoding stage and also one week later. The results show that accuracy is reliably increased when participants view pictures from their own domain of expertise. This pattern is stable over time, as there are no reliable differences in accuracy (for each group) between immediate and delayed test. This supports previous studies that have reported good visual long term memory (Castelhano and Henderson, 2005; Hollingworth & Henderson, 2002; Hollingworth, 2005) and also those that have found evidence of better long-term memory for domain-specific material (Hayes-Roth, 1983; Postal, 2004; Drai-Zerbib and Baccin, 2004) than irrelevant material. This domain-specific memory advantage is also stable over time.

The number of fixations to salient regions of interest at encoding was analysed. The results showed that participants did look to areas of high saliency but this was reliably reduced when viewing pictures from their own domain. This suggests a bottom-up cognitive override of saliency by top-down domain expertise – participants looked at the parts of the picture that were semantically interesting or meaningful, even if they were not the most visually salient areas. Findlay and Walker's (1999) model of saccade generation refers to this as "intrinsic saliency", whereby top-down influences at level 4 and 5 of the model can override bottom-up low-level influences at level 2. This analysis provides evidence that visual saliency attracts attention, and

thus eye fixations, but doesn't take account of the order or pattern of fixations (i.e. how quickly salient regions attract attention). To measure this, the participants' actual scanpaths on each picture were compared to the scanpaths predicted by Itti and Koch's (2000) saliency algorithm. The results showed that similarity was reliably decreased when participants viewed pictures from their own domain of expertise. This provides more evidence in favour of a cognitive override of low-level visual saliency.

Scanpaths at encoding were also compared to scanpaths at immediate recognition test and those at delayed recognition test one week later. When comparing encoding and immediate recognition, scanpaths for Engineers were most similar when viewing engineering stimuli. There was not a reliable difference between the groups for Civil War pictures. However, when scanpaths at encoding were compared to those at delayed imagery, similarity scores were highest when viewing domain-specific pictures, for both Engineers and American Studies students. The difference between the groups for Civil War pictures was due to a large decrease in similarity for Engineers (when viewing Civil War pictures). The similarity score for American Studies students was not reliably different (from encoding v immediate test). This suggests that Engineers used short-term memory to guide their eye movements around the Civil War pictures at immediate test, but due to the lack of semantically relevant visual information, the stimuli provided no advantage in encoding to long term memory. Therefore, when viewing the Civil War pictures again after a week delay, Engineers had little memory of where they looked previously to guide their eye movements. When viewing domain-specific stimuli, on the other hand, participants are likely to remember semantically interesting features of the pictures, resulting in enhanced long-term visual memory (and thus higher recognition accuracy) and increased scanpath similarity. Alternatively, it could have been that Civil War pictures contained distinctive objects (e.g. artifacts, medals, etc.) that were easy for Engineers to recall from short-term memory. Engineering pictures, on the other hand were more complex and harder for American Studies participants to recall only using short-term memory with no semantic relevance, hense not mirroring the advantage demonstrated by engineers at initial recognition test.

One outstanding question is *why* were scanpaths similar at encoding and subsequent recognition memory tests? Scanpath Theory suggests that producing similar eye movements helps to recreate the internal mental image or 'map' created at encoding, which in turn help the viewer to recognise the picture. The relationship between the increase in scanpath similarity and the increase in accuracy can been seen as evidence in support of this theory. However, it could be argued that scanpaths are reproduced on non-domain pictures due to low-level visual saliency and on domain-specific pictures due to top-down knowledge and semantically interesting visual features. Evidence for this comes from both the number of fixations in salient regions and the override of the model-predicted saliency scanpath by domain expertise. Either way, scanpaths could rely on visual information being present at both encoding and recognition. The imagery experiments in chapter 2 provide evidence for the reproduction of scanpaths in the absence of visual information. However, they did not investigate the effect of expertise during imagery. Scanpath similarity scores were found to be highest when encoding and retrieval processes were most similar, resulting in reliably reduced scanpath similarity between encoding and imagery. This reduction in similarity might eliminate the domain-specific advantage seen in the current study, thus suggesting that the advantage (in accuracy and scanpath similarities) is due to the presence of semantically informative regions rather than a superior memory for domain-specific material. This could be tested by taking away the visual information and repeating the experiment with an imagery condition.

In conclusion, this experiment has shown that scanpaths are stable over time and over multiple viewings and that accuracy and scanpath similarities increased when experts viewed domain-specific stimuli. This effect was also stable over time.

3.4 Experiment 6: Expertise vs. saliency on imagery, verbalisation and accuracy.

3.4.1 Introduction
The amount and organization of experts' domain knowledge has previously been linked with the speed of construction of mental images (Egan and Schwartz, 1979). It is argued that experts use 'chunking' methods (Gobet et al., 2001) which reduce working memory load so that more cognitive capacity can be devoted to reasoning. These chunking processes play an important role in blindfold chess (Saariluoma and Kalakoski, 1997), where imagery is used to retrieve information about game positions from long-term working memory and transform them in visual working memory. The link between memory, imagery and expertise is also highlighted by Hishitani (1988) who found that expert abacus users can encode long verbally presented numbers by encoding them into an imagined abacus. Similarly, artists have been found to perform reliably better on imagery and drawing tasks than non-artists, which Calabrese and Marucci (2006) conclude is due to superior cognitive abilities of artists for spatial imagery. This spatial expertise has, however, been shown to be highly domain-specific and not easily transferable to other domains. For example, skilled Tetris players were shown to have a mental rotation advantage when the shapes were either identical or very similar to Tetris shapes. However, no such advantage existed on other tasks of spatial ability (Sims & Mayer, 2002).

As mentioned in the previous chapter, verbalization can be combined with imagery to increase the similarity of eye movements at encoding and 'recognition'. Since the initial pilot experiment has shown this technique to be successful, verbalisation can now be explored further to investigate the influences of top-down (expertise) and bottom-up (visual saliency) factors on scanpaths and other eye measures. It would be interesting to see if the previously observed advantage of domain-expertise on scanpaths (sections 3.2 and 3.3) still exists when the visual information is taken away. If these advantages are still seen during imagery then they cannot be purely due to salient or semantically interesting visual features – there must be some kind of memory involved.

Verbalising a memory of a previously viewed scene has been known to affect recognition accuracy, which Schooler et al., (1993) found was true for both written and spoken verbalizations. Spoken verbalizations have the advantage of allowing eye movements to be recorded at the same time; therefore this was chosen as the most suitable method for the following experiment. This deterioration in memory due to verbalization is called verbal overshadowing and has been claimed to be due to the disruption of configural processing (e.g. Fiore and Schooler, 2002), especially in examining domain-specific stimuli (Diamond and Carey, 1986). The association between configural processing and expertise is well documented, (e.g. Fallshore and Schooler, 1995) with non-expertise linked to featural processing. Verbal overshadowing has previously been found in situations where participants' perceptual expertise exceeded their verbal expertise (e.g. Ryan and Schooler, 1995; Melcher and Schooler, 1996). For example, untrained wine drinkers (assumed to have low verbal expertise but high perceptual expertise), who described previously tasted wines showed a decreased discrimination performance compared to those who did not describe it. On the other hand, for the non-drinkers (assumed to be low on both types of expertise) and the trained

wine drinkers (assumed to be high on both types of expertise) verbalization was actually slightly helpful. Melcher and Schooler (1996) suggest that verbalization reduces one's ability to draw on perceptual expertise. It could also be that verbalization causes *transfer-inappropriate processing* in which, during recognition, participants inappropriately emphasize the verbal-featural information that they considered during verbalization rather than the nonverbal-configural information typically emphasized during the encoding. Under situations in which the default encoding involves a proportionately greater reliance on featural information, or recognition requires featural processing, the featural processing encouraged by verbalization is no longer inappropriate, and consequently no interference is observed. This performance enhancing effect of similar encoding and retrieval processes can be likened to the pure- and mixed-processing theory suggested in the previous chapter.

Many researchers have been unable to obtain evidence of the verbal overshadowing effect or have found that verbalization has a facilitating effect (e.g. Chance & Goldstein, 1976; Kitagami, Sato, & Yoshikawa, 2002; McKelvie, 1976; Meissner, Brigham, & Kelley, 2001; Read, 1979; Yu & Geiselman, 1993). For example, Read (1979) found that verbal rehearsal improved accuracy and decreased response time in recognition of faces, and McKelvie (1976) demonstrated that hard-to-label faces benefited from verbalisation, whereas easy-to-label faces were not significantly affected. Furthermore, when distracter and target items are highly similar, verbalization has been found to increase recognition accuracy by allowing the participant to distinguish between them (Bartlett, Till, and Levy, 1980; Kitagami, Sato, and Yoshikawa, 2002). The type of instructions given with respect to verbalization, and the delay between verbalization and the subsequent recognition test affect the direction and size of the verbalization effect. If participants are instructed to give a detailed verbal account of the features in the previously seen picture, a large overshadowing effect is observed. On the other hand, when a long delay elapses between verbalization and the recognition test then a facilitating effect of verbalization is observed (Meissner and Brigham, 2001). Verbalisation has also been found to enhance recognition memory if visual memory of the target is poor (Itoh, 2005).

Verbal overshadowing has previously been explained by the theory that post-stimulus verbalization induces a second, less detailed representation, which is subsequently drawn upon in the recognition test. This second representation then leads to recoding interference or source confusion, thereby decreasing recognition accuracy. However, if the initial representation of the visual stimulus is itself coarse and lacking in detail (Itoh, 2005), or if the verbal description is itself sufficient in discriminating between target and distracter items (Bartlett et al., 1980; Kitagami et al., 2002), the verbal overshadowing effect disappears or even reverses, and improved recognition is observed under conditions of verbalization (Huff and Schwan, 2008).

The following experiment aims to test whether recognition accuracy is affected by giving a description of a previously seen picture, compared to when simply imagining it, and whether this is affected by domain expertise. Eye movements will be recorded at imagery, verbalization and recognition and compared to eye movements at encoding to further investigate the similarity of scanpaths between these encoding and retrieval processes. The stimuli contain both visually salient regions and domain specific regions that are of semantic interest to the expert participants. Eye movements (average number of fixations and fixation durations) in these regions of interest, scanpaths over the whole picture and the content of the verbal descriptions will be analyses for both expert and control participants to investigate how expertise and saliency affect imagery, verbalisation and recognition accuracy.

3.4.2 Methodology

Participants

Thirty-one participants took part in the experiments, all of whom were students at Nottingham University. Sixteen were computer specialists and fifteen were control participants. The age range was 18-42 and the mean age was 22. The sample comprised 14 females and 17 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and the participants being naïve to eye movements being recorded.

Materials and apparatus

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker (for specifications, see section 1.5.5) A set of 52 high-resolution digital photographs were prepared as stimuli, taken using an 8MP digital camera. Half the stimuli were designated "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test (see *figure 44* for a diagrammatic breakdown of the stimuli). All stimuli were pictures of junk and clutter, such as the inside of a draw or a messy shelf. Half the new and half the old stimuli contained a computer-specialist item somewhere in the scene. The content of each picture was carefully chosen and arranged so that the computer-specialist part would not stand out to a non-specialist more than other objects in the picture or defy the gist of the scene. Due to the 26 test stimuli not splitting exactly into four equal groups, half the participants imagined 7 computer stimuli and described 6 computer stimuli and the other half of the participants imagined 6 computer stimuli and described 7 computer stimuli. This counterbalancing also applied to the control stimuli, and the pictures that were imagined or described were randomised. New and old pictures were matched for complexity, semantic and emotional content, so that the only thing in the scene that was different at recognition was the presence or absence of a computer-specialist item. Examples of the stimuli are shown in *figure 45*. Pictures were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. The monitor measured 43.5cm by 32.5cm, and a fixed viewing distance of 98cm gave an image that subtended 25.03 by 18.83 degrees of visual angle.

Figure 44: _diagrammatic breakdown of the stimuli._
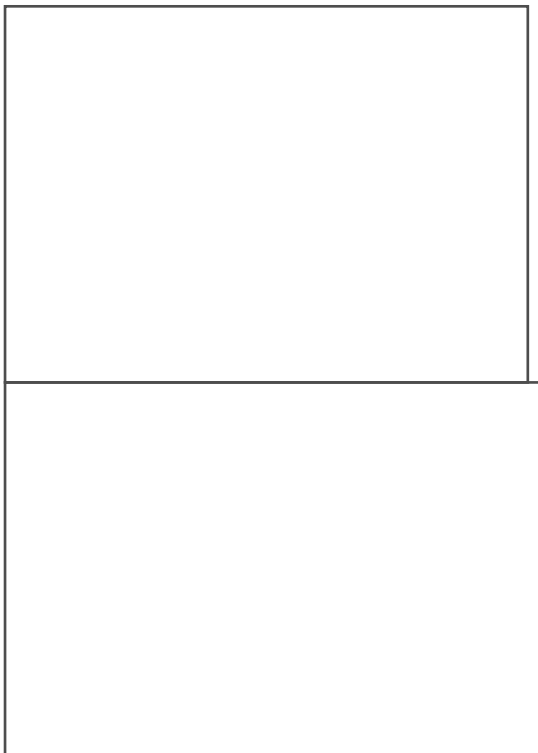

Figure 45: _Examples of stimuli. On the left: a computer-specialist picture, and on the right: a matched control picture._

All stimuli were processed through the Itti & Koch (2000) computer algorithm to make sure that the domain-specialist Region of Interest in each picture was not one of the five most visually salient areas of the picture.

A pilot study was conducted prior to the experiment to make sure the computer specialists had true expertise in their domain. Potential participants had to complete a short questionnaire requiring them to name and describe

pictures of computer parts. Only those who scored 100% were invited to take part in the experiment as domain specialists. Control participants also had to complete the same questionnaire and were considered if they scored 5% or below.

Design

The experiment used a mixed design, with one between groups IV with two levels (specialist or non-specialist participant) and one within groups IV with two levels (condition – imagery or describe). The dependent variable measures were: accuracy in deciding whether a picture was old or new, verbal feedback (the average number of references to visually salient and specialist Regions of Interest) the average number of fixations in salient and specialist RoIs, the average fixation duration and total amount of time fixated in salient and specialist RoIs, where the first fixation fell and the similarities of scan patterns compared at encoding and visualisation, at encoding and description and at encoding and recognition.

Procedure

Participants were told that their pupil size was being measured in relation to mental workload. They were informed that although their eye movements were not being recorded, it was important to keep their eyes open so pupil size could be reliably measured.

Following a 9-point calibration procedure, participants were shown written instructions on the experimental procedure and given a short practice. The main experimental procedure was split into three parts with breaks between each section. This made sure that participants understood what they had to do for each part of the experiment, gave them the opportunity to ask the experimenter any questions and also allowed for recalibration on the eye tracker. The first stage ('visualisation') involved seeing a picture for 7000 milliseconds (preceded by a 1000 millisecond central fixation cross) then a brightly coloured mask for 1000 milliseconds. The word "visualize" then appeared on a white background for 1000 milliseconds before disappearing and leaving a blank screen for 10,000 milliseconds, in which time the participants visualized the last picture they had seen. After this time, a fixation cross reappeared for 1000 milliseconds to ensure that fixation at picture onset was in the centre of the screen. Once 13 stimuli had been presented and subsequently visualised, the participant took a short break before starting the second stage – 'description'. The procedure for the second stage was identical to the first apart from instead of 'visualising' the last picture they saw, participants were asked to describe it. The results of the pilot study (chapter 2.4) indicate that the participants need time to think what to say, therefore after the word 'describe' appeared for 1000 milliseconds, the numbers '3', '2', and '1' appeared for 1000 milliseconds each in a countdown style. A head-mounted microphone was worn and verbal descriptions were recorded digitally. See figures 46a and 46b for diagrams of the experimental procedure. The final stage was a recognition memory test, where participants had to decide if they had seen each picture before by making a button press on the keyboard ('O' for 'old' and 'N' for 'new'). Each picture at recognition was presented for 7000 milliseconds and the computer would only accept a response after this time. This was to encourage similar eye movement patterns at recognition as at encoding (stages 1 and 2) so that fair string comparisons

could be made at analysis.
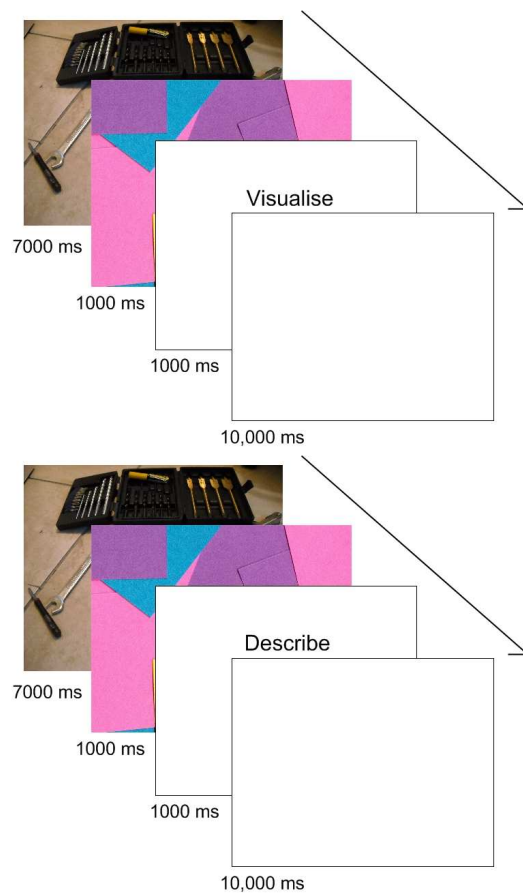
Figures 46a (left) & 46b (right): *diagrams of the visualise (stage 1) and description (stage 2) procedures. Each picture was either visualised or described, but was never in both conditions.*

3.4.4. Results

Accuracy
An independent samples T-test was performed to compare the overall accuracy of specialists and control participants (see appendices 28a and 28b). There was a reliable difference: $t(29) = 2.829$, SED $= 0.0253$, $p<0.01$. From this result and the descriptive statistics, it can be concluded that overall specialists were more accurate than controls. See figure 47.

Figure 47: *A bar chart to illustrate the reliable difference in accuracy between specialists and control participants.*

Accuracy scores were broken down further to take into account whether the stimuli were old or new and whether they contained computer parts or not. A 2x2x2 split-plot ANOVA was performed (see appendices 28c, and 28d) and found a 2-way interaction between new/old and comp/non comp stimuli: $F(1,29) = 5.582$, MSE = 0.030, $p<0.05$ (see figure 48); and a 3-way interaction between group, old/new and comp/non comp stimuli: $F(1) = 7.140$, MS = 0.213, $p<0.05$ (see figure 49). There was also a main effect of group, but this as already illustrated above.

Figure 48: *a line graph to illustrate the reliable 2-way interaction between new/old and comp/non comp stimuli.*

The 2-way interaction shows that if a picture is new, it is recognised more accurately if it is a computer picture, and if a picture is old, it is recognised more accurately if it is a non-computer picture. Post-hoc paired-samples T-tests (appendices 28e and 28f) showed that the difference between old computer stimuli and old non computer stimuli was statistically significant: $t(30) = 2.156$, *SEM* = 0.03913, $p<0.05$. No other comparisons were statistically reliable.

Figure 49: *A line graph to show the 3-way interaction between group, old/new and comp/no comp stimuli.*

Post-hoc independent samples T-tests were performed to help explain the 3-way interaction (appendices 28g and 28h) and found a reliable difference between the groups for new non computer stimuli: $t(29) = 0.265$, SED = 0.0633, $p<0.05$; and for old computer stimuli; $t(19) = 2.163$, SED = 0.0647, $p<0.05$. From these results and the descriptive statistics, it can be concluded that Specialists are more accurate at new non-comp stimuli than controls and specialists are more accurate at old comp than controls.

Post-hoc paired samples T-tests were also performed to help explain the 3-way interaction (appendices 28i and 28j) and found a reliable difference between old computer pictures and old non computer pictures for control participants: $t(14) = -5.104$, SEM = -0.19487, $p<0.001$; and between new non computer pictures and old non computer pictures for control participants: $t(14) = -3.543$, SEM = -0.200, $p<0.05$. No other comparisons were statistically reliable. From These results and the descriptive statistics, it can be concluded that if the picture is old, controls are more accurate if the stimuli do not contain computer parts, and controls are more accurate at new

computer stimuli than old computer stimuli.

Accuracy was also analysed taking into consideration whether the picture was imagined or described. A 2x2x2 split-plot ANOVA was performed (group x imagine/describe x computer/non computer pictures. See appendices 29a, 29b and 29c) and found reliable main effect of condition (imagery or describe): $F(1, 29) = 5.907$, $MSE = 568.334$, $p<0.05$ (see figure 50); and a main effect of stimulus type (computer or non computer picture): $F(1, 29) = 5.342$, $MSE = 369.237$, $p<0.05$. There was no main effect of group: $F(1,29) = 0.536$, $MSE = 711.191$, $p = 0.470$. From these results and the descriptive statistics, it can be concluded that accuracy is also reliably higher when pictures were previously described than when they were only imagined, and accuracy is overall higher for non computer pictures.
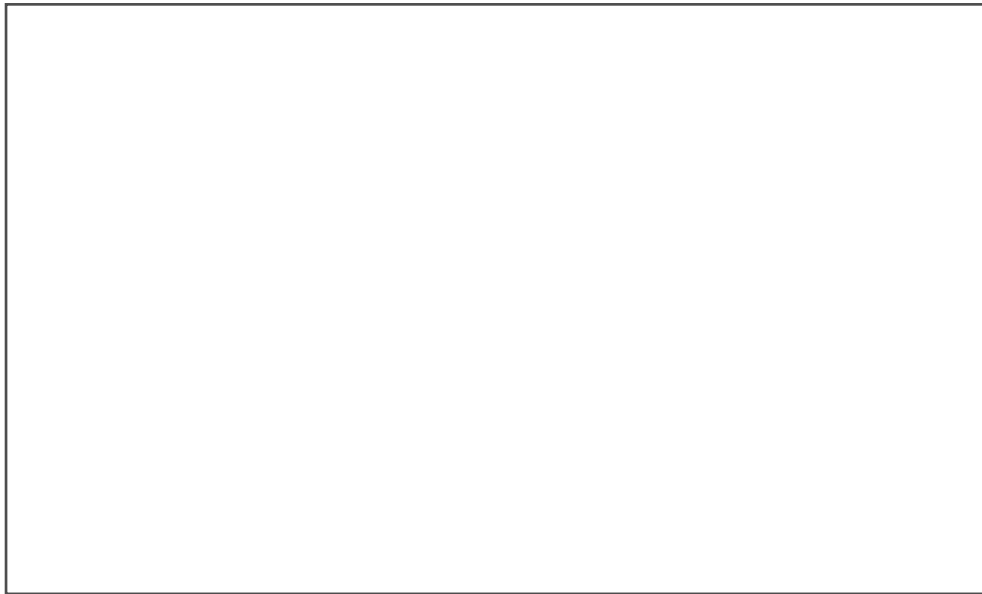
Figure 50: *A bar graph to show the reliable difference in accuracy between pictures that were previously imagines and those that were described.*

There was no reliable interaction between condition and group: $F(1) = 0.347$, $p = 0.560$; or between condition and stimulus type: $F(1) = 9.56$, $p = 0.336$. There was however a reliable interaction between stimulus type of group: $F(1) = 9.266$, $p<0.01$ (see figure 51). There was no reliable 3-way interaction between the variables.

Figure 51: *A line graph to show the interaction in accuracy between group and stimulus type.*

Post-hoc paired samples T-tests showed no statistically reliable differences (see appendices 29d and 29e), but an independent-samples T-test showed a reliable difference between the groups for computer stimuli: $t(29) = 2.611$, $SED = 3.66475$, $p<0.01$ (see appendices 29f and 29g). From these results and the descriptive statistics, it can be concluded that specialists are reliably more accurate at computer pictures than control participants.

Verbal Feedback
Verbal feedback during the description phase was recorded and analysed. The number of references to the salient regions of interest was calculated and a 2x2 repeated measures ANOVA was performed (appendices 30a and 30b). Reliable main effects were found for picture type (computer or non computer): $F(1,14) = 11.718$, $MSE = 3.995$, $p<0.01$; and for group (specialist or control participant): $F(1,14) = 6.351$, $MSE = 4.852$, $p<0.05$. See figure 52.
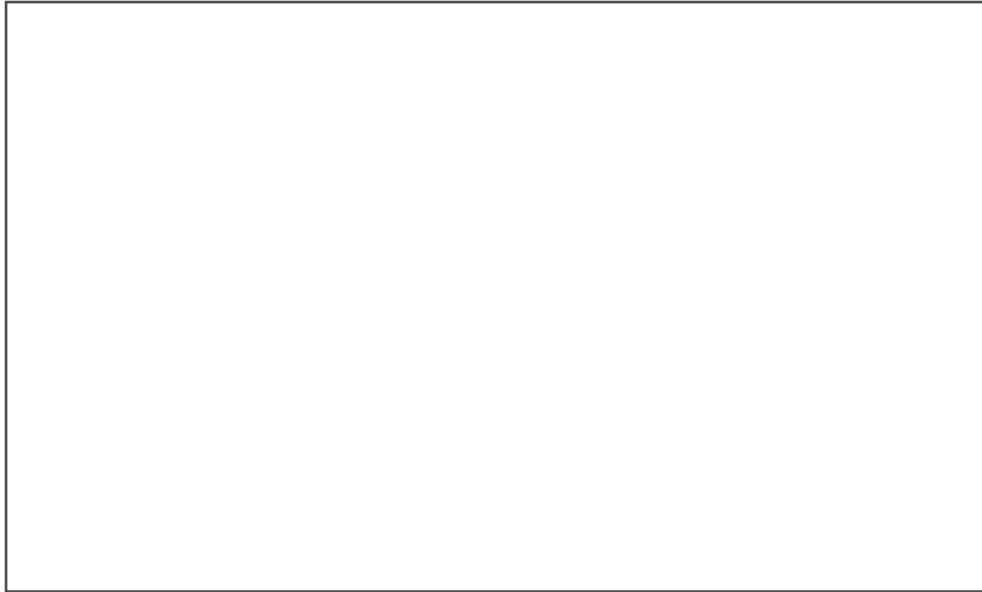
Figure 52: *A bar chart to illustrate the average number of references to visually salient areas in computer and non computer pictures by specialist and control participants.*

Post-hoc paired samples T-tests were conducted (appendices 30c and 30d) and found reliable differences between specialists and controls for computer pictures: $t(14) = 3.587$, SEM = 0.46462, $p<0.01$; between computer and non computer pictures for control participants: $t(14) = -2.553$, SEM = 0.6005, $p<0.05$; and between computer and non computer pictures for specialists: $t(14) = 2.197$, *SEM* = 0.91026, $p< 0.05$. There was no reliable difference between specialists and control participants for non computer stimuli: $t(14) = 1.151$, *SEM* = 1.04289, $p = 0.269$.

A paired samples T-test was also carried out to compare the number of references to specialist Regions of Interests when viewing computer stimuli (appendices 30e and 30f). There was a reliable difference between specialist and control participants: $t(14) = 7.429$, SEM = 0.54743, $p<0.001$. See figure 53.

Figure 53: *A bar graph to illustrate the average number of references to specialist RoIs by the two groups of participants.*

From the analysis of the verbal feedback data, it can be concluded that specialists make reliably more references to specialist RoIs than control participants do, and when viewing computer stimuli, specialists make reliably fewer references to visually salient areas than control participants**.** On non-computer stimuli, there is no reliable difference in the number of references to visually salient areas between specialists and controls.

Number of fixations in RoIs
Inclusion in the eye movement analyses was dependant on good eye tracking and four participants (two computer specialist and two controls) had to be excluded due to calibrations issues.

The number of fixations that fell inside visually salient Regions of Interest and specialist Regions of Interest at recognition were calculated. The salient RoIs were defined using Itti & Koch's (2000) computer algorithm, which highlighted the five most salient regions. Each picture therefore had 5 salient RoIs, each spanning 2 degrees of visual angle. There was only 1 specialist RoI in each computer picture and it was generally larger that the salient ones. When the average number of fixations was calculated, adjustments were made to balance the size of the RoIs. The number of fixations that fell inside visually salient Regions of Interest and specialist Regions of Interest were calculated. The salient RoIs were defined using Itti & Koch's (2000) computer algorithm, which highlighted the five most salient regions. Each picture therefore had 5 salient RoIs, each spanning 2 degrees of visual angle. There was only 1 specialist RoI in each computer picture and it was generally larger that the salient ones. When the average number of fixations was calculated, adjustments were made to balance the size of the RoIs. The average size of the specialist RoI was 1.75 times greater than one salient RoI, so over all five salient RoIs were summed and then divided by five (creating an average value), then multiplied by 1.75 to balance the size of the salient and specialist RoIs.

An independent samples T-test was carried out (see appendices 31a and 31b) and found a reliable difference between the groups for the number of fixations in specialist RoIs: $t(25) = 5.093$, *SED* $= 0.22141$, $p<0.05$. There were no reliable differences between the groups for the number of fixations in salient RoIs for either computer pictures ($p=0.838$) or non computer pictures ($p=0.084$). Paired samples T-Tests were also carried out (see appendices 31c and 31d) and found a reliable difference between the number of fixations in salient RoIs made by specialist participants, between computer and non computer stimuli: $t(13) = 3.653$, *SEM* $= 0.09647$, $p<0.05$. There was no reliable difference for control participants between computer and non computer stimuli. From the results of the T-Tests and the descriptive statistics (appendices 31a and 31c) it can be concluded that specialists make reliably more fixations to specialist RoIs than control participants, and that specialists make reliably fewer fixations to salient RoIs in computer pictures compared to non computer pictures. See figures 54 and 55.

Figure 54: A bar chart illustrating the number of fixations to specialist and salient RoIs in computer and non computer stimuli.
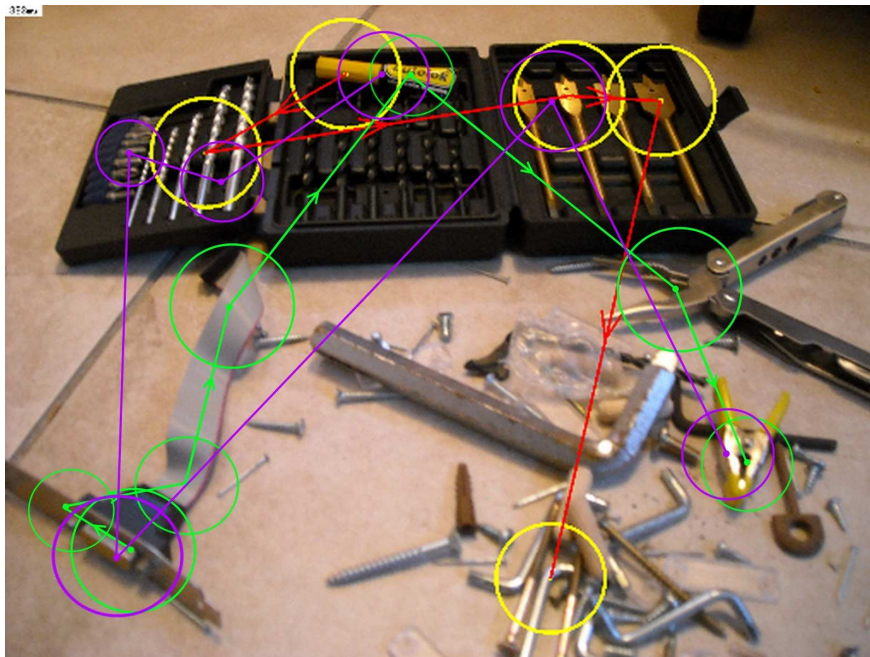

Figure 55: A computer stimulus with a model-predicted saliency scanpath (yellow and red), a specialist's scanpath (green) and a control participant's scanpath (purple) superimposed on top of it. Note that the specialist's scanpath fixates more on the specialist RoI than the control participant's scanpath.

The number of fixations to specialist RoIs was further analysed by dividing the pictures at

recognition into four categories: New 'added in' (computer part added in), New 'taken away' (computer part taken away), Old 'same present' (computer part present at encoding and recognition), Old 'same absent' (no computer part at encoding or recognition). For specialists (see appendices 31v and 31w), there was no reliable difference between computer pictures at encoding and identical pictures at recognition: $t(13) = 1.939$, $SEM = 0.31909$, $p = 0.075$; or at recognition between 'same present' and 'added in': $t(13) = 0.256$, $SEM = 0.27846$, $p = 0.802$; or at recognition between 'same absent' and 'same taken away': $t(13) = 1.098$, $SEM = 0.01085$, $p = 0.292$. There were reliable differences between: computer pictures at encoding and 'taken away' at recognition ($p<0.001$); non computer pictures at encoding and 'added in' at recognition ($p<0.001$); 'same present' at recognition and 'same absent' at recognition ($p<0.001$); 'same present' at recognition and 'taken away' at recognition ($p<0.001$); 'added in' at recognition and 'same absent' at recognition ($p<0.001$); and between 'added in' at recognition and 'taken away' at recognition ($p<0.001$).

This analysis was repeated for control participants (see appendices 31x and 31y) and found that there was no reliable difference between computer pictures at encoding and identical pictures at recognition: $t(12) = 2.128$, $SEM = 0.34563$, $p = 0.055$; or at recognition between 'same present' and 'added in': $t(12) = 1.751$, $SEM = 0.22582$, $p = 0.105$; or at recognition between 'same absent' and 'same taken away': $t(12) = 0.826$, $SEM = 0.00861$, $p = 0.425$. There were reliable differences between: computer pictures at encoding and 'taken away' at recognition ($p<0.001$); non computer pictures at encoding and 'added in' at recognition ($p<0.001$); 'same present' at recognition and 'same absent' at recognition ($p<0.001$); 'same present' at recognition and 'taken away' at recognition ($p<0.001$); 'added in' at recognition and 'same absent' at recognition ($p<0.001$); and between 'added in' at recognition and 'taken away' at recognition ($p<0.001$).

Total fixation duration in RoIs

The amount of time fixated in salient and specialist RoIs (per picture) at recognition was calculated (see *figure 56*).

Figure 56: A bar chart illustrating the average total time fixated in specialist and salient RoIs in computer and non computer stimuli at recognition.

An independent samples T-test was carried out (see appendices 32a and 32b) and found a reliable difference between the groups for the total time fixated in specialist RoIs: $t(25) = 5.362$, *SED* = 65.70062, $p<0.001$. There were no reliable differences between the groups for the amount of time ficated in salient RoIs for either computer pictures ($p=0.987$) or non computer pictures ($p=0.098$). Paired samples T-Tests were also carried out (see appendices 32c and 32d) and found a reliable difference between the average total amount of time fixated in salient RoIs made by specialist participants, between computer and non computer stimuli: $t(13) = 2.679$, *SEM* = 56.84325, $p<0.05$. There was no reliable difference for control participants between computer and non computer stimuli. From the results of the T-Tests and the descriptive statistics (appendices 32a and 32c) it can be concluded that specialists spend reliably longer fixated in specialist RoIs than control participants, and that specialists spend reliably less time fixated in salient RoIs of computer pictures compared to non computer pictures.

First Fixation
The position of the first fixation was calculated and the number of first fixations that fell in the different RoIs was compared using T-Tests (see *figure 57*).
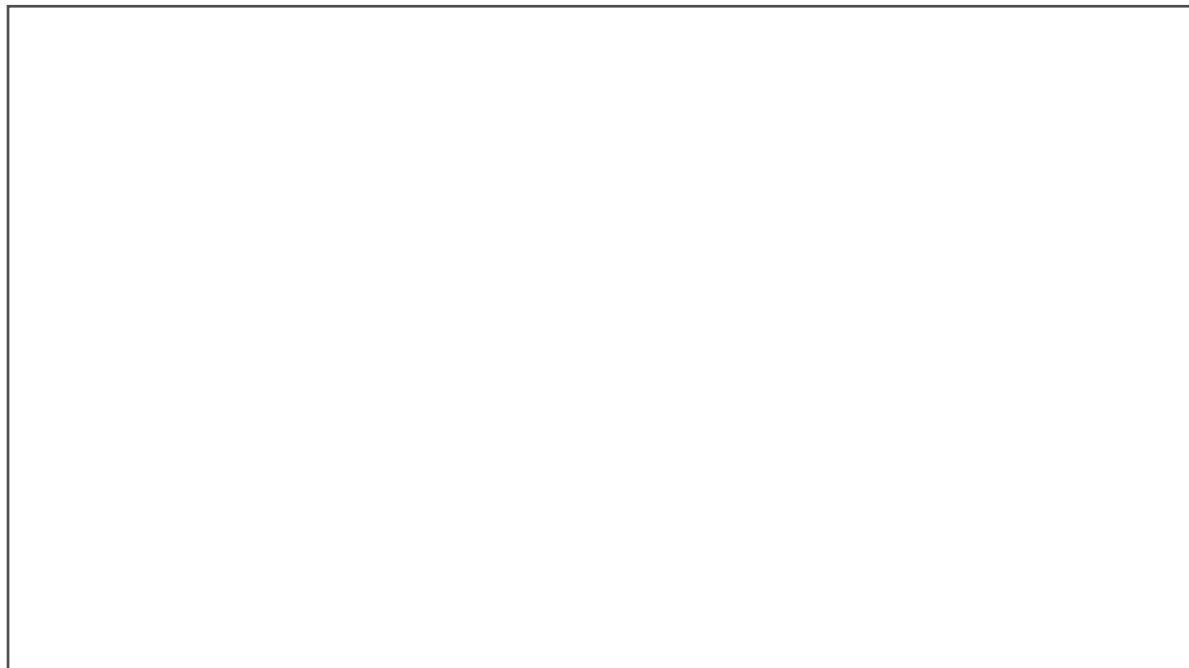

Figure 57: A bar chart to illustrate the percentage of first fixations to fall in salient and specialist RoIs.

An independent samples T-test was carried out (appendicies 33a and 33b) and found no reliable difference between the groups for salient RoIs of computer pictures ($p=0.244$), for salient RoIs of non computer pictures ($p=0.243$) or for specialist RoIs of computer pictures ($p=0.061$). Paired samples T-tests were also carried out (appendices 33c and 33d) and found a reliable

difference in the number of salient RoIs first fixated by specialists, between computer and non computer pictures. It can be concluded that specialist participants are reliably less likely to first fixate a salient region if a specialist (semantically informative) region is also present.

String Comparisons

A string editing algorithm was used to compare eye movements at visualisation, encoding and recognition, and a similarity score was calculated. A trial was excluded if there were more than 25% data loss or the participant did not move their eyes away from the centre. When comparing Encoding and Imagery, an average of 17.2% of trials had to be excluded. When comparing Encoding and Description, an average of 16.85% of trials had to be excluded. When comparing Encoding and Recognition, an average of 7% of trials had to be excluded. If a participant had more than 75% or more trial exclusions over all, that participant was discarded from the analysis. Because paired samples T-tests compared averages across the above three comparison groups (EvI, EvD, EvR), if a participant was excluded in one group, they had to be taken out of the other groups as well. In total, 12 participants out of 31 were excluded from the string comparison analysis.

A split-plot ANOVA was performed and found a main effect of comparison: $F(2, 34) = 17.053$, $MSE = 0.001$, $p<.001$ (appendix 34a). There was no main effect of group: $F(1,17) = 3.737$, $MSE = 0.001$, $p = 0.070$ (appendix 34b); and there was no reliable interaction between comparison and group: $F(2) = 1.253$, $MS = 0.001$, $p = 0.298$. See figure 58.

Figure 58: *A Bar chart to illustrate the similarity of eye patterns at different stages in the experiment.*

Post-hoc paired samples T-tests were carried out to further investigate the main effect of comparison (appendices 34c and 34d). There was a reliable difference between encoding v visualisation and encoding v describe: $t(18) = 2.511$, $SEM = 0.01197$, $p<0.05$; a reliable difference between encoding v visualisation and encoding v recognition: $t(18) = 6.568$, $SEM = 0.00962$, $p<0.001$; and a reliable difference between encoding v describe and encoding v recognition: $t(18)$

= 2.978, *SEM* = 0.01081, *p*<0.01. Each comparison was also compared to chance. There was a reliable difference between encoding v describe and chance: *t*(18) = 2.477, *SEM* = 0.00936, *p*<0.05; and a reliable difference between encoding v recognition and chance: *t*(18) = 5.870, *SEM* = 0.00943, *p*<0.001. There was no reliable difference between encoding v visualisation and chance: *t*(18) = 0.803, *SEM* = 0.00854, *p* = 0.432.

The data were then split by stimulus type and group to see whether scanpath similarities differed with domain (see figure 59).

Figure 59: *A Bar chart to illustrate the similarity of eye patterns at encoding compared to the other stages in the experiment. The data has also been split by stimulus type and participant group.*

Independent-samples T-tests were conducted (appendices 34e and 34f) and found that scanpaths at encoding and verbalisation for computer pictures were reliably more similar for specialists than control participants: *t*(17) = 2.552, *SED* = 0.01965, *p*<0.05; and that scanpaths at encoding and recognition for computer pictures were reliably more similar for specialists than control participants: *t*(17) = 2.128, *SED* = 0.02481, *p*<0.05. No other between-groups comparisons were statistically reliable. Scanpaths for a computer specialist viewing a computer picture at different stages in the experiment can be seen in figure 60.
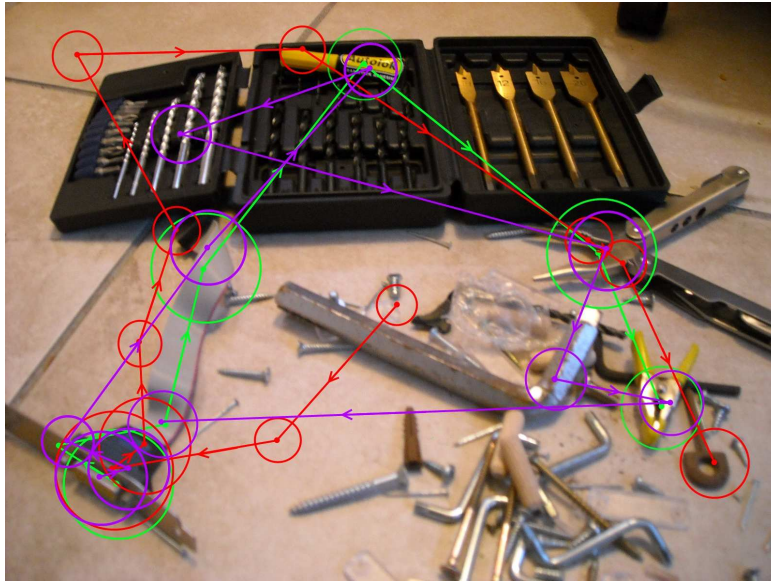
Figure 60: *A computer stimulus with a specialist's scanpaths superimposed on top of it. The green scanpath is from encoding, the red from description and the purple from verbalisation. Each picture at encoding was either described from memory or simply imagined, but never both, therefore an imagery scanpath does not exist for this stimulus.*

3.4.4. Discussion

This experiment aimed to investigate the effect of top-down expertise and bottom-up saliency on imagery, verbalisation and accuracy. Recognition accuracy was measured and found that overall specialists were more accurate than controls. This was most likely because half the pictures were computer pictures, which specialists were better at recognising, and therefore increased the overall average accuracy. A 3-way interaction was found and post-hoc analyses revealed that specialists were more accurate at new non-computer stimuli and old computer stimuli than control participants were. Pictures were matched i.e. old pictures were the same as new ones except for containing either a computer part or a matched neutral object. Specialists' domain expertise means they were more likely to have attended to the computer parts at encoding, which helped them distinguish old comp pictures at recognition and gave them an advantage over control participants. Similarly, specialists' high accuracy for new non-comp pictures could be due to them attending to a computer parts at encoding and noticing that they were missing at recognition. The post-hoc analyses also revealed that if pictures were old, control participants were more accurate if the stimuli did not contain computer parts. It could be that all the computer parts look the same to control participants and therefore they are more likely to incorrectly identify computer pictures than non-computer pictures that contain more every-day objects. Lastly, it was found that control participants were more accurate at identifying old computer pictures than new computer pictures. This could be because all the computer parts look the same to controls; therefore there is a response bias to say 'old' when they notice a computer part. When the picture is actually old, the average accuracy for 'old' increases, and if the picture is new then the average accuracy for 'new' decreases, thus explaining the reliable difference in accuracy. Although it is not feasible to reanalyse the accuracy data for the purposes of this thesis, future research could consider replacing the accuracy analyses with signal-detection measures to better distinguish response bias from actual memory.

Accuracy was also analysed to take into account whether the picture was previously imaged or described and whether this differed with stimulus type or participant group. Specialists were found to be reliably more accurate at computer pictures than control participants, which implies that they are truly specialists of their domain. Overall, accuracy was reliably greater for non computer pictures than computer pictures. This is due to control participants being particularly inaccurate on computer pictures, thus bringing the average down (see figure 51). Most interestingly, accuracy was reliably greater when the picture had been previously described compared to when it was simply imagined and this did not reliably differ with stimulus type of group. This is in contrast to previous studies that have found a negative effect of verbalisation on recognition memory (verbal overshadowing). Itoh (2005) found that verbalization interferes with subsequent face recognition when memory for the target person is good, whereas verbalization enhances recognition when memory is poor. So, the findings of the current experiment could be due to poor recognition memory for the stimuli at encoding, which can be seen in the low accuracy scores for pictures that were imagined only. One reason for poor memory could be the similarity between pictures at encoding and recognition, since the only thing that differed was the specialist region of interest. Verbalisation could have therefore helped participants to distinguish between old and new pictures, thereby increasing recognition accuracy. This supports previous finding of Bartlett et al (1980), and Kitagami et al (2002) who also found a facilitating effect of verbalisation when distracter and target items were highly similar. Another reason why stimuli were difficult to remember was because they were made up of many small features (e.g. a drawer full of junk) and thus could not be easily summarised in a word or descriptive sentence. According to the processing shift hypothesis, people encode visual stimuli using a configural/holistic process style, but when they verbalise the memory of these pictures, they focus on visual details of the stimulus, resulting in a feature-based cognitive style. This results in *transfer-inappropriate processing* and can lead to decreased recognition accuracy. However, due to the complex feature-based nature of the stimuli in the current experiment, it is unlikely that the participants would have been able to encode them holistically. Verbalising memories of the picture details in this case facilitated recognition at test. Furthermore, this did not differ with stimulus type. Previous studies have found that verbalization may produce a rather generalized disruption of perceptual memory processing, not just for the pictures that were verbalised (Dodson et al., 1997; Westerman and Larsen, 1997). Therefore, it is reasonable to suggest that the same generalisation effect could occur when verbalising facilitates recognition. This would explain the high accuracy across stimulus types.

Descriptions given during verbalisation were recorded and analysed for the number of references to salient and specialist regions of interest (RoIs). On computer pictures, specialists made fewer references to salient areas than control participants did, and specialists also made more references to specialist RoIs. Moreover, specialists refer to salient areas less in computer pictures (when a specialist RoI is present) than in non computer pictures. Firstly, this shows that specialists were attending to the semantically interesting regions and secondly implies an overriding effect of domain knowledge on saliency. The fact that controls refer more to the salient areas in computer pictures than specialists and less to specialists RoIs, supports previous findings of Lowe (1999) who found that novices mentioned more often irrelevant but perceptually salient features after the inspection of dynamic weather maps, suggesting that they attended more to these features. Control participants made more references to salient areas for non computer pictures than computer pictures. It could be that the specialist RoIs attract control participants' attention because they are unusual (unfamiliar), even though they are not semantically meaningful or

informative (and thus not easily describable). Control participants then had less time to describe the other (more salient) parts of the picture. There was no reliable difference in the number of references to salient areas between specialists and control participants for non computer pictures. This shows that control participants do not make more references to salient areas in general and reinforces that the difference in references to salient areas on computer pictures between specialists and controls is due to a cognitive override.

To test whether the verbal feedback related to where the participants actually looked, the number of fixations in salient and specialist RoIs at recognition was analysed. Independent samples and paired samples T-Tests were carried out and found that specialists make reliably more fixations to specialist RoIs than control participants, and that specialists make reliably fewer fixations to salient RoIs in computer pictures compared to non computer pictures. This backs up the verbal description data and implies that semantically relevant/interesting parts of a picture attract attention. Furthermore, the decrease in fixations to salient RoIs when a specialist RoI is present suggests a cognitive override of low-level visual saliency by top-down cognitive knowledge. For controls, there was no reliable difference in the number of fixations to salient RoIs between computer and non computer pictures. This too backs up the verbalisation data, whereby specialists' reduced number of fixations to salient RoIs on computer pictures is due to a cognitive override and not because control participants generally made more fixations to salient RoIs. Control participants did fixate the specialist RoIs in computer pictures (despite these areas offering no memory advantage), possibly due to them being unusual or unexpected. The fact that controls make reliably fewer fixations to these specialist RoIs than specialist participants do suggests that even though these areas may capture attention, control participants disengage their attention sooner, making fewer refixations and inspecting them less thoroughly.

Previous research has found that viewers often fixate an empty scene region when that region previously contained a task-relevant object (Altmann, 2004; Richardson & Spivey, 2000) and that individuals view the region of a familiar scene in which a change has been introduced more than they view a matched region of another scene that has not been changed (Ryan et al., 2000; Smith et al., 2006). Pictures at recognition (in which fixations to specialist RoIs could be measured) could belong to one of four categories: New 'added in' (computer part added in), New 'taken away' (computer part taken away), Old 'same present' (computer part present at encoding and recognition), Old 'same absent' (no computer part at encoding or recognition). So, in order to establish whether the increased number of fixations to specialist RoIs at recognition was due to participants looking at the area that used to contain a computer part, or whether they simply looked more to computer parts present at recognition to facilitate memory for the picture, further post-hoc tests were carried out for specialists and control participants. The results for specialists showed no reliable difference in the number of fixations between when the specialist RoI was present at encoding and when the identical picture was present at recognition. The overall result was that specialists fixated computer parts present at recognition, regardless of whether they were present at encoding or not. This suggests that these domain-specific features automatically attract attention, rather than participants fixating them due to a memory effect. This is supported by the finding that there was no reliable difference between specialists RoIs that appear at encoding (old 'same present' stimuli) and those that have been added in to pictures that originally contained a neutral object (new 'added in'). So far this has established that attention is drawn towards semantically interesting information. However, to test whether memory for location of previously seen specialist RoIs plays a role in this, stimuli where the specialist RoI had been taken away (new 'taken away') were compared to stimuli that never contained a computer part (old 'same

absent'). If memory played a role in guiding eye movements to the specialist RoI then there should be more fixations to RoIs in new 'taken away' stimuli than old 'same absent' stimuli. There was no reliable difference, suggesting that memory does not play a significant role in guiding eye movements to specialist RoIs, and rather attention is drawn automatically due to the underlying semantic relevance. These post hoc analyses were repeated for control participants and the same results were found apart from a slight increase in the number of fixations to specialist RoIs in old 'same present' pictures compared to the identical pictures at encoding. It could be that all computer pictures look the same to control participants and they therefore try to use the specialist RoIs to help identify them at recognition, (even though these don't actually provide any memory advantage). This difference was nearing statistical significance ($p = 0.055$) and could be the reason for the overall increase in fixations to specialist RoIs at recognition.

The average total amount of time spent in salient and specialist RoIs was analysed for computer and non computer pictures at recognition for each participant group. Computer specialists spent reliably more time fixated in specialist RoIs of computer pictures than control participants did. Computer specialists also spent reliably less time fixated in the salient RoIs when viewing computer stimuli than when viewing non computer stimuli but there was no such reliable difference for control participants. These results support the verbal data and number of fixation analyses, suggesting that participants do look at visually salient areas, but this is decreased when a competing semantically informative area is present. This relates to Findlay and Walker's (1999) concept of "intrinsic saliency" whereby top-down cognitive knowledge can override bottom-up influences such as visual saliency. Both participant groups also spent reliably more time fixated in specialist RoIs on computer stimuli than the same areas on non computer stimuli (i.e. chance), which supports the aforementioned suggestion that attention is automatically drawn to semantically informative areas rather than relying on memory of where the specialist RoI was originally at encoding.

The location of the first fixation was analysed and found that on computer pictures, specialists made reliably more first fixations to specialist RoIs than to salient RoIs and fewer first fixations to salient RoIs on computer pictures than on non computer pictures. This reinforces the argument for a cognitive override effect of saliency by domain-specific expertise. The increased number of first fixations to domain-specific RoIs, coupled with the reliably longer time spent fixated in these RoIs supports Evans and Treisman (2005) who found evidence for initial orienting towards RoIs of semantic relevance but once fixated, further processing was needed.

Scanpaths at each condition (encoding, imagery, description and recognition) were compared and string similarity scores were calculated. Encoding and recognition were most similar (reliably more similar than all other comparisons and reliably greater than chance). This is due to it being a 'pure' comparison (see chapter 2.4), with visual information present in both conditions. More interestingly, scanpaths at encoding and description were more similar than those at encoding and visualisation. These results show that describing a complex picture from memory produces more similar eye movements to encoding than are produced when simply imagining it. The string similarity analyses were split by stimulus type and participant group to see if a domain advantage still existed when no visual information was present. As well as replicating the previously seen advantage for domain-specific pictures at recognition, the results showed that for computer pictures at description (verbalisation), scanpaths produced by computer specialists were reliably more similar to encoding than scanpaths produced by control participants. There was no reliable difference for non domain-specific pictures in any of the conditions. This suggests that the increased scanpath similarity for domain pictures is not purely due to bottom-up

visual information (as none was present) but is likely to involve some sort of top-down expert memory. The lack of difference at imagery suggests that describing the previously seen picture played an important role in increasing scanpath similarity. It could be that, due to the complex nature of the pictures and the similarity between them, verbalisation helped participants to remember the contents of the picture and to guide their eye movements around the screen. Imagery alone may have allowed participants to forget some of the picture features or even encouraged them to rely more on 'peripheral' vision.

In conclusion, this experiment has established that recognition accuracy of complex visual pictures can be increased by post-stimuli verbalisation. This effect is due to the featural (rather than configural) processing that is required at encoding and the similarity of the old and new 'matched' stimuli. Specialists also have a recognition memory advantage for stimuli that contain domain-relevant information. Analyses of verbalisation found that participants refer to salient regions of interest, but this is reduced for specialists when semantically informative domain-specific regions of interest are also present. This cognitive override of saliency is also evident in the number of fixations, total fixation durations and first fixations to salient and specialists regions of interest. No reliable difference in the number of fixations to specialist RoIs was found between stimuli at recognition where the specialist RoI had been taken away (new 'taken away') and those where a computer part never existed (old 'same absent'). This suggests that memory does not play a significant role in guiding eye movements to specialist RoIs, and rather attention is drawn automatically due to the underlying semantic relevance. Finally, describing a complex picture from memory produces more similar eye movements to encoding than are produced when simply imagining it.

3.5. Experiment 7: Expertise and congruency in comparative visual search

3.5.1. Introduction

As seen in sections 3.3 and 3.4, bottom-up saliency can be overridden by domain expertise. This occurs because the top-down knowledge that the experts possess makes certain areas of the picture more semantically meaningful to them, even though these areas are not necessarily visually salient. This cognitive override has also been shown to occur in non-experts when parts of the scene are more semantically informative. For example, Einhauser, Spain & Perona (2008) found that saliency map features do not need to drive attention (Carmi & Itti, 2006; Einhauser & Konig, 2003; Tatler, 2007) despite saliency's high correlation with fixation locations during free viewing (Peters et al., 2005). They argued that saliency drives attention indirectly through predicting interesting objects and so saliency map features causally drive object recognition, with allocation of attention being preceded by some pre-attentive understanding of the scene. However, Joubert, Rousselet, Fize, and Fabre-Thorpe (2007) found that participants were impaired at categorising the context of a scene when salient objects were present, especially if the objects were incongruent with the context. They suggest that scene processing might thus involve parallel extraction of the global gist of the scene and the concurrent object processing leading to categorization. High-level scene interpretation is quickly available to the visual system (Li, VanRullen, Koch & Perona, 2002; Rousselet, Fabre-Thorpe & Thorpe, 2002; Thorpe, Fize & Marlot, 1996), with less effort or more quickly than low-level concepts (Hochstein & Ahissar, 2002; Li et al., 2002). This suggests that eye movements or spatial attention are by-products of object based attention or object recognition.

An early study that looks at the effect of a scene's semantic content upon scanpaths comes

from Mackworth and Morandi (1967) who recorded eye movements while participants judged which of two pictures they preferred. Fixations were more frequent on regions of the pictures that were regarded subjectively as being most informative, with non informative regions often not fixated at all. This experiment was later repeated with a larger sample of pictures (Antes, 1974), and as well as confirming previous results, also found that the first fixation tended to be on an 'informative' region.

Loftus and Mackworth (1978) defined semantic informativeness as the degree to which an object was predictable within the scene. An unpredictable object would therefore be considered more informative than a predictable one. Their results confirmed the effects of top-down processes upon the early inspection of scenes. Fixation density for the semantically informative regions was found to be greater than for the uninformative regions. A tendency for participants to fixate upon incongruous objects earlier than congruous counterparts after the first fixation on the scene was also found. Longer saccades of greater than 7º of visual angle were made to unpredictable objects. This would suggest the site for the next fixation could be based upon a semantic analysis of scene areas that are relatively far away in the visual periphery. Results therefore provide evidence for the role of peripheral vision in the early comprehension of the scenes gist and in the detection of objects or regions that violate the gist. These objects or regions may attract attention because the current schema needs to be revised, or because the identity of the incongruent object needs to be confirmed.

These results have not gone unchallenged however. Friedman (Friedman, 1979; Friedman and Liebelt, 1981) found no relationship between fixation density and semantic informativeness. One explanation for this is that objects in Freidman's study ranged in consistency, from highly consistent to somewhat inconsistent (e.g., hockey sticks in a kitchen scene). If viewers distributed their fixations in a graded fashion across the multiple, somewhat inconsistent objects that were present, the overall effect of inconsistency on fixation density would be smaller and therefore more difficult to detect.

Further evidence against the effects of inconsistency of eye movements comes from De Graef, Christiaens and d'Ydewalle (1990) and Henderson and Hollingworth (1999) found no evidence for informative regions being fixated earlier than uninformative regions. The latter conducted a systematic study whereby participants viewed line drawings containing congruous or incongruous objects in preparation for a later memory test. Eye movement recordings revealed that saccades to incongruous objects were short, being on average around 3º visual angle. This object was also fixated no earlier than a congruous counterpart positioned in the same location.

Discrepancy therefore exists regarding whether the initial saccade to an object is affected by a semantic analysis of peripheral information. Why has such discrepancy been found? Henderson et al., (1999) suggest that both the size and level of complexity of the scenes has an influence upon whether the effect is seen. They criticised Loftus and Mackworth for using stimuli that were not only semantically anomalous, but also anomalous in their low-level visual properties. They argued that this confounded their results, since these low level discrepancies were more likely to explain why long saccades were attracted early to the target objects.

The study by Henderson et al., (1999) can also be criticised. Their negative results might not be generalisable to scenes viewed in the real world, for the following reasons. Their study used line drawings that were relatively crowded. Hilz & Cavonius (1974) suggest that the recognition of line drawings is reliant upon high spatial frequency contour information and that this is subject to severe loss in the periphery. A further criticism stems from the use of crowded stimuli (Becker at al., 2007). This increase in crowding further diminishes the ability for identity

of a target (Bouma, 1970; Ehlers, 1936; Flom, Heath & Takahashi, 1963).

The findings of more recent research seem to corroborate the findings of Loftus and Mackworth. Comparative visual search tasks have been used in eye movement studies whereby participants have to decide whether two real world images side by side are the same (Galpin & Underwood, 2005; Underwood, Templeman, Lamming, and Foulsham, 2008). When the difference between the two pictures involved an object being taken from another scene, and therefore violating the gist of the scene in which it was put into, it was found that eye fixations were attracted to this incongruous object earlier than the congruous counterpart. This effect was apparent following several seconds after initial presentation of the pictures. Such comparative visual search tasks have also been found to be affected by task difficulty and divided attention (Pomplun, Reingold, and Shen, 2001).

Becker, Pashler and Lubin, (2007) and Harris, Kaplan, and Pashler (2008) improved upon previous studies by transforming an item in such as way that it violated the stored canonical form of the object, thereby making it incongruous. Their results did find that participants fixated upon the incongruous objects (e.g. a green stop sign) in the periphery sooner than control objects (e.g. the red stop sign). This offers further support for peripheral detection of areas that go against what is expected, with this increasing the chance that such areas are fixated upon. Furthermore, Rayner, Castelhano, and Yang (2009) found that the eyes were drawn to the weird parts of the scene earlier than when the weird aspect was missing.

However, Gareze and Findlay (2007) considered different methodologies and scene types including simple line drawings, complex naturalistic photographs and line drawings and found no effect of incongruence on scene processing within a single fixation or on subsequent eye movements prior to target fixation.

It is apparent that the identity and appearance of objects within a scene undoubtedly influence eye movements. In addition, individual differences exist regarding preferences in terms of where in a scene to attend to. Certain individuals show biases towards objects within a scene that are personally relevant, reflecting their own personal knowledge or experience. A recent study by Verena & Gordon (2008) used two groups of participants, smokers and non-smokers. The task was a passive viewing task and involved some objects being altered to make them incongruous with the gist. Some of these objects were smoking related whilst others were not. Results showed that all participants preferentially allocated their attention to the incongruous objects, thereby supporting the incongruence effect. Also, they found that smokers selectively attended to smoking-related objects.

This effect of personal experience can be related back to previous sections of this Thesis regarding expertise, which has been shown to affect eye movements during scene viewing and visual search. For example, experienced football players have higher search rates, involving more fixations of shorter duration and fixated for longer on the hip region of the opposing player (Williams and Davids, 1998). Conversely, expert gymnastic coaches made longer and fewer fixations than novices (Moreno, Reina, Luis, and Sabido, 2002). It could be that eye movements depend on the domain of expertise and the nature of the scene being search. Furthermore, the advantage of expertise in visual search is reliant on the stimuli being domain specific (Nodine, and Krupinski, 1998).

The present experiment firstly explores the relationship between semantically interesting regions within images of natural scenes, and the effect such regions have upon the pattern of eye movements. The semantically interesting areas being in the form of unexpected or strange objects created in a similar way to one of the methods used by Becker et al, (2007). This was achieved

by introducing an object-intrinsic anomaly via a colour change, for example, altering the colour of a person's bare arm to blue. This method is an improvement on previous studies that have used object-context manipulation, e.g. putting a cow in an office scene. In such a case, the semantic context isn't the only thing that is changed, the visual attributes of the object are also altered, as well as the objects integration with the features that surround it. Object-intrinsic oddity differs from object-context oddity regarding the level of processing required to detect the incongruous object. This is due to the fact that in object-intrinsic oddity, the oddity is likely to be detected at an intermediate level of processing. It requires only activation of the canonical representation of the object, followed by a comparison of this with what is visibly present. Detecting objects made incongruous via object-context oddity requires identification of the gist of the scene, identification of the incongruous object, activation of semantic information providing ideas of what would be consistent, and the same mechanism to then compare this information with the object.

The task employed was that of a comparative visual search, so that participants actively scanned the pictures with the aim of finding something. The second element of the study looked at the effect of domain proficiency upon the incongruence effect. The domain specialists were chemistry students and therefore half of the stimuli used were that of chemistry laboratory scenes. The aim of the study was therefore not only to contribute to the debate regarding the existence of the incongruence effect upon eye movements, but also to look at the effect domain proficiency has on this. Is there an effect of attraction to incongruous objects in a visual scene? If so, is this effect reduced, increased or unaffected when participants view pictures that are from their area of expertise?

3.5.2 Method

Participants

Participants were undergraduate students attending the University of Nottingham, who volunteered to participate in return for £4. They consisted of 15 chemistry students and 15 non-chemistry students; the non-chemistry students comprising the control group. All participants were required to have no background or keen interest in music and had normal or corrected-to-normal vision. The age of the participants ranged from 19-23. The sample contained 17 males and 13 females. Participation in the study was reliant upon reliable eye tracking calibration.

Materials and Apparatus

An SMI iVIEW X Hi-Speed eye tracker was used to record eye position. The stimuli consisted of 120 pairs of colour photographs sourced from the internet and manipulated using Photoshop 7.0 software. Sixty of these were chemistry specific, depicting scenes from a chemistry laboratory, and sixty were music specific, depicting music related scenes, such as an orchestra. In each trial, pairs of stimuli were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. They were arranged in portrait orientation side by side (790 pixels x 1200 pixels) with a white background and separated by a gap of 20 pixels between them.

Fifteen stimuli from each of the two categories were considered fillers as they contained no difference, fifteen in each category contained a contextually incongruous object and a

difference, fifteen contained an incongruous object and no difference and fifteen contained a difference only. A diagrammatic breakdown of the stimuli can be seen in figure 61. Incongruous objects were created by altering the colour of an object to one that is unexpected, a green arm for example, or a blue violin. Examples of stimuli are given in figures 62a, 62b, 62c, 62d, 62e and 62f.
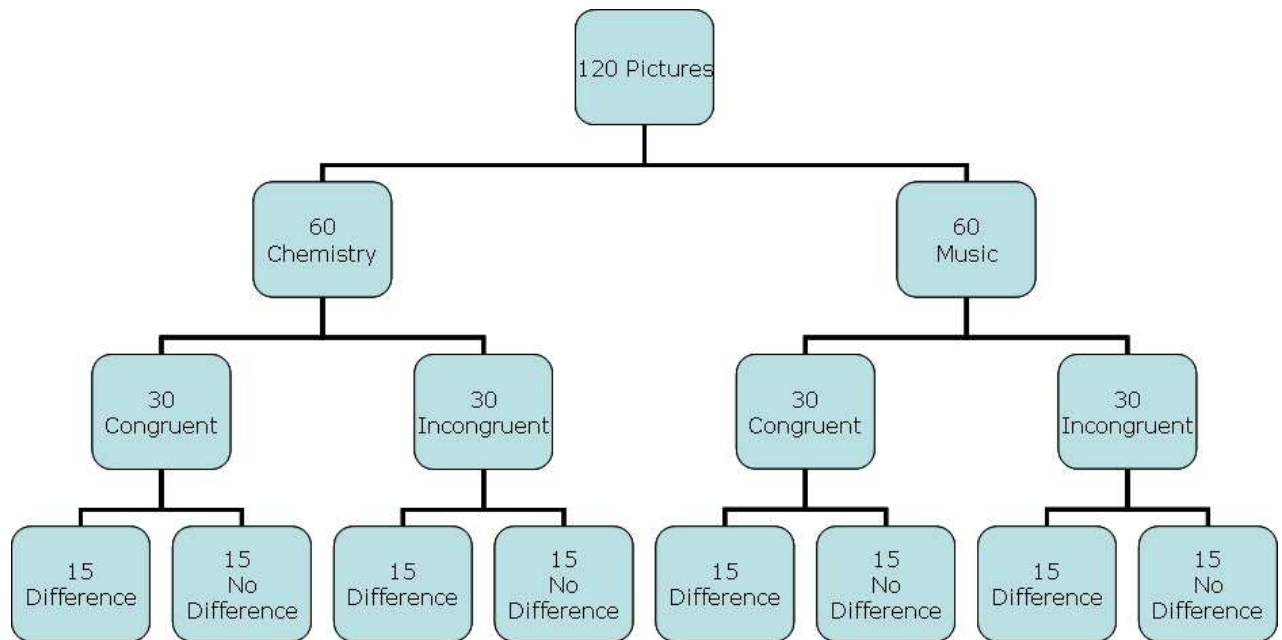

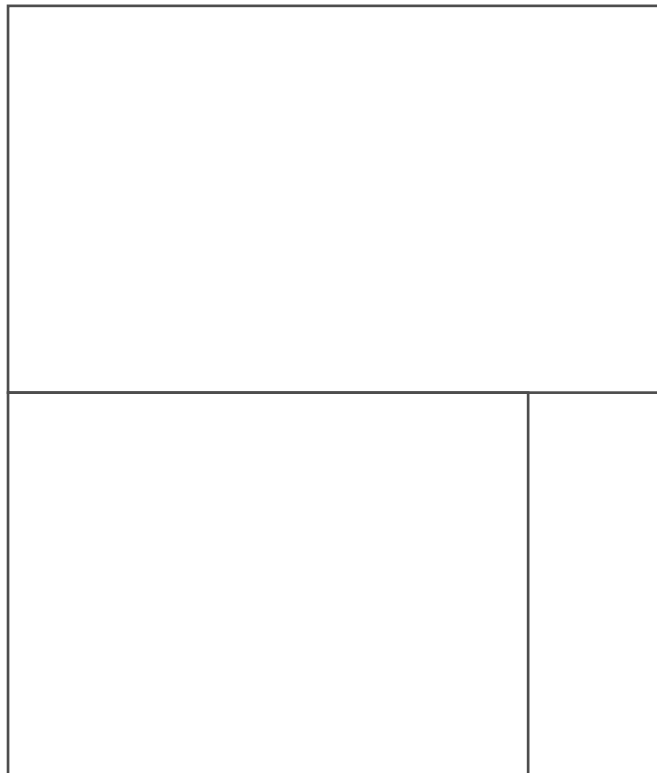*Figure 61:* *A diagrammatic breakdown of the stimuli.*

Figure 62a (left): Chemistry picture with a difference (unlabelled test tubes on the right) and an incongruous region (red eyebrows); and figure 62b (right): Music picture with a difference (hair band colour) and incongruous region (green hand).
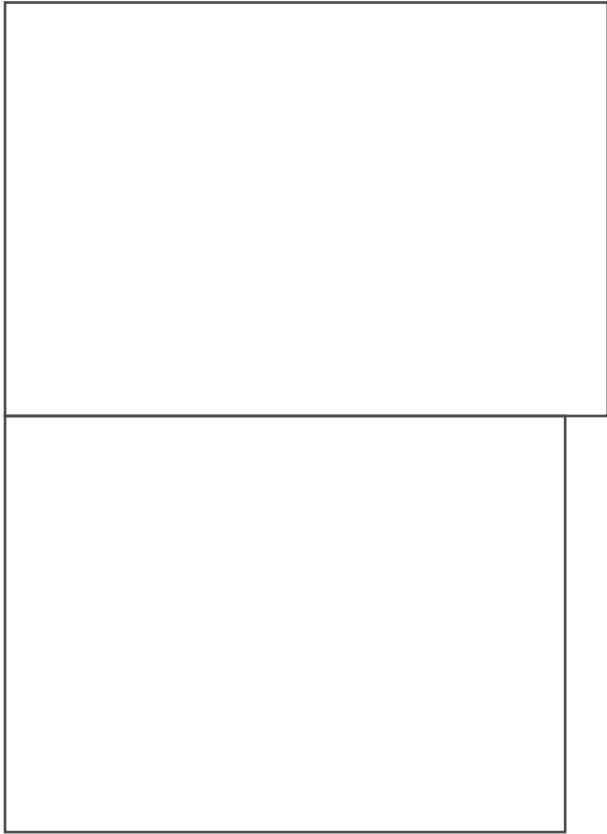
Figure 62c (left): Chemistry picture with an incongruous region (red eyes) and no difference; and figure 62d (right): Music picture with an incongruous region and no difference.

Figure 62e (left): *Chemistry picture with only a difference (test tube fluid); and* **figure 62f (right):** *Music picture with only a difference (conductor's baton missing).*

Design

A two-by-two mixed design was adopted with two groups, specialists and non-specialists. These were chemistry students and non-chemistry students respectively. The independent variables were the group to which the participant belonged, and whether the stimuli contained an incongruous object. The dependent variable measures taken included participants response time and accuracy in identifying a difference, the total number of fixations on the incongruous and difference regions, the average total time spent fixating on the incongruous and difference regions and the number of first fixations on the incongruous regions and difference regions

Procedure

Participants sat at a fixed viewing distance of 98cm with their head resting on a chinrest. Firstly, a nine point calibration was applied. As soon as this was successful, participants were presented with written instructions on the computer monitor that explained what was being asked of them. These stated that they would be presented with a fixation cross on which they were to remain focussed until a pair of pictures side by side appeared. The fixation cross remained on the screen for 1000ms. They were informed that their task was to decide whether the pictures were the same or different. If they thought the pictures were the same, they were to respond by pressing 1 on their keyboard, and if they thought the pictures were different, they to press 2. They were asked to make their responses as quickly and as accurately as possible. A practice phase consisting of the presentation of eight pairs of stimuli was initially performed. Here participants were given feedback on whether their responses were correct or incorrect. The aim of the practice was to familiarise participants with the procedure and ensure they were clear of what was expected of them. Following the practice phase, the experimental phase began. This was split into two parts. The first part consisted of a presentation of 60 pairs of stimuli in a randomised order. There was no time limit for the presentation of the pictures, enabling participants to inspect the pictures for as long as they wished before they gave a response. No feedback was given in the experimental phase. The participants were told the first part of the experiment was over. This break allowed for re-calibration if necessary. The second part of the experiment then began which also consisted

of sixty pairs of stimuli presented in a randomised order. Participants were then thanked for their participation.

3.5.3 Results

Before measures were extracted from the raw data, the percentage of calibration loss for each trial was determined. This resulted in one trial being excluded due to having over 20% loss. The measures taken from the acquired data were response time (RT) and accuracy in deciding whether the presented images were the same or different, as well as eye tracking measures. These included the total number of fixations on the incongruous object and the difference regions, the average total time fixating on the incongruous object and the difference regions, and the number of first fixations on both the incongruous and difference regions. The varying sizes of the regions of interest (ROI) were accounted for by calculating the average size of the RoI and adjusting the analyses appropriately, for example, if the RoI on one picture was 1.5 times larger as the average RoI, then the data for that picture (e.g. number of fixations would be divided by 1.5.

Accuracy

Accuracy was measured by the number of responses that correctly identified the presented stimuli as being either the same or different (See figure 63).
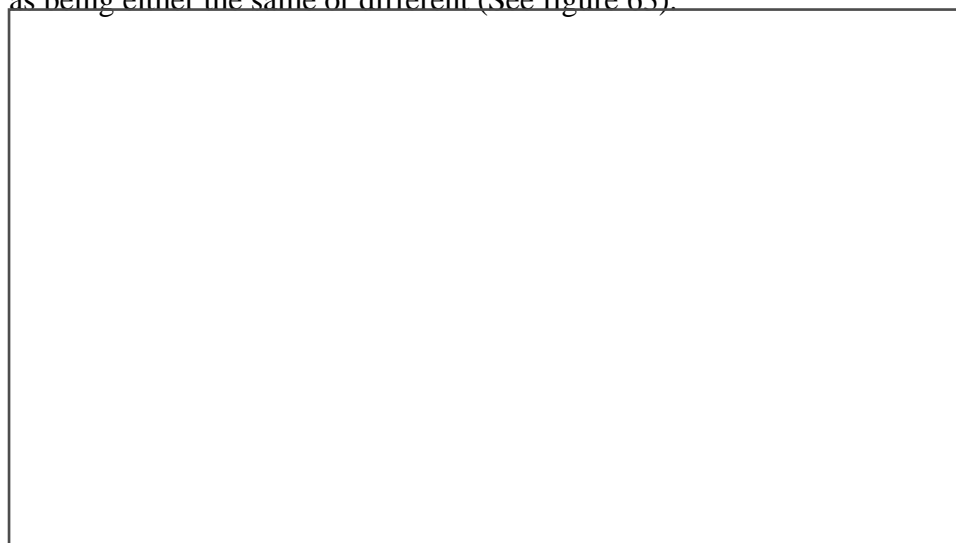
Figure 63: A bar chart to illustrate the average accuracy scores for identifying whether a picture contained a difference or not.

The picture type (chemistry picture or music picture), the participant group, and the congruency, were all taken into consideration. A 2x2x2 mixed design ANOVA (appendices 35a, 35b and 35c) revealed a main effect of picture type: $F(1,26) = 26.331$, $MSe = 35.157$, $p<0.001$. This result taken with the descriptive statistics suggests that there was a higher overall accuracy for chemistry pictures than music (control) pictures. There was no main effect of congruency: $F(1,26) = 1.518$, $MSe = 40.636$, $p = 0.229$. A main effect of group was found, $F(1,26) = 6.967$, $MSe = 179.924$, $p<0.05$, which taken with the descriptive statistics suggest that chemists were

overall more accurate than control participants.

Post-hoc independent samples T-tests (appendices 35d and 35e) revealed that Chemists were reliably more accurate on chemistry pictures than control participants, $t(26) = 3.220$, $p<0.01$. There was no reliable difference between the groups for music pictures: $t(26) = 1.788$, $p = 0.085$. Post-hoc paired samples T-tests (appendices 35f and 35g) revealed that chemists were more accurate when looking at chemistry pictures than when looking at control pictures: $t(12) = 5.381$, $SEM = 1.302$, $p<0.001$. However, control participants were also found to be more accurate when looking at chemistry pictures than when looking at control pictures, $t(14) = 2.572$, $SEM = 6.81207$, $p<0.05$.

Reaction Time (RT)

The average time taken to identify whether the picture contained a difference or not was calculated. A 2x2 between-groups ANOVA was conducted on the reaction time data (appendices 36a, 36b and 36c), taking the picture type and group into consideration. This revealed no main effect of picture type, [$F(1,27) = 0.076$, MSe $= 3164342.267$, $p = 0.785$], or group, [$F(1,27) = 0.269$, $MSe = 7282630.449$, $p = 0.608$], but a reliable interaction between picture type and group, $F(1) = 13.868$, $MS = 43882562.10$, $p < 0.05$, (see figure 64).
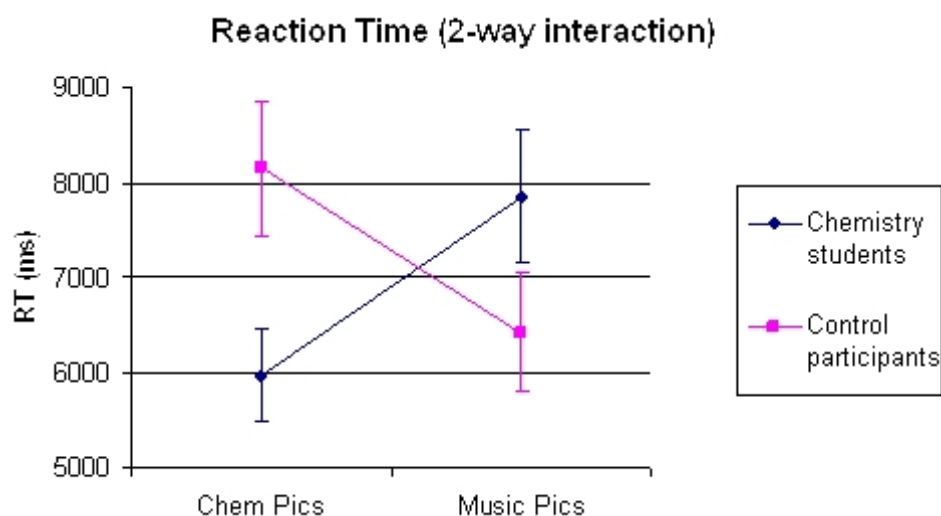
Figure 64: *A line graph to illustrate the 2-way interaction in reaction time between picture type and group.*

Post-hoc independent samples T-tests (appendices 36d and 36e) revealed a reliable difference in RT to chemistry pictures between the groups, with Chemists being reliably faster (smaller RT) at chemistry pictures than control participants: $t(27) = 2.579$, $p<0.05$. There was no reliable difference in RT between the groups for control pictures, [$t(27) = 1.560$, $p=0.130$]. Post-hoc paired samples T-tests (appendices 36f and 36g) revealed reliable differences between picture types within both groups. For chemists, $t(12) = 2.521$, $SEM = 745.121$, $p<0.05$. For control participants, $t(15) = 2.734$, $SEM = 592.531$, $p<0.05$. These results show that Chemists were

reliably faster at chemistry pictures than control pictures and that control participants were reliably faster at control pictures than chemistry pictures.

A 2x2x2x2 split-plot ANOVA (appendices 36h, 36i and 36j) was conducted to take into account group, picture type, congruency and difference. This revealed a main effect of congruency: $F(1,27) = 4.593$, $MSe = 9802583.511$, $p<0.05$; a main effect of difference: $F(1,27) = 4.290$, $MSe = 24414488.01$, $p<0.05$; a 3-way interaction between picture type, congruency and difference: $F(1,27) = 5.723$, $MSe = 12803222.11$, $p<0.05$ (see figure 65); and a 4-way interaction between the picture type, congruency, difference and group: $F(1) = 5.843$, $MS = 74804209.13$, $p<0.05$. There was no main effect of group: $F(1,27) = 406.335$, $MSe = 28266816.310$, $p = 0.636$, and no main effect of picture type: $F(1,27) = 0.004$, $MSe = 12251994.20$, $p = 0.953$.
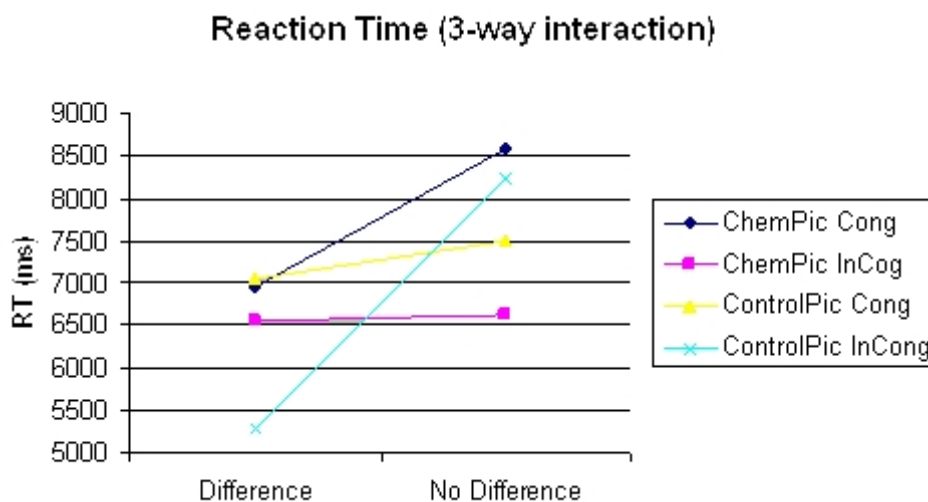
Figure 65: *A line graph to illustrate the 3-way interaction between picture type, congruency and difference.*

Post-hoc independent samples T-tests (36k and 36L) revealed a reliable difference between the groups for chemistry pictures when congruent and different, $t(27) = 2.898$, $p<0.01$. Post-hoc paired samples T-tests (36m and 36n) help explain the 3-way interaction, which when considered alongside the graph in figure 3, shows increased RT for control pictures with an incongruous region, when there was a difference compared to when there was no difference: $t(28) = 3.818$, $SEM = 767.81527$, $p<0.01$. Furthermore, for control pictures that contain a difference, RT was increased when there was no incongruous region compared to when an incongruous region was present, $t(28) = 2.863$, $SEM = 608.82374$, $p<0.01$.

Post-hoc paired samples T-tests (36o and 36p) were conducted to help explain the 4-way interaction. These revealed that for chemists: for chemistry pictures with no difference, RT was reliably decreased when and incongruous region was present: $t(12) = 3.426$, $SEM = 1275.145$, $p<0.01$; and for chemistry pictures that are congruent, RT was reliably decreased when there was a difference, $t(12) = 3.723$, $SEM = 1480.180$, $p<0.01$. For Chemists, for incongruous pictures with no difference, RT was reliably decreased when the picture was a chemistry related picture rather than a control picture: $t(12) = 2.345$, $SEM = 1640.79$, $p<0.05$; and for Chemists for congruent pictures with a difference, RT was reliably decreased also when the picture was a chemistry related picture rather than a control picture, $t(12) = 3.475$, $SEM = 1281.36$, $p>0.01$. For Chemists, for control pictures that were incongruent, RT was reliably decreased if there was a difference,

$t(12) = 2.923$, *SEM* = 1185.87, $p<0.05$; and finally, for Chemists for control pictures that had a difference, RT was reliably decreased if it was incongruous, $t(12) = 2.666$, *SEM* = 1098.92, $p<0.05$. For control participants, these post-hoc paired samples T-tests revealed firstly that for incongruous pictures with a difference, RT was reliably increased if the pictures were chemistry pictures: $t(15) = 2.316$, *SEM* = 1126.209, $p<0.05$; and secondly, that control pictures that were incongruent, RT was reliably decreased if there was a difference present, $t(15) = 2.434$, *SEM* = 1025.878, $p<0.05$.

Eye-tracking measures

Total Time in 'Incongruence' RoIs
A 2x2x2 mixed design ANOVA (appendices 37a, 37b and 37c) was carried out to analyse the effects of group, picture type and difference on the average total time fixated in incongruence RoIs (incongruent regions). The results of the ANOVA revealed no main effect of group: $F(1,22) = 0.005$, *MSe* = 5902448.173, $p = 0.945$; and no main effect of difference: $F(1,22) = 0.889$, *MSe* = 906695.894, $p = 3.56$. However, there was a main effect of picture type, $F(1,22) = 23.729$, *MSe* = 638321.640, $p<0.001$, which alongside the descriptive statistics shows that overall, more time was spent looking at incongruous regions in control pictures than in chemistry pictures. The interactions between picture type and difference and between picture type, difference and group were not reliable ($p = 0.269$ and $p = 0.892$, respectively). A reliable interaction was found between picture type and group, $F(1,22) = 21.669$, $p<0.001$. Post-hoc paired samples T-tests (appendices 37d and 37e) found that Chemists spent less time fixated on incongruous regions when viewing chemistry pictures than when viewing control pictures: $t(15) = 4.862$, *SEM* = 320.67069, $p<0.01$ (see figure 66).

Figure 66: *Average time fixated in incongruence RoIs. This line graphs illustrates the two-way interaction between participant group and picture type.*

A reliable interaction was also found between group and difference, $F(1,22) = 7.603$, $p<0.05$. Paired samples T-test (appendices 37f and 37g) revealed that control participants spent reliably less time fixated in incongruence RoIs when there was a difference between the pictures, compared to when no difference existed: $t(10) = 2.858$, $SEM = 237.39514$, $p<0.05$. See figure 67.
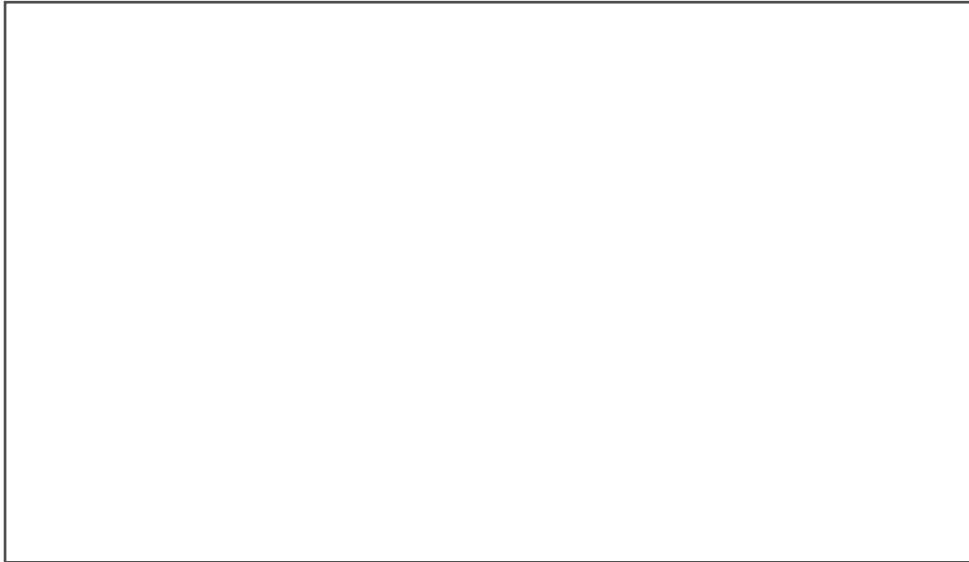

Figure 67: *A line graph to illustrate the reliable two-way interaction between 'difference' and group. There was a reliable difference in the average total time fixated in incongruence RoIs between pictures that contained a difference and those that did not, but this was only statistically reliable for control participants.*

Total Time in Difference RoIs

A 2x2x2 mixed design ANOVA (appendices was carried out to analyse the effects of group, picture type and congruency on the average total time fixated in 'difference' RoIs (where the difference between the two pictures was located). Results of the ANOVA revealed a main effect of congruency: $F(1,22) = 7.091$, $MSe = 361976.173$, $p<0.05$, which combined with the descriptive statistics shows that when the pictures were congruent, participants spent more time fixated within the regions where there was a difference, (see figure 68). There was no main effect of picture type: [$F(1,22) = 2.292$, $MSe = 1199871.643$, $p = 0.144$]; and no main effect of group: [$F(1,22) = 2.08$, $MSe = 8692090.414$, $p = 0.653$]. No reliable interactions were uncovered.

Figure 68: *A bar chart to show the average total time fixated in areas that contained the difference. Overall, participants spent longer fixating 'difference' RoIs when the picture was congruent.*

Number of Fixations in Incongruence RoIs

A 2x2x2 mixed design ANOVA (appendices 39a, 39b and 39c) was carried out to analyse the effects of group, picture type and difference on the average number of fixations in incongruence RoIs (incongruent regions). The results of the ANOVA revealed a main effect of difference, $F(1,22) = 12.758$, $MSe = 48.218$, $p<0.01$ (see figure 69), which alongside the descriptive statistics (appendix 39d) suggests reliably more fixations were made to incongruous regions when the picture contained no difference. There was no main effect of picture type: [$F(1,22) = 0.005$, $MSe = 66.25$, $p = 0.944$]; or group: [$F(1,22) = 0.248$, $MSe = 537.736$, $p = 0.624$]. No reliable interactions were revealed.

Figure 69: *The average number of fixations in incongruence RoIs was reliably increased when there was no difference between the pictures.*

Number of Fixations in Difference RoIs

A 2x2x2 mixed design ANOVA (appendices 40a, 40b and 40c) was carried out to analyse the effects of group, picture type and congruency on the average number of fixations in 'difference' RoIs (where the difference between the two pictures was located). The results of the ANOVA revealed a main effect of congruency, $F(1,22) = 5.045$, $MSe = 35.772$, $p<0.05$, which in conjunction with the descriptive statistics (appendix 40d) shows that when the pictures contained an incongruous region, less fixations were made to the difference regions (see figure 70). There was no main effect of picture type: [$F(1,22) = 3.176$, $MSe = 55.244$, $p = 0.089$]; and no main effect of group, [$F(1,22) = 0.380$, $p=0.491$].
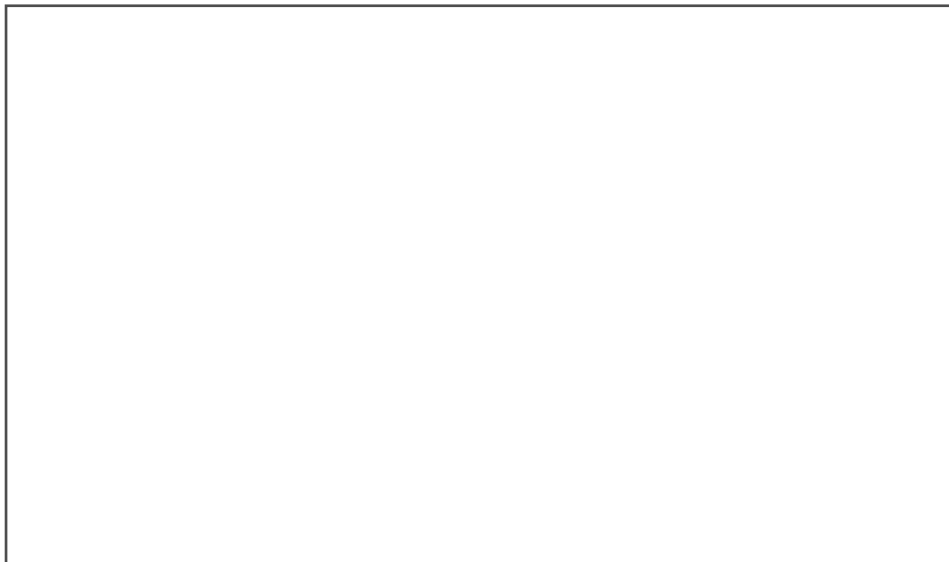
Figure 70: *A bar chart to illustrate the number of fixations to 'difference' RoIs in pictures that did or did not contain an incongruous region.*

A reliable interaction was found between picture type and group, $F(1) = 6.699$, $p<0.05$. No other interactions were statistically reliable. Post-hoc paired samples T-tests (appendices 40e and 40f) revealed that control participants made reliably more fixations to difference RoIs in chemistry pictures than in music pictures: $t(10) = 2.587$, $SEM = 2.618$, $p<0.05$.

Percentage of First Fixations to Incongruence RoIs

The average number of first fixations that fell in incongruence RoIs was calculated. An independent samples T-test (appendices 41a and 41b) revealed a reliable difference between the groups regarding the number of first fixations in the incongruous regions of the chemistry pictures, $t(22) = 3.063$, p<0.01. When viewing chemistry pictures, chemists were not as likely as

control participants to initially fixate the incongruous regions. No reliable difference was found between the groups regarding the number of first fixations upon incongruous regions in control pictures [$p = 0.454$].

Paired samples T-tests (appendices 41c and 41d) revealed that chemists were reliably less likely to initially fixate the incongruous regions when viewing chemistry pictures than when viewing control pictures, $t(10) = 2.656$, $SEM = 0.38356$, $p<0.05$. There was no reliable difference between chemistry and control pictures for control participants, [$t(12) = 0.806$, $SEM = 2.85674$, $p = 0.436$]. See figure 71.
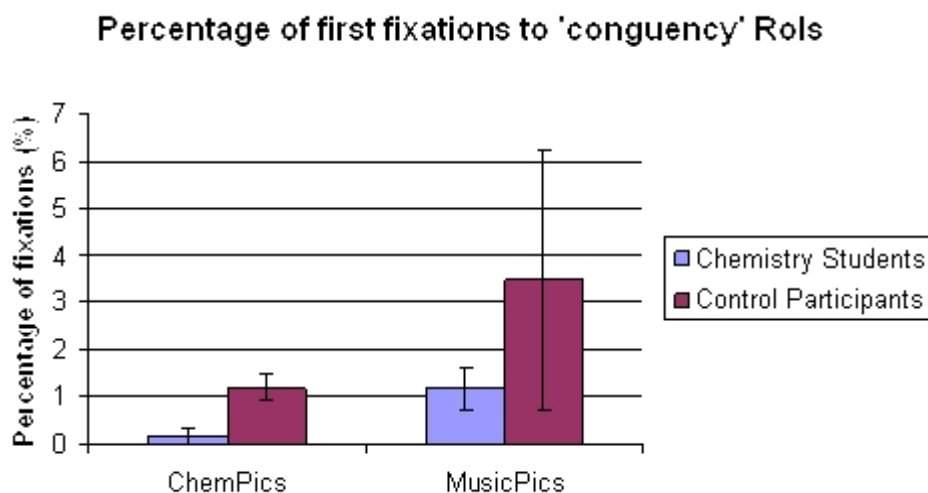
Figure 71: Percentage of all first fixations that fell into incongruence RoIs. When viewing chemistry pictures, chemists were not as likely as control participants to initially fixate the incongruous regions. Chemists were reliably less likely to initially fixate the incongruous regions when viewing chemistry pictures than when viewing control pictures

Percentage of First Fixations to Difference RoIs
The average number of first fixations on each picture that fell in 'difference' RoIs was calculated. An independent samples T-test (appendices 42a and 42b) revealed no reliable difference between the groups [for chemistry pictures, $t(22) = 1.151$, $p = 0.262$; for control pictures, $t(22) = 0.17$, $p = 0.987$]. Paired samples T-tests (appendices 42c and 42d) revealed no reliable difference between the stimuli for each group [for chemists: $t(10) = 0.602$, $SEM = 1.23494$, $p = 0.561$; for controls, $t(12) = 1.003$, $SEM = 0.48410$, $p = 0.335$].

Time to first fixate Incongruence and Difference RoIs
As well as the likelihood to first fixate, the time it took for participants to first fixate incongruous and different regions of the pictures. A 2x2x2 mixed-design ANOVA was carried out (appendices 42f and 42g) and found no reliable main effects of picture type or RoI [$F(1,22) = 0.050$, $MSE = 1906586$, $p = 0.825$; and $F(1,22) = 2.400$, $MSE = 1116874$, $p = 0.136$, respectively]. There was a main effect of group: $F(1) = 8.047$, $MSE = 6406393$, $p<0.05$, which taken with the descriptive

statistics (appendix 42e) shows that overall, chemists took longer to fixate the regions of interest. There were no reliable interactions.

3.5.4. Discussion

The first aim of the current study was to identify whether an effect of incongruence upon eye movements exists. Do regions of a scene that are incongruous attract early fixations, as well as a greater density and duration of fixations? Secondly, the study aimed to look at the effect of domain knowledge. In particular, what effect this has upon the incongruence effect. Does cognitive knowledge in the form of domain proficiency override the incongruence effect upon eye movements?

Previous evidence in the literature for an incongruence effect is inconsistent. For example Loftus and Mackworth (1978) found that not only was there a greater fixation density for semantically informative regions, but also there was a tendency for viewers to fixate upon these incongruous regions sooner than congruous counterparts. This latter finding has also recently been corroborated by Underwood and Foulsham (2006) and Becker et al., (2007). However, contradictory evidence comes from De Graef et al., (1990), Henderson et al., (1999) and Gareze and Findlay (2007) who found that semantically informative regions were fixated no sooner than uninformative regions.

Finding an effect of incongruence provides evidence for the role of peripheral vision in the early comprehension of objects or regions that violate what would normally be expected in a scene. In other words, an effect would provide evidence that foveal inspection is in fact not necessary to identify a region or object that is strange or unpredictable, and these incongruous regions identified in the periphery are processed to the point that the representation is then available to the mechanism that guides eye movements and therefore attract eye fixations.

As the present study was particularly interested in the effect of cognitive knowledge upon the incongruence effect, it is important to firstly determine whether in fact specialists, these being the chemistry students in the present study, can in fact be considered experts within their field. The analysis of accuracy and reaction time (RT) data are able to confirm this. Chemists were found to be reliably faster and more accurate when viewing chemistry pictures compared to control (music) pictures and compared to control participants. Such results are consistent with previous findings, such as that by Walker (1987), McCarley at al., (2004) and Lansdale et al., (2009) who found that accuracy increased with level of domain proficiency.

Chemistry participants were also found to be more accurate overall, which could indicate that they are generally better at comparative visual search than control participants. This could be due to the nature of their domain environment (e.g. crowded/complex chemistry labs) requiring them to adapt a more efficient visual scanning method. The control participants, on the other hand, were not a homogenous group and therefore had mixed levels of visual search abilities (this can be seen in the increased variation of accuracy scores). If this experiment was repeated, it would be a good idea to include two homogenous groups of participants. If two expert groups and two domain specialist stimuli types were used, it is predicted that a cross-over would be found, similar to that found in section 3.1.3.

An overall accuracy advantage for chemistry pictures was also found. This was mostly due

to Chemists being reliably more accurate on chemistry pictures, which increased the average score overall. Chemistry pictures could also have been easier to recognise overall than music pictures, possibly due to participants being more familiar with chemistry scenes than music scenes. For example, most schools enforce science as a compulsory subject (at least until after GCSE level), but a lot of the time music is optional. Furthermore, many popular television programs are based around science labs (e.g. CSI and similar forensic dramas) but few are based around classical orchestras or marching bands. This increased exposure could have facilitated recognition of the chemistry stimuli.

Reaction time analyses showed that for Chemists, for chemistry pictures with no difference, RT was reliably decreased when an incongruous region was present compared to when no incongruous region was present. This result is unexpected as one would predict an increased RT due to the distraction of an incongruous region. It could be that when pictures are domain-specific, specialists first use top-down knowledge to find the difference. However, if no difference is found, they then attend to incongruent regions, to confirm that they are the same on both left and right pictures, in order to make a 'same' response. Alternatively, there may have been a problem with the participants' understanding the test, whereby they incorrectly believed that the the incongruent RoI and the difference RoI were the same. Because incongruent regions attract attention sooner in scene perception, participants may have been quick to realise that an incongruous region was present on both sides of the picture, therefore incorrectly concluding that there was no difference.

For chemists viewing chemistry pictures that are congruent, RT was reliably decreased when there was a difference. Similar results were found for chemists viewing congruent control pictures and controls viewing incongruent control pictures. This facilitating effect of a difference present is due to search ceasing once the difference is found (resulting in lower RT) whereas if there is no difference, search continues for longer before the participant can confirm that the picture is the same.

For Chemists viewing incongruous pictures with no difference, RT was reliably decreased when the picture was a chemistry-related picture rather than a control picture, and for Chemists for congruent pictures with a difference, RT was reliably decreased also when the picture was a chemistry-related picture rather than a control picture. These two findings demonstrate that domain expertise facilitates decision in otherwise difficult situations.

For Chemists viewing control pictures that had a difference, RT was reliably decreased if it contained an incongruous region compared to when it did not. This is another unexpected result, as an incongruent region would be predicted to act as a distraction, therefore increasing RT. A possible explanation could be that fixating the incongruent regions focused participants' attention, making them more aware of changes and thus facilitating identification on the difference. This is of course just a speculative elucidation and further research would be needed to justify this suggestion.

Lastly, for controls viewing incongruous pictures with a difference, RT was reliably increased if the pictures were chemistry-related rather than music-related. Control participants' attention is attracted by regions of incongruence. However, chemistry pictures may be more complex than music pictures in that they contain more items to search, which, due to lack of semantic relevance for control participants (e.g. recognising that something domain-specific has been changed), therefore increases RT.

Both the density of fixations and the duration of fixations can be considered as eye movement measures that indicate an effect of object incongruence. This is because incongruous

regions attracting both a reliably large number of fixations for a reliably large duration, suggests that the regions were considered by observers to be semantically interesting due to attracting a greater deal of attention. Analysis of the total time spent within incongruous regions revealed that chemists spent reliably less time fixated upon incongruous regions of chemistry pictures than incongruous regions of the control pictures. It can be concluded from this that domain proficiency caused an overriding effect of incongruence, which is confirmed by control participants showing no reliable difference between the total time spent in the incongruous regions of control and chemistry pictures. This is as expected; the non-specialists were equally distracted by incongruous regions in the two types of pictures as neither was from their area of expertise. Analyses also showed that more time was spent fixated in difference RoIs (regions where the differences existed) when the picture was congruent. This makes sense, as there were no regions of incongruence to distract attention away from the difference.

The number of fixations in regions of incongruence was also calculated. More fixations fell in this region when the picture did not contain a difference. This can be explained due to the longer inspection required before a decision is made, resulting in more fixations being made by the viewer and therefore more being likely to fall within the incongruous regions. Taking these results in conjunction with the data on overall average time spent in the incongruous regions, it can be concluded that although participant groups don't differ reliably in terms of the number of fixations in incongruous regions, the amount of time fixated on these regions does differ. Control participants spent longer fixated on incongruous regions in chemistry pictures compared to chemists. This suggests that firstly, their fixations were longer on the incongruous regions (than chemists), which has previously been associated with more extensive processing (Rayner, 1998). The increase in fixation durations could be indicative of object recognition processes for unexpected objects requiring more time, or it could be because memory consolidation for unexpected objects takes longer (Henderson, 1992).

Secondly, they had trouble moving their attention away from the incongruous regions and thirdly, there was a cognitive override effect of expertise that allowed chemists to direct their attention away from the incongruous regions. Findlay and Walker's (1999) model of information processing can be applied to the present findings. Perhaps, the Chemists domain knowledge (or "intrinsic saliency") is able to override the incongruence effect by enabling them to move their eyes away from these regions, resulting in both shorter fixations on the incongruous regions as well as for a shorter duration.

The number of fixations in difference RoIs was calculated and results showed that when an incongruent region existed, the number of fixations to difference RoIs was reliably reduced. This is due to the distracting effect of the incongruence. Control participants were found to make more fixations to difference RoIs in chemistry pictures than control pictures. Due to chemistry pictures being semantically uninformative to control participants, they may have needed to refixate the difference region more frequently in order to confirm whether the pictures were different.

Analyses of the location of first fixation revealed when viewing chemistry pictures, Chemists were reliably less likely than control participants to first fixate on incongruous regions. They were also reliably less likely to first fixate on incongruent regions on chemistry pictures compared to control pictures. These findings and the fixation data above all support the theory of a cognitive override of saliency by domain expertise. These results could also be interpreted in terms of Haider and Frensch's (1999) information-reduction hypothesis, which states that with experience, people learn to ignore task-redundant information and limit their processing to task-

relevant information. Further evidence in the literature for the information-reduction hypothesis comes from Charness, Reingold, Pomplun, and Stampe, (2001) who found that chess experts fixated proportionally more on relevant pieces than non-expert players.

Interestingly, although chemists were less likely to make their *first* fixation to an incongruous region in chemistry pictures, when they did fixate the incongruous regions, there was a large variance as to when in the scanpath sequence they did so. Consequently, this resulted in a large variance in times to fixate (see appendix 42e), and when all the trials were added together in a 'time to first fixate' analysis, the cognitive override effect seen in the first fixation analysis is lost. One interpretation of this variance could be that incongruous regions attract attention *early* in viewing (even before the first fixation), and this is the effect than can be overcome by top-down cognitive knowledge, rather than a *complete* override of incongruence. This is supported by the lack of reliable difference between the numbers of fixations to incongruous RoIs between the participant groups. Relating this theory to Findlay and Walker's (1999) model of saccade generation, there could be competing activation in level 5 between top-down incongruence and top-down domain knowledge. Parts of the picture that were particularly relevant or interesting to the specialist may have caused a greater activation of the 'cognitive knowledge' and were thus fixated first, consequently decreasing first fixations to incongruent RoIs. However, there may have been individual differences as to how many domain-specific parts of the picture were interesting to each specialist, thus varying the amount of time before the next most informative thing in the picture was the region of incongruence.

Analysis of the first fixations that fell within the difference regions revealed, as expected, no reliable difference between the two groups. There is no hypothesised reason why one group would fixate on the difference sooner than the other.

The present findings corroborate to a certain extent with the findings from previous research such as that by Loftus and Mackworth, (1978) and Underwood and Foulsham, (2006). However, it is important to note that due to differences regarding how the stimuli were manipulated in order to make them incongruous, it must be with caution that results are said to agree with such previous studies, and disagree with others. This difference in manipulation may offer one explanation for the variability of results that has been uncovered from previous studies. Most previous research has used object-context oddity. This involves an object being placed in a scene where it would not normally be expected. This procedure is limited due to the fact that the visual saliency of an object is defined in part by what it is surrounded by. For example, putting a round object into a scene of a grid-like nature in which straight lines are common will result in the object standing out more than if it were placed into a scene in which rounded features are common. This results in the incongruous object being more visually salient than a congruous counterpart and could cause masking of the incongruence effect. Manipulation of stimuli in the present study instead involved object-intrinsic oddity whereby the object was altered in such a way that it violated the canonical form stored of that object. These two types of manipulation differ in terms of the level of processing required for detection. An incongruous object created by object-intrinsic oddity is likely to be detected at an intermediate level of processing. It requires only activation of the canonical representation of the object, followed by a comparison of this with what is visibly present. In other words, the ability to detect the anomaly is likely to rely on identification of the object (Ballaz, Boutsen, Peyrin, Humphreys, Marendaz, 2005), enabling the stored canonical representation of the object to be accessed (Corballis, 1988). Detecting objects made incongruous via object-context oddity requires identification of the gist of the scene, identification of the incongruous object, activation of semantic information providing ideas of

what would be consistent, and the same mechanism to then compare this information with the object. Previous research indeed suggests that detecting object-intrinsic oddities and object-context oddities operate independently (Riddoch and Humphreys, 1987). From neuropsychological research, they found that a patient, J.B, was unable to detect object-context oddities however was able to successfully detect object-intrinsic oddities. To date Becker et al., (2007) are the only researchers that have used object-intrinsic anomalies whilst looking at the incongruence effect. Their results too suggest that such oddities are indeed processed in the periphery, although these objects were not immediately fixated. Further research could confirm the argument for object-intrinsic oddities being less visually salient than context anomalies but running the stimuli through the Itti and Koch saliency algorithm.

Previous studies have found domain knowledge to have an overriding effect upon visual saliency in guiding eye movements. Humphrey and Underwood (2009) found that visual saliency was a strong predictor of fixation location, however, when specialists viewed pictures taken from their specialist domain, the effect was reduced. Domain knowledge was therefore found to moderate the influence of visual saliency in scene recognition. Lansdale et al., (2009) also found that general domain expertise can override the influence of saliency on the guidance of fixations. These studies are supported by the present findings, which highlight the powerful role cognitive knowledge has upon guiding our attention and therefore eye movements when looking at natural scenes. The regions of incongruence in the current experiment attracted attention due to their unexpected nature, and the domain proficiency of the specialist participants allowed them to disengage from the incongruous regions and fixate on more semantically relevant areas of domain-specific stimuli. Future research could investigate the influence of incongruous regions that are also semantically relevant for the specialist group. One would predict that these incongruous yet semantically informative regions would be fixated sooner by the chemistry students than control students, and should attract attention away from the 'difference'. Alternatively, if the difference was always domain specific, reaction time would decrease for chemists viewing chemistry pictures with a difference. In the current experiment, the difference RoIs were sometimes domain specific and sometimes not, therefore this effect wasd masked.

In conclusion, the present study provides additional evidence to the debate concerning the existence of the incongruence effect upon eye movements and shows that a cognitive override of the effect comes into play when domain-specific knowledge is taken into consideration.

3.6. Chapter 3 conclusions

In support of previous findings such as Wells (1987), this chapter has provided evidence for a domain specialist advantage of recognition memory, whereby American Studies students and Engineers were reliably more accurate when viewing stimuli from their domain of expertise (chapter 3.2), as were computer specialists when viewing pictures containing a computer component (chapter 3.4). Repetition of the memory test after a delay of one week (chapter 3.3) showed that this advantage is stable over time, corroborating previous studies that have reported good visual long term memory (Castelhano and Henderson, 2005; Henderson and Hollingworth,

2003; Hollingworth & Henderson, 2002; Hollingworth, 2005) and also those that have found evidence of better long-term memory for domain-specific material (Hayes-Roth, 1983; Postal, 2004; Drai-Zerbib and Baccin, 2004) than irrelevant material. Furthermore, when stimuli at encoding were visually complex and hard to differentiate, verbalising memories of previously seen pictures improved subsequent performance at recognition test (chapter 3.4).

This chapter has also shown that saliency helps guide eye movements, lending support for the saliency map theory (Koch & Ullman, 1985), but plays less of a role when top-down knowledge is utilised (Parkhurst et al, 2002). This cognitive override of saliency can be seen in the reduced number of fixations to salient regions when specialists viewed domain specific pictures (chapters 3.2, 3.3 and 3.4) and the decrease in scanpath similarity between actual and model-predicted saliency scanpaths when viewing domain specific pictures (chapters 3.2 and 3.3). It is also evident by the decrease in verbal references to salient areas when semantically interesting areas are present (chapter 3.4). Participants may have learnt to ignore task-redundant information and limit their processing to task-relevant information (Haider and Frensch, 1999), or, as suggested by Torralba, Oliva, Castelhano and Henderson (2006), image saliency and global-context features may be computed in parallel, in a feed-forward manner and integrated at an early stage of visual processing (i.e. before initiating image exploration).

As in the Chapter 2, the sequences of eye movements at encoding and recognition (scanpaths) were found to be more similar than would be expected by chance (chapters 3.2, 3.3 and 3.4) and more so when describing previously seen pictures than when simply imagining them. Furthermore, the advantage of expertise for domain-specific stimuli still existed when no visual information was present, with specialists' scanpaths for computer pictures at verbalisation being more similar to encoding than control participants'. This suggests that some sort of expert memory is involved when no visual information is available, and that describing the previously seen picture helps both to remember the details and to guide eye movements around the blank screen. When visual information *is* available, on the other hand, eye movements are drawn to areas of semantic relevance, rather than participants relying on memories of previous target locations.

Lastly, this chapter investigated whether the advantage of domain expertise exists in a comparative visual task situation, and how this is affected by the presence of incongruent regions (chapter 3.5). Consistent with the previous findings of Walker (1987), McCarley at al., (2004) and Lansdale et al., (2009), specialists were reliably faster and more accurate at identifying the presence/absence of a difference between two pictures when stimuli were domain specific. Incongruous regions attracted attention early on in viewing, indicated by the increased number of first fixations to these areas, and, once fixated, retained attention as if participants had difficulty disengaging (indicated by the longer fixation durations). However, when viewing pictures from their domain of expertise, specialists were reliably less likely to first fixate the incongruent regions and spent reliably less time overall fixated within them, compared to both control stimuli and control participants. These results suggest a cognitive override due to top-down expertise that allowed the specialists to disengage attention from the incongruent regions. This experiment confirmed that domain expertise is advantageous in a comparative visual task situation.

Chapter 4: The effects of people and emotion and eye movements and memory

4.1. Introduction

Emotional faces have long been used to explore the effect of emotion on attention, and research suggests that these stimuli are processed automatically (e.g. Wells & Matthews, 1994), even, in some cases, without conscious awareness (Ohman, Esteves, & Soares, 1995). In a visual search paradigm, threat-related faces are detected faster than friendly or sad faces among an array of neutral faces (Fox et al., 2000; Ohman, Lundqvist, et al., 2001; and Tipples, Atkinson, and Young , 2002). Furthermore, increasing the number of distracters increases the reaction time to find a happy face, but not an angry face, which is still detected faster (Eastwood, Smilek, and Merikle, 2001). These claims of automatic processing are supported by neuroimaging studies that have found amygdala activation not only when participants viewed angry faces, but also when these faces were masked and participants appeared to be unaware of them (Morris, Ohman, & Dolan, 1998; Whalen, Rauch, Etcoff, McInerney, Lee, & Jenike, 1998). Furthermore, physiological studies have found that despite very short exposure (30 ms) to a masked stimulus, participants still make reliable facial muscle reactions to emotional expressions (Dimberg, Thunberg, & Elmehed, 2000). However, there is evidence that some top-down control of emotional faces may exist. Pessoa, Kastner & Ungerleider (2002) found that brain regions activated by emotional faces only responded when attentional resources were available to process the faces. When the participants focused their attention on another task, brain activation to emotional expression was eradicated.

The effects of emotional faces on overt attention have also been measured, by analysing eye movements. When viewing threat-related faces (anger and fear), participants produced an "extended scanning pattern" characterised by increased saccadic amplitude and more frequent, longer fixations on the eyes and mouth (Green, Williams and Davidson, 2003). It is suggested that this "vigilant" style of scanning is due to participants 'searching for threat (increased saccadic amplitudes) and then focusing on the areas that are high in threat information (eyes and mouth) to facilitate appraisal of the situation (e.g. direction of threat).

Other studies have found, however, that although emotive faces can capture attention, angry distracters are no more powerful in doing so than happy distracters (Hunt, Cooper and Kingstone, 2007). Hunt et al found that eye movements were not directed toward angry distracters any more often than toward happy distracters, and saccades to angry face targets were no faster than to other targets. Furthermore, there is evidence that attention is actually actively directed away from threat-related (fearful and angry) faces (Becker & Detweiler-Bedell, 2009). Using Ekman faces, (e.g. Matsumoto and Ekman, 2004), Becker et al demonstrated that negative emotional expressions are rapidly analysed in the early stages of processing, but are then actively avoided.

Despite this evidence for distinct scanning patterns of emotional faces, this can be influenced by, for example, the nature of the task. When asked to judge faces on similarity of identity to a target face, participants scanned the upper face more, but when asked to judge the similarity of expressions, they scanned the lower face more (Malcolm, Lanyon, Fugard and Barton, 2008). The age of the viewer can also affect inspection of emotional stimuli – older adults have shown an attentional preference towards happy faces and away from angry ones, whereas the only preference shown by young adults was towards fearful faces (Isaacowitz, Wadlinger, Goren and Wilson, 2006). It is possible that the preferential processing of fear stimuli in young adults is not present in older adults. Furthermore, eye patterns differ depending on the gender of the viewer. For example, although both males and females have been found to focus on the eye regions for faces, males make longer and more frequent fixations to the nose area than females do (Vassallo, Cooper and Douglas, 2009). Females, however, outperform males in correctly

identifying facial expressions (Vassallo et al, 2009), regardless of whether the viewed faces are presented in static (Hall and Matsumoto, 2004; Kirouac and Dore, 1985; Nowicki and Hartigan, 1988; Scholten, Aleman, Montagne, and Kahn, 2005; Thayer and Johnsen, 2000) or dynamic form (Biele and Grabowska, 2006; Montagne, Kessels, Frigerio, de Haan, and Perrett, 2005). This accuracy advantage has also been found to vary depending on the facial expression depicted, for example Mandal and Palchoudhury (1985) found that females were more accurate at recognising sad faces, while males were better at angry faces. It has been suggested that this is due to males being more frequently aggressive and having a more aggressive social role (Biele & Grabowska, 2006). This trend has also been noted in other studies (Biele & Grabowska, 2006; Montagne et al., 2005; Nowicki & Hartigan, 1988; Rotter & Rotter, 1988; Wagner, MacDonald, & Manstead, 1986), while Goos and Silverman (2002) reported that females were more sensitive to angry and sad expressions specifically when the poser was female. On the other hand, other recent studies such as Calvo and Lundqvist (2008) found no reliable differences in accuracy between male and female viewers.

As well as in normal participants, the effect of emotional faces has also been extensively researched in patients with psychological abnormalities, such as Social Anxiety, Autism, Schizophrenia and patients prone to delusions.

Social Anxiety is characterised by heightened awareness and increased anxiety in social situations and has been shown to affect attention to stimuli that convey social information (e.g. faces). Modified Stroop studies suggest that socially anxious individuals have an attentional bias for negative social information (e.g., Maidenberg, Chen, Craske, Bohn, & Bystritsky, 1996; Mattia, Heimberg, & Hope, 1993) and visual search studies have shown enhanced detection of angry relative to happy faces (Gilboa-Schechtman, Foa, & Amir, 1999). Bradley, Mogg & Millar (2000) found that as anxiety levels increased, reaction time to respond to targets replacing threat faces decreased. This was not found for happy or neutral faces, and to a lesser extent for sad faces. Not all patients made overt eye movements, but those who did showed preferential initial orienting to negative faces.

Garner, Mogg & Bradley (2006) also found that high Social Anxiety individuals initially fixated on emotional faces quicker than neutral faces, but looked at emotional faces for less time, compared to low Social Anxiety individuals. Anxious individuals may strategically reduce attention to threatening stimuli (as reflected by subsequent overt disengagement of gaze) in an attempt to alleviate their discomfort and discourage further social interaction. This initial orienting to then subsequent avoidance of emotional faces has been termed the "vigilance-avoidance" pattern of attentional bias (Heinrichs & Hofmann, 2001). The fact that this was true of happy faces as well as angry faces suggests that anxious individuals may evaluate happy expressions in a negative fashion (e.g., as a sign of mocking) and thus allocate attentional resources toward them along with other stimuli evaluated as threatening (Bradley, Mogg, White, Groom, & de Bono, 1999).

When viewed close up (e.g. in a free viewing rather then a search task), emotional faces provoke a unique type of eye movement pattern called "Hyperscanning" by individuals with Social Anxiety. This is characterised by fewer and shorter fixations, increased saccadic amplitude and avoidance of the main facial features, especially the eyes (Horley, Williams, Gonsalvez, & Gordon, 2003). This is most prominent for negative faces and is suggested to be a defensive strategy to avoid evaluation by others (Salkovskis, 1991; Eysenck, 1992; Clark and Wells, 1995). Negative facial stimuli have more threatening connotations associated with the fear of negative evaluation, thus these stimuli promote more hyper-scanning than other emotional faces. The eyes

carry more social information than other facial features (e.g. it is easier to distinguish an angry face from the eyes alone than from the mouth alone) so it is not surprising that the eyes are avoided, especially in situations where they convey negative emotion.

Another group of individuals known for lack of eye contact are those with Autism, or Autistic Spectrum Disorder (ASD). ASD is a pervasive neurodevelopmental disorder that is characterized by impairments in social interaction, communication, and repetitive/stereotypic behaviours (American Psychiatric Association, 2000). Individuals with ASD show heightened scanning of the mouth, bodies and objects in a picture, at the expense of scanning of the eyes (Klin, Jones, Schultz, Volkmar, and Cohen, 2002), produce 'disorganized' scanpaths (Pelphrey, Sasson, Reznick, Paul, Goldman, & Piven 2002) and are less accurate in identifying emotions. During emotion judgment, those with ASD generally rely more on the mouth region than matched controls do (Spezio, Adolphs, Hurley, & Piven 2007), however this varies depending on the emotional expression of the stimuli. Negative expressions have been shown to promote more extensive scanning (Green et al, 2003) and increased scanning of the eye area (De Wit, Falck-Ytter and Hofsten, 2008). Conversely, Rutherford & Towns (2008) found no reliable difference in scanning between individuals with ASD and controls, but those with ASD did look at the eyes less than the control group when viewing *complex* emotions.

Difficulty in recognising emotions has also been associated with Schizophrenia and has been found with both positive (Archer, Hay, and Young, 1994; Schneider, Gur, and Shtasel, 1995) and negative emotions (Dougherty, Bartlett., & Izard 1974; Mandal and Rai, 1987; Muzekari and Bates, 1977). Individuals with Schizophrenia have been found to make atypical scanpaths to happy faces, represented by fewer fixations, a longer median fixation duration shorter scanpath length and distance between fixations, and shorter and fewer fixations to facial features such as the eyes and mouth (Loughland, Williams and Gordon, 2002). This scanpath style has been termed "restricted" and suggests an abnormal reliance on a sequential visual search strategy; possibly a compensatory mechanism arising from earlier problems in Gestalt processing (Schwartz, Rossee, Johri, & Deutsch 1999). Conversely, when viewing sad faces, Schizophrenic participants showed greater attention to facial features (i.e. eyes, nose and mouth) and they were equally accurate as controls at identifying the emotion. This effect might represent Schizophrenics' bias towards emotionally negative stimuli, as previously found in emotional Stroop studies (Bentall and Kaney, 1989; Kinderman, 1994) and experiments using happy and sad faces (David, 1989). Caution must me taken, however, to take into account individual differences, as different Schizophrenic symptoms have been associated with different scanpaths to emotional faces. Positive symptoms (e.g. lack of apathy, poor social functioning) have been associated with 'extensive' scanpaths and negative symptoms (e.g. hallucinations, delusions) with 'restricted' scanpaths (Gaebel, Ulrich, & Frick 1987; Kojima, Matsumisha, Nakajima, Shiraishi, Ando, Anod, & Shimazono, 1990; Streit, Wolwer, & Gaebel, 1997), both strategies reflecting reduced attention to facial features such as the eyes and mouth. Delusion-prone individuals also spend more time looking at threat-related stimuli (Freeman, Garety, & Phillips, 2000)

In summary, faces showing emotional expression such as happiness, sadness, anger and fear, have a strong effect on attention and eye movements, both in normal participants and individuals with psychological abnormalities. However, to further investigate the effect of emotional stimuli on eye movements, the stimulus type has to be carefully considered. The presence of people (and especially faces) in pictures, have been found to hold a unique 'natural' ability to attract attention – a 'social saliency'. Therefore, using emotional faces or people as stimuli poses the difficulty of whether fixations on areas of interest are due to the emotional

valance of the stimuli or simply because they contain people. For this reason, the effect of the presence of people on eye movements will first be investigated before further examining emotional valance. Furthermore, many previous studies have used unrealistic (e.g. schematic: Eastwood, Smilek, and Merikle, 2001) faces, which hold little ecological validity. The research in this Thesis is interested in whether these effects can be found in using less artificial stimuli, and thus the following studies will use more naturalistic complex visual scenes.

4.2. Experiment 8: The effects of the presence of people in pictures, on eye movements and memory

4.2.1 Introduction

Does the presence of people in a natural scene affect the way we move our eyes? Previous research suggests that we have a natural tendency to look at the social information before other items in a scene. Yarbus (1967) showed participants a picture of the Repin painting 'An Unexpected Visitor' and found that there was a tendency to look at the heads and faces of the people. This focus on heads and faces could be to try and work out where the people are attending (e.g. Baron-Cohen, 1994), and has been shown to occur in children as young as three months (Hood, Willen & Driver, 1998), in 3-5 year old children (Ristic, Friesen & Kingstone, 2002) and in adults (Friesen & Kingstone, 1998; Langton & Bruce, 1999). These studies suggest that we do indeed look at social information in a scene. However, Birmingham, Bischof, & Kingstone, (2008) argue that the effect of social information/cuing in these (and other similar) studies might be so strong because the stimuli show only a face, therefore restricting what the participant looks at. In response to this, Birmingham et al used complex real-world scenes containing people and found that participants still fixated on the eyes more frequently then other objects/regions. Similarly, Smilek, Birmingham, Cameron, Bischof, and Kingstone (2006) recorded participants' eye movements while viewing natural scenes and found that they tended to make saccades towards the eyes of people rather than to other objects or background regions of the scenes.

This repeated finding that participants focus on people in a scene (and more specifically faces), has lead to the suggestion that faces have a biological significance that attracts attention (e.g. Ro, Russell and Lavie, 2001). Ro et al found that, using a flicker paradigm, changes to faces were noticed both more accurately and more quickly than changes to other objects. Participants have also been shown to detect a change made to a scene sooner when an individual appearing in the scene was gazing at the changing object than when the individual was absent, gazing straight ahead, or gazing at a non-changing object (Langton, O'Donnell, Riby & Ballantyne, 2006). Similarly, Fletcher-Watson, Findlay, Leekam, and Benson (2007) showed participants two scenes next to each other, one of which contained a person. Many more initial fixations were made to the pictures containing the person. This was true during free-viewing and during a gender discrimination task

Despite the abundance of evidence for people being fixated more often than other areas, few studies have looked at the effect that the presence of people has on measurements such as saccadic amplitude, accuracy, or scanpaths. Previous research has found similarities between scanpaths at encoding and recognition (e.g. Humphrey & Underwood, 2009), are these similarities affected by the presence of people? It would be also be interesting to find out whether scanpaths differ at depending on accuracy, i.e. whether participants correctly or incorrectly identify the stimuli at recognition test. According to Scanpath Theory, reproducing the same eye movements

should help to recognise the stimulus, and thus it follows that less scanpath similarity should lead to a decrease in accuracy. If the presence of people in a scene creates very similar eye movements at encoding and recognition due to attention being drawn towards social cues, it is predicted that accuracy should increase for pictures containing people.

Previous research has suggested that accuracy of recognition and recall is affected by eye movements within a scene. For example, Underwood, Chapman, Berger and Crundall (2003) recorded eye movements while participants watched video recordings taken from a moving vehicle. Hazardous events (e.g. a pedestrian stepping out into the road), were fixated more often than non-hazardous objects/events. When memory was tested immediately after a hazardous event, there was evidence of attentional focussing and reduced availability of details about incidental objects. This demonstrates that where people look in a scene can affect how accurate they are at recognising or recalling details from the scene at a later time. The aim of this experiment is to investigate if eye movements (including fixation duration, saccadic amplitude and scanpaths) differ when participants correctly and incorrectly identify stimuli at recognition, and whether this is affected by the presence of people in the natural scene.

4.2.2 Methodology

Participants

Fifteen participants took part in the experiment, all of whom were students at Nottingham University (undergraduates and postgraduates). The age range was 18-39 and the mean age was 21.5. The sample comprised 10 females and 5 males. All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and datum from one participant had to be excluded.

Materials and apparatus

Eye position was recorded using an SMI iVIEW X Hi-Speed eye tracker (see section 1.5.1 for specifications). A set of 200 high-resolution digital photographs were prepared as stimuli, sourced from a commercially available CD-ROM collection and taken using a 5MP digital camera. The photos were of agricultural scenes and of this set of 200, 100 contained people and 100 did not (see figures 72 and 73). The regions of interest were defined around the person or persons inside the picture, meaning that every 'people picture' had its own unique RoIs. Half the pictures contained one person and half contained more than one person. When the scenes contained more than one person, the people were generally in the same region of the picture and the RoI was defined around them. However, when the scene contained multiple people that were not in the same region, RoIs were defined around each individual and an average RoI size and distance from the centre was calculated. In 54% of the pictures, the RoIs were less than 10 degrees of visual arc from the centre, with the remaining 46% being 10 degrees or more from the centre of the picture. In half the pictures, the 'people' RoIs were less than 20 pixels and in half they were 20 pixels or more.

Half of each stimulus category (people/no people) were designated "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test. Pictures were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. The monitor measured 43.5cm by 32.5cm, and a fixed viewing distance of 98cm gave an image that subtended 25.03 by 18.83 degrees of visual angle.
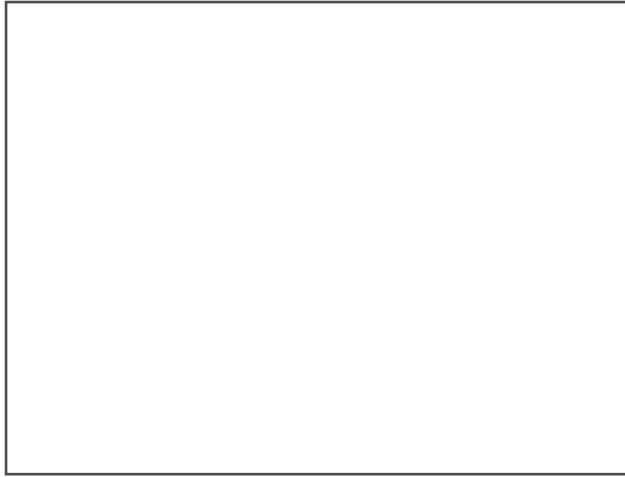
Figure 72: *An example of a 'no people' stimulus.*

Figure 73: *An example of a 'people' stimulus.*

Procedure

Following a 9-point calibration procedure, participants were shown written instructions asking them to inspect the following pictures in preparation for a memory test.

In a practice phase designed to familiarise participants with the equipment, the displays, and the task, they were shown a set of five photographs that were similar to the ones in the experimental set. Participants were not told to look for anything in particular in any of the pictures but were asked to look at them in preparation for a memory test. Following the practice phase, the first stage of the experiment began. One hundred stimuli (50 with people, 50 without) were presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 2000 milliseconds, during which time participants moved their eyes freely around the screen. A presentation time of 2000 milliseconds long enough to get an average of 5 fixations but short enough to make the task quite challenging. The task was designed to be difficult in order to decrease the accuracy rate, so that eye movements from correct responses at recognition could be compared to incorrect responses. This was achieved by shortening the presentation time and also using a large number of

pictures (200 pictures at recognition).

After all 100 stimuli had been presented, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was new (never seen before) or old (from the previous set of pictures). Participants were instructed to press "N" on the keyboard if the picture was new, and "O" on the keyboard of the picture was old. During this phase, 200 stimuli were presented in a random order; 100 of these were old and 100 new (though the participants were not informed of this fact). In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 2000 milliseconds and participants could only make a response after this time. Accuracy was emphasised over speed. This was to encourage scanning of the whole picture so that scanpaths from the first and second phases of the experiment could be compared. At the start of the second phase, participants were given a practice of the task, using 10 photographs that were similar to the ones in the experimental set, 5 of which were the practice photographs from the first part of the experiment. Feedback was given in the practice phase as to whether or not the participant gave the correct response of "old" or "new". No feedback was given in the experimental phase.

4.2.3. Results

There were 2 main types of data, recognition memory data (accuracy); and eye tracking measures – average fixation durations, average saccadic amplitude, and string analyses (encoding compared to second viewing).

Trials were excluded where the fixation at picture onset was not within the central region (the central square around the fixation cross when the picture was split into a 5x5 grid at analysis), when participants looked away from the screen (e.g. to the keyboard), or when calibration was temporarily interrupted (e.g. if the participant sneezed, therefore removing their head from the eye tracker). Overall, 7% of trials were excluded.

4.2.3.1. Recognition Memory

Accuracy

Participants were more accurate when the stimuli contained people (76%) compared to when no people were present in the stimuli (71%). Accuracy was measured by the percentage of pictures correctly identified at recognition (see figure 74).

Figure 74: *A bar chart showing the accuracy rates for pictures containing people and those containing no people. Participants are more accurate when pictures contain people.*

A paired samples T-test was conducted (appendices 43a and 43b) and found the difference to be statistically reliable, $t(13) = 3.322$, *SEM* $= 0.01333$, $p<0.001$. This analysis was repeated using a 2x2 repeated measures ANOVA (appendix 43c), taking into account whether stimuli were old (from the previous set) or new (distracters). See figure 75.
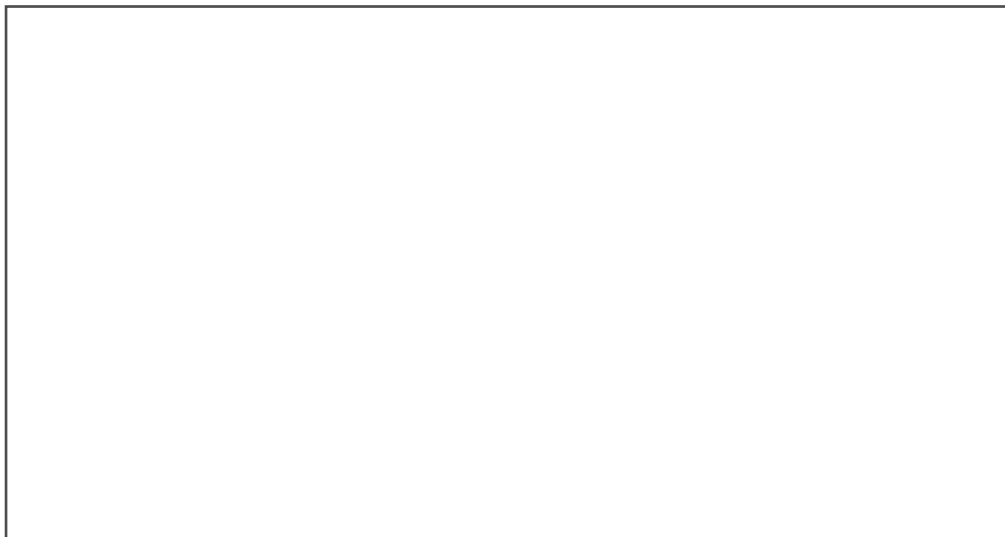
Figure 75: *A bar chart showing the accuracy rates for old and new pictures containing people or no people.*

There was a reliable effect of people $F(1,13) = 6.403$, *MSE* $= 23.604$, $p<0.05$, and also a reliable effect of old/new stimuli $F(1,13) = 13.401$, *MSE* $= 217.495$, $p<0.05$. There was no reliable

interaction [$F(1,13) = 1.001$, $MSE = 125.846$, $p = 0.335$].

      Three further one-way ANOVAs were carried out (appendices 43d, 43e and 43f) to take into account other variables of the stimuli that could affect recognition accuracy, namely the size of the Region of Interest (RoI), the number of people in the picture, and the distance of the RoI from the centre of the screen (see figures 76a, 76b and 76c).
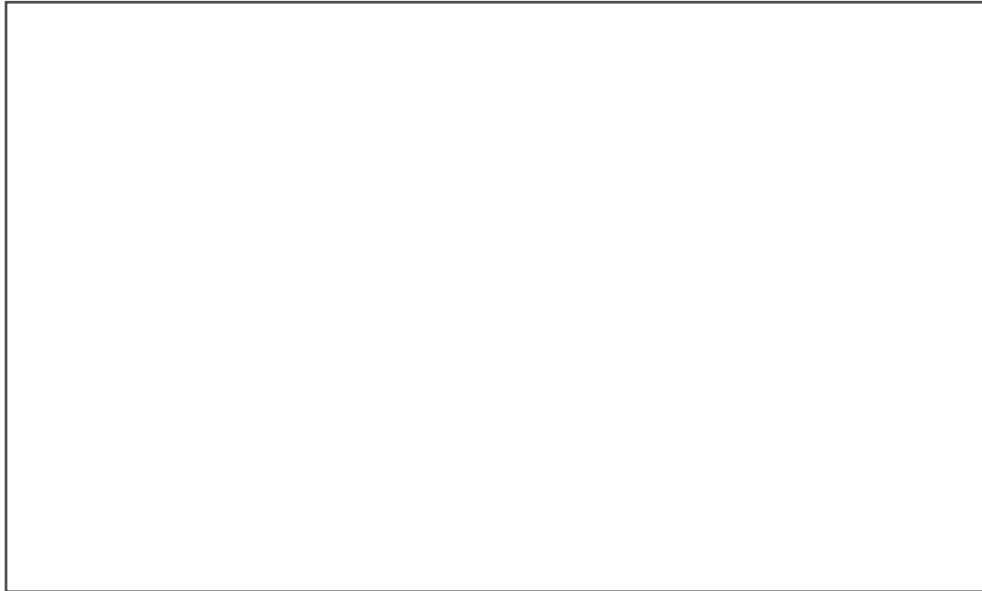
Figure 76a. *A bar chart illustrating the accuracy rates for pictures with a Region of Interest of less than 20 pixels and for pictures with a Region of Interest of 20 pixels or more.*
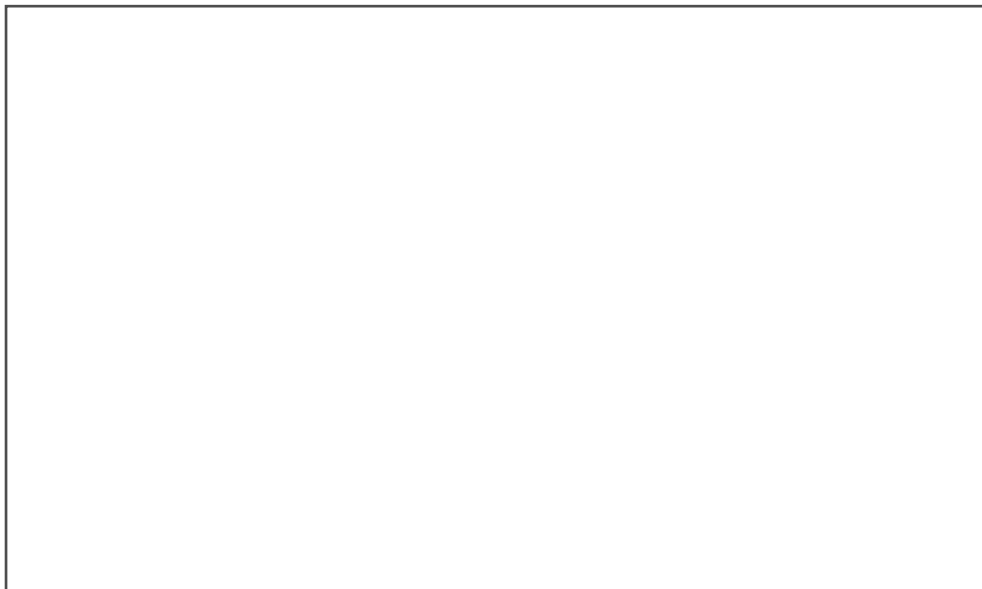
Figure 76b: *A bar chart illustrating the accuracy rates for pictures containing one person and for pictures containing more than one person.*

Figure 76c: A bar chart illustrating the accuracy rates for pictures where the RoI is less than 10 degrees of visual arc from the centre of the screen (initial fixation) and where the RoI is 10 degrees of visual arc or more from the centre of the screen.

There was no reliable difference in accuracy due to the number of people ($p = 0.780$), the size of the RoI ($p = 0.206$), or the distance of the RoI from the centre of the screen ($p = 0.089$).

4.2.3.2 Eye-tracking measures

Average Fixation Duration
Fixations under 70ms were counted as corrective fixations and were not included in the analyses. The mean average fixation durations are shown in figure 77.
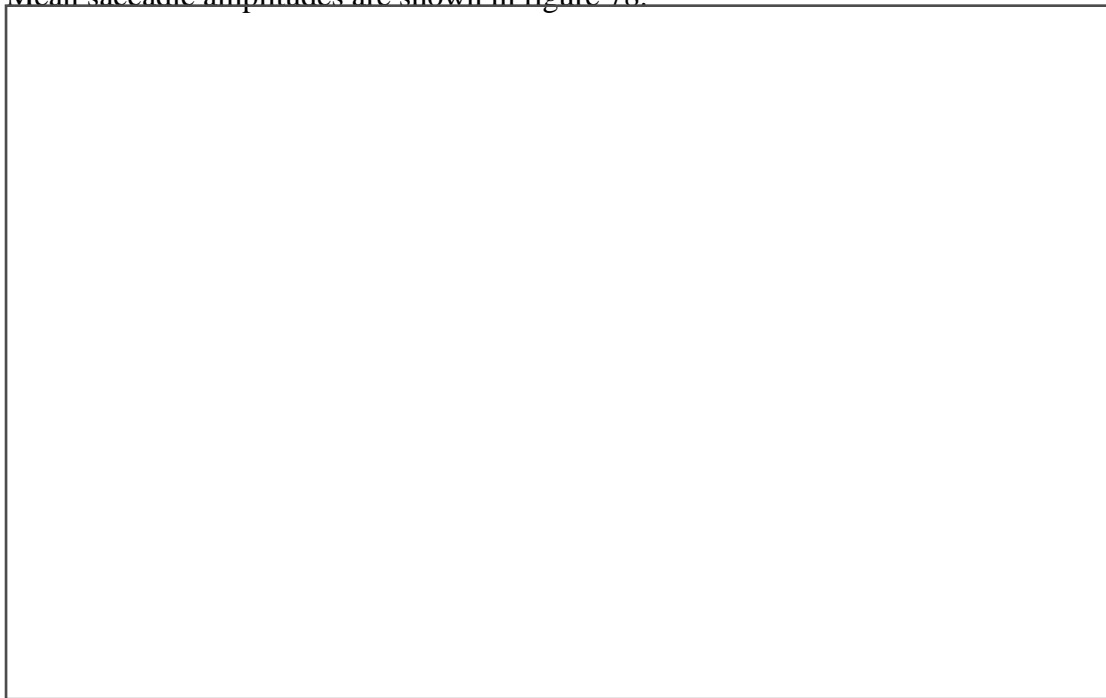
Figure 77: A bar chart showing the average fixation durations for pictures shown at the recognition memory test. Stimuli have been separated into old or new, people or no people and correctly or incorrectly identified.

A 2x2x2 repeated measures ANOVA (appendix 44) found no reliable difference in fixation duration between correctly an incorrectly identified stimuli [$F(1,13) = 0.297$, *MSE* = 671.642, *p* = 0.595], or between people and no people stimuli [$F(1,13) = 2.242$, *MSE* = 786.249, *p* = 0.158]. There was a reliable difference in fixation duration between old stimuli and new stimuli, $F(1,13) = 14.223$, MSE = 518.773, *p*<0.05. There were no reliable interactions.

Regions of Interest Analysis
Regions of Interest (RoI) analyses were conducted using a toolbox in Matlab called 'ilab' (Gitelman, 2002). This software allows the user to define an area of interest and calculate, for example, how many fixations fell inside this area, how long (per trial) was spent fixated within this area, etc. The RoIs are defined using a graphic user interface that allows the user to draw a box around the region of interest. The x-y coordinates are calculated and saved, so only have to be defined once and can be applied to many data sets. Blink removal and filtering (smoothing) was applied to allow the removal of certain types of artifacts in the data. For example, during eye blinks, the eye tracker loses track of the pupil center and corneal reflection. Because eye blinks are not instantaneous, the amount of data loss can be variable. Therefore, a combination of filtering the data by incorrect position information (outside the boundaries of the computer screen) and filtering by 0 pupil size (eliminates those data points at which the pupil size is 0) was used to 'clean up' the data.

On average, the RoI occupied 19.9% of the 'people picture', and 58% of fixations fell inside the RoI (i.e. were focused on the person/persons in the picture). Additionally, an average of 45% of all the time spent on each 'people picture' was focused on the RoI. In order to calculate a chance baseline, each 'people picture' was randomly paired with a 'no people picture'. The RoI from the 'people picture' was applied to the 'no people picture' and the number of fixations and

time spent in that area was calculated. Paired samples T-tests were carried out (appendices 45a and 45b) and found that a reliably greater number of fixations fell in the RoIs in the people pictures than in the same areas on the randomly assigned 'no people pictures'. Furthermore, a reliably longer amount of time was spent in the RoIs in the people pictures than in the same areas on the randomly assigned 'no people pictures'. It can be concluded that more fixations fell inside the RoIs of the people pictures than would be expected by chance: $t(13) = 9.709$, $p<0.05$; and that more time was spent fixating within the RoIs of the people pictures than would be expected by chance: $t(13) = 12.755$, $p<0.05$.

Average Saccadic Amplitude
Mean saccadic amplitudes are shown in figure 78.

Figure 78: A bar chart showing the average saccadic amplitude for pictures shown at the recognition memory test. Stimuli have been separated into old or new, people or no people and correctly or incorrectly identified. Saccadic amplitude is measured in degrees of visual ark.

A 2x2x2 repeated measures ANOVA (appendix 46a) found no reliable difference in saccadic amplitude between correctly and incorrectly identified stimuli [$F(1,13) = 1.799$, $MSE = 0.105$, $p = 0.203$], people and no people stimuli [$F(1,13) = 4.150$, $MSE = 0.542$, $p = 0.063$], or between old and new stimuli [$F(1,13) = 3.608$, $MSE = 0.326$, $p = 0.080$]. There were no reliable two-way interactions, although the interaction between people/no people and old/new stimuli was nearing significance [$F(1,13) = 4.433$, $MSE = 0.267$, $p = 0.055$]. There was a reliable 3-way interaction, $F(1,13) = 10.454$, $MSE = 0.317$, $p<0.05$ (see figure 79).

Figure 79: *A Line graph to illustrate the 3-way interaction between the factors people/no people, correct/incorrect, and old/new pictures.*

Post-hoc T-tests (appendices 46b and 46c) showed a reliable difference between: 'incorrect old no people' and 'incorrect new no people', $t(13) = 2.747$, *SEM =* 0.32824, $p<0.05$; 'incorrect old no people' and 'correct old no people', $t(13) = 2.708$, *SEM =* 0.21263, $p<0.05$; and between 'incorrect old people' and 'correct old people', $t(13) = 2.424$, *SEM =* 0.21592, $p<0.05$.

4.2.3.3. Scanpaths: String Editing
String editing was used to analyse the similarity between scanpaths produced on encoding and second viewing. Strings were cropped to seven letters, and were computed for each subject viewing each stimulus in the experiment. Seven letters were used because this was the mean number of fixations made on each stimulus. This gave a more standardised and manageable data set, and was long enough to display any emerging similarity. In those trials where fewer than seven fixations remained after condensing gazes, any comparison strings were trimmed to the same length. Once the strings had been produced for all trials, they were compared using the editing algorithm and an average string similarity was produced across trials.

Chance was calculated by comparing eye movements on each picture a participant viewed, to eye movements that the participant produced on another randomly selected picture. This analysis was split into two categories - pictures that contained people and pictures that did not contain people. This was repeated for all 14 participants and average similarity scores were calculated; 0.20969 for pictures that contained people and 0.232419 for pictures that did not contain people.

Encoding vs. Recognition
The scanpaths generated from encoding of a picture were compared to those on second viewing during the recognition test. When participants were incorrect, scanpaths were more similar at encoding and recognition if the pictures contained people.
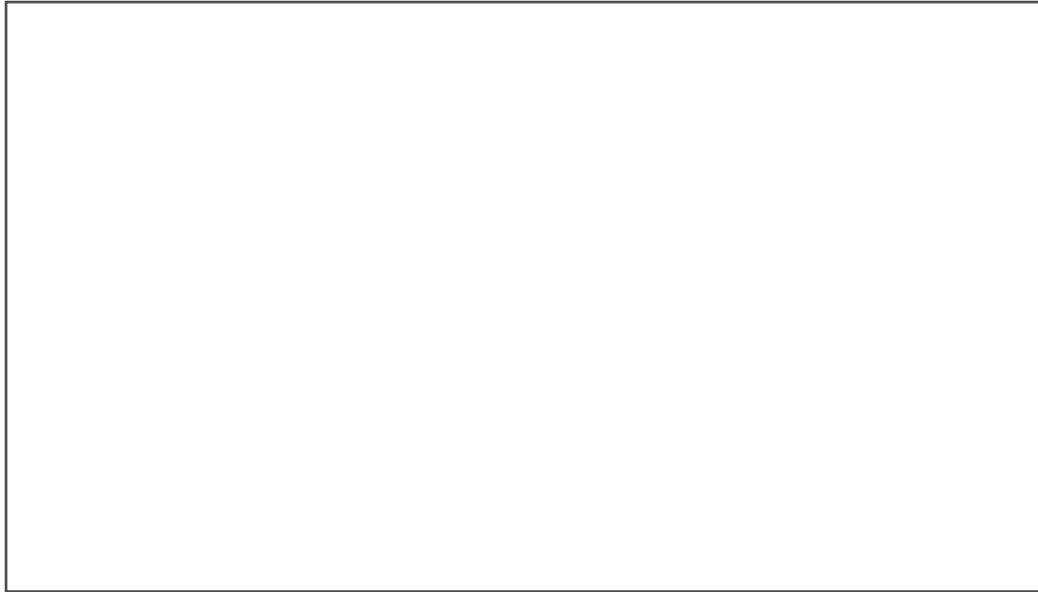
Figure 80: *A bar chart showing the average string similarities at encoding and recognition for each condition.*

A 2x2 repeated measures ANOVA (appendix 47a) found no reliable difference in string similarity scores between correct and incorrect stimuli [$F(1,13) = 1.964$, *MSE* $= 0.004$, $p = 0.185$], and no reliable interaction [$F(1,13) = 0.853$, *MSE* $= 0.002$, $p = 0.372$] between correct/incorrect and people/no people. However, there was a reliable difference in string similarity scores between people and no people pictures $F(1,13) = 7.184$, *MSE* $= 0.002$, *p<0.05*. Alongside the statistics in appendix 47b, it can be concluded that scanpaths are reliably more similar at encoding and recognition when pictures contain people.

For three of the conditions, scanpaths at encoding and recognition were reliably more similar than would be expected by chance (appendices 47c and 47d): for Correct People vs. chance, $t(13) = 4.462$, *p<0.001*; for Incorrect People vs. chance, $t(13) = 5.071$, *p<0.001*; for Correct No People vs. chance, $t(13) = 2.517$, *p<0.05*. The Incorrect No People condition was not statistically reliably greater than chance (*p=0.475*).

4.2.4 Discussion

Does the presence of people in a picture affect recognition memory? Our findings show that it does. Participants were reliably more accurate when the stimuli contained people compared to when no people were present in the stimuli. Henderson et al's (1999) study suggests that if participants had to remember scenes in a later memory test, they looked at more semantically informative scene items and consistent with Birmingham et al (2008), people in the scene could act as socially informative items. When encoding the pictures, participants may use these social cues to form conclusions such as 'picking bananas' or driving the tractor', which may aid memory at recognition. These conclusions would be harder to form in 'no people' pictures, consisting of fields, hay bails, farm houses, cattle, tractors etc.

This said, it should be noted that the task, with 200 pictures at recognition, was far from easy (mean accuracy was 70% for people pictures and 61% for no people pictures). Both the 'people' and 'no people' pictures had others in the set that were very similar – in some cases it

was the same scene taken from a different camera angle (participants were warned of these possibilities and were told to treat them as 'new' unless they were *identical* to previously seen pictures). The 'no people' pictures had other inanimate objects of a similar size to persons in the 'people' pictures, and in both cases objects/people were sometimes in the foreground and sometimes the background. The task was deliberately designed to be difficult in order to compare the differences in eye data on the old pictures that participants correctly identified and those that they incorrectly identified. The increased accuracy for 'new' pictures could be due to participants not recognising some pictures due to the difficult nature of the task and subsequently responding 'new'. If the picture was in fact new, then this response would increase the accuracy for new pictures, but if the picture was in fact old and the participant simply did not recognise it, then this would decrease the accuracy for old pictures. The task was difficult, but it did not overshadow the fact that participants were reliably more accurate when the stimuli contained people compared to when no people were present in the stimuli. One possible limitation worth concidering for future research is that some 'people' pictures were paired with pictures that had regions of interest of a similar size and visual complexity but which contained inanimate objects such as tractors. If this study was repeated, a more controlled method could pair the people RoIs with animal RoIs, e.g. horses, as animals have also been found to attract attention in a similar manner to the presence of people in a scene.

The large and varied nature of the stimulus set helped to make the task difficult and the stimuli unpredictable. Further analysis of the data revealed that accuracy was not affected by the number of people in the pictures, the size of the Region of Interest, or the distance of the RoI from the centre. This suggests that differences in fixation duration, saccadic amplitude and scanpaths are not due to variations within the stimulus set. More importantly it suggests that no matter how many people there are in a picture, how large or small those people are or how far away they are from initial fixation, the presence of people in a natural scene affects eye movements and recognition memory.

Average Fixation Duration analyses showed that when participants viewed old stimuli, they made longer fixations. One possible explanation is that the participants fixated on an area/object that looked familiar and looked at it for longer to be sure that they had seen it before. On new pictures, participants may have looked around the stimuli trying to find familiar areas, but since there weren't any; the duration of each fixation was shorter. This coincides with the accuracy data that suggest 'old' pictures are harder to identify than 'new' ones.

The Region of Interest analyses revealed that more fixations fell inside the RoIs of the people pictures than would be expected by chance, and that more time was spent fixating within the RoIs of the people pictures than would be expected by chance. In other words, participants did look at the people in the pictures, providing evidence that the presence of people in natural scenes does affect the way we move or eyes. We understand that there is no agreed way to calculate a chance baseline and many different methods have been considered, (e.g. Fletcher-Watson Findlay, Leekam, and Benson, 2008, paired a person present scene with a person absent scene in the same display). The problem lies in the visual complexity of people. Even if this study was repeated using the same scenes with and without people, there is no objective way of measuring the complexity of the object(s) that would replace the person in the paired scene. We therefore maintain that the method of randomly pairing people and no people pictures and comparing RoIs was a fair and reliable way of measuring chance.

Saccadic amplitude analyses revealed a 3-way interaction. Firstly, for old pictures that were *incorrectly identified*, when no people were present in the pictures saccadic amplitude was

reliably smaller, and when people were present, saccadic amplitude was reliably greater. Secondly, when pictures were old and contained *no people* an incorrect response was related to reliably smaller saccadic amplitude and a correct response was related to reliably greater saccadic amplitude. Thirdly, when pictures were old and *contained people*, an incorrect response was related to reliably greater saccadic amplitude. It seems that when the picture contains people, if participants ignore this social information but instead search more widely (i.e. increased saccadic amplitude), they are more likely to incorrectly identify the picture. One possible explanation is that when no social information is present, participants have to search more widely to find familiar objects/areas. If they fail to do this (i.e. decreased saccadic amplitude), then they are more likely to incorrectly identify the stimuli. The accuracy data & fixation duration analyses suggest that it is harder to correctly identify an old picture than a new one, which could explain why these interactions are found only on old pictures.

Comparisons of scanpaths at encoding and recognition showed that similarity between the two was high in all conditions. According to Scanpath Theory, the scanpaths at encoding were similar to those at recognition because they were stored and recalled top down, to determine the scanning sequence. Birmingham et al (2007) suggest another reason for the similarity in scanpaths at encoding and recognition for pictures containing people. They found that when participants were asked to encode scenes for a memory test, they fixated on the eye area within 'people scenes' more frequently than when they were simply asked to freely view the pictures. This was true at both encoding and recognition and suggests that the eyes are scanned strategically by observers who are aware that they will have to remember the scenes. Participants who were not told of the memory test fixated the eyes more strongly in the (surprise) test session than in the free-viewing study session. Thus, Birmingham concludes, the eyes appear to be informative for both deliberately encoding scenes and for spontaneously trying to recognize them.

An aim of this study was to investigate whether scanpaths differed at encoding and recognition depending on accuracy. The results suggest some evidence for the replication of eye movement sequences over multiple viewings, however, the lack of relationship between string similarity and accuracy challenges the idea that the reproduction of eye movements alone is enough to create a memory advantage. People pictures could have been easier to recognise because participants formed conclusions about the semantic content/gist, assigning mental labels to these pictures, e.g. 'driving the tractor'. Therefore, even when eye movements were not perfectly reproduced at recognition, people pictures were still easier to identify.

An interesting finding of this study was that when participants incorrectly identified 'old' pictures as 'new', scanpaths at encoding and recognition were reliably more similar when the pictures contained people. This suggests that even though participants were making very similar eye movements at encoding and recognition on the 'people' pictures, they were still incorrectly identifying them. This could be because participants did not use the social information in the people pictures at either encoding or recognition, making it harder to correctly identify the picture. This coincides with the greater saccadic amplitudes on incorrectly identified people pictures, suggesting that they were not looking at the people, but rather at the wider scene. For the incorrectly identified pictures that contained no people, lower scanpath similarities imply that participants did not always look in the same places at encoding and recognition, which could have impeded successful recognition.

In conclusion, the presence of people in natural scenes increases recognition accuracy regardless of the size of the RoI, its distance from initial fixation or the number of people. Old pictures are harder to correctly identify and produce larger fixation durations. When people were

present in the scene, increased saccadic amplitude was related to a reliable decrease in accuracy, possibly due to participants ignoring important social cues. Scanpath analyses showed a high similarity between encoding and recognition, but the lack of relationship between string similarity and accuracy challenges the idea that the reproduction of eye movements alone is enough to create a memory advantage.

4.3. Experiment 9: Emotion vs. Saliency

Foreword

As can be seen from the last experiment and from previous research in the literature, there is a bias for participants to look preferentially at people than non social cues (e.g. Garner, Mogg & Bradley, 2006). The presence of people in an experiment designed to test the effect of emotion may reduce and differences between positive and negative complex scenes and make it difficult to differentiate whether participants look at the stimuli because they are emotionally salient, or because they are 'socially salient'. It is the decision of the author, therefore, not to include people in the stimuli for the following experiment.

4.3.1 Introduction

Our ability to rapidly detect emotive stimuli has been studied extensively. It is suggested that the automaticity of processing has evolutionary links, in that the fast and reliable detection of positive and negative reinforcers facilitates adaptive behaviour, finally promoting survival and reproductive success (Cacioppo et al., 1999; Lang et al., 1997; Öhman, Flykt, & Lundqvist, 2000). Evidence for such rapid detection comes from autonomic measures (electrodermal responses and heart rate), facial EMG responses (corrugator and zygomaticus), startle blink reflex (cf. Bradley, 2000; Hamm, Schupp, & Weike, 2003), and Event Related Potential studies (Schupp, Junghofer, Weike, Hamm, 2004).

Much of the previous literature has found that our attention is automatically drawn to negative information more strongly than it is automatically drawn to positive information (Hansen and Hansen, 1988; Pratto and John, 1991). This has been termed a 'negativity effect' (or 'negativity hypothesis') and can be seen in Event Related Potentials (Ito et al, 1998; Smith, Cacioppo, Larsen & Chartrand, 2003), faster reaction to subliminally presented negative stimuli (Dijksterhuis and Aarts, 2003) and increased skin conductance responses even when presented subliminally (30 ms of exposure) and backwardly masked (see Öhman & Mineka, 2001). According to the aforementioned evolutionary theory, the faster detection of negative stimuli may have developed as a survival mechanism. Negative stimuli such as a deadly poisonous spider require more immediate action (and thus faster processing) than, for example, a cute cuddly kitten, which poses little threat. In this sense, evaluation of threat may be a key underlying component of this evolutionary system (Öhman, 1993; Öhman, Flykt, & Esteves, 2001). Öhman (1993) suggested that threat information is first processed by a feature detection system, which 'tags' the stimuli as ancestrally or behaviourally relevant and passes the information to the organism's arousal system, which optimizes selective attention and orienting. Individuals then compare the stimuli with earlier memories and respond appropriately. However, many fears are irrational and

although they may have evolutional relevance, the individual may never have had an actual bad experience with the feared stimulus and thus have no associated negative memories to compare it to. Despite this criticism, evidence for rapid processing of fear stimuli has been found. Fear-relevant targets amongst neutral distracters are detected faster than neutral targets amongst fear-relevant distracters (Öhman, Flykt, & Esteves, 2001). However, this was only true if the target was feared by that individual – no effect was present for generic fear-relevant stimuli if they were not feared by the viewer. This raises the question of why there are so many individual differences in feared stimuli - if this system is based on evolutionary dangers, everyone should hold the same fears. Furthermore, these individual differences cannot be completely explained by fears developed through bad experience because many phobias are irrational.

As well as covert attention, negative emotional stimuli also affect overt attention, measured by eye movements. Viewers fixate more on unpleasant than on neutral stimuli (Christianson, Loftus, Hoffman, and Loftus, 1991) and when presented in the left visual field negative stimuli increase the 'leftward bias' (a tendency for people to first look into the left visual field) whereas if presented in the right visual field, leftward bias is decreased (LaBar et al., 2000). Harris et al. (2008) showed participants neutral scenes such as people throwing a beach ball, and scenes where the neutral object is replaced, giving the picture an emotional content, e.g. the beach ball is replaced with a baby. They found that viewers looked earlier at the emotional aspect of the scene.

Eye movement analysis has also provided evidence for an advantage of emotional stimuli irrespective of whether it is negative or positive. For example, Alpers (2008) found a left visual field bias regardless of valance and Lang et al. (1993) found that increased viewing duration for both pleasant and unpleasant scenes relative to neutral scenes. This equal advantage for positive and negative stimuli over neutral stimuli is referred to as the 'emotionality effect' or 'emotionality hypothesis' and is an alternative to the 'negativity hypothesis'. Calvo & Lang (2004) found that when emotional and neutral pictures were presented simultaneously, the probability of placement of the first fixation and the proportion of viewing time during the first 500 ms were higher for both pleasant and unpleasant pictures than for neutral pictures, suggesting emotional meaning captures initial overt orienting and engages attention early. However, one could argue that because each emotional picture was an entirely positive or negative scene (rather than a scene with a positive or negative region of interest), participants may have used the gist to process the stimuli top-down. Therefore one aim of the current experiment is to see if the emotionality effect still exists in complex scenes with an emotional region of interest but without an overall obvious emotional gist.

The emotionality effect has also been found when pictures are presented in the periphery (Nummenmaa, Hyona and Calvo, 2006), suggesting that the processing of emotional stimuli occurs at a very early stage, even before the eyes move to focus on the target. Nummenmaa et al found that the probability of first fixating an emotional picture, and the frequency of subsequent fixations, were greater than those for neutral pictures. Even when participants were instructed to avoid looking at the emotional pictures, first fixations were still more likely to fall on emotional stimuli than neutral stimuli.

Some studies have even found a 'positivity effect', that is, an advantage for positive pictures. When instructed to make a saccade towards a picture in the right peripheral visual field, facilitation occurred only for pleasant pictures and saccadic reaction times towards unpleasant pictures were slowed (Kissler & Keil, 2008).

Like emotional faces, the effect of other emotional stimuli has been studied in groups with

psychological abnormalities such as depression and Arachnophobia (fear of spiders). Initial orienting to negative stimuli does not seem to be affected by depression, but once attention is focused on negative information, depressed individuals may have greater difficulty disengaging attention (Bradley at el, 1997). This can be seen in increased gaze duration to negative stimuli by depressed individuals than control participants (Eizenman et al., 2003; Caseras, Garner, Bradley & Mogg, 2007). The difficulty in disengaging attention from negative stimuli is also seen in individuals with Arachnophobia, but for different reasons. Negative stimuli attended to by depressed individuals hold a valance that they can relate their feelings to, e.g. sadness. Attention is focused on these stimuli to reinforce existing negative emotions. In contrast, negative stimuli focused on by Arachnophobic individuals tend to be threat-related and difficulty in disengaging from these negative stimuli (Miltner, Krieschel, Hecht, Trippe, and Weiss, 2004; Rinck, Reinecke, Ellwart, Heuer, & Becker, 2005; Gerdes, Alpers, & Pauli, 2008) may be a defensive mechanism so as not to loose sight of the feared object. As well as difficulty disengaging attention from spider stimuli, individuals also seem to exhibit an attentional bias towards it in the first place (Mogg & Bradley, 2006). First fixation is more likely to be on spider pictures in a free viewing task (Rinck and Becker, 2006) and the more visually similar a distracter is to the feared object, the slower the participant is to respond to a target. However, there are individual differences in scanning behaviour depending on individual strategies of coping with threat (Pflugshaupt, Mosimann, Schmitt, Wartburg, Wurtz, Luthi, Nyffeler, Hess & Muri, 2007), and some studies have even found an initial orienting towards the threat stimulus followed by an avoidance of it (Mogg and Bradley, 2002; Mogg, Bradley, Miles, & Dixon, 2004). An aim of the current experiment is to see if there is evidence of increased likeliness of first fixating to and difficulty in disengaging attention from negative stimuli in participants without psychological abnormalities.

The emotional valance of a stimulus can also affect memory for that picture, neurophysiological evidence for which is extensive. Amygdale activity during encoding, which is modulated by hippocampal activity, has been found to be most correlated with memory for arousing events (Cahill and McGaugh's, 1998; Hamann, Ely, Grafton, and Kilts, 1999; Canli et al., 1999; and Canli, Zhao, Brewer, Gabrieli, and Cahill, 2000). Increased amygdala activity at encoding for negative words is significantly correlated with enhanced recognition memory for negative words tested after a short delay (Hamann and Mao, 2001). Furthermore, patients with amygdala lesions fail to show normal emotional enhancement of memory for positive and negative emotional pictures (Cahill et al., 1995; Hamann, Lee, and Adolphs, 1999). Other brain structures found to be related to emotional valance and memory include the right middle frontal gyrus and the superior anterior cingulate cortex (visual recognition memory for both negative and neutral stimuli) and greater activation in the lingual gyrus during recognition of negative images (Taylor, Liberzon, Fig, Decker, Minoshima, and Koeppe, 1998). Activation of right anterior parahippocampal and extrastriate visual brain areas have been related to recall of positive stimuli (Erk, Kiefer, Grothe, Wunderlich, Spitzer, Waltera, 2003), and when participants visualized previously viewed negative stimuli, activation in the occipital-parietal areas increased compared to visualized neutral stimuli (Kosslyn et al., 1996). Lastly, it has been proposed that the left hemisphere is primarily associated with processing of pleasant emotions, whereas the right hemisphere is primarily associated with processing of unpleasant emotions. This is referred to as the 'valence hypothesis' and is supported by electrophysiological (Davidson & Tomarken, 1989) and functional neuroimaging (Davidson, 1995; Davidson & Irwin, 1999) evidence.

The memory advantage for emotional regions of interest in a picture (especially negative ones) may be accompanied by diminished memory for other details in the picture; a phenomena

known as attentional narrowing. An example of this is the weapon focus effect, whereby the presence of a weapon captures attention at the expense of memory for other details such as the perpetrator's facial characteristics and clothing. When participants were asked to identify a target from a line up, they were less accurate if the target had previously held a gun compared to if they had previously held a cheque book (Loftus, Loftus, and Messo, 1987). Eye tracking analysis showed that whilst watching the original slide show containing the target person, participants fixated longer and more frequently on the gun than the cheque book. Similarly, Hope and Wright (2007) found that if the critical object was a gun, participants were more accurate and detailed in their descriptions of the object but recognition accuracy of the person was impaired. Participants were also less confident when questioned about the target person. The memory advantage for the critical object cannot be simply explained by increased fixations upon it, as the effect was still found when participants were only allowed to make one fixation (Christianson, Loftus, Hoffman & Loftus, 1991). Another aim of the current experiment is to see if there is any evidence of attentional narrowing. It is predicted that participants will accurately recall fewer neutral peripheral details of the scene when an emotionally negative region of interest exists.

Enhanced memory for negative images exists even when attention is divided or taken up by a competing task (Kern, Libkuman, Otani and Holmes, 2005), and even after an interval of 1-2 days (Kensinger and Corkin, 2003a). There is a small amount of evidence for a memory advantage for positive information (e.g. Erk, Kiefer, Grothe, Wunderlich, Spitzer, Waltera, 2003) and this positivity effect seems to increase with age (Quackenbush and Barnett, 2001; Mather and Carstensen, 2003; Charles, Mather, & Carstensen, 2003). Charles et al showed images on a computer screen and, after a distraction task, participants were asked first to recall as many as they could and then to identify previously shown images from a set of old and new ones. Both recall and recognition memory for negative images decreased with age and the proportion of positive images accurately recalled or recognized increased with age.

In summary, previous research has shown that emotion can act as a top-down influence on eye movements, attention and memory. However, many of these studies have not controlled for visual saliency, which, as we have seen in previous experiments (see section 3) and in the literature (e.g. Underwood, Foulsham, van Loon, Humphreys, and Bloyce, 2006), can also affect the way in which we inspect a visual scene. There is little research into the effect of emotional stimuli on saliency (or indeed the effect of saliency on emotional stimuli). One could argue that emotional content might be fixated because it is also visually salient. For example, a picture of a dead body may attract attention because of the bright red blood. Therefore, to determine which factor (bottom-up or top-down) most influences eye movements when viewing emotional pictures, we must control for both visual and emotional saliency. It would be interesting to see which of these variables has the strongest effect on eye movements, attention and memory when they are put in a competing situation. This will be tested in the current experiment by manipulating visual scenes so that the emotive region of interest is not the most visually salient region. This experiment also aims to find out whether an emotionality effect or a negativity effect exists in the guidance of eye movements and memory for emotional pictures.

4.3.2. Methodology

Participants

Forty-five participants originally took part in the experiment, all of whom were students at Auckland University (undergraduates and postgraduates). All participants had normal or corrected-to-normal vision. Inclusion in the study was contingent on reliable eye tracking calibration and data from five participants had to be excluded.

Materials and apparatus

Eye movements were recorded using an 'Eyegaze' eye tracker and 'Trace Suite' software. The Eyegaze System uses the Pupil-Centre/Corneal-Reflection method to determine the eye's gaze direction. A video camera located below the computer screen remotely and unobtrusively observes the subject's eye. No attachments to the head are required. A small, low power, infrared light emitting diode (LED) located at the centre of the camera lens illuminates the eye. The LED generates the corneal reflection and causes the bright pupil effect, which enhances the camera's image of the pupil.

Specialized image-processing software in the Eyegaze computer identifies and locates the centres of both the pupil and corneal reflection. Trigonometric calculations project the person's gaze point based on the positions of the pupil centre and the corneal reflection within the video image. The Eyegaze System generates raw gaze point location data at the camera field rate of 60 Hz. At each camera image sample, the Eyegaze System generates: an eye-found flag indicating whether or not the eye is visible to the camera and thus whether or not a valid gaze point is calculated; x-y coordinates of the subject's gaze point on the computer screen; pupil diameter; 3-dimensional location of the eyeball centre within the camera field-of-view, an indicator of head location and movement; and fixation and saccade analysis.

The Eyegaze System is robust and extremely easy to calibrate. RMS tracking errors are typically 0.25 inch (0.63 cm) or less and the advanced image processing algorithms in the Eyegaze System explicitly accommodate several common sources of gaze point tracking error, including head range variation and pupil diameter variation.

A set of 90 high resolution digital photographs were prepared as stimuli. All of the pictures were run through a saliency program, using Itti and Koch's (2000) model with standard parameters. Saliency maps were produced for the first four simulated shifts and thus indicated the first five most salient regions for each picture. The pictures were edited in Photoshop so that the most visually salient feature was not the most emotionally salient one. The pictures were also edited so that the emotionally salient object and the most visually salient object were always equidistant from the initial fixation cross prior to the picture onset.

Some of the stimuli were chosen from the International Affective Picture System (IAPS), which had been previously reliably rated as having positive, negative or neutral emotive content. Some of the stimuli were created using a 6MP digital camera and were subject to a pilot study to obtain objective ratings of how positively, negatively, or neutrally emotive they were. For continuity purposes, the chosen IAPS pictures were also included in the pilot rating experiment. The pilot study used the same rating scales as those used by Lang, Bradley, and Cuthbert, (1997) – see figure 81. Approximately 100 saliency-controlled pictures were used in the pilot study and 15 participants took part.
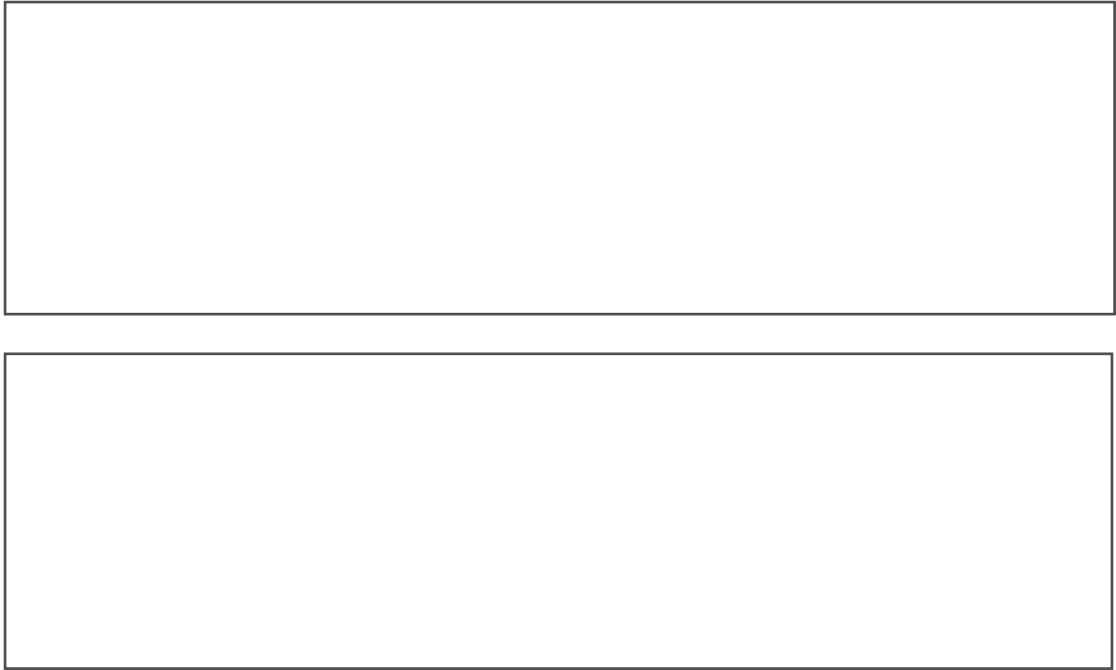
Figure 81: Examples of the happy-unhappy scale (top) and excited-calm scale (bottom) used in this experiment and by Lang et al (1997).

Participants were instructed to look at each picture and rate how happy-unhappy and how excited-calm it made them feel, by putting a cross ('x') in the box that most represented their feelings. It was explained that excitement could be positive or negative excitement, and could include feelings such as anxiety, fear, arousal, and stimulation. If participants wished to make a more finely-tuned rating, then they were allowed to put a cross *between* the boxes.

From the results of the pilot study, 60 of the most unambiguously rated pictures were chosen and categorised as positive, negative, or neutral (20 pictures in each category). Half of each category were designated as "old" and shown in both encoding and test phases, while the other half were labelled "new" and were shown only as fillers at test. Mirror images of the "old" stimuli were produced and were also used as fillers at test. As these were not identical to the original, participants would have to identify them as "new" stimuli.

Pictures were presented on a colour computer monitor at a resolution of 1600 by 1200 pixels. Figures 82a, 82b, and 82c are examples of the stimuli.

Figure 82a: *An example of an emotionally positive stimulus. The emotional area of interest here is the basket of kittens.*
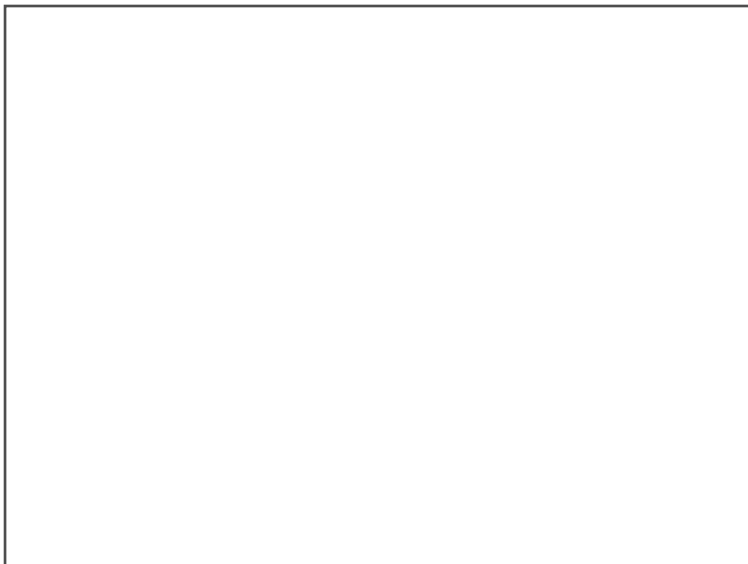
Figure 82b: *An example of an emotionally negative stimulus. The emotional area of interest here is the dead animal on the road.*

Figure 82c: *An example of an emotionally neutral stimulus. There is no emotional area of interest in this picture.*

Design

The experiment used a within groups design. The independent variable was the type of stimulus (positive, negative or neutral) and the dependant variable measures were: accuracy in deciding whether a picture was old or new; average fixation duration, average saccadic amplitude, average number of fixations, and scan pattern analyses (encoding compared to recognition and actual scan patterns compared to those predicted by the saliency model).

Procedure

Following a 9-point calibration procedure, participants took part in a practice phase designed to familiarise them with the equipment, the displays, and the task. The practice consisted of 10 pictures appearing for 3000 milliseconds each and then a recognition memory test with 20 pictures, where participants has to decide whether each one was old or new.

Following the practice phase, the first stage of the experiment began. Participants were not told to look for anything in particular in any of the pictures but were asked to look at them in preparation for a memory test. Thirty stimuli (10 positive, 10 negative and 10 neutral) were presented in a randomised order. Each picture was preceded by a fixation cross, which ensured that fixation at picture onset was in the centre of the screen. Each picture was presented for 3000 milliseconds, during which time participants moved their eyes freely around the screen.

After all 30 stimuli had been presented, participants were given a distracter task to do. They played a computerised version of 'spot the difference' which had levels, timed rounds and a points system. They were told that it was part of the experiment and that their scores would be recorded. After approximately 5 minutes, participants took a break from the game and were given 10 minutes to fill out a recall memory test. They were instructed to write a brief sentence about each of the pictures that they could remember from the first part of the experiment. After this, participants played the distracter game for another 10 minutes. The difficulty levels were set according to how they coped with the game prior to the recall test, so that each participant would play for approximately the same amount of time (before they ran out of 'lives' and the game

automatically quit).

When the game had finished, participants were informed that they were going to see a second set of pictures and had to decide whether each picture was old (from the first set of pictures) or new (never seen before). They were warned that mirror images were counted as new pictures. Ninety stimuli were presented in a random order; 30 of these were old and 60 were new (30 of which were mirror images of old pictures). Participants were not aware of the ratio of old to new pictures. In order to facilitate an ideal comparison between encoding and test phases, each picture was again shown for 3000 milliseconds and participants could only make a response after this time. This was to encourage scanning of the whole picture so that eye movements from the first and second phases of the experiment could be compared. Each picture was preceded by a fixation cross and after each picture disappeared, participants had to respond verbally "old" or "new" and the experimenter entered their response into a separate computer.

The last stage of the experiment was a rating questionnaire. Participants looked through a book of the "old" stimuli and rated each picture on how happy-unhappy and excited-calm it made them feel. The questionnaire used the same rating scales as those used in the pilot study (see figure 2). This was carried out in case any possible noise in the data could be explained by individual differences in emotive ratings of the stimuli.

4.3.3. Results

Two main types of data were collected, memory data (both recognition and recall memory), and eye tracking measures – the average number of fixations in salient and emotive RoIs, the average fixation duration in salient and emotive RoIs, and the position of the first fixation. Trials were excluded where the fixation at picture onset was not central (within the central 2 degrees of visual arc), or when calibration was temporarily interrupted (e.g. if the participant sneezed).

Recognition Accuracy
A repeated measures ANOVA (appendix 48a) was conducted and found a reliable difference in accuracy between the picture types: $F(2,46) = 6.790$, $MSE = 19.022$, $p<0.01$ (see figure 83)

Figure 83: A bar chart to illustrate accuracy at correctly recognising 'old' negative, positive and neutral stimuli at test.

Post-hoc paired samples T-tests (appendices 48b and 48c) revealed a reliable difference in accuracy between negative and neutral stimuli: $t(23) = 3.817$, *SEM* $= 1.20072$, $p<0.01$. There was no reliable difference between positive and neutral stimuli [$t(23) = 1.696$, *SEM* $= 0.98295$, $p = 0.103$] or between negative and positive stimuli [$t(23) = 1.804$, *SEM* $= 1.53216$, $p = 0.070$].

Recall Memory

Written responses from participants during the recall test were analysed and the average number of references (per stimulus) to salient and emotive RoIs was calculated for negative, neutral and positive pictures.

The descriptive statistics indicated more references to salient RoIs for neutral stimuli than for positive or negative stimuli. A repeated measures ANOVA was conducted (appendix 49a) and found a main effect of picture type: $F(2,46) = 742.543$, *MSE* $= 0.010$, $p<0.001$ (see figure 84)
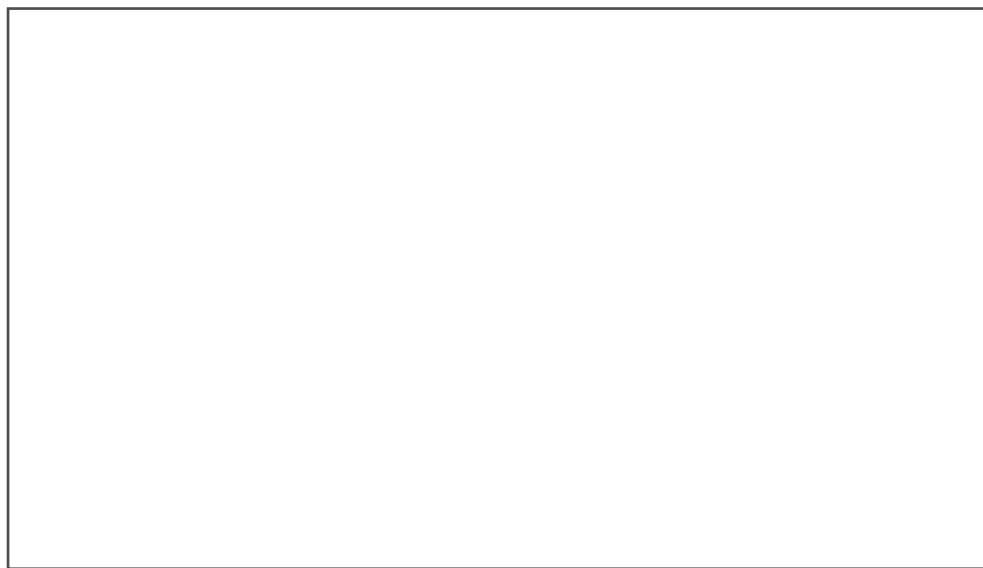
Figure 84: A bar chart to illustrate the average number of references (per picture) to salient Regions of Interest in Negative, Neutral and Positive stimuli.

Post-hoc paired-samples T-tests were carried out (appendices 49b and 49c) and found a reliable difference between negative and neutral stimuli: $t(23) = 27.024$, *SEM* $= 0.03685$, $p<0.001$; and a reliable difference between positive and neutral stimuli: $t(23) = 31.099$, *SEM* $= 0.03108$, $p<0.001$. There was no statistically reliable difference in the number of references to salient RoIs between positive and negative stimuli ($p = 0.09$).

The number of references to emotive RoIs in was also analysed using a paired-samples T-test (appendices 49d and 49e). The results showed that reliably more references were made to emotive RoIs for negative stimuli than positive ones: $t(23) = 3.233$, *SEM* $= 0.03866$, $p<0.05$ (see

figure 85)
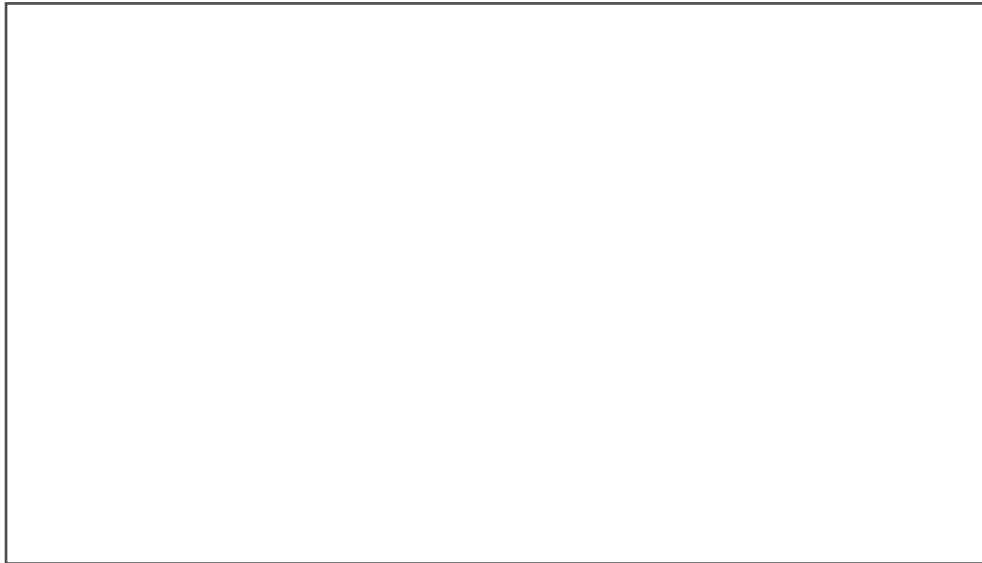
Figure 85: *A bar chart to illustrate the average number of references (per picture) to emotive Regions of Interest in negative and positive stimuli (there were no emotive RoIs in neutral pictures).*

The average number of references to salient RoIs was compared to the average number of references to emotive RoIs, for positive and negative stimuli. Paired-samples T-tests (appendices 49f and 49g) found reliable differences between salient and emotive RoIs for both positive stimuli: $t(23) = 30.753$, *SEM* $= 0.3184$, $p<0.001$; and negative stimuli: $t(23) = 31.978$, *SEM* $= 0.03544$. See figure 86.

Figure 86: *A bar chart to illustrate the average number of references (per picture) to visually salient and emotive Regions of Interest in negative and positive stimuli.*

Average percentage of fixations in RoIs

A repeated measures ANOVA was carried out to compare the percentage of fixations that fell inside the visually salient RoIs (appendix 50a). There was a reliable difference between the picture types: $F(2,46) = 43.063$, $MSE = 2.868$, $p<0.001$. Post-hoc paired samples T-tests (appendix 50c) found that there was a reliable difference between negative and neutral pictures: $t(23) = 8.071$, $SEM = 0.46519$, $p<0.001$; and a reliable difference between positive and neutral pictures: $t(23) = 7.999$, $SEM = 0.51046$, $p<0.001$. There was no reliable difference between negative and positive pictures ($p = 0.509$). From these results and the descriptive statistics (appendix 50b), it can be concluded that reliably less fixations fell in salient RoIs in positive and negative pictures than in neutral pictures. See figure 87.
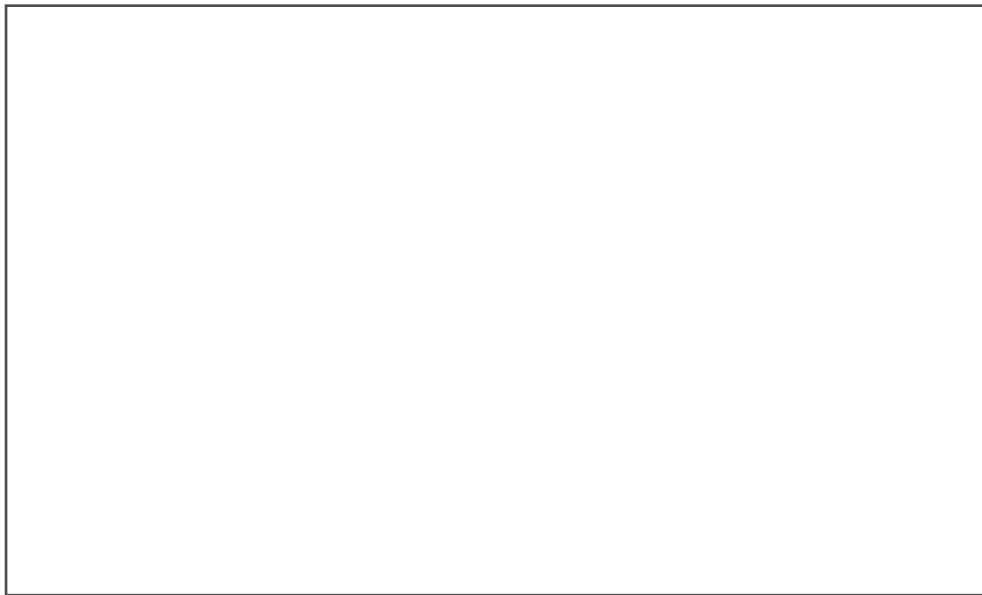
Figure 87: *A bar chart to show the average percentage of fixations that fell in the salient RoI of each picture type.*

A paired samples T-test (appendix 50e) was carried out to compare the percentage of fixations that fell in emotive RoIs on positive and negative pictures. A reliable difference was found: $t(23) = 4.135$, $SEM = 1.009$, $p<0.001$. From these results and the descriptive statistics (appendix 50d), it can be concluded that reliably more fixations fell in emotive RoIs on negative pictures than on positive pictures. See figure 88.

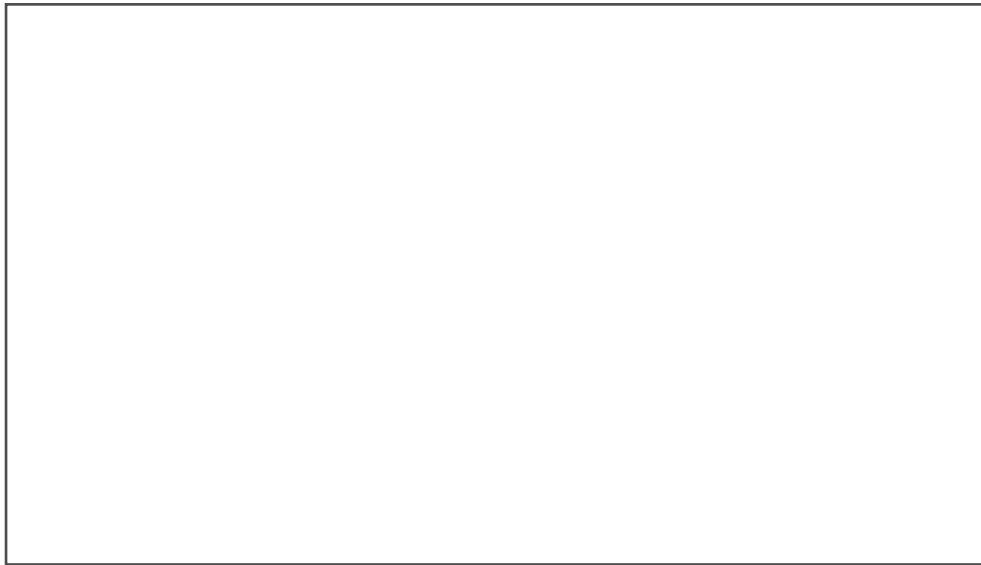*Figure 88: A bar chart to show the average percentage of fixations that fell in emotive RoIs of negative and positive pictures.*

Paired samples T-tests were carried out to compare the percentage of fixations that fell in salient RoIs and emotive RoIs (appendices 50f and 50g). For both negative and positive pictures, reliably more fixations were made in emotive RoIs then in salient RoIs: $t(23) = -8.181$, *SEM* = 1.504, $p<0.001$; and $t(23) = 6.174$, *SEM* = 1.369, $p<0.001$, respectively. See figure 89.

Figure 89: A bar chart to show the percentage of fixations in salient and emotive RoIs for negative and positive pictures.

Average Fixation Duration

A repeated measures ANOVA was conducted (appendix 51a) to compare the average duration of fixations that fell inside salient RoIs. There was a reliable difference between the picture types:

$F(2,46) = 3.762$, *MSE* = 0.053, *p*<0.05. Post-hoc paired samples T-tests (appendix 51c) found a reliable difference between negative and neutral pictures: $t(23) = 4.146$, *SEM* = 0.0438, *p*<0.001. This result and the descriptive statistics (appendix 51b) show that fixation durations in salient RoIs are reliably longer on neutral pictures than negative pictures. There were no reliable differences between negative and positive (*p* = 0.143) or between positive and neutral (*p* = 0.334). See figure 90.
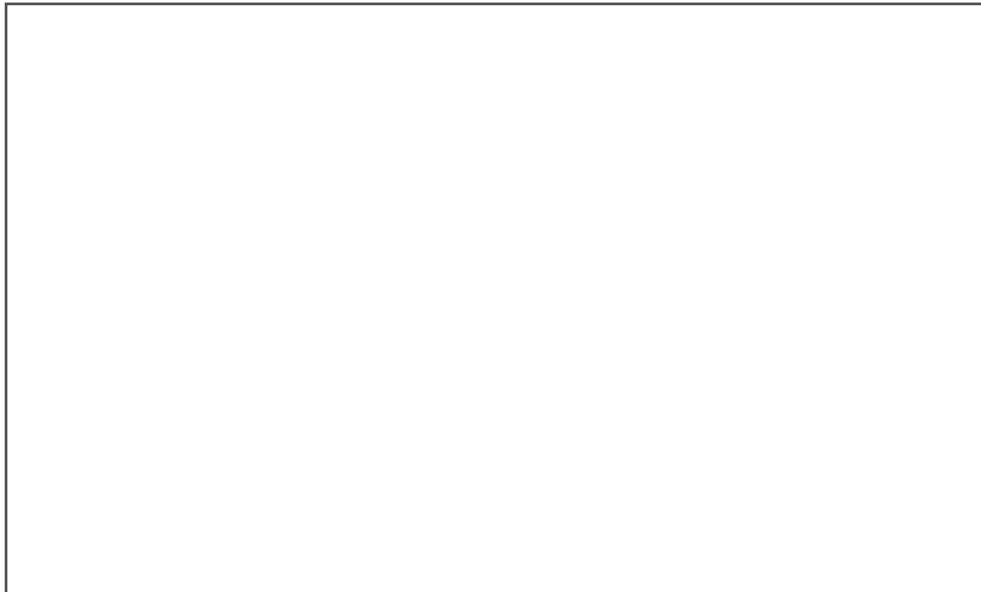
Figure 90: A bar chart to show the average fixation duration in salient RoIs for each picture type.

A paired samples T-test (appendices 51d and 51e) was conducted to compare the average duration of fixations in emotive RoIs of negative and positive pictures. No statistically reliable difference was found (*p* = 0.107).

Paired samples T-tests were carried out to compare the average duration of fixations that fell in salient RoIs and emotive RoIs (appendix 51g). A reliable difference was found between salient and emotive RoIs in negative pictures: $t(23) = -2.792$, *SEM* = 0.03640, *p*<0.05. From this result and the descriptive statistics (appendix 51f), it can be concluded that for negative pictures, fixation duration is reliably longer in emotive RoIs than salient RoIs. No statistically reliable difference was found for positive pictures (*p* = 0.461). See figure 91.
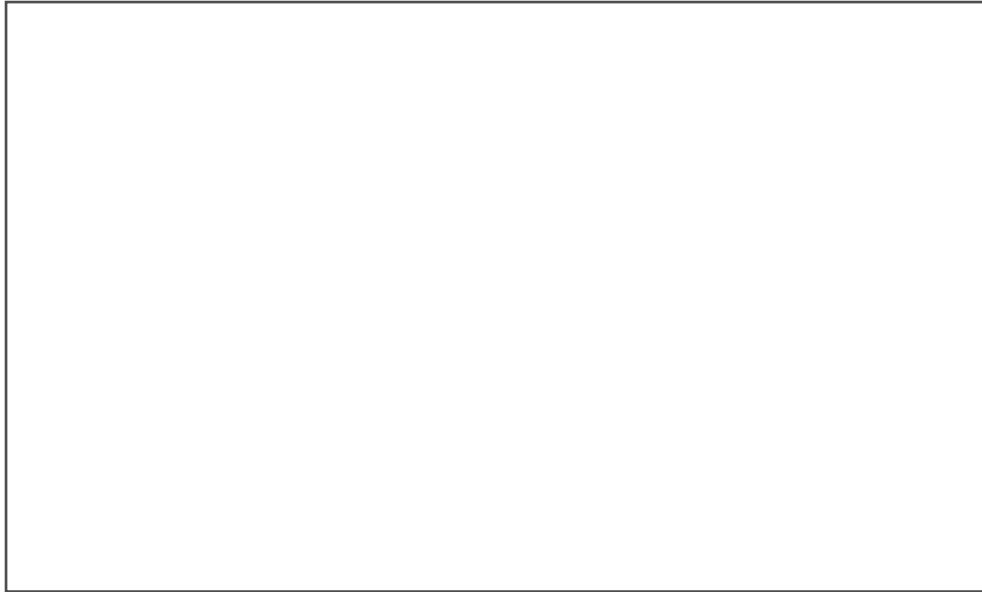
Figure 91: *A bar chart to show the average fixation duration in salient and emotive RoIs for negative and positive pictures.*

First fixation

The position of the first fixation on each stimulus was analysed and the percentage of these that fell inside salient and emotive RoIs was calculated for each picture type.

Figure 92 indicates a greater number of first fixations falling in salient RoIs on neutral pictures than on positive or negative pictures. A repeated measures ANOVA was carried out (appendix 52a) and found a reliable main effect of picture type: $F(2,46) = 4.965$, $MSE = 54.831$, $p<0.05$.
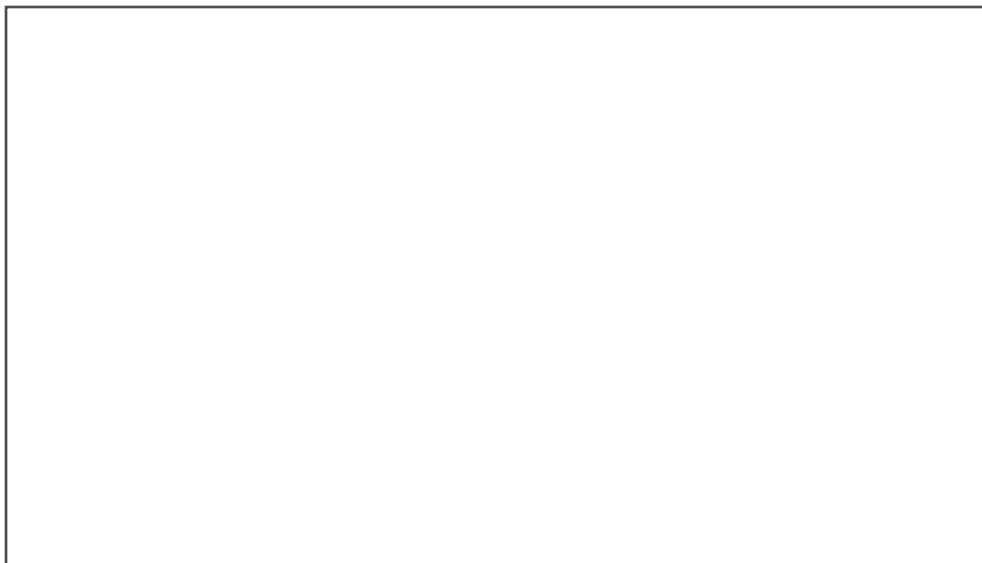
Figure 92: *A bar chart to illustrate the percentage of first fixations that fell in salient RoIs*

Post-hoc paired samples T-tests (appendices 52b and 52c) revealed a reliable difference between

negative and neutral pictures: $t(23) = 3.444$, *SEM* = 1.694, $p<0.01$; and between positive and neutral: $t(23) = 2.429$, *SEM* = 2.401, $p<0.05$. There was no reliable difference in the number of first fixations to positive and negative stimuli ($p = 1$).

A paired samples T-test (appendices 52d and 52e) was carried out to compare the percentage of first fixations that fell in emotive RoIs on positive and negative pictures. No statistically reliable difference was found ($p = 328$).

The average percentage of first fixations to salient RoIs was compared to the average percentage of first fixations to emotive RoIs. Paired-samples T-tests (appendices 52f and 52g) revealed reliable differences for negative stimuli: $t(23) = 5.042$, *SEM* = 3.471, $p<0.001$; and for positive stimuli: $t(23) = 3.091$, *SEM* = 4.853, $p<0.01$ (See figure 93)

Figure 93: *A bar chart to show the average percentage of first fixations in salient and emotive RoIs for negative and positive pictures.*

Scanpaths Similarities

Scanpaths produced when viewing stimuli at encoding were compared to model-predicted saliency scanpaths, based on the Itti & Koch saliency algorithm. This analysis was split to take into account the different stimulus types. Paired samples T-tests were carried out (appendices 53a and 53b) and found that scanpaths when viewing both negative and positive pictures were less similar to saliency scanpaths than when viewing neutral pictures: $t(23) = 2.106$, *SEM* = 0.01085, $p<0.05$; and $t(23) = 3.216$, *SEM* = 0.00750, $p<0.01$, respectively. See figure 94a. Examples of these actual and model-predicted saliency scanpaths on neutral and emotive stimuli are seen in figures 94b and 94c.
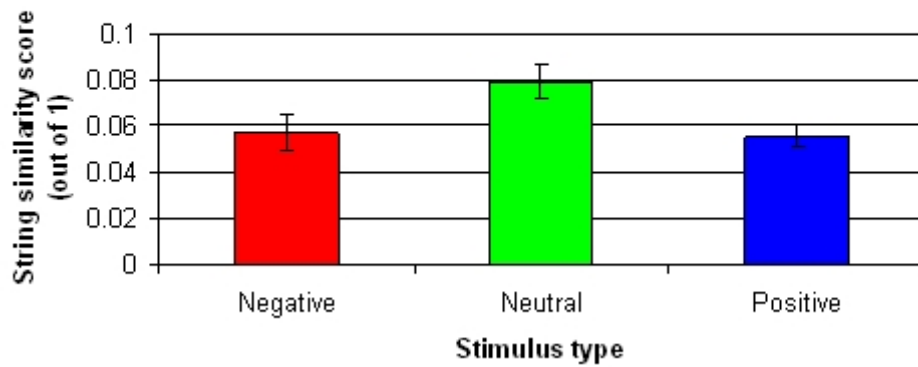
Figure 94a: *Bar chart showing the similarity of scanpaths at encoding compared to model-predicted saliency scanpaths, for each stimulus type.*

Figure 94b: *An example of a neutral picture with the saliency scanpath superimposed in yellow (and red) and an actual participant scanpath superimposed in green.*
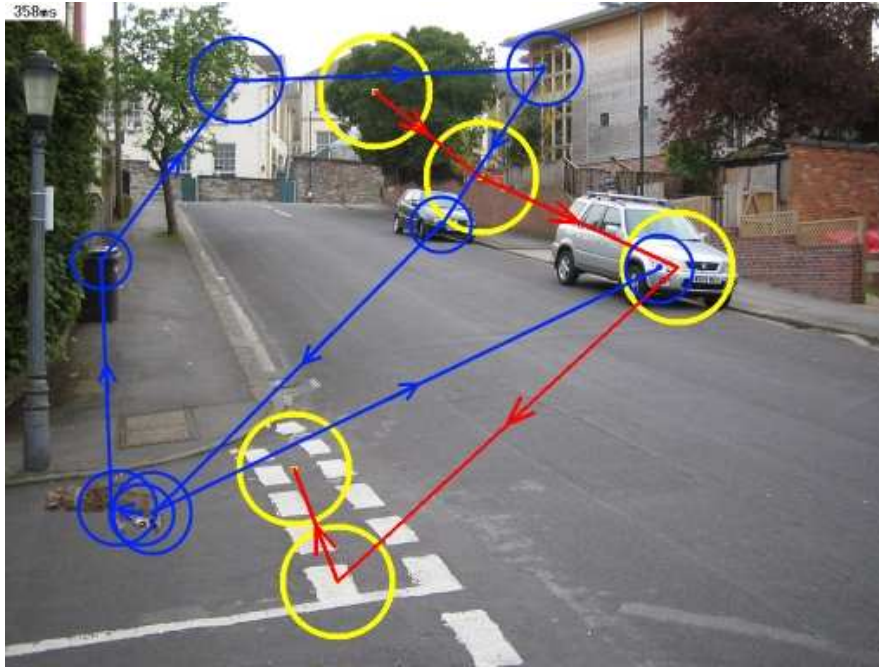
Figure 94c: An example of an emotionally negative picture with the saliency scanpath superimposed in yellow (and red) and an actual participant scanpath superimposed in blue. Note the decreased number of fixations in salient areas compared to figure 94b and also the multiple fixations of the negative RoI.

4.3.4. Discussion

This experiment aimed to determine how eye movements and memory are affected when emotional saliency and low-level visual saliency are in direct competition. Furthermore, it aimed to find out whether evidence for the emotionality or negativity hypothesis exists.

At a recognition test, participants had to decide whether they had seen a picture before or not. The results show a reliable memory advantage for negative pictures, which were remembered more accurately than neutral pictures. Although there was no reliable difference in accuracy between negative and positive pictures, indicating that recognition memory was also high for positive pictures (thus suggesting an emotionality effect), these stimuli did not have the advantage over neutral pictures that negative stimuli did. Overall, the results suggest a negativity effect on recognition memory.

A recall memory test was also carried out; where participants had to write down everything they could remember about each picture. The results show that reliably more references were made to visually salient Regions of Interest on neutral stimuli than on positive or negative stimuli. When an emotive RoI existed, the number of references to visually salient regions significantly decreased. The increase in recall for emotional details at the expense of recall for other details could be interpreted as evidence of attentional narrowing (Easterbrook, 1959). However, this did not affect recognition memory for emotional pictures, as reported in some previous experiments (Loftus et al., 1987; Hope and Wright, 2007). The memory advantage for emotional RoIs was reliably greater for negative stimuli than positive stimuli, suggesting a negativity effect.

The advantage for negative pictures at recognition and recall tests could be due to a tendency for the viewer to 'personalise' the emotive stimuli at encoding, relating them to

memories of similar emotional events that individuals may have experienced in the past (Heuer and Reisberg, 1990). The emotional memory may then serve as a cue at retrieval, (Kensinger & Corkin, 2003b). The advantage of negative over positive stimuli could be due to the infrequency of negative events relative to positive ones and thus negative memories may be more specific and distinct. For example, one may see cute kittens every other day but have only ever been bitten by a spider once, thus making the negative memory more distinct. The negatively emotive stimulus that the participant relates to that memory may therefore be more accurately recognised or recalled at a later time.

The negativity effect shown in the recall memory data is reflected in the greater average percentage of fixations (per picture) to negative RoIs than to positive RoIs and suggests a difficulty in disengaging attention from negative stimuli (Bradley et al, 1997; Miltner et al, 2004; Rinck et al, 2005; Gerdes et al, 2008). It could be that looking at threat is a beneficial strategy aimed at (a) monitoring the fear-relevant stimulus and (b) increasing the probability of appropriate responses in case the threat situation changes (Clark, 1999). This difficulty in disengaging can be related to increased activation of the 'fixate' centre in Findlay and Walker's (1999) model of saccade generation. The increase in activation in the 'fixate' centre outweighs the activation in the 'move' centre and thus the threshold needed to saccade is not surpassed.

The increased number of references to visually salient RoIs in neutral pictures is accompanied by a reliably higher percentage of fixations in visually salient RoIs for neutral pictures than for positive or negative pictures. When emotive RoIs were present, the increased number of references to these areas was accompanied by a reliable increase in the percentage of fixations to emotive RoIs at the expense of fixations to salient RoIs. The attentional capture by these emotive areas could be interpreted as a cognitive override of saliency, similar to that seen with domain expertise. This is supported by the differences in scanpath similarities between picture types. Scanpaths produced when viewing neutral pictures were reliably more similar to model-predicted saliency scanpaths than when viewing negative or positive pictures, suggesting that the presence of emotional RoIs attracts attention more than the visually salient Regions.

Furthermore, the average fixation duration in salient RoIs was reliably decreased when a negative RoI was present. This suggests attentional capture by the negative feature, which overrides the effect of visual saliency. However, the lack of reliable difference in fixation durations between neutral and positive stimuli suggests that the presence of positive RoIs is not strong enough to produce this overriding effect. This partially supports previous findings of increased number of fixations to emotional pictures (Nummenmaa et al, 2006) but suggests that it is specific to negative stimuli. These results also confirm that previous negativity effects found in depressed individuals (Eizenman et al, 2003; Caseras, 2007) can be replicated in participants with no psychological abnormalities. No reliable difference in fixation duration was found between emotionally positive and negative RoIs, which reinforces the argument that the increased number of fixations to negative RoIs is due to difficulty in disengaging attention. If participants made longer fixations to positive RoIs, this would explain why they only had time to make a smaller number of fixations. However, this was not the case.

Many studies have found that difficulty in disengaging is not necessarily accompanied by an initial orienting bias to negative stimuli (Calvo and Lang, 2004; Nummenmaa et al, 2006). Analysis of the position of first fixation supports this, in that no advantage was found for negative RoIs over positive RoIs. There was, however, an emotionality effect, whereby the presence of an emotive RoI (positive or negative) decreased the likelihood of first fixation to a visually salient RoI. This suggests an automatic processing of emotional stimuli, irrespective of whether they hold

positive or negative valence. An advantage of the current study over Calvo and Lang's is that only one picture was present therefore attention was not divided. The comparison of two simultaneous pictures in Calvo and Lang's experiment may have hindered a more spontaneous allocation of attention after initial orienting to one of the pictures. Another advantage is that each picture in the current study was predominantly neutral but with an emotional region of interest (rather than an entirely emotionally positive or negative scene), therefore the 'gist' of the scene did not influence eye movements. Furthermore, no emotive RoIs defied the gist of the scene, which is important because inconsistencies have been found to attract attention, even in the parafoveal (Loftus and Mackworth, 1978).

Other methodological advantages of the current study include the use of a wide variety of emotional stimuli, and the recruitment of non-selected participants (i.e. not all socially anxious patients or Schizophrenics), therefore making the results more generalisable. It has previously been suggested that informing participants (for ethical reasons) that they will be presented with emotional pictures may bias them to deliberately search for emotive RoIs (Calvo and Lang, 2005). However, emotive stimuli in the current study were not predictable. Firstly, positive, negative and neutral stimuli were presented in a random order and secondly the RoIs were in unique locations in every picture. Negative, positive and neutral stimuli were also (on average) equally distanced from the central fixation point.

The results of this study show that low-level visual saliency does have an effect on eye movements when inspecting a visual scene, but this can be overridden by 'emotional saliency' (especially negative). This cognitive override has also been found with domain specific knowledge. Future research could investigate which of these top-down variables (emotion or expertise) is most dominant, by presenting them both in a visual scene and analysing which has the biggest effect on eye movements.

In summary, attentional narrowing for negative regions of interest was demonstrated in the recall memory test. A negativity effect was found for recognition and recall memory, the percentage of fixations in RoIs and the average fixation duration in RoIs. In contrast, an emotionality effect was found for the first fixation, implying that the processing of emotional features occurs at a very early stage of perception (even when they are still in the parafoveal), but once fixated only emotionally negative features hold attention. It is suggested that participants have difficulty in disengaging attention from negative stimuli, which could be a defence mechanism to fear- or threat-related material. Lastly, the effect of negative RoIs was so robust that it reliably decreased the effect of visual saliency on eye movements.

4.4. Chapter 4 Conclusions

This chapter investigated the effects of the presence of people and emotional regions in pictures, on eye movements and memory. The presence of people in a picture was found to reliably increase recognition memory. A possible explanation for this is that people in the scene could act as socially informative items, which viewers use to form conclusions such as 'picking bananas' or driving the tractor' that aid memory at recognition (Henderson et al, 1999; Birmingham et al, 2008). The presence of people also attracts overt attention, indicated by the longer and more frequent fixations in 'people' RoIs than would be expected by chance. Saccadic amplitude analyses suggest that if participants ignore this social information but instead search more widely,

they are more likely to incorrectly identify the picture. When no social information is present, participants have to search more widely to find familiar objects/areas and if they fail to do this, then they are more likely to incorrectly identify the stimuli.

The similarity of scanpaths at encoding and recognition is consistent with previous chapters but the lack of relationship between string similarity and accuracy challenges the idea that the reproduction of eye movements alone is enough to create a memory advantage. People pictures could have been easier to recognise because participants formed conclusions about the semantic content/gist, assigning mental labels to these pictures. Therefore, even when eye movements were not perfectly reproduced at recognition, people pictures were still easier to identify.

Both recognition and recall memory were also found to be enhanced by the presence of a negatively emotional RoI, the effect of which was so pronounced that it overrode the influence of visual saliency on eye movements. Participants fixated on and made verbal references to areas of high visual saliency in neutral pictures, but this was reliably reduced when emotional regions were present. When viewing negative and positive stimuli, participants made more references to and reliably longer and more frequent fixations to emotional RoIs than salient areas, demonstrating a cognitive override of saliency similar to that seen with domain expertise. This effect was more prominent for negative regions, whereby the average fixation duration in salient RoIs was reliably decreased when a negative RoI was present, but not a positive RoI, suggesting that the presence of positive RoIs is not strong enough to produce this overriding effect.

The 'negativity effect' of recall memory was mirrored in the greater average percentage of fixations (per picture) to negative RoIs than to positive RoIs and suggests a difficulty in disengaging attention from negative stimuli (Bradley et al, 1997; Miltner et al, 2004; Rinck et al, 2005; Gerdes et al, 2008). However, in support of Calvo and Lang (2004) and Nummenmaa et al (2006), this difficulty in disengaging was not accompanied by an initial orienting bias to negative stimuli. There was, however, an 'emotionality effect', whereby the presence of an emotive RoI (positive or negative) decreased the likelihood of first fixation to a visually salient RoI. Combined, these results suggest that the processing of emotional features occurs at a very early stage of perception (even when they are still in the parafoveal), but once fixated, only emotionally negative features hold attention.

In summary, both the presence of people and emotional regions in pictures can enhance recognition memory and attract attention (thus influencing eye movements). This is more evident for negative than positive emotional stimuli.

Chapter 5 - The importance of the order of fixation

Experiment 10: Fixation sequence v semantic content v saliency

5.1 Introduction
Scanpath theory (Noton and Stark, 1971) emphasises the importance of the order of a fixation sequence and that moving our eyes in a similar pattern at recognition as when first encoding the picture, helps to identify that picture. An interesting example of how eye movement sequences and memory are linked is demonstrated by Brockmole and Henderson (2006)b. They found that

when previously-seen pictures were mirrored, participants initially partly reproduced the scanpath they made to the original picture, even though the target position had now changed.

However, as seen in previous chapters and in the literature, other factors also influence our eye movements and memory for visual scenes such as low-level visual saliency, domain expertise, the presence of people and emotional regions of interest. Therefore, to directly test how important the order of fixation actually is, fixation sequence at encoding and test must be manipulated. If scanpath theory holds true, then changing this pattern of eye movements should hinder identification of the picture. By forcing participants to make certain eye movement patterns at encoding and then manipulating the order of the fixation pattern at recognition, differences in accuracy and reaction time can help us separate the underlying theories of memory and eye movements. For example, if the order of fixation is paramount (as suggested by Scanpath Theory), then reaction times should decrease at recognition when stimuli are fixated in the same order as at encoding, and if the order of fixation is reversed, reaction times should increase. If, however, the content of the scene is most important, then the order of fixation at recognition should not make a reliable difference. The content of a scene could be memorable due to semantically interesting features, for example Mackworth and Morandi (1967) found that the density of fixations on a picture was linearly related to the semantic informativeness of the fixated region. The content of a scene could also be memorable due to visually salient properties; therefore measures of both visual saliency and semantic information will be recorded. Visually salient properties can be objectively measured using the Itti & Koch (2000) computer algorithm. Identifying semantically interesting properties, on the other hand, is more subjective, and therefore individual participant ratings will be taken and then compared to the corresponding accuracy and reaction time data to define whether a relationship is present.

A memory advantage for objects fixated later in a scanpath (recency effect) has been found in previous research (Wyer and Srull, 1986; Wedel and Pieters, 2000; Irwin and Zelinski, 2002; Hollingworth, 2004; Zelinski and Loschky, 2005; Tatler, Gilchrist and Land, 2005). A control condition will therefore be employed in the current study, using a random fixation order at recognition. This should still highlight any change in reaction time due to a change in the order of fixation compared to encoding, but will also control for recency effects.

Finally, the effect of 'gist' needs to be considered. The importance of fixation sequence may be overshadowed by participants recognising the picture due to the whole scene gist, which can occur from a single brief exposure (Rayner, 1998) and even before the eyes begin to move (De Graef, 2005). Target pictures during rapid serial visual presentation are identified even when subsequent recognition memory is poor (Potter, 1976; Potter and Levy, 1969; Potter et al., 2002), and when the gist of the scene is disturbed, target objects in individually presented pictures are recognised with more difficulty (Biederman, 1972; Biederman et al., 1974). Therefore, to decrease participants' reliance on gist, parafoveal information in the current experiment is minimised by showing one square of the picture at a time, with some parts of the picture missing. This also allows the order and sequence of fixation to be controlled and manipulated.

5.2 Methodology

Participants

Sixteen participants took part in the experiments, all of whom were students at Nottingham University. The age range was 18-24 and the mean age was 21. The sample comprised 10 females

and 6 males. All participants had normal or corrected-to-normal vision.

An Eprime experiment was run on a PC with a colour computer monitor measuring 43.5cm by 32.5cm. Stimuli were presented at a resolution of 1600 by 1200 pixels and participants sat at a fixed viewing distance of 98cm, giving an image that subtended 25.03 by 18.83 degrees of visual angle. No eye tracking was recorded in this experiment.

A set of 60 high resolution digital photographs were prepared as stimuli, sourced from a commercially available CD-ROM collection and taken using an 8MP digital camera. Half the stimuli were designated "old" and shown in both encoding and recognition, while the other half were labelled "new" and were shown only as fillers during the recognition test. Stimuli were photographs of landscapes and agricultural scenes (e.g. farm houses) that had been divided into 25 equally sized squares. Each stimulus revealed itself square by square, building up the picture. Some parts of this picture were always missing – this design aimed to decrease the participants' reliance on the gist of the whole picture or the use of parafoveal vision, and instead encouraged them to make their decision on a square-by-square basis. Examples of the stimuli can be seen in figures 95a and 95b.
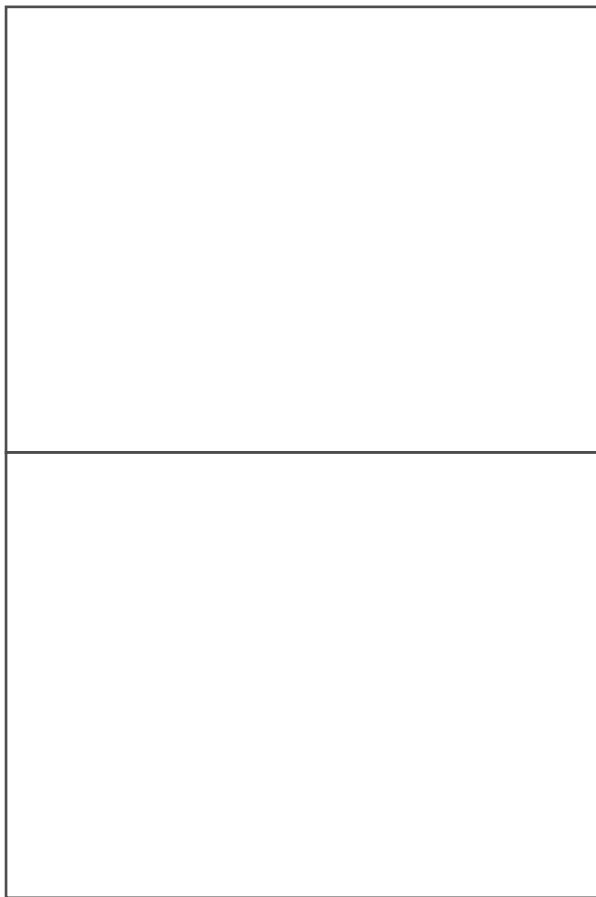
Figure 95a (left): An example of one square of the stimulus. Figure 95b (right): An example of all the squares of a stimulus present.

Design

The experiment used a within groups design. The Independent Variable was order of presentation of the squares at recognition, and had 3 levels (same as encoding, reverse order, or random). The Dependant Variable measures were: Accuracy at recognition test, reaction time at recognition, and participants' ratings for semantically interesting features.

Procedure

Participants were shown written instructions on the experimental procedure and given a short practice prior to starting the experiment. Feedback was given in the practice, but not in the real thing. The first stage of the experiment was the 'encoding' stage, where participants were shown the stimuli and instructed to remember as much as possible. Each picture revealed itself one square at a time, each square appearing for 800 milliseconds, with a total of 12 squares, resulting in each stimulus being present for a total of 9600 milliseconds. The order that the squares of each stimulus appeared was pre-determined using a random number generator. Each stimulus was preceded by a 1000 millisecond fixation cross to ensure that gaze was central at picture onset, and the first square of any stimulus was never the central square. Participants were told that it was important to move their eyes to look at each square as it appeared. Once all 30 pictures had been viewed, the participants took a short break before starting the recognition memory test. In this second stage, pictures appeared in the same fashion as before and participants had to make a button press as soon as they could confidently identify the picture as old or new. If they believed the picture to be old (i.e. from the previous set) the 'O' button was pressed, and if they believed the picture to be new (i.e. never seen before) the 'N' button was pressed. That stimulus sequence was then terminated and a new picture started to reveal itself square by square. Participants were instructed to be as fast and as accurate as possible.

When participants had finished the recognition memory test, they were asked to complete a short computerised questionnaire that presented some of the pictures in the experiment. All the squares were present and a grid was superimposed over the top, splitting the picture into squares, labelled with the letters a-y. Participants were asked to choose which square was most semantically interesting and responded by pressing the corresponding letter on the keyboard. An example of these pictures is shown in figure 96:

Figure 96: *An example of the stimuli used in the post-experiment questionnaire.*

5.3 Results

Accuracy
Accuracy at identifying pictures as 'old' at recognition was calculated for pictures shown in the same, reverse and random orders. See figure 97. Paired-samples T-tests (appendices 54a and 54b) found that there were reliable differences in accuracy between: 'same' and 'reverse': $t(15) = 4.521$, $SEM = 4.45470$, $p<0.001$; and between 'reverse' and 'random': $t(15) = 4.024$, $SEM = 4.16797$, $p<0.01$. There was no reliable difference in accuracy between 'same' and 'random': $t(15) = 1.004$, $SEM = 3.35428$, $p = 0.331$.

Figure 97: *A bar chart illustrating accuracy at identifying stimuli as 'old' at recognition.*

Reaction Time
Reaction time to identify pictures as 'old' at recognition was calculated for pictures shown in the same, reverse and random orders. See figure 98. Paired-samples T-tests (appendices 55a and 55b) found that there was a reliable difference in reaction time between: 'same' and 'reverse': $t(15) = 2.503$, $SEM = 207.22561$, $p<0.05$. There were no reliable differences in reaction time between 'same' and 'random': $t(15) = 0.061$, $SEM = 207.90596$, $p = 0.952$; or between 'reverse' and 'random': $t(15) = 1.923$, $SEM = 276.20701$, $p<0.074$.
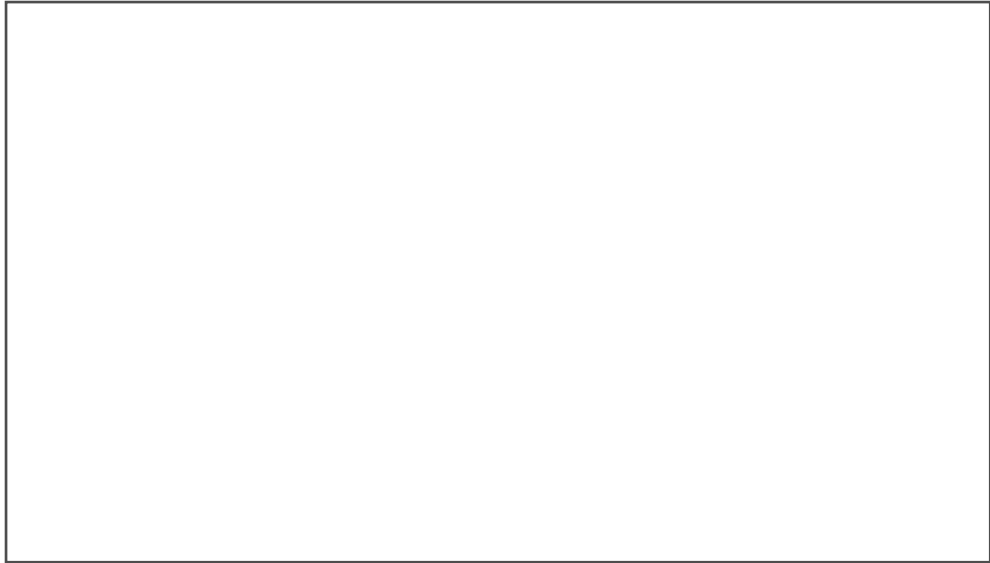
Figure 98: A bar chart illustrating reaction time to identify stimuli as 'old' at recognition.

Visually Salient Features

The position in the fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average accuracy for that picture (see appendices 56a, 56b and 56c). There was a reliable positive correlation between accuracy in identifying pictures at recognition and the position of the visually salient feature for 'same' order sequences: $r(10) = 0.687$, $p<0.05$. See figure 99. There were no reliable correlations for 'reverse' or 'random' sequences.
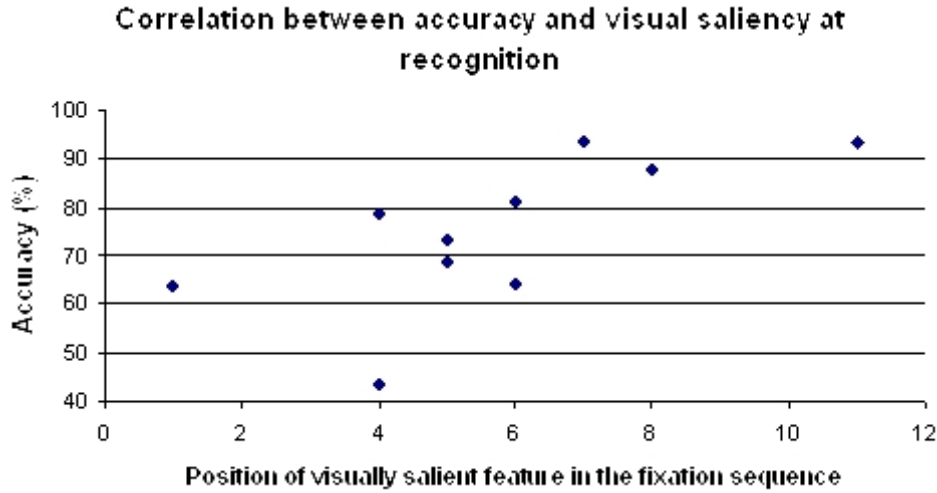
Figure 99: A scatter graph to illustrate the correlation between accuracy and the position of the salient feature when the stimuli were presented in the same order at recognition.

The position in the fixation sequence of the most visually salient feature was also correlated with the average reaction time for that picture (appendices 56d, 56e and 56f). There was a reliable negative correlation between reaction time to identify pictures at recognition and the position of

the visually salient feature for 'same' order sequences: $r(10) = 0.740$, $p<0.05$. See figure 100. There were no reliable correlations for 'reverse' or 'random' sequences.

Correlation between reaction time and visual saliency at recognition
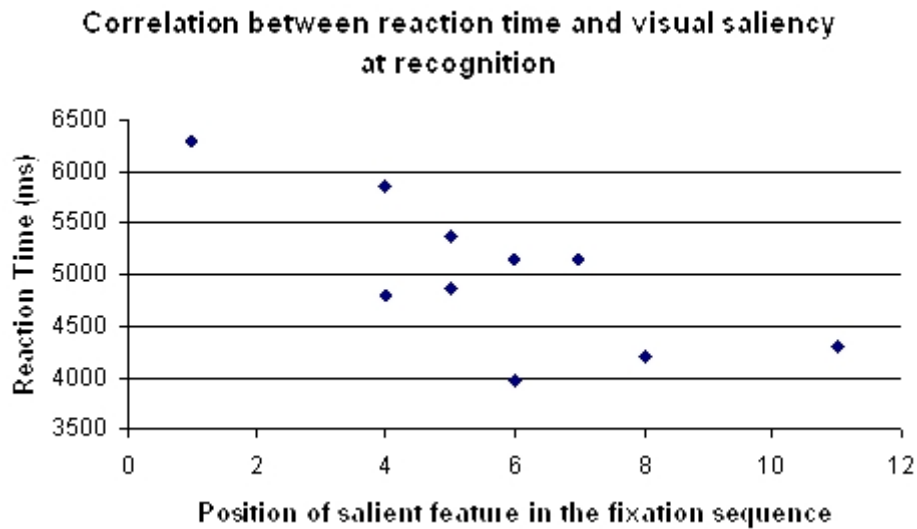
Figure 100: A scatter graph to illustrate the correlation between reaction time and the position of the salient feature when the stimuli were presented in the same order at recognition.

Semantically Interesting Features

The position in the fixation sequence of the most semantically interesting feature identified by each individual participant was correlated with that individual's corresponding accuracy for that picture (appendices 57a, 57b and 57c). There was a reliable negative correlation between reaction time to identify pictures at recognition and the position of the semantically interesting feature for 'reverse' order sequences: $r(10) = 0.747$, $p<0.01$. See figure 100. There were no reliable correlations for 'same' or 'random' sequences.

A corellation between accuracy and the position of the semantically interesting feature
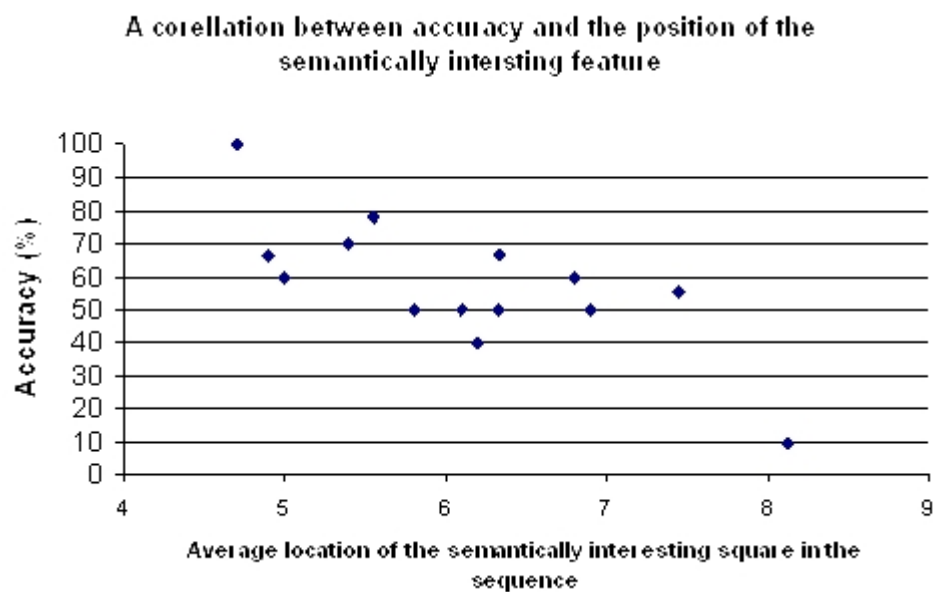
Figure 101: *A scatter graph to illustrate the correlation between accuracy and the position of the semantically interesting feature when the stimuli were presented in the reverse order at recognition.*

The position in the fixation sequence of the semantically interesting features was also correlated with the average reaction time for that picture (appendices 57d, 57e and 57f). There were no reliable correlations.

5.4. Discussion

This experiment aimed to test the importance of fixation order on recognition memory of visual scenes. It was predicted that if the fixation sequence was in the same order at encoding and recognition then recognition memory would be most accurate and reaction time to respond would be fastest.

Results showed that participants were reliably more accurate at recognising 'old' pictures if the fixation sequence was the same (at recognition and encoding) than if it was reversed. In fact, reversing the order of the sequence produced the lowest accuracy, even lower than the control condition (random order), suggesting that the order of fixation at encoding and recognition is important to memory. However, this theory is questioned by the lack of reliable difference in accuracy between 'same' and 'random' conditions. It could be that the fixation sequence in the random condition is more similar to the original order than the reverse condition and as long as the sequence is similar, recognition is facilitated. The less similar the fixation sequence, the worse recognition accuracy is.

Analyses of reaction time revealed an opposite effect to that predicted, in that participants took longer to respond in the 'same' condition than the 'reverse' condition. It could be that a certain amount of the scanpath has to be reproduced before an accurate decision can be made, thus explaining the increased reaction time for 'same' sequences. The shorter reaction time for reverse sequences may be due to a speed-accuracy trade-off, leading to the decrease in accuracy described above.

The increased accuracy for 'same' and 'random' sequences compared to 'reverse' sequences and the shorter RT of 'random' compared 'reverse' sequences opposes the idea of a recency effect purely based the most recently fixated squares. For example, if such a recency effect existed, accuracy when the fixation sequence is reversed would be greater than if it were the same or randomly ordered. However, it may not just be the order of fixation that influences accuracy and reaction time. The positions in the sequence of the most visually salient square and the most semantically interesting square were calculated for each participant and then correlated with accuracy and reaction time. There was a reliable positive correlation between saliency and accuracy for 'same' sequences, and a reliable negative correlation between reaction time and saliency for 'same' sequences. This implies that the closer to the end of the sequence the most salient feature is, the faster and more accurate the participants are. This increased memory for recently fixated objects suggests a recency effect sensitive to saliency. However, if this was true, one would expect increased accuracy and decreased RT for the reverse sequence, where the recently fixated salient object at encoding appears sooner at recognition. It could be that the fixation sequence is important and is linked to the saliency map, but salient features can only play a role after a certain amount of the picture has been fixated (and thus after a certain amount of the

fixation sequence has been reproduced), due to the similarity between pictures and the lack of parafoveal information.

There was also a negative correlation between semantically interesting features and accuracy for the reverse order sequence, implying that the nearer to the beginning of the sequence the semantically interesting feature is, the more accurate the participants. This could be evidence for a 'semantically-sensitive' recency effect, as the semantically interesting features closer to the beginning of the reverse sequence appeared near to the end of the sequence at encoding. This coincides with previous findings of a reliable memory advantage for more recently fixated objects (Irwin and Zelinski, 2002) and a reliable decrease in performance as the number of intervening fixations between the last fixation of the object and the end of the trial increased (Tatler, Gilchrist and Land, 2005). It is fascinating though that this is the case for semantically interesting features but not for visually salient features. Usually, visual saliency attracts attention, initiating the next saccade. However, due to is limited parafoveal information in the current experiment, the salient regions of the picture may not be visible, and thus more of the pre-determined fixation sequence has to be followed before either a visually salient or semantically interesting feature appear and facilitate recognition. Saliency may therefore aid recognition memory but is denied the role of actively guiding eye movements in this experiment. This bottom-up process might also be overridden by semantic features that could hold 'top down' significance, similar to the overriding effect of expertise on saliency (see chapter 4). It is predicted that in real world scenes were parafoveal information is not restricted, saliency and semantic relevance play a more significant role in attracting attention and guiding eye movements, and to an extent decrease the importance if the fixation sequence.

This experiment could be modified by showing a full scene at encoding and recording eye movements. Then, the eye dats would be analysed to see which parts of the picture were fixated early in viewing and which were fiaxated later. Some of the stimuli would be replaced with pre-prepared edited stimuli, some of which contain a change/edit to an object/area fixated early in the scanpath sequence and some contain a change/edit to an object/area fixated late in the scanpath sequence. Edits to the pre-prepared stimuli could be predicted by changing the most salient thing (should be fixated early on) and the least salient thing (should be fixated later in the scanpath). The participant would then return later that day and have to decide whether the picture has been changed, and then choose which quadrant of the picture the change has occurred in. If the sequence and order of fixation is important to recognising pictures at a later point, then it would be predicted that changes to RoIs fixated early on at encoding should be identified reliably quicker than RoIs fixated later at encoding, due to the similarity in scanpaths over multiple viewings.

In conclusion, the order of fixation is important and the less similar the patterns of eye movements are at encoding and recognition, the less accurate recognition memory is. It is suggested that due to the lack of parafoveal information (prohibiting recognition based on the gist of the scene) and thus the difficult nature of the task, a certain amount of the fixation sequence had to be followed before other factors such as saliency and semantics could enhance recognition memory. No reliable evidence was found for recency effects based purely on the most recently fixated squares, but the position of the most visually salient and most semantically interesting features did affect accuracy. So, the fixation sequence *is* important for recognition memory, but the influence of visually salient and semantically relevant parafoveal cues in real-world scenes may decrease the necessity to fixate in the same order. Further research could include a replication if this experiment but using eye tracking, to see whether participants did follow the predetermined sequence and whether they made refixations to previous squares.

Chapter 6 – Main Discussion

This chapter summarises and discusses the main findings of the experiments in this Thesis in relation to the theories and possible applications mentioned in Chapter 1. Limitations of methodolgies and the extent to which conclusions can be drawn are also discussed. The following table provides an overview of the main results from each experiment.

Summary of results

Exp	Chapter	Brief overview of experimental findings
1	2.2	Highest scanpath similarities when the encoding and retrieval processes are most similar The similarity of scanpaths at encoding and imagery show that eye movement sequences can be, to an extent, replicated at a later time and argues against the fixation selection being based on low level factors. The re-sizing and reorganising of mental images is suggested to occur during imagery. The high scanpath similarity scores when comparing imagery and delayed imagery suggests that much less resizing happens once the mental images have been formed and that these scanpaths are relatively stable over time.
2	2.3	Reinforced findings from experiment 1 Participants fixated outside the picture boundaries during imagery, despite the presence of a distinct border.
3	2.4	Increased scanpath similarity to encoding for pictures that were described post-presentation compared to those that were only imagined
4	3.2	Recognition memory advantage for specialists when viewing domain-specific stimuli Low-level visual saliency does influence eye movements, at both encoding and recognition Cognitive override of saliency by domain proficiency: decreased percentage of fixations to salient areas and decreased similarity between actual and model-predicted saliency scanpaths when viewing domain-specific stimuli Evidence for the replication of scanpaths over multiple viewings
5	3.3	Reinforced findings from experiment 4 Evidence for the replication of scanpaths even after a substantial time delay (1 week) between encoding and recognition.
6	3.4	Recognition accuracy of complex visual pictures can be increased by post-stimuli verbalisation. Specialists have a recognition memory advantage for stimuli that contain domain-relevant information. Participants make verbal reference to salient regions of previously seen pictures, but this is reduced for specialists when semantically informative domain-specific regions of interest were also previously present. Evidence of cognitive override of saliency in the number of fixations, total fixation durations and first fixations to salient and specialists regions of interest.

		It is suggested that memory does not play a significant role in guiding eye movements to specialist RoIs, but rather attention is drawn automatically due to the underlying semantic relevance.
		Describing a complex picture from memory produces more similar eye movements to encoding than are produced when simply imagining it.
7	3.5	Incongruous regions of a scene attract attention, and do so early on in perception, even before the first fixation
		Participants seem to have a difficulty in disengaging, once fixated on incongruous regions
		Domain proficiency acts as a cognitive override of this incongruence effect, seen in the decreased number of fixations to incongruous regions and the decreased percentage of first fixations to incongruous regions when viewing domain-specific stimuli.
8	4.2	The presence of people in natural scenes increases recognition accuracy regardless of the size of the RoI, its distance from initial fixation or the number of people.
		Old pictures are harder to correctly identify and produce larger fixation durations.
		When people were present in the scene, increased saccadic amplitude was related to a reliable decrease in accuracy, possibly due to participants ignoring important social cues.
		Scanpath analyses showed a high similarity between encoding and recognition, but the lack of relationship between string similarity and accuracy challenges the idea that the reproduction of eye movements alone is enough to create a memory advantage.
9	4.3	Negativity effect for recognition and recall memory, the percentage of fixations in RoIs and the average fixation duration in RoIs.
		Emotionality effect location of first fixation
		It is suggested that the processing of emotional features occurs at a very early stage of perception (even when they are still in the parafoveal), but once fixated only emotionally negative features hold attention.
		Participants seem to have difficulty in disengaging attention from negative stimuli
		The presence of negative RoIs reliably decreased the effect of visual saliency on eye movements.
10	5	The less similar the patterns of eye movements are at encoding and recognition, the less accurate recognition memory is.
		The influence of saliency in guiding eye movements was limited in this experiment due to the lack of parafoveal information.
		No reliable evidence was found for recency effects based purely on the most recently fixated squares, but the position of the most visually salient and most semantically interesting features did affect accuracy.

6.1. Relating the experimental results to underlying theories of eye movements

This Thesis has presented 10 experiments testing theories of eye movement control and the top-down and bottom-up factors that can influence eye movements. The reproduction of scanpaths over experimental sessions and after extended periods of time (24 hours and 1 week) suggests that these sequences of eye movements are relatively stable over time. These findings support previous research that has found high similarity between eye movements at encoding and recognition (Walker-Smith, 1977; Stark and Ellis, 1981; Foulsham and Underwood, 2008). Results from chapter 5 indicate that the sequence of eye movements is important for scene recognition when limited peripheral information is available.

The question then remains, what guides eye movements in the first place? Scanpath Theory offers an explanation for how eye movements and memory may be linked but doesn't specify why a person initially chooses to fixate on one region over another. A bottom-up account is put forward by Koch and Ullman (1985) and Itti and Koch (2000) who proposed that eye movements are guided by low-level visual saliency. Their 'Saliency Map' model argues that eye fixations are attracted to regions of greatest brightness, colour, contrast and orientation change, in a winner-takes-all situation. Experiment 10 (chapter 5) found that although the lack of peripheral information limited the role of saliency in guiding eye movements, the position of the salient object in the forced fixation sequence still affected recognition memory. Experiments in chapters 3 and 4 show that fixations do fall in areas of high visual saliency and that scanpaths are reliably more similar to saliency-model-produced scanpaths than would be expected by chance. However, these same experiments have shown that the influence of saliency on eye movements is reduced by top-down factors such as expertise and emotive regions of interest. The Saliency Map theory does not account for this.

An alternative to the Saliency Map is the Area Activation Model (Pomplun, Reingold, Shen, & Williams, 2000) is a computational model that predicts the statistical distribution of saccadic endpoints in visual search tasks. It proposes that fixations are drawn towards areas of task-relevant information and like 'saliency peaks' in the Saliency Map model, 'activation peaks' are task-relevant areas and are therefore more likely to be fixated. However, this model only applies to visual search, and also fails to account for the role of memory (see Findlay, Brown, & Gilchrist, 2001; Frens, van der Geest, & Hooge, 2000; Gilchrist, North, & Hood, 2001; Horowitz & Wolfe, 1998, 2001, 2003; Peterson, Kramer, Wang, Irwin, & McCarley, 2000) consequently leaving it unable to explain refixations on the same items.

Another model of eye movements that takes into consideration top-down knowledge is the cognitive relevance hypothesis. This proposes that fixation sites are selected based on the needs of the cognitive system in relation to an understanding of scene meaning (i.e., based on cognitive knowledge structures in memory) interacting with the goal of the current task (Henderson, 2003, 2007; Henderson, Brockmole, Castelhano, & Mack, 2007; Henderson & Ferreira, 2004; Navalpakkam & Itti, 2005; Torralba, Oliva, Castelhano, & Henderson, 2006; see also Land & Hayhoe, 2001; Foulsham & Underwood, 2007). This theory still acknowledges the role of saliency in guiding eye movements (be it a reduced one) but puts emphasis on the role of cognitive knowledge. However, it does not offer a full explanation as to how these two processes (top-down and bottom-up) interact in situations where competing influences are present.

The more fitting model to explain interactions between bottom-up and top-down influences is Findlay and Walker's (1999) model of saccade generation. At level 2, low-level visual saliency plays a role in guiding eye movements, but can be overridden by top-down "intrinsic saliency" at levels 4 and 5. This also allows for overriding of top-down knowledge by

other top-down knowledge. For example, it has previously been found that objects that defy the gist of the scene play more of a role in guiding eye movements than other more salient objects. However, it was demonstrated in experiment 7 (section 3.5) that semantic incongruence (top-down scene knowledge) is less distracting in comparative visual search when the pictures are specific to an expert's domain (top-down domain knowledge). According to Findlay and Walker's model, top-down and bottom-up influences are combined at different stages before an eye movement is produced. This is similar to the Contextual Guidance model put forward by Torralba, Oliva, Castelhano and Henderson (2006) which suggests that image saliency and global-context features are computed in parallel, in a feed-forward manner and are integrated at an early stage of visual processing (i.e. before initiating image exploration).

6.2 Implications for Content-Based Image Retrieval (CBIR) and Human-Computer Interaction (HCI)

The experiments in chapter 2 (and section 3.4) show that scanpaths produced whilst visualising pictures are similar to scanpaths produced when first encoding the pictures. Furthermore, these scanpaths at imagery are reliably similar over time. This is an encouraging result for creating eye-movement controlled content based image retrieval programs. Josephson and Holmes (2002) recorded eye movements whilst participants viewed multiple web pages and found that scanpaths for the same individual became more similar over time. They also found some similarities in scanpaths between participants. Both of these are promising findings with regards to creating an image library controlled by eye movements. However, Josephson and Holmes also found that scanpaths when viewing graphics were less similar than those when viewing text, which suggests that the similarity could simply be an unavoidable outcome of participants reading the text. The decreased similarity for graphics poses a problem for possible eye-movement-controlled CBIR.

The finding that eye movements can be predicted to an extent by the saliency algorithm designed by Itti & Koch (2000) could be applied to CBIR. If we had the saliency coordinates for a picture that we wanted to find in a large image database, the saliency map algorithm could be used to match the coordinates and retrieve the matching picture from the database. However, this would mean computing and storing x-y coordinates for a whole database of pictures and the problem would be finding the coordinates for the picture you want to search for. Sets of coordinates would have to be labelled contextually, and this brings us back to the problem of context-based image searches – it's a 'catch 22' situation.

Even if eye movements were applied directly to an image library, using an eye tracking camera to record scanpaths and a computer algorithm to match eye patterns, several problems remain. Firstly, section 2.3 provides evidence of boundary extension during imagery, which leads to a decrease in scanpath similarity between the encoding and recognition. For a CBIR database to be controlled by eye movements, it would first have to be 'trained' by recording eye movements made when looking at the picture, which would then be compared to eye movements made when retrieving the picture. Increased saccadic amplitudes and more extensive scanpaths due to boundary extension would cause problems for this pattern matching process.

Secondly, the different influences on eye guidance when first viewing pictures pose a problem for inter-participant usability. This is especially prominent for expertise, which seems to override both bottom up features of saliency and other top-down influences such as incongruence.

This suggests that different people are interested in different picture features and that depending on their cognitive knowledge of the domain, different parts of the picture could hold different semantic meaning. This could affect the way they move their eyes, meaning that even if scanpaths at encoding and recognition for that participant were highly similar, scanpaths between participants may be unique, thus limiting their generalisability for use in a CBIR model. For example, Chua, Boland, and Nisbett (2005) found cultural differences in eye movements during scene perception. American participants fixated more on focal objects than did Chinese participants, who in tern made more saccades to the background than did Americans. However, it has been suggested that similar viewing patterns occur when different subjects view different images in the same semantic category (Jaimes, Pelz, Grabowski, Babcock and Chang, 2001). Therefore, commonly fixated areas could indicate shared semantic interest. This information could be used to create a computer algorithm predictive of gaze based on semantically informative areas. Furthermore, it has been suggested that differences in scanpaths, especially those produced by domain-experts, could be used to train the eye movements of novices, encouraging them to look at and remember the same features as experts (Hembrooke, Feusner and Gay, 2006; Jarodzka, Scheiter, Gerjets and Van Gog, 2009, *in press*).

However, human perceptual subjectivity even exists at low levels of perception, when searching for a picture based on its visual properties, for example, one person may be more interested in an image's colour feature while another may be more interested in the texture feature. Even if both people are interested in texture, the way how they perceive the similarity of texture may be quite different.

Despite these potential problems with individual subjectivity, there is some evidence of inter-participant scanpath similarity (e.g. Stelmach, Tam and Hearty, 1991; Stelmach and Tam, 1994; Tosi, Mecacci and Pasquali, 1997). Stelmach, Tam and Hearty (1991) recorded eye movements of 24 subjects viewing 15 forty-five second movie clips to determine if viewing behaviour could be incorporated into video coding schemes. They found reliable similarity in eye movements and fixation locations between participants and even used eye movements from a follow-up study to create a model of predicted gaze position (Stelmach and Tam, 1994). This would not have been possible if substantial individual differences in scanpaths existed. Tosi, Mecacci and Pasquali (1997) also recorded the eye movements whilst participants viewed video clips and reported that, qualitatively, individual differences in scanpaths were relatively small.

A third drawback to creating a computer program based on eye movements is that computers do not have the ability to remember associations like humans do, unless specifically programmed to, and even then they do not form the same semantic links. However, some studies have tried to tackle this problem. Oyekoya and Stentiford (2004) created a computer prototype that, on the assumption that viewers look at the parts of the picture that they are interested in, 'remembers' where the viewer looks and uses this information to retrieve plausible candidate images for the user. Such images will contain regions that possess similarity links with the gazed regions. Similarly, Ban, and Lee (2006) modelled a novelty-detection computer program on the human hippocampus. The novelty of a picture was determined by comparing the description for the current scene with that for the experienced scene. The model combined visual scanpaths based on top-down attention and those based on the bottom-up saliency map model.

A forth major problem for using scanpaths in CBIR is that even though the similarity between encoding and imagery within the same participants is in most cases above chance, it is nonetheless still rather low, and nowhere near perfect. If a CBIR program was built, it is doubtful that it would be able to accommodate that level of inaccuracy. The scanpaths at imagery in the

experiments in this thesis were recorded without the participants being told to move their eyes. Therefore, one way to increase scanpath similarity at imagery, to a level that could be used by pattern matching algorithms, would be to give participants specific instructions to move their eyes. For example, participants could be pre-warned to remember the contents of the picture and where they looked, as they will be asked to reproduce those scanpaths at imagery. This intentionality has also been found to improve spatial memory (Naveh-Benjamin, 1987), in that participants who learn information intentionally remember more accurately both the absolute positioning of the objects and their spatial relations than do those who incidentally learn the information. Naveh-Benjamin also reported two other interesting findings, one that could be detrimental to eye-controlled CBIR, and one that could improve it. Firstly, older participants were found to be worse at encoding spatial location than younger participants, which means that their eye movements when trying to retrieve a picture may be less similar to the scanpath produced at encoding and thus cause more inaccurate retrieval. This reduces the generalisability of the computer program. Secondly, practice was found to significantly improve performance in memorising absolute spatial location information. This suggests that participants can be trained to remember pictures and the associated scanpaths.

To design an eye-controlled retrieval system for an image library, computer algorithms need to be based and tested on a large number of pictures and scanpaths associated with those pictures. These databases are time consuming and often economically unviable to compile, making it difficult to reliably create such algorithms. However, recently Van der Linde, Rajashekhar, Bovik, and Cormack, (2008) created DOVES, a database of visual eye movements recorded from 29 observers as they viewed 101 natural calibrated images. Recorded using a high precision dual-Purkinje eye tracker, the database consists of around 30,000 fixation points, and is believed to be the first large-scale database of eye movements to be made available to the vision research community. DOVES can be used to quantitatively compare eye movements with salient regions identified by gaze modelling algorithms and can also be used directly as input to novel saliency *learning* algorithms (Kienzle et al., 2006). Furthermore, it presents a source of generic eye movement data for calculating probability distributions (such as those relating to fixation centrality (Tatler, 2007), fixation duration, saccade angle and saccade length) and sequential eye movement characteristics, enabling common *image-independent* eye movement statistics to be established. This could be of great use when designing an image retrieval system based on scanpaths, as it could reduce the amount of 'training' of the program needed to compare scanpaths at retrieval to the computer predicted scanpaths.

Despite possible methodological difficulties with designing an image library controlled by scanpaths, there have been other successful applications of eye movements in human-computer-interaction, proving that eye movements and computer technology can be integrated. Several eye tracking systems have been designed to assist people with motor impairments that result from conditions such as cerebral palsy or Amyotrophic Lateral Sclerosis (ALS). Systems include Quick Glance (eyetechds.com), VisionKey (eyecan.ca), and the Eyegaze Communication System (eyegaze.com). Hundreds of people use these systems to communicate by looking at letters to type with their eyes. The Eyegaze Communication System offers perhaps the most functionality, with software for uttering phrases via a speech synthesizer, making telephone calls, controlling lights and appliances, and turning pages in electronic books. 'Dasher' (Ward and MacKay, 2002) is a method for text entry that relies purely on gaze direction. EyeDraw (Hornof and Cavender, 2005) is a software program that, when run on a computer with an eye tracking device, enables children with severe motor disabilities to draw pictures by just moving their eyes. More recently,

Wei-Gang, Huang, and Hwang (2007) created an eye-wink control human computer interaction system that allows the severely handicapped people to control the appliances by using eye winks. It has even been shown possible to carry out these commands without being near a computer, by attaching a detector unit to a normal pair of glasses (Shaw et al, 2000). The long term aim is to develop a wheelchair unit that can be driven by commands issued through a portable eyewink system. Developing a pictorial database controlled by eye movements would be an excellent addition to the innovative technology designed to help people with motor impairments.

As well as aiding the disabled, such eye-controlled technology has been applied as an alternative method for the activation of controls and functions in aircraft (Schnell and Wu, 2000), to investigate eye movements of neurological disease patients (Iijima et al, 2003), and to detect driver fatigue by measuring the number of pixels of the eye image available (Gan et al, 2006). If the driver starts to fall asleep, the computer program detects a decreased number of eye pixels and alerts the driver.

The design of eye-controlled computer programs has highlighted general problems with using eye movements, the most famous of all being the Midas Touch problem. Essentially, if the eyes are used in a manner similar to a mouse, a difficulty arises in determining intended activation of foveated features (the eyes do not register button clicks!). Moreover, gaze location is not as precise as with a mouse since the fovea limits the accuracy of the measured point of regard. Another significant problem is accuracy of the eye tracker. Following initial calibration, eye tracker accuracy may exhibit significant drift, where the measured point of regard gradually falls off from the actual point of gaze. All of these potential methodological problems should be considered if an eye-movement-controlled image retrieval system was to be designed.

6.3 Methodological Limitations

6.3.1 Problems with measuring chance

Chance was measured in the experiments in this thesis by comparing the scanpath produced for each picture a participant viewed to a scanpath that the participant produced on another a randomly selected picture. This was repeated for every participant and an average similarity score was calculated. This score differed for each experiment as eye movements differed depending on task. However, the first limitation to this is that there is no agreed way to create a chance score for trials where the encoding and retrieval processes differ. For example, in chapter 2 (and section 3.4) scanpaths at encoding (where visual information is present) were compared to scanpaths at imagery (where no visual information is present). The above calculation of chance is not possible – are scanpaths at encoding randomly compared to scanpaths at imagery? This would be very unfair, as scanpaths in the two conditions differ in length and spatial variation. Or is it more accurate to compare each scanpath at encoding to another randomly selected picture at encoding and also compare each scanpath at imagery to another randomly selected scanpaths it imagery, and then take an average of them both? The chance rates in this thesis were all based on encoding v encoding, to remain consistent, but this discussion highlights the difficulty in deciding how chance should be defined. A second potential problem is that chance values change depending on the length of the strings being compared, as seen in section 2.4.

6.3.2 Problems with measuring saliency

Saliency was measured in the experiments in this Thesis using the Itti and Koch (2000) algorithm which predicts fixation to the most salient region in an image first, followed by the second most salient region then the third most salient region, and so on. However, one has to acknowledge the limitations of this model. For example, the predicted saliency scanpaths used in this Thesis have relied on standard parameters, but the results of some of the experiments might change if these parameters (brightness, colour contrast and orientation change) were adjusted. On a similar line of argument, the standard two degrees of visual angle were used to define the area of saliency, but adjusting the size of this region could also affect experimental results. Furthermore, some of the most salient regions identified according to the algorithm parameters are boundaries such as horizons or when a house roof meets the sky. However, it is unlikely that these are the most informative regions in the picture and are therefore not fixated as often as the model would predict. Navalpakkam and Itti (2005) attempt to deal with this problem by proposing a new version of the saliency map model that includes "task relevant maps". These allow search for specific objects on the basis of low-level features that have been learnt, biasing attention towards objects that share these features. The saliency map that is built is then able to represent task-relevant features. This version of the model accounts for why attention is not captured by a bright sky, as would be predicted based on a simple saliency peak. However, the computer algorithm does not take this into account.

The underlying Saliency Map model has been questioned as to how many times the map is generated for a given scene. It could be that just one map is computed across the entire scene during the first fixation and is retained over multiple fixations. However, this is unlikely due to evidence that metrically precise sensory information about a scene is not retained across saccades (Irwin, 1992; Henderson, and Hollingworth, 2003). Alternatively, a new saliency map could be computed in each fixation, meaning that the saliency map would not need to be retained across fixations. Consequently though, computational load would increase due to the constant generation of new maps. Tatler et al (2005) suggest that the saliency map does not change during the period of inspection, but the influence of cognitive strategic factors does change. They call this the Strategic Divergence model and propose that saliency has a role in guiding eye movements, but the actual movement of the eyes is due to task demands and individual cognitive knowledge. Evidence for this model comes from early consistency of fixations but with deviation between the viewers increasing linearly with inspection time (Tatler et al, 2005). The strategic divergence model accounts for variations between cognitive tasks and between individuals. It also accommodates the processes of spatial selection, search selection, and intrinsic saliency proposed by Findlay and Walker's (1999).

Some recent studies have attempted to create new algorithms for predicting eye movements, but taking into account top-down cognitive influences as well as bottom-up saliency. Kienzle, Wichmann, Scholkopf, and Franz (2006) present a saliency model that is *learned* directly from human eye movements, with minimal emphasis on visual saliency. Similarly, Zelinsky, Zhang, Yu, Chen, Samaras (2006) found that top-down information was most predictive of eye movements.

6.4 Evaluation of string editing

There is not yet a universally agreed way of best analysing eye movements and scanpath data and one issue frequently raised is how to divide up the target into viewing areas. Brandt & Stark

(1997) divided their stimuli up into 7 areas or 'zones', however, others have divided stimuli up into 25 zones (e.g. Foulsham and Underwood, 2008), and how fine-grained it should actually be is not agreed upon. This could make a difference to the data analysis. Take for example Figure 102.
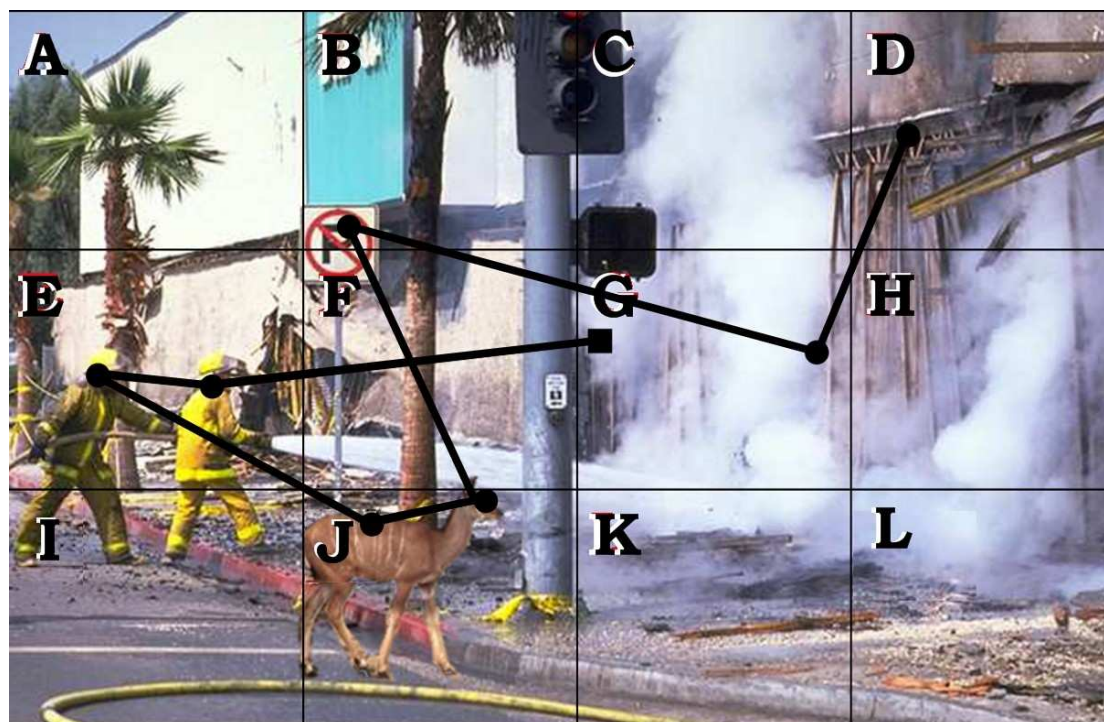
Figure 102: An Example picture divided up into 12 regions, with a hypothetical scanpath superimposed on it.

If the zones were smaller, the two fixations in E could be counted as being in different areas. Or similarly, if the zones were bigger, the fixations in K and E could be counted as being in the same area. This could cause a different outcome from the algorithm. A 5 x 5 grid was chosen for scanpath analyses in this Thesis (resulting in 25 zones) as it achieved a balance between spatial resolution and simplicity of computation.

A related problem is whether to analyse zones or objects. For example, two fixations may fall close to each other in the same zone, but fixate on two different objects. Should that be counted as one fixation (one zone) or two (two different objects)? Similarly, what if two fixations were extremely close to each other, but the borderline meant that they fell into two different zones?

Furthermore, strings are cropped to a specified length or the length of the shortest string in the comparison, which could discard valuable scanpath information. It could be that scanpaths are globally similar, but due to, for example, one string diverging in the middle of the sequence, later more similar parts of the scanpath are lost due to cropping.

Lastly, a big problem with string editing is the dubious assumption that all edits have equal weight e.g. deletions vs additions, as well as neglecting the location of the fixation. For example, a viewer may fixate the sky in a picture during encoding (when initially 'exploring' the picture) but not during recognition because they then know that nothing of interest exits in that region. This situation would be treated as equal, by the string editing method, to fixating the animal in the above example picture at encoding but not recognition. These two examples are clearly not equal.

These methodological points are accepted as limitations to the use of string-editing as a method of computing scanpath similarities, however as yet there is no universally agreed solution. A possible suggestion could be to repeat the analyses using a technique developed by Mannan et al (1995) which computes the mean linear distance between fixations in one scanpath and their nearest neighbour in the other set. Henderson et al (2007) refined this method slightly by adding the constraint that each fixation in a scanpath is assigned to only one other in the comparison scanpath. This "Unique-assignment" (UA) version ensures that the scanpath comparison is not disproportionately affected by differences in overall distribution of fixations and therefore also requires the scanpaths to have an equal number of fixations. Foulsham and Underwood (2008) have compared these three methods for scanpath computation and have found that they produce very similar results in practice. String editing was used in this Thesis because it is the most conservative of the three methods.

An alternative method of measuring scanpath similarity was proposed by Gareze, Tadmor, Barenghi and Harris (2004). They presented a new mathematical tool, based on knot theory and modern geometry, which characterises the number of 'crossings' in a scanpath. Crossings occur when the scan path folds over itself in space, forming a 2-D 'loop', providing a quantitative measure of how 'tangled' a scan path is. Saccades were associated with larger loops and fixations were associated with smaller ones. Similarly, West, Haake, Rozanski and Karn (2006) created a string similarity tool called "eyePatterns". This computer model integrates many currently used similarity techniques (e.g., string-editing, transition frequency analysis, clustering) with analyses that are not usually applied to eye movement research (e.g., sequence alignment, pattern discovery). Unknown patterns can be found through the 'discovery method' which allows users to input length and content criteria, and returns all patterns matching those criteria. For example, a user can elect to discover only patterns that are at least five characters long, which contain the AOI labelled "B," and are found in at least four sequences. Any instance of a substring that meets those criteria will be returned. Specified patterns can be found through pattern matching, which searches for an exact pattern, or provide a regular expression - a string that uses certain syntax rules to describe a set of strings.

6.5 Conclusion

This thesis aimed to investigate eye movements during the viewing of real-world scenes and the top-down and bottom-up factors that affect them. Low-level visual saliency was found to attract attention and guide eye movements, but this was reliably reduced by top-down influences such as domain proficiency and the presence of emotional features. Furthermore, recognition memory advantages were found for domain-specific pictures, negatively emotional pictures and complex pictures that were described post-presentation. Scanpath analyses showed that sequences of fixations are, to an extent, replicable over multiple viewings and over extended periods of time. Higher similarity was found when the encoding and retrieval processes were most similar, and the (part) reproduction of scanpaths during imagery firstly argues against a purely bottom-up explanation for this similarity and, secondly, suggests a link between eye movements and memory. In conclusion, eye movements are neither influenced by purely top-down nor bottom-up factors, but instead a combination of both, which interact to guide attention to the most relevant parts of the picture.

7. References

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data. *American Journal of Sociology, 16,* 144-185.

Abernethy, B., & Russell, D. G. (1987)a. The relationship between expertise and visual search strategies in a racquet sport. *Human Movement Science, 6,* 283–319.

Abernethy, B., & Russell, D. G. (1987)b. Expert novice differences in an applied selective attention task. *Journal of Sport Psychology, 9,* 326-45.

Allen, R., McGeorge, P., Pearson, D., & Milne, A. B. (2004). Attention and expertise in multiple-target tracking. *Applied Cognitive Psychology, 18(3),* 337–347.

Allopenna, P. D., Magnuson, J. S. and Tanenhaus, M. K.. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language, 38(4),* 419-439

Alpers, G. W. (2008). Eye-catching: Right hemisphere attentional bias for emotional pictures. *Laterality: Asymmetries of Body, Brain and Cognition, 13(2),* 158-178.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: a reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory & Cognition, 25,* 997-1010.

Altmann, G. T.M. (2004). Language-mediated eye-movements in the absence of a visual world: The "blank screen paradigm." *Cognition, 93,* 79-87.

Amadasun, M., & King, R. A. (1998). Low-level segmentation of multispectral images via agglomerative clustering of uniform neighborhoods. *Pattern Recognition, 21(3),* 261-268.

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology, 103,* 62-70.

Archer, J., Hay, D. C., & Young, A. W. (1994). Movement, face processing and Schizophrenia: evidence of a differential deficit in expression analysis. *British Journal of Clinical psychology, 33,* 517-528.

Aula, A., Majaranta, P., & Räihä, K-J. (2005). Eye-Tracking Reveals the Personal Styles for Search Result Evaluation. *LNCS, 3585,* 1058-1061.

Ballaz, C., Boutsen, L., Peyrin, C., Humphreys, G. W., & Marendaz, C. (2005). Visual search for object orientation can be modulated by canonical orientation. Journal of Experimental Psychology: Human Perception and Performance, 31, 20-39.

Ban, S.-W. & Lee, M. (2006). Selective attention-based novelty scene detection in dynamic environments. *Neurocomputing, 69,* 1723–1727.

Baron-Cohen S, (1994). How to build a baby that can read minds: cognitive mechanisms in mindreading. *Cahiers de Psychologie Cognitive, 13,* 513–552.

Bartlett, J. C., Till, R. E., & Levy, J. C. (1980). Retrieval characteristics of complex pictures: Effect of verbal encoding. *Journal of Verbal Learning and Verbal Behaviour, 19,* 430 4~9.

Bean, K. L, (1938). An experimental approach to the reading of music. *Psychological monographs, 50,* (whole no. 50).

Becker, M. W., & Detweiler-Bedell, B. (2009). Early detection and avoidance of threatening faces during passive viewing. *The Quarterly Journal of Experimental Psychology, 62(7),* 1257–1264.

Becker, M.W., Pashler, H., & Lubin, J., (2007). Object-intrinsic oddities draw early saccades. *Journal of Experimental Psychology, 3,* 20-30.

Behtmann, M., Winocur, G. and Moseovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature, 359,* 636-637.

Bentall, R. P., Kanely, S. (1989). Content specific processing and persecutory delusions: an investigation using the emotional Stroop test. *British Journal Medical Psychology, 62,* 355-364.

Bichot, N.P. and Schall, J.D. (1999). Effects of similarity and history on neural mechanisms of visual selection. *Nature Neuroscience, 2,* 549-554.

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177,* 77–80.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143–177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103,* 597–600.

Biele, C., & Grabowska, A. (2006). Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research, 171,* 1–6.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2007). Why do we look at people's eyes? *Journal of Eye Movement Research, 1(1),* 1-6.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008) Gaze selection in complex social scenes. *Visual Cognition,* 341-355.

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science, 299,* 81-86

Blangero, A., Khan, A. Z., Salemme, R., Deubel, H., Schneider, W. X., Rode, G., Vighetto, A., Rossetti, Y., & Pisella, L. (2009) Pre-saccadic perceptual facilitation can occur without covert orienting of attention, *Cortex (in press).*

Bouma, M., (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177-178.

Bradley, B. P., Mogg, K., & Lee, S. C. (1997). Attentional biases for negative information in induced and naturally occurring dysphoria. *Behaviour Research and Therapy, 35,* 911-927.

Bradley, B. P., Mogg, K., & Millar, N. H. (2000). Covert and overt orienting of attention to emotional faces in anxiety. *Cognition and Emotion, 14(6),* 789 808.

Bradley, B. P., Mogg, K., White, J., Groom, C., & de Bono, J. (1999). Attentional bias for emotional faces in generalized anxiety disorder. *British Journal of Clinical Psychology, 38,* 267-278.

Bradley, M. M. (2000). Emotion and motivation. In Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.), *Handbook of psychophysiology*, 602-642. New York: Cambridge University Press.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*, 27-38.

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13(1),* 99-108.

Brockmole, J. R., & Henderson, J. M. (2006)b. Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology, 59(7),* 1177–1187

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. Journal *of Personality and Social Psychology, 76,* 839–855.

Cahill L, McGaugh JL. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences, 21*(7), 294–299.

Cahill, L., Babinsky, R., Markowitsch, H., McGaugh, J. L. (1995). The amygdale and emotional memory. *Nature, 377,* 295-6.

Calabrese, L., & Marucci, F. S. (2006). The influence of expertise level on the visuo-spatial ability: differences between experts and novices in imagery and drawing abilities. *Cognitive Processes, 7(1),* 118–120.

Calvo, M. G., & Lang, P. J., (2004). Gaze Patterns When Looking at Emotional Pictures: Motivationally Biased Attention. *Motivation and Emotion, 28(3),* 221-243.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behaviour Research Methods, 40,* 109–115.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience. 20(19),* 99.

Canli, T., Zhao, Z., Desmond, J. E., Glover, G., & Gabrieli, J. D. E. (1999). fMRI identifies a network of structures correlated with retention of positive and negative emotional memory. *Psychobiology, 27,* 441–452.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46(26),* 4333-4345.

Carter, C. J., & Laya, O. (1998). Drivers' visual search in a field situation and in a driving simulator. In A. G. Gale, I. D. Brown, C. M. Haslegrave, & S. P. Taylor (Eds.), *Vision in vehicles VI*, Amsterdam: Elsevier, pp. 21–31.

Caseras, X., Garner, M., Bradley, B. P., & Mogg, K., (2007). Biases in Visual Orienting to Negative and Positive Scenes in Dysphoria: An Eye Movement Study. *Journal of Abnormal Psychology 116(3),* 491–497.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12,* 1017-1040.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009) Viewing task influences eye movement control during active scene perception. *Journal of Vision, 9(3):6,* 1–15

Cerf, M., Harel, J., Huth, A., Einhauser, W., & Koch, C. (2008) Decoding what people see from where they look: predicting visual stimuli from scanpaths. *Proceedings of the Fifth International Workshop on Attention in Cognitive Systems*

Chance, J., & Goldstein, A. G. (1976). Recognition of faces and verbal labels. *Bulletin of the Psychonomic Society, 7,* 384–386.

Chapman, P., & Underwood, G. (1998). Visual search of driving situations. *Perception, 27,* 951–964.

Charles, S. T., Mather, M., & Carstensen, L. L. (2003). Aging and emotional memory: The forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General, 132,* 310–324.

Charlot, V., Tzourio, N., Zilbovicius, M., Mazoyer, B. & Denis, M. (1992). Different mental imagery abilities result in different regional cerebral blood flow activation patterns during cognitive tasks. *Neuropsychologia, 30,* 565-580.

Chase, W. G., & Erikson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, 16,* 1-58. New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing,* 215-281. New York: Academic Press.

Chase, W. G., & Simon, H. A. (1973b). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chi, M. T. H., Gaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.). *Advances in Psychology of human intelligence, 1,* 7-75. Hilldale, NJ: Laurence Erlbaum Associates, Inc.

Choi, Y. S., Mosley, A. D., & Stark, L. (1995). Sting editing analysis of human visual search. *Optometry and Vision Science, 72,* 439-451.

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J., & Moore, J. G. (1981). The effect of search time on perception. Radiology, 138, 361–365.

Christianson, S. A., Loftus, E. F, Hoffman, H., & Loftus, G. R. (1991). Eye Fixations and Memory for Emotional Events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(4),* 693-701.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceeding of the National Academy of Science, 102(35),* 12629-12633.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28_71.

Clark, D. M. (1999). Anxiety disorders: why they persist and how to treat them. *Behaviour Research and Therapy, 37,* 5–27.

Clark, D. M., Wells, A. A., (1995). A cognitive model of social phobia. In: Heimberg, R.G., Liebowitz, M.R., Hope, D.A., Schneier, F.R. (Eds.), *Social Phobia: Diagnosis, Assessment and Treatment*. Guilford, New York.

Clifton, J. V. (1986). *Cognitive components in music reading and sight reading performance.* Unpublished doctoral dissertation, University of Waterloo.

Cocchi, L., Bosisio, F., Berchtold, A., Orita, A., Debbane, M., Wood, S. J., Schenk, F. (2009) **Visuospatial encoding deficits and compensatory strategies in schizophrenia revealed by eye movement analysis during a working memory task.** *Acta Neuropsychiatrica, 21(2),* 75-83.

Cooke, N. J., Atlas, R. S., Lane, D. M., & Berger, R. C. (1993). Role of high-level knowledge in memory for chess positions. *American Journal of Psychology, 106,* 321-351.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95,* 115-123.

Cox, I. J., Miller, M. L., Minka, T. P., & Yianilos, P. N. (1998). An optimized interaction strategy for Bayesian relevance feedback. *CVPR'98, Santa Barbara, CA*, 553-558.

Croft, J. L., Pittman, D. J., & Scialfa, C. T., (2005). Gaze behaviour of experienced and novice spotters during air-to-ground search. *Perception, 34,* ECVP Abstract Supplement.

Crundall, D. E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, *41,* 448–458.

Crundall, D., Chapman, P., France, E., Underwood G., & Phelps, N. (2005). What attracts attention during police pursuit driving? *Applied Cognitive Psychology, 19,* 409-420.

David, A. S. (1989). Perceptual asymmetry for happy-sad chimeric faces: effects of mood. *Neuropsychologia, 27,* 1289-1300.

Davidson, R. J. (1995). Cerebral asymmetry, emotion and affective style. In Davidson, R. J. H., & Hugdahl, K. (Eds.), *Brain asymmetry,* 361-387. Cambridge, MA: MIT Press.

Davidson, R. J., & Irwin, W. (1999). The functional neuroanatomy of emotion and affective style. *Trends in Cognitive Sciences*, *3,* 11-20.

Davidson, R. J., & Tomarken, A. J. (1989). Laterality and emotion: An electrophysiological approach. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology, 3,* 419-441. New York: Elsevier.

De Graef, P. (2005). Semantic effects on object selection in real-world scene perception. In Underwood, G. (Ed.), Cognitive processes in eye guidance, 189-211. Oxford, UK: Oxford University Press.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research, 52,* 317-329.

De Lisi, R., & Cammarano, D. M. (1996). Computer experience and gender differences in undergraduate mental rotation performance. *Computers in Human Behaviour, 12,* 351-361.

de Valk, J. P. J., & Eijkman, E. G. J. (1984). Analysis of eye fixations during the diagnostic interpretation of chest radiographs. *Medical & Biological Engineering & Computing, 22, 353-360.*

De Wit, T. C. J., Falck-Ytter, T., von Hofsten, C. (2008) Young children with Autism spectrum Disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders, 2,* 651-659.

DeAngelus, M., & Pelz, J., B. (*in press*) Top-down control of eye movements: Yarbus revisited. *Visual Cognition.*

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*, 1827-1837.

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14,* 14-18.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11,* 86–89.

Dodson, C. S., Johnson, M. K. and Schooler, J. W. (1997). The verbal overshadowing effect: Source confusion or strategy shift? *Memory and Cognition, 25,* 129-139.

Dougherty, P. E., Bartlett, E. S., & Izard, C. E., (1974). Responses of schizophrenics to expressions of fundamental emotions. *Journal of Clinical Psychology, 30,* 243-246.

Dowe, J. (1993). Content-based retrieval in multimedia imaging. *Proc. SPIE Storage and Retrieval for*

Image and Video Databases

Drai-Zerbib, V., & Baccin, T. (2005). *Année Psychologique, 105(3),* 387-422

Draper, B. A., Baek, K., & Boody, J. (2004). Implementing the expert object recognition pathway. *Machine Vision and Applications 16*, 27–32.

Eakins, J.P and Graham, M.E. (1999). Content-Based Image Retrieval. A report to the JISC Technology Applications Programme. Institute for Image Data research, Northumbria Univ. (www.unn.ac.uk/report).

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behaviour. *Psychological Review, 66,* 183–201.

Eastwood, J. D., Smilek, D., & Merikle, P, M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception & Psychophysics*, *63*(6), *1004-1013.*

Egan, D. E., Schwartz, E. J. (1979). Chunking in recall of symbolic drawings. *Memory and Cognition 7,* 149-158.

Ehlers, J. T. (1936). The movements of the eyes during reading. *Acta Ophthalmologica Kjobenhavn, 14,* 56-63.

Einhauser, W., & Konig, P. (2003). Does luminance-contrast contribute to a saliency map for overt visual attention? *European Journal of Neuroscience, 17,* 1089-1097.

Einhauser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8,* 1-19.

Einhauser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8,* 1-26.

Eizenman, M., Yu, L. H., Grupp, L., Eizenman, E., Ellenbogen, M., Gemar, M., & Levitan, R. D. (2003). A naturalistic visual scanning approach to assess selective attention in major depressive disorder. *Psychiatry Research, 118,* 117-128.

Enser, P. (2002). Separating the Eiffel Tower from the Statue of Liberty. *Lib & Info Update, 1(4),* 34-36.

Eom, K., & Park, J. (1991). Shape recognition by a scale-invariant model. *Journal of Systems Integration, 1(2),* 215-233.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211 245.

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spizer, M., & Waltera, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage 18,* 439–447.

Evans, A., Sikorski, J., Thomas, P., Cha, S-H., Tappert, C., Zou, J., Gattani, A., & Nagy, G. (2005) Computer Assisted Visual Interactive Recognition (CAVIAR) Technology, *Procs. IEEE International Conference on Electro-Information Technology,* Lincoln, NE.

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476–1492.

Eysenck, M. W. (1992). Anxiety: The Cognitive Perspective. Erlbaum Ltd, Hove, UK.

Fallshore, M., & Schooler, J. W. (1995). Verbal Vulnerability of Perceptual Expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(6),* 1608-1623.

Faloutsos, C., Flickner M., Niblack, W., Petkovic, D., Equitz, W., & Barber, R. (1993). Efficient and effective querying by image content. *Tech Report, IBM Research Report.*

Farah, M. J., Weisberg, L. L., Monheit, M. & Peronnet, F. (1989). Brain activity underlying mental imagery: Event-related potentials during mental image generation. Journal *of Cognitive Neuroscience, 1,* 302-316.

Fecteau, J. H., & Munoz, D. P. (2003). Exploring the consequences of the previous trial. *Nature Reviews: Neuroscience, 4,* 435-443.

Fecteau, J. H. & Munoz, D. P. (2005). Correlates of capture of attention and inhibition of return across stages of visual processing. *Journal of Cognitive Neuroscience, 17,* 1714-1727.

Fecteau, J. H., Bell, A. H., & Munoz, D. P. (2004). Neural correlates of automatic and goal-driven biases in orienting spatial attention. *Journal of Neurophysiology, 92,* 1728-1737.

Findlay, J. M., Brown,V., &Gilchrist, I. D. (2001). Saccade target selection in visual search: The

effect of information from the previous fixation. *Vision Research, 41*, 87-95.

Findlay, J.M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioural and Brain Sciences, 22,* 661-674.

Finke, R. A. & Pinker, S. (1982). Spontaneous imagery scanning in mental extrapolation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 2,* 142-147.

Finke, R., & Sehmidt, M. (1977). Orientation-specific after effects following imagination. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 599-606.

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

Fiore, S. M., & Schooler, J. W. (2002). How did you get here from there? Verbal overshadowing of spatial mental models. *Applied Cognitive Psychology, 16,* 897-909.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson. V. (2008). Rapid detection of person information in a naturalistic scene. *Perception,* 571-583.

Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafne, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: The QBIC system. *IEEE Computer.*

Flom, M. C., Heath, O., & Takahaski, E. (1963). Contour interaction and visual resolution: contralateral effect. *Science, 142,* 979-980.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36,* 1123–1138.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye-movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8(2):*6, 1-17.

Fournier, J., Cord, M., & Philipp-Foliguet, S. (2001). RETIN: A Content-Based Image Indexing and Retrieval System. *Pattern Analysis & Applications, 4,* 153-173.

Fox, E., Lester, V., Russo, R., Bowles, R., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61-92.

Freeman, D., Garety, P. A., & Philips, M. L. (2000). An examination of hypervigilance for external threat in individuals with Generalised Anxiety Disorder and individuals with persecutory delusions using visual scanpaths. *Human Experimental Psychology, 53(2),* 549-567.

Frens, M., van der Geest, J., & Hooge, I. T. C. (2000). Inhibition of saccade return. *Perception.* CVP abstract, S1.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General,108,* 316–355.

Friedman, A., & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In Fisher, D. F., Monty, R. A., & Senders J. W. (Eds.), *Eye movements: Cognition and visual perception*, 137–155. Hillsdale, NJ: Erlbaum.

Friesen, C. K., Kingstone, A. (1998). The eyes have it!: Reflexive orienting is triggered by non-predictive gaze. *Psychonomic Bulletin & Review, 5,* 490-495.

Gaebel, W., Ulrich, G., Frick, K. (1987). Visiomotor performance of schizophrenic patients and normal controls in a picture-viewing task. *Biological Psychiatry, 22,* 1227-1237.

Galpin, A. J., & Underwood, G. (2005). Eye movements during search and detection in comparative visual search. *Perception & Psychophysics, 67(8),* 1313-1331.

Gan, L., Cui, B., & Wang, W. (2006). Driver fatigue detection based on eye tracking. In *Proceedings of the 6th World Congress on Intelligent Control and Automation, 2,* 5341–5344.

Gareze, L., & Findlay, J. M. (2007). Absence of scene context effects in object detection and eye gaze capture. In van Gompel, R. P. G., Fischer, M. H., Murray, W. S. & Hill, R. L. (Eds.), *Eye movements: A window on mind and brain*, 617–637.

Gareze, L., Tadmor, Y., Barenghi, C. F., & Harris, J. M. (2004). Tangled eye movements: A new method to quantify human scan paths. *Perception, 33,* ECVP Abstract Supplement.

Garner, M., Mogg, K., & Bradley, B. P. (2006). Orienting and Maintenance of Gaze to Facial Expressions in Social Anxiety. *Journal of Abnormal Psychology, 115(4)*, 760–770

Gbadamosi, J., Oechsner, U., & Zangemeister, W. H. (1997). Quantitative analysis of gaze movements during visual imagery in hemianopic patients and control subjects. *Journal of Neurological Rehabilitation, 3*, 165-172.

Gerdes, A. B. M., Alpers, G. W. & Pauli, P. (2008). When spiders appear suddenly: Spider phobic patients are distracted by task-irrelevant spiders. *Behaviour Research and Therapy, 46*, 174-187.

Gilboa-Schechtman, E., Foa, E. B., & Amir, N. (1999). Attentional biases for facial expressions in social phobia: The face-in-the-crowd paradigm. *Cognition & Emotion, 13*, 305-318.

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition, 14(4),* 704-715.

Gilchrist, I. D., North, A., & Hood, B. (2001). Is visual search really like foraging? *Perception, 30,* 1459-1464.

Gilman, E., & Underwood, G. (2003). Restricting the field of view to investigate the perceptual spans of pianists. *Visual Cognition, 10,* 201-232.

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology, 31*, 1-40.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *TRENDS In Cognitive Sciences*, 5, 236–243.

Goldberg, M.E. et al. (2002) The role of the lateral intraparietal area of the monkey in the generation of saccades and visuospatial attention. *Ann. New York Academy of Sciences, 956,* 205–215

Goldenberg, G., Miillbacher, W., and Nowak, A., (1995). Imagery without perception – a case study of agnosagnosia for cortical blindness. *Neuropsychologica, 33,* 1373-1382.

Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Suess, E. and Deecke, L. (2989). Regional cerebral blood flow patterns in visual imagery. *Neuropsychologia,* 27, 641-664.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine 37*, 957 – 964.

Goolsby, T. W. (1994)a. Eye-movement in music reading - Effects of reading ability, notational complexity, and encounters. *Music Perception, 12,* 77-96.

Goolsby, T. W. (1994)b. Profiles of processing - Eye-movements during sight reading. *Music Perception, 12,* 97-123.

Goos, L. M., & Silverman, I. (2002). Sex related factors in the perception of threatening facial expressions. *Journal of Nonverbal Behaviour, 26,* 27–41.

Gopher, D., Weil, M., & Bareket, T. (1994). Transfer of skill from a computer game trainer to flight. *Human Factors, 36,* 387-405.

Gottlieb, J. (2002). Parietal mechanisms of target representation. *Current Opinion in Neurobiology, 12,* 134–140

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423,* 534-537.

Green, C. S., & Bavelier, D. (2006)a. Effect of action video games on the spatial distribution of visuospatial attention. Journal of Experimental Psychology: *Human Perception and Performance, 32,* 1465-1478.

Green, C. S., & Bavelier, D. (2006)b. Enumeration versus multiple object tracking: the case of action video game players. *Cognition, 101,* 217-245.

Green, C. S., & Bavelier, D. (2006)c. The cognitive neuroscience of video games. In Messaris, P. & Humphreys, L. (Eds.), *Digital Media: Transformations in Human Communication*, 211-224. New York: Peter Lang.

Green, C. S., & Bavelier, D. (2007). Action video game experience alters the spatial resolution of

vision. Psychological Science, 18, 88-94.

Green, M. J., Williams, L. M., & Davidson, D. (2003). In the face of danger: Specific viewing strategies for facial expressions of threat? *Cognition and Emotion, 17*, 779–786.

Groner, R. et al. (1984) Looking at face: local and global aspects of scanpaths. In Gale, A.G., & Johnson, F. (Eds.) *Theoretical and Applied Aspects of Eye Movement Research*, 523-533, Elsevier

Guaraglia, C., Padovani, A., Pantano, P. and Pizzamaglio, L. (1993). Unilateral neglect restricted to visual imagery. *Nature, 364,* 235-237.

Gudivada, V. N., & Raghavan, J. V. (1995). Special issue on content-based image retrieval systems. *IEEE Computer Magazine, 18(9).*

Hacisalihzade, S. S., Allen, J. S., & Stark, L. (1992). Visual perception and sequences of eye movement fixations: A stochastic modelling approach. *IEEE Transactions on Systems Man And Cybernetics, 22*, 474-481.

Haider, H., & Frensch, P. A. (1999). Eye movements during skill acquisition: more evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 172-190.

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion, 4,* 201–206.

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology, 95,* 31-50.

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, S. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience, 2,* 289-293.

Hamann, S.B. & Mao, H. (2001). Neural correlates of emotional memory and reactions to verbal and nonverbal emotional stimuli: an fMRI study. *Journal of Cognitive Neuroscience Abstract,* 30.

Hamann, S.B., Lee, G. P., & Adolphs, R. (1999). Impaired declarative emotional memory but intact emotional responses following human bilateral amygdalotomy. *Society for Neuroscience Abstract, 25*, 99.

Hamm, A. O., Schupp, H. T., & Weike, A. I. (2003). Motivational organization of emotions: Autonomic changes, cortical responses, and reflex modulation. In Davidson, R. J., Scherer, K., & Goldsmith H. H. (Eds.), *Handbook of affective sciences*, 188–211. Oxford: Oxford University Press.

Hansen, C. H., Hansen, R. D., (1988). Finding the face in the crowd: an anger superiority effect. *Journal of Personality and Social Psychology, 54*, 917–24.

Harris, C. R., Kaplan, R. L., & Pashler, H. (2008). Alarming events in the corner of your eye: Do they trigger early saccades? *Manuscript submitted for publication.*

Hayes-Roth, B. (1983). The blackboard architecture: a general framework for problem solving? *Technical report, no. HPP 83-30,* Stanford, Calif: Stanford University.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition 7,* 43–64.

Hayhoe, M. M., Shrivastava, A., Mruczek, R. & Pelz, J.B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3,* 49–63.

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

Heinrichs, N., & Hofmann, S. G. (2001). Information processing in social phobia: A critical review. *Clinical Psychology Review, 21,* 751–770.

Hembrooke, H., Feusner, M., & Gay, G. (2006). Averaging scan patterns and what they can tell us. *ETRA, San Diego, CA.*

Henderson, J. M. (1992). Visual attention and eye movement control during reading and scene perception. In Rayner, K. (Ed.), Eye movements and visual cognition: Scene perception and reading, 260-283. New York: Springer-Verlag.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498-504.

Henderson, J. M. (2007). Regarding scenes. Current Directions in Psychological Science, 16, 219-222.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In Henderson J. M.,

& Ferreira F. (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*, 1-58. New York: Psychology Press.

Henderson, J. M., & Hollingworth, A. (1998). Eye Movements during Scene Viewing: An Overview. In Underwood, G. (Ed.), *Eye Guidance while Reading and While Watching Dynamic Scenes,* 269-293. Oxford: Elsevier.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243-271.

Henderson, J. M., & Hollingworth, A. (2003). Global trans-saccadic change blindness during scene perception. *Psychological Science, 14,* 493-497.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M., (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In van Gompel, R., Fischer, M., Murray, W., & Hill, R. (Eds.), Eye movements: A window on mind and brain, 537-562.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the Dark: Cognitive Relevance Drives Attention in Real-World Scenes. *Psychonomic Bulletin & Review, in press.*

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25(1),* 210-228.

Hering, E. (1879). Über Muskelgeräusche des Auges. *Sitzungsberichte der Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse. Abt. III, 79*, 137-154.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18,* 496-506.

Hilz, R., & Cavonius, C.R. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research, 14,* 1333-1337.

Hishitani, S. (1988) Imagery experts: How do expert abacus operators process imagery? *Applied Cognitive Psychology, 4(1),* **33-46**

Hochstein, S., & Ahissar, M., 2002. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron,* 36, 791-804.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and Long term Scene Memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31(3),* 396-411.

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

Holsanova, J., Hedberg, B., & Nilsson, N. (1998). Visual and verbal focus patterns when describing pictures. In Becker, W., Deubel, H. & Mergner, T. (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects,* 303-304. New York: Plenum.

Hood, B. M., Willen, J. D., Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9(2),* 131-134.

Hope, L., & Wright, D. (2007). Beyond Unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology, 21*, 951–961.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye: a visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders, 17*, 33-44.

Hornof A. J. & Cavender, A. (2005). EyeDraw: Enabling Children with Severe Motor Impairments to Draw with Their Eyes. In *Proceedings of Computer-Human Interface*. ACM Press.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*, 575–577.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception & Psychophysics, 63*, 272–285.

Horowitz, T. S., & Wolfe, J. M. (2003). Memory for rejected distracters in visual search? *Visual*

Cognition, 10, 257-298.

Howarth, P., & Ruger, S. (2004). Evaluation of Texture Features for Content-Based Image Retrieval. *LNCS, 3115*, 326-334.

Howarth, P., Yavlinsky, A., Heesch, D., & Ruger, S. (2005). Medical Image Retrieval Using Texture, Locality and Colour. *LNCS, 3491*, 740-749

Huang, T. S., Mehrotra, S., & Ramchandran, K. (1996) Multimedia analysis and retrieval system (MARS) project. *Proc of 33rd Annual Clinic on Library Application of Data Processing Digital Image Access and Retrieval.*

Huff, M., & Schwan, S. (2008). Verbalizing events: Overshadowing or facilitation? *Memory & Cognition, 36(2), 392-402.*

Humphrey, K. & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. *British Journal of Psychology,100(2),* 377-398

Hunt, A. R., Cooper, R. M., Hungr, C., & Kingstone, A. (2007). The effect of emotional faces on eye movements and attention, *Visual Cognition, 15(5),* 513-531.

Iijima, A., Haida, M., Ishikawa, N., Minamitani, H., & Shinohara, Y. (2003). Head mounted goggle system with liquid crystal display for evaluation of eye tracking functions on neurological disease patients. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 4,* 3225-3228.

Intraub, H., & Bodamer, J. L., (1993). Boundary extension: Fundamental aspect of pictorial representation or encoding artefact?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1387-1397.

Intraub, H., & Richardson, M., (1989). Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 179-187.

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J., (1996). Boundary Extension for Briefly Glimpsed Photographs: Do Common Perceptual Processes Result in Unexpected Memory Distortions? *Journal of Memory and Language, 35*(2), 118-134.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *18,* 307-317.

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

Isaacowitz, D. M., Wadlinger, H. A., Goren D., & Wilson, H. R. (2006) Selective Preference in Visual Fixation Away From Negative Images in Old Age? An Eye-Tracking Study. *Psychology and Aging*, *21(1),* 40-48.

Ishai, A., and Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science, 268*, 1772-1774

Ito, T. A., Larsen, J. T., Smith, N. K., Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*, 887-900.

Itoh, Y. (2005). The Facilitating Effect of Verbalization on the Recognition Memory of Incidentally Learned Faces. *Applied Cognitive Psychology, 19*, 421-433.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40 (10-12),* 1489-1506.

Jaimes, A., Pelz, J. B., Grabowski, T., Babcock, J., & Chang, S.-F. (2001). Using Human Observers' Eye Movements in Automatic Image Classifiers. In *proceedings of SPIE Human Vision and Electronic Imaging VI*, San Jose, CA.

Jain, R. (1993) Workshop report: NSF workshop on visual information management systems. *Proc. SPIE Storage and Retrieval for Image and Video Databases.*

Jain, R., Pentland, A., & Petkovic, D. (1995) NSF-ARPA workshop on visual information management systems. Cambridge, MA.

Jarodzka, H., Scheiter, K., Gerjets P., & Van Gog, T. (2009). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and instruction, in press.*

Ji, L., Peng, K. & Nisbett, R. E. (2000) *Journal of Personal and Social Psychology, 78,* 943–955.

Johansson, R., Holsanova, J., Holmqvist, K. (2006) Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science: A Multidisciplinary Journal, 30(6),* 1053-1079.

Jose, J. M., Furner, J., & Harper, D, J. (1998). Spatial querying for image retrieval: a user-oriented evaluation. In *Proceedings of ACM SIGIR '98*, 232–240.

Josephson, S., & Holmes, M. E. (2002). Attention to repeated images on the World-Wide Web: Another look at scanpath theory. *Behaviour Research Methods, Instruments, & Computers, 34,* 539-548.

Joubert, O. R., Rousselet, G. A., Fize, D., Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47,* 3286-3297.

Kensinger, E. A., Corkin, S., (2003)a. Effect of Negative Emotional Content on Working Memory and Long-Term Memory. *Emotion, 3(4),* 378-393

Kensinger, E. A., Corkin, S., (2003)b. Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory & Cognition*, *31(8),* 1169-1180.

Kern RP, Libkuman TM, Otani H, Holmes K. (2005). Emotional stimuli, divided attention, and memory. *Emotion, 5(4),* 408-417.

Kerne, A. (2000). Collage machine: An interactive agent of web recombination. *Leonardo, 33(5),* 347 350.

Kerne, A. (2001). Collage machine: Interest-driven browsing through streaming collage. In *Proceedings of Cast: Living in Mixed Realities. Bonn, Germany,* 241-244.

Kienzle, W., Wichmann, F. A., Schölkopf, B. and Franz, M. O. (2006). A nonparametric approach to bottom-up visual saliency. In *Advances in Neural Information Processing Systems, 19,* 689-696. MIT Press, Cambridge, MA, USA.

Kirouac, G., & Dore´, F. Y. (1985). Accuracy of the judgement of facial expression of emotions as a function of sex and level of education. *Journal of Nonverbal Behaviour, 91,* 3-7.

Kissler, J., & Keil, A., (2008). Look–don't look! How emotional pictures affect pro and anti-saccades. *Experimental brain research, 188*, 215–222.

Kitagami, S., Sato, W., & Yoshikawa, S. (2002). The influence of test–set similarity in verbal overshadowing. *Applied Cognitive Psychology, 16,* 963–972.

Kitayama, S., Duffy, S., Kawamura, T. & Larsen, J. T. (2003). Perceiving an abject and its context in different cultures: A cultural look at new look. *Psychological Science, 14,* 201-206.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*, 809-816.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*, 219-227.

Kojima, T., Matsumisha, E., Nakajima, K., Shiraishi, H., Ando, K., Anod, H., Shimazono, Y. (1990). Eye movements in acute, chronic and remitted schizophrenics. *Biological Psychiatry, 27,* 975-989.

Kosslyn, S. (1973). Scanning visual images: Some structural implications. *Perception & Psychophysics, 14,* 90-94.

Kosslyn, S. M. (1994). *Image and Brain: The Resolution of the Imagery Debate.* MIT Press, Cambridge.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., Hamilton, S. E., Rauch, S. L. and Buonanno, F. S. (1993). Visual mental imagery activates topographically organized visual cortex: PET investigations. *Journal of cognitive Neuroscience, 5,* 263-287.

Kosslyn, S. M., Shin, L. M., Thompson, W. L., McNally, R. J., Rauch, S. L., Pitman R. K., & Alpert, N. M. (1996) Neural effects of visualizing and perceiving aversive stimuli: A PET investigation. *NeuroReport, 7,* 1569-1576.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the

programming of saccades. *Vision Research, 35*, 1897-1916.

Kundel, H. L., & La Follette, P. S. (1972). Visual search patterns and experience with radiological images. *Radiology, 103*, 523-528.

Kundel, H. L., Nodine, C. F., & Toto, L. (1984). Eye movements and the detection of lung tumours in chest images. In: Gale, A. G. & Johnson, F. (Eds.), *Theoretical and applied aspects of eye movement research.* Amsterdam: North-Holland.

LaBar, K. S., Mesulam, M., Gitelman, D. R., Weintraub, S., (2000) Emotional curiosity: modulation of visuospatial attention by arousal is preserved in aging and early-stage Alzheimer's disease. *Neuropsychologia, 38*, 1734-1740.

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

Land, M. F., & Hayhoe, M., 2001. In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559-3565.

Land, M.F. & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3,* 1340-1345.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. (1993). Looking at pictures: Affective, facial, visceral, and behavioural reactions. *Psychophysiology*, *30,* 261-273.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): instruction manual and affective ratings. *Technical Report A4*, The Centre for Research in Psychophysiology, University of Florida.

Langton, S. R. H., Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6,* 541-568.

Langton, S. R. H., O'Donnell, C., Riby, D. M., & Ballantyne, C. J. (2006). Gaze cues influence the allocation of attention in natural scene viewing. *The Quarterly Journal of Experimental Psychology, 59(12),* 2056-2064.

Lansdale, M., Underwood, G., & Davies, C. (2009). Something overlooked? How experts in Change Detection Use Visual Saliency. *Applied Cognitive Psychology, in press.*

Latour, P., L. (1962) Visual threshold during eye movements. *Vision Research, 2,* 261-262.

Li, F. F., VanRullen, R., Koch, C., & Perona, P., (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596-9601.

Li, Z. (2002) A saliency map in primary visual cortex. *Trends in Cognitive Science, 6,* 9-16

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about 'weapon focus'. *Law and Human Behaviour, 11,* 55–62.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance* 4, 565-572.

Loughland, C. M., Williams, L. M., & Gordon, E. (2002). Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. *Schizophrenia Research, 55,* 159-170.

Lowe, R. (2004). Interrogation of a dynamic visualization during learning. *Learning and Instruction, 14,* 257-274.

Ma, W. Y., & Manjunath, B. S. (1997) Netra: A toolbox for navigating large image databases. *Proc. IEEE International Conference on Image Procedures.*

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. Perception & Psychophysics, 2, 547-552.

Maidenberg, E., Chen, E., Craske, M., Bohn, P., & Bystritsky, A. (1996). Specificity of attentional bias in panic disorder and social phobia. *Journal of Anxiety Disorders, 10,* 529-541.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., Barton, J. J. S. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-

driven effects. *Journal of Vision, 8*, 1-9.

Mandal, M. K., & Palchoudhury, S. (1985). Perceptual skill in decoding facial affect. *Perceptual and Motor Skills, 60,* 96-98.

Mandal, M. K., & Rai, A. (1987). Responses to facial emotion and Psychopathology. *Psychiatric Research, 20,* 317-323.

Mandal, M. K., Aboulnasr, T., & Panchanathan, S. (1996). Image indexing using moments and wavelets. *IEEE Transactions on Consumer Electronics, 42*(3), 557-565.

Mannan, S. K., Ruddock, K., & Wooding, D. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11,* 157-178.

Mannan, S., Ruddock, K., & Wooding, D. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images. *Spatial Vision, 9(3),* 363-368.

Manning, D., Ethell, S., Donovan, T., Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography* 12, 134-142.

Mast, F. W., & Kosslyn, S. M. (2002). Eye movements during visual mental imagery. *Trends in Cognitive Science, 6,* 271-272.

Masuda, T. & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personal and Social Psychology, 81,* 922-934.

Mather, M., & Carstensen, L. L. (2003). Aging and attentional biases for emotional faces. *Psychological Science*, 14, 409-415.

Matsumoto, D., & Ekman, P. (2004). *Japanese and Caucasian facial expressions of emotion and neutral faces.* Berkeley, CA: Paul Ekman & Associates.

Mattia, J. I., Heimberg, R. G., & Hope, D. A. (1993). The revised Stroop colour naming task in social phobics. *Behaviour Research and Therapy, 31,* 305-313.

Mazer, J. A. & Gallant, J. L. (2003) Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron 40,* 1241-1250

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., Boot, W. R. (2004). Visual skills in airport security screening. *Psychological Science, 15,* 302-306.

McKelvie, S. J., (1976). The effects of verbal labelling on recognition memory for schematic faces. *Quarterly Journal of Experimental Psychology, 28,* 459-474.

McPeek, R. M. & Keller, E. L. (2002). Saccade target selection in the superior colliculus during a visual search task. *Journal of Neurophysiology, 88,* 2019-2034.

McPeek, R. M., & Keller, E. L. (2004). Deficits in saccade target selection after inactivation of superior colliculus. Nature Neuroscience7, 757-763.

Meissner C. A., & Brigham J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15(6),* 603-616.

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition, 29,* 176–186.

Melcher, J. M. & Schooler, J. W. (1996). The mis-rememberance of wines past: Verbal and perceptual expertise differentially mediate verbal overshadowing of taste memory. *Journal of Memory and Language, 35,* 231-245.

Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1995). A positron emission tomography study of visual and mental exploration. *Journal of Cognitive Neuroscience, 7,* 433-445.

Miltner, W., H., R., Krieschel, S., Hecht, H., Trippe, R., & Weiss, T. (2004). Eye Movements and Behavioural Responses to Threatening and Nonthreatening Stimuli During Visual Search in Phobic and Non phobic Subjects. *Emotion, 4(4),* 323-339

Minka, T. P., & Picard, R. W. (1997) Interactive learning with a "society of models". *Pattern Recognition, 30,* 565-581

Mogg, K., & Bradley, B. P. (2002). Selective orienting of attention to masked threat faces in social anxiety. *Behaviour Research and Therapy, 40*, 1403-1414.

Mogg, K., & Bradley, B. P. (2006). Time course of attentional bias for fear-relevant pictures in spider-fearful individuals. *Behaviour Research and Therapy, 44,* 1241-1250.

Mogg, K., Bradley, B. P., Miles, F., & Dixon, R. (2004). Time course of attentional bias for threat scenes: testing the vigilance-avoidance hypothesis. *Cognition & Emotion, 18*(5), 689-700.

Montagne, B., Kessels, R. P., Frigerio, E., de Haan, E. H., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack sensitivity? Cognitive *Processing, 6,* 136–141.

Moreno, F. J., Reina, R., Luis, V., and Sabido, R. (2002). Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual and Motor Skills, 95(3 pt 1),* 901-902.

Morris, J., Ohman, A. & Dolan, R. J. (1998) *Nature (London), 393,* 467-470.

Muzekari, L. H., & Bates, M. E. (1977). Judgement of emotion among chronic schizophrenics. *Journal of Clinical Psychology, 32,* 662-666.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology: Learning Memory & Cognition, 14,* 553-57.

Narasimhalu, A. D. (1995). Special section on content-based retrieval. *Multimedia Systems*

Navalpakkam, V., & Itti, L. (2005). Modelling the influence of task on attention. *Vision Research, 45*, 205-231.

Naveh-Benjamin, M., (1987). Coding of spatial location information: an automatic process? *Journal of Experimental Psychology: Learning, Memory & Cognition, 13,* 595-605.

Niblack, W., & Barber, R. (1994). The QBIC project: Querying images by content using colour, texture and shape. *Proc SPIE Storage and Retrieval for Image and Video Databases*.

Nodine, C. F., Krupinski, E. A. (1998). Perceptual skill, radiology expertise, and visual test performance with NINA and WALDO. *Academic Radiology, 5*(9), 603-612.

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. American Journal of Roentgenology, 179, 917-923.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11* (9), 929.

Nowicki, S., Jr., & Hartigan, M. (1988). Accuracy of facial affect recognition as a function of locus of control orientation and anticipated interpersonal interaction. *Journal of Social Psychology, 128,* 363-372.

Nummenmaa, L., Hyona, J., & Calvo, M. G. (2006). Eye Movement Assessment of Selective Attentional Capture by Emotional Pictures. *Emotion, 6(2)*, 257-268

Öhman, A. (1993). Stimulus prepotency and fear learning: Data and theory. In N. Birbaumer & N. Öhman (Eds.), *The structure of emotion,* 218-239. Seattle, Gottingen: Hogrefe & Huber.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108,* 483-522.

Ohman, A., Esteves, F. & Soares, J. J. F. (1995). Preparedness and preattentive associative learning: Electrodermal conditioning to masked stimuli. *Journal of Psychophysiology,* 9, 99-108.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130,* 466 478.

Öhman, A., Flykt, A., & Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data and neuropsychological mechanisms. In Lane, R. D. & Nadel, L. (Eds.), *Cognitive neuroscience of emotion*, 296-327. New York: Oxford University Press.

Ohman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80,* 381-396.

Ottati,W. L., Hickox, J. C., & Richter, J. (1999). Eye Scan Patterns of Experienced and Novice Pilots During Visual Flight Rules (VFR) Navigation. In *Proceedings of the Human Factors and Ergonomics Society, 43rd Annual Meeting.* Minneapolis, MN.

Oyekoya, O. K., & Stentiford, F. W. M. (2004). Eye tracking as a new interface for image retrieval. *British Telecommunications Technology Journal, 22(3).*

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review, 76,* 241 263.

Paivio, A. (1971). *Imagery and verbal processes.* Hillsdale, N.J.: Erlbaum.

Parkhurst, D., Law, K. & Niebur, E. (2002). Modelling the role of salience in the allocation of overt visual attention. *Vision Research, 42,* 107-123.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders, 32,* 249-261.

Pentland, A., & Picard, R. (1996) Special issue on digital libraries. *IEEE Transactions on Pattern Recognition and Machine Intelligence*

Pentland, A., Picard, R. W., & Sclaro, S. (1996) Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision, 18*(3), 233-254.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neutral and emotional stimuli. *Cognitive Brain Research, 15,* 31-45.

Peters, R.J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision Research* 45, 2397-2416

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2000). Visual search has memory. *Psychological Science*, *12*, 287-292.

Pflugshaupt, T., Mosimann, U. P., Schmitt, W. J., von Wartburg, R., Wurtz, P., Luthi, M., Nyffeler, T., Hess, C. W., & Muri, R. M. (2007). To look or not to look at threat? *Journal of Anxiety Disorders 21,* 353-366.

Pieters, R., Rosbergen, & Wedel, M. (1999). Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research, 36,* 424-438.

Pomplun, M., Reingold, E. M., & Shen, J. (2001). Investigating the visual span in comparative search: the effects of task difficulty and divided attention. *Cognition, 81,* 57-67

Pomplun, M., Reingold, E. M., Shen, J., & Williams, D. E. (2000). The area activation model of saccadic selectivity in visual search. In Gleitman, L. R. & Joshi, A. K. (Eds.), *Proceedings of the twenty-second annual conference of the cognitive science society,* 375-380. Mahwah, NJ: Erlbaum.

Postal, V. (2004). Expertise in cognitive psychology: testing the hypothesis of long-term working memory in a study of soccer players. *Perceptual and motor skills, 99(2)*, 403-420

Postle, B.R., Idzikowski, C., Della Sala, S., Logie, R.H., & Baddeley, A.D. (2006). The selective disruption of spatial working memory by eye movements. *The Quarterly Journal of Experimental Psychology, 59,* 100-120.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Perception & Performance, 2,* 509-522.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81,* 10-15.

Potter, M. C., Staub, A., Rado, J., & O'Connor, D. H. (2002). Recognition memory for briefly presented pictures: The time course of rapid forgetting. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1163-1175.

Potter, M.C. (1999) Understanding sentences and scenes: the role of conceptual short-term memory. In Coltheart, V. (Eds). *Fleeting Memories,* 13-46, MIT Press

Pratto, F, & John, O. P. (1991). Automatic vigilance: the attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-91.

Privitera, C. M., Stark, L. W. & Zangemeister, W. H. (2007). Bonnard's representation of the perception of substance. *Journal of Eye Movement Research, 1(1):3,* 1-6.

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

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

127-158.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory, *Behavioural and Brain Sciences, 25(2),* 157-238.

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

Quackenbush, S. W., & Barnett, M. A. (2001). Recollection and evaluation of critical experiences in moral development: A cross-sectional examination. *Basic and Applied Social Psychology*, 23, 55-64.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K., Castelhano, M. S., & Yang, J. (2009). Eye movements when looking at unusual/weird scenes: Are there cultural differences? Journal of Experimental Psychology: Learning, *Memory & Cognition, 35,* 254-259.

Read, J. D. (1979). Rehearsal and recognition of human faces. American Journal of Psychology, 92, 71-85.

Reingold, E. M., Charness, N., Pomplun, M. & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science, 12(1),* 48-55

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

Riddoch, M. J., & Humphreys, G. W. (1987). Visual object processing in optic aphasia: A case of semantic access agnosia. *Cognitive Neuropsychology, 4,* 131-185.

Rinck, M., & Becker, E. S. (2006). Spider fearful individuals attend to threat, then quickly avoid it: Evidence from eye movements. *Journal of Abnormal Psychology, 115,* 231-238.

Rinck, M., Reinecke, A., Ellwart, T., Heuer, K., & Becker, E.S. (2005). Speeded detection and increased distraction in fear of spiders: Evidence from eye movements. *Journal of Abnormal Psychology, 114,* 235-248.

Ristic, J., Friesen, C. K., & Kingstone, A. (2002). Are eyes special? It depends on how you look at it. *Psychonomic Bulletin & Review, 9,* 507-513.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12,* 94-99.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and decoding of negative facial emotions. *Journal of Nonverbal Behaviour, 12,* 139-148.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J., (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience* 5, 629-630.

Rui, Y., Huang, T. S, Ortega, M., & Mehrotra, S. (1998) Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Trans on Circuits and Video Technology,* 1-13.

Rutherford, M. D., & Towns, A. M. (2008). Scan Path Differences and Similarities during Emotion Perception in those With and Without Autism Spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 1371-1381 .

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J., (2000) Amnesia is a deficit in relational memory. *Psychological Science, 11,* 454-461.

Ryan, R. S., & Schooler, J. W. (1998) Whom Do Words Hurt? Individual Differences in Susceptibility to Verbal Overshadowing. *Applied Cognitive Psychology, 12*, 105-125.

Saariluoma, P., & Kalakoski, V. (1997) Skilled Imagery and Long-Term Working Memory. *American Journal of Psychology, 110(2),* 177-201.

Salis, D. L., (1980). Laterality effects with visual perception of musical chords and dot patterns. *Perception and psychophysics, 28,* 284-292.

Salkovskis, P. M., (1991). The importance of behaviour in the maintenance of anxiety and panic:

a cognitive account. *Behavioural Psychotherapy, 19*, 6-19.

Schatz, B., & Chen, H. (1996). Building large-scale digital libraries. *Computer.*

Schau, H. C. (1992) Shape recognition with scale and rotation invariance. *Optical Engineering, 31,* 268.

Schneider, F., Gur, R. C., & Shtasel, D. L. (1995). Emotional processing in Schizophrenia: neurobehavioural probes in relation to psychopathy. S*chizophrenia Research, 17,* 67-75.

Schneider, W. X., & Deubel, H. (2002). Selection-for-perception and selection-for-spatial motor-action are coupled by visual attention: a review of recent findings and new evidence from stimulus driven saccade control. In Prinz, W., & Hommel, B. (Eds.), *Attention and performance XIX: common mechanisms in perception and action,* 609–627. Oxford: Oxford University Press.

Schnell, T., & Wu, T. (2000). Applying eye tracking as alternative approach for activation of controls and functions in aircraft. *Proceedings of the 5th International Conference on Human Interaction with Complex Systems (HICS),* 113.

Scholten, M. R., Aleman, A., Montagne, B., & Kahn, R. S. (2005). Schizophrenia and processing of facial emotions: Sex matters. *Schizophrenia Research, 78,* 61-67.

Schooler JW, Ohlsson S, & Brooks K. (1993). Thoughts beyond words: when language overshadows insight. *Journal of Experimental Psychology: General, 122,* 166-183.

Schroder M, Rehrauer H, Seidel K, & Datcu M. (2000) Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Trans Geoscience and Remote Sensing*, *38,* 2288-2298

Schupp, H., T., Junghofer, M., Weike, A. I., & Hamm, A. (2004) The selective processing of briefly presented affective pictures: An ERP analysis. *Psychophysiology, 41*, 441-449

Schwartz, B. L., Rossee, R. B., Johri, S., & Deutsch, S. I. (1999). Visual scanning of facial expressions in schizophrenia. *Journal of Neuro-Psychiatry and Clinical Neuroscience, 11*, 103-106.

Schyns, P. & Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale dependent scene recognition. *Psychological Science, 5,* 195-200

Seamon, J. G., Schlegal, S. E., Hiester, P. M., Landau, S. M., & Blumenthal, B. F. (2002). Misremembering Pictured Objects: People of All Ages Demonstrate the Boundary Extension Illusion. The American Journal of Psychology, *115*(2), 151-167.

Servos, P., & Goodale, M. A., (1995). Preserved visual imagery in visual form agnosia, *Neuropsychologica, 33,* 1383-1394.

Shaw, R., Crisman, E., Loomis, A., & Laszewski, Z. (1990) The eye wink control interface: using the computer to provide the severely disabled with increased flexibility and comfort. In *Proceedings of the 3rd Annual IEEE Symposium on Computer-Based Medical Systems (CBMS '90)*, 105-111.

Shepard, R. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning & Verbal Behaviour, 6(1),* 156-163

Shepard, R. N., & **Metzler, J. (1971).** Mental Rotation of Three-Dimensional Objects, Science, *171*, 701-703.

Sheth, B. R. & Shimojo, S. (2001). Compression of space in visual memory. *Vision Research, 41,* 328 341.

Siegler, R. S., & Richards, D. D. (1982). The development of intelligence. In R J Steinberg (End.) *Handbook of Human Intelligence,* 897-971. Cambridge, England: Cambridge University Press.

Simons, D. J. & Rensink, R. A. (2005) *Trends in Cognitive Science, 9,* 16–20.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16,* 97-115.

Sloboda, J. A. (1978). Perception of contour in music reading. *Psychology of music, 6(2),* 3-20.

Sloboda, J. A. (1976). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology, 28,* 1-16.

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive ethology and exploring attention in real-world scenes. *Brain Research*, *1080*, 101-119.

Smith, C. N., Hopkins, R. O., & Squire, L. R. (2006). Experience-dependent eye movements,

awareness, and hippocampus-dependent memory. *Journal of Neuroscience, 26,* 11304-11312.

Smith, C. N. & Squire, L, R. (2008). Experience-Dependent Eye Movements Reflect Hippocampus Dependent (Aware) Memory. *The Journal of Neuroscience*, *28(48),* 12825–12833.

Smith, J. R., & Chang, S. F. (1997). Visually searching the web for content. *IEEE Multimedia Magazine, 4(3),* 12-20.

Smith, N. K., Cacioppo, J. T., Larsen, J. T. & Chartrand, T. L. (2003) May I have your attention, please: electrocortical responses to positive and negative stimuli. Neuropsychologia, **41,** 171-183

Sohn, Y. W., & Doane, S. M. (2004). Memory Processes of Flight Situation Awareness: Interactive Roles of Working Memory Capacity, Long-Term Working Memory, and Expertise**.** *The Journal of the Human Factors and Ergonomics Society, 46(3),* 461-475.

Spezio, M. L., Adolphs, R., Hurley, R. S. E., & Piven, J. (2007). Analysis of face gaze in autism with ''Bubbles''. *Neuropsychologia, 45*, 144-151.

Standing, L., Conezio, J., & Haber, R. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19(2),* 73-74.

Stark, L., & Ellis, S. R. (1981). Scanpaths revisited: cognitive models direct active looking. In Fisher, D. F., Monty, R. A., & Senders, J. W. (Eds.), *Eye movements: cognition and visual perception*, 193-227. Hillsdale, NJ: Lawrence Erlbaum.

Staszewski, J. J., (1990). Exceptional memory: the influence of practice and knowledge on the development of elaborative encoding strategies. In Schneider, W., & Weinert F. E. (Eds.). *Interactions among aptitudes, strategies and knowledge in cognitive performance*, 252-285. Springer, New York.

Stelmach, L., & Tam, W., J. (1994). Processing image sequences based on eye movements. In: Rogowitz, B.E., & Allebach, J.P. (Eds.), *Human Vision, Visual Processing and Digital Display V, 2179,* Proceedings of SPIE, SPIE, Bellingham, WA, 90-98.

Stelmach, L., Tam, W. J., & Hearty, P. (1991). Static and dynamic spatial resolution in image coding: an investigation of eye movements, in: *Human Vision, Visual Processing and Digital Display II, 1453,* SPIE, Bellingham, WA, 147-152.

Streit, M., Wolwer, W., & Gaebel, W. (1997) Facial affect recognition and visual scanning behaviour in the course of schizophrenia. *Schizophrenia research, 24*, 311-317.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Neuroscience, 3(8),* 764–769.

Tatler, B. W. (2007). The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14), 1-17.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*(5), 643-659.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology, 58*(5), 931-960.

Taylor, S. F., Liberzon, I., Fig L. I., Decker, L. R., Minoshima, S., & Koeppe, R. A. (1998). The Effect of Emotional Content on Visual Recognition Memory: A PET Activation Study. *Neuroimage, 8,* 188-197.

Thayer, J. F., & Johnsen, B. H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. Scandinavian Journal of Psychology, 41, 243-246

Thompson, K.G. & Bichot, N.P. (2005) A visual salience map in the primate frontal eye field. *Progress in Brain Research, 147,* 251-262.

Thompson, W. B. (1987). Music sight-reading skill in flute players. *Journal of General Psychology*, 114, 345-352.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520-522.

Thorpe, S. J., Fize, D., & Marlot, C. (1996) Speed of processing in the human visual system. *Nature 381,* 520-522

Tipples, J., Atkinson, A. P., & Young, A. W. (2002). The eyebrow frown: A salient social signal. *Emotion, 2,* 288-296.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766-786.

Tosi, V., Mecacci, L., Pasquali, E. (1997) Scanning eye movements made when viewing film: preliminary observations. *International Journal of Neuroscience, 92,* 47-52.

Underwood, G., Chapman, P., Berger, Z., & Crundall, D. (2003). Driving experience, attentional focusing, and the recall of recently inspected events. *Transportation Research Part F,* 6, 289-304

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology, 59 (11),* 1931-1949

Underwood, G., Foulsham, T., & Humphrey, K., (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17(6),* 812-834.

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L., & Bloyce, J. (2006). Eye movements during scene inspection: A test of the saliency map hypothesis. *European Journal of Cognitive Psychology*, *18 (3)*, 321-342.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology, 57A (1),* 165–182.

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary for object identification? Evidence from eye movements during the inspection of real-word scenes. *Consciousness & Cognition, 17(1),* 159-170.

Van der Linde, I., Rajashekhar, U., Bovik, A. C., & Cormack, L. K. (2008) *DOVES: A Database of Visual Eye Movements.* Available: http://live.ece.utexas.edu/research/doves

Vassallo, S., Cooper, S. L., & Douglas, J. M. (2009). Visual scanning in the recognition of facial affect: Is there an observer sex difference? *Journal of Vision, 9(3):11,* 1-10

Veltcamp, R.C. & Tanase, M. (2002). Content-Based Image Retrieval Systems: A Survey. *Technical Report UU-CS-2000-34;* University of Utrecht, Netherlands.

Verena, S.B., & Gordon., R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica* 129, 255-263.

Voss, J. F., Vesonder, G. T., & Spilich, G. J. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning and Verbal Behaviour, 19,* 651-667.

Wagner, H. L., MacDonald, C. J., & Manstead, A. S. R. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology, 50,* 737-743.

Walker, C., H. (1987). Relative importance of domain knowledge and overall aptitude on acquisition of domain-related information. *Cognition and Instruction 4 (1)*, 25-42

Walker-Smith, G. J., Gale, A. G., Findlay, J., M. (1977). Eye movement strategies involved in face perception. *Perception 6(3),* 313-326.

Ward, D. J., & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction, *Nature 418,* 838.

Waters, A. J., & Underwood, G. (1998). Eye movements in a simple music reading task: A study of expert and novice musicians. *Psychology of Music*, *26,* 46-60.

Waters, A., Underwood, G., & Findlay, J. (1997) Studying expertise in music reading: use of a pattern matching paradigm. *Perceptual Psychophysics, 59,* 477-488.

Wedel, M., Pieters, R. (2000). Eye Fixations on Advertisements and Memory for Brands: A Model and Findings. Marketing Science, *19*(4), 297-312

Wei-Gang, C., . Huang, C-L., & Hwang, W-L. (2007). Automatic EyeWinks Interpretation System for Human-Machine Interface. *Journal of Image and Video Processing, 2007.*

Wells, A. & Matthews, G. (1994) *Attention and Emotion: A Clinical Perspective* (Lawrence Erlbaum, Hove, U.K.).

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). EyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences. *ETRA 2006, San Diego, CA.*

Westerman, D. L. & Larsen, J. D. (1997). Verbal-overshadowing effect: Evidence for a general shift in processing. *American Journal of Psychology, 110(3),* 417-428.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B.& Jenike, M. A. (1998) Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge. *Journal of Neuroscience, 18,* 411-418.

Williams, A. M., Davids, K., & Williams, J. G. (1999). *Visual perception and action in sport.* London: E & FN Spon.

Williams, A., M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise & Sport*, *69(2)*, 111-128

Wyer, R. S., T. K. Srull. 1986. Human cognition in its social context. *Psychology Review,* 93, 322–359.

Yarbus, A. (1967). *Eye Movements and Vision.* New York: Plenum Press.

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice & Behaviour, 20,* 280–292.

Zangemeister, W. H., Oechsner, U., & Freksa, C. (1995). Short-term adaptation of eye movements in patients with visual hemifield defects indicates high level control of human scan path. *Optometry & Vision Science,* 72, 467-478.

Zelinsky, G., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. In Weiss, Y., Scholkopf, B., & Platt, J. *Advances in Neural Information Processing Systems, 18,* 1569–1576. MIT Press, Cambridge, MA, 2006.

8. Appendix
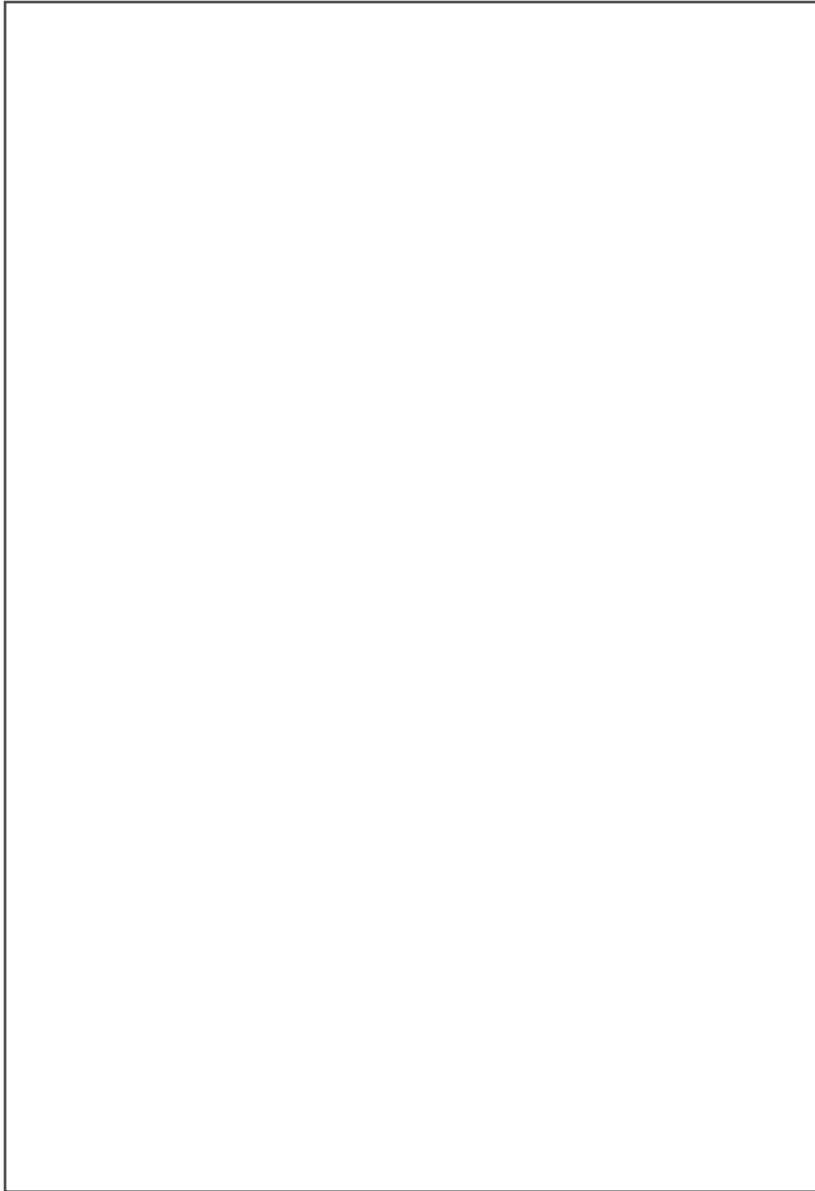
Appendix 1 (chapter 1): A diagram of Findlay & Walker's (1999) model of saccade generation.
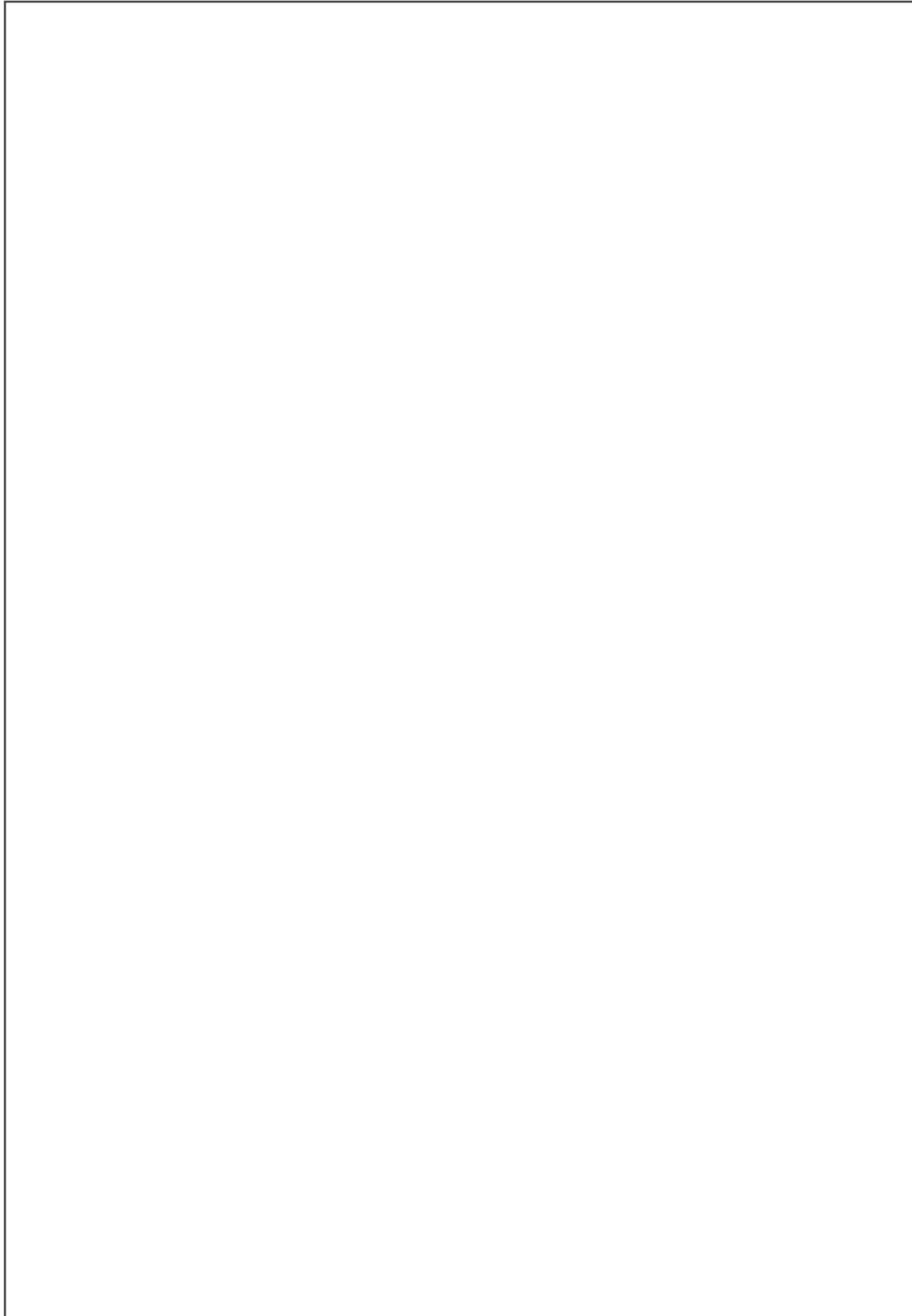
Appendix 1b: An overview of the saliency map model described by Itti and Koch (2000), as applied to a natural image. The image provides the input at the bottom of the model. Linear filtering extracts variations in colour, intensity and orientation across several scales using Gaussian pyramids. These are combined to give centre-surround contrast within each feature, and the features are then summed into a single saliency map. Thanks goes to Dr Tom Foulsham for the original creation of the above diagram.
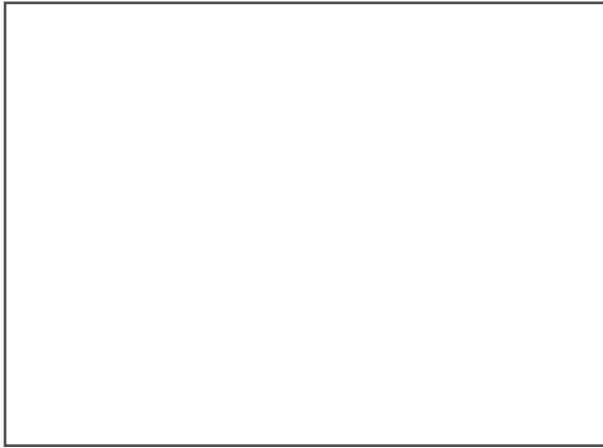
Appendix 1c: The algorithm used by Brandt & Stark (1997) to calculate string similarity.

Java code for string edit distance

```java
public double stringEditDistance()
{
//***************************
// Compute Levenshtein distance
//*****************************
//based on code at http://www.merriampark.com/ld.htm
int d[ ][ ]; // matrix
String a; // string from scanpath a
String b; // string from scanpath b
int n; // length of scanpath a
int m; // length of scanpath b
int i; // iterates through a
int j; // iterates through b
char a_i; // ith character of a
char b_j; // jth character of b
int cost; // cost
double sd; // the edit distance
double sdNorm; // the normalized edit distance
double sdSim; // the normalized similarity
// if either string is empty, distance is length of the
other
n = a.length ();
m = b.length ();
if (n == 0)
{
sd = m;
return sd;
}
if (m == 0)
{
sd = n;
return sd;
}
d = new int[n+1][m+1];
// set the first row/column to integers ascending from 1
for (i = 0; i <= n; i++)
{
d[i][0] = i;
}
for (j = 0; j <= m; j++)
{
d[0][j] = j;
}
// loop through the first string
for (i = 1; i <= n; i++)
{
s_i = a.charAt (i - 1);
// loop through the second string
for (j = 1; j <= m; j++)
{
```

```java
t_j = b.charAt (j - 1);
// compare the two characters
if (s_i == t_j)
{
cost = 0;
}
else
{
cost = 1; //cost for unequal
characters is 1
}
// set the current cell
d[i][j]
= findlowest (d[i-1][j]+1, d[i][j-1]+1, d[i-1][j-1]
+ cost);
}
}
// lowest distance is bottom left cell
sd = d[n][m];
//normalise the distance over the length of the longer
string
if (n>=m)
{
sdNorm = sd/n;
}
else
{
sdNorm = sd/m;
}
//similarity is 1 minus the normalised distance
sdSim = 1-sdNorm;
return sd;
}
public int findlowest(int int1,int int2,int int3)
//method to find the lowest of three integers
{
if (int2<int1)
{
int1=int2;
}
if (int3<int1)
{
int1=int3;
}
return int1;
}
```

Appendix 1d: The javacode used to calculate string editing distance to compare string similarity.

Appendix 2a (Exp 1): Group statistics for the accuracy T-test. 'Group 1' represents the imagery first task and 'group 2' represents the recognition first task.
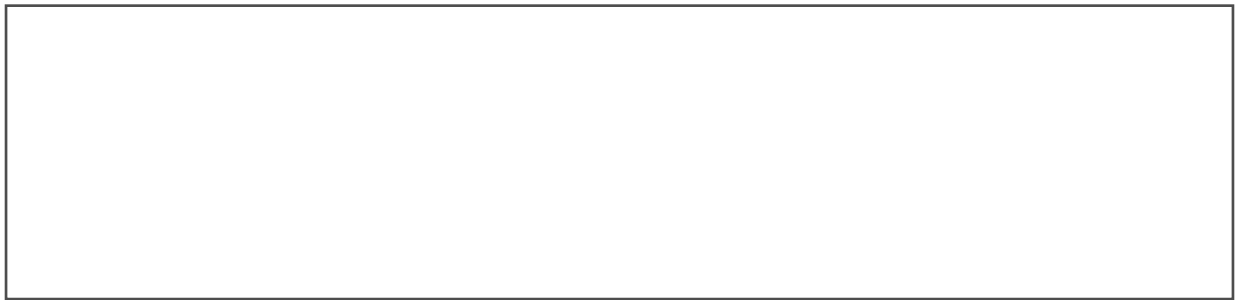
Appendix 2b (Exp 1): The independent samples T-test showing no reliable difference in accuracy between the groups.
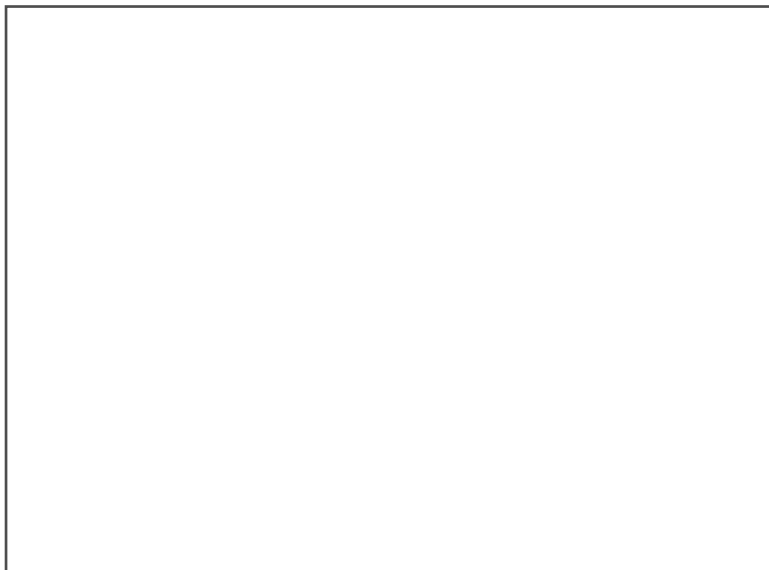
Appendix 3a (Exp 1): Descriptive statistics for the average fixation duration for each experimental condition, for each group. In both Tasks, participants made shorter fixations at encoding than at imagery or delayed imagery. Participants also made shorter fixations at recognition (old and new pictures) than at imagery or delayed imagery.

Appendix 3b (Exp 1): Results from a mixed-design ANOVA showing a reliable effect of group.
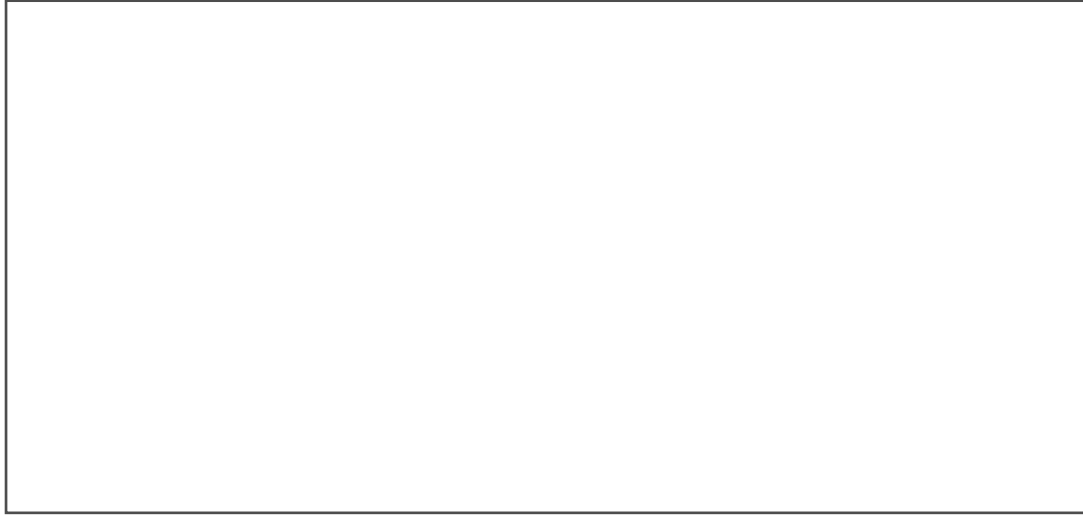
Appendix 3c (Exp 1): Results from a mixed-design ANOVA showing a reliable effect of test phase.

Appendix 3d (Exp 1): Descriptive statistics for 'group'. Overall, the Imagery first group make longer fixations than the Recognition first group.
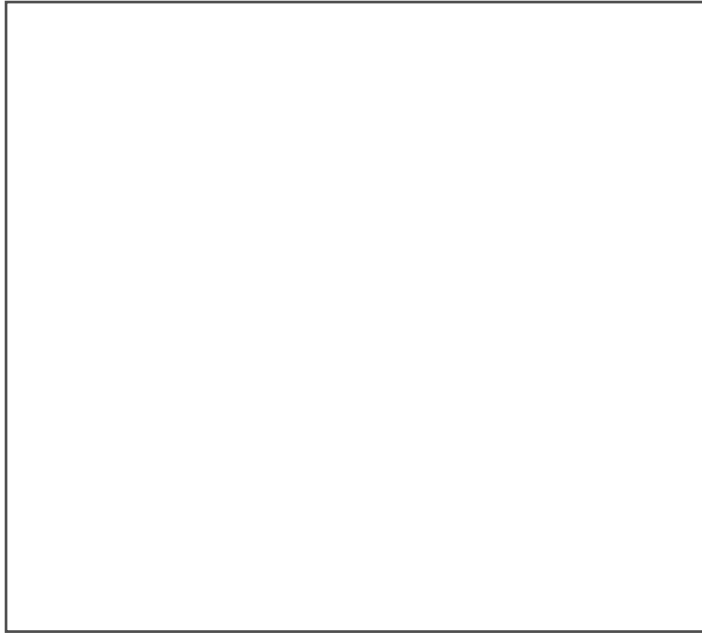
Appendix 3e (Exp 1): The statistics for the post-hoc paired samples comparisons

Appendix 3f (Exp 1): Post-hoc paired samples T-tests to further investigate the main effect of condition (test phase) on fixation duration.

Appendix 4a (Exp 1): Descriptive statistics for the average number of fixations made per picture in each condition for each group. Participants made more fixations at encoding then at imagery or delayed imagery. Participants also made fewer fixations at imagery and at delayed imagery than at recognition (old and new).

Appendix 4b (Exp 1): The results from a mixed-design ANOVA, revealing no reliable main effect of group.

Appendix 4c (Exp 1): A mixed-design ANOVA, revealing a reliable main effect of condition (test phase).
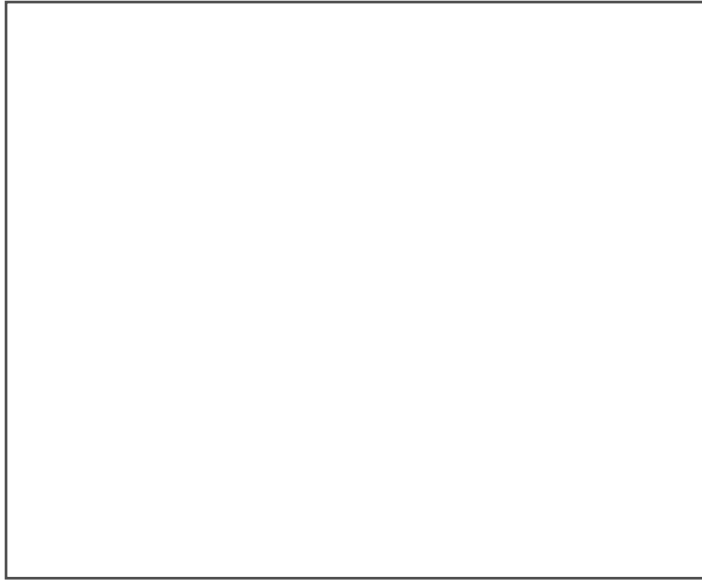
Appendix 4d (Exp 1): The statistics for the post-hoc paired samples comparisons

Appendix 4e (Exp 1): Post-hoc paired samples T-tests to further investigate the main effect of condition (test phase) on the average number of fixations per picture.

Appendix 5a(Exp 1): Descriptive statistics for the average saccadic amplitudes per picture in each condition for each group. Participants in the Recognition First group produced greater saccadic amplitudes than participants in the Imagery First group. Participants also produced greater saccadic amplitudes at imagery than at encoding and at delayed imagery than at encoding.

Appendix 5b (Exp 1): The results from a mixed-design ANOVA, revealing a reliable main effect of group.
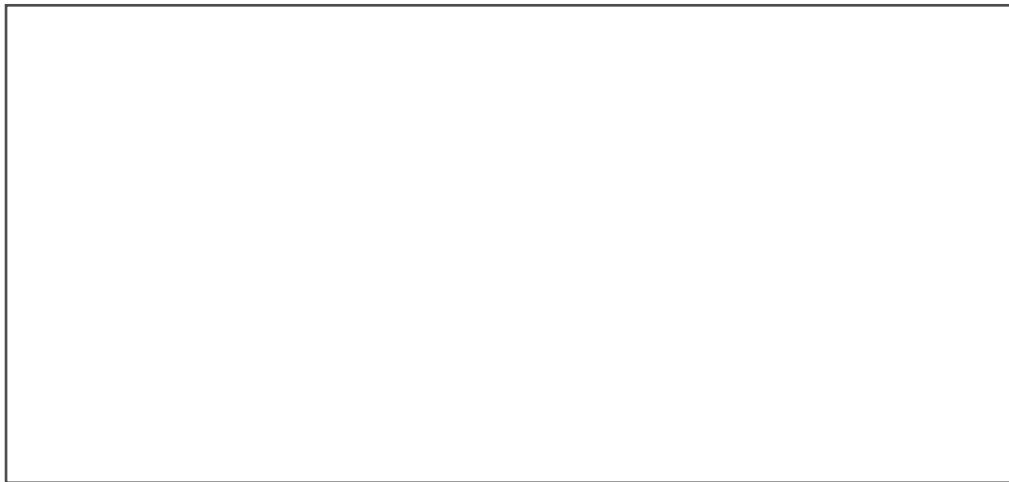
Appendix 5c(Exp 1): A mixed-design ANOVA, revealing a reliable main effect of condition (test phase).

Appendix 5d (Exp 1): Descriptive statistics for 'group'. Overall, the Imagery first group make longer fixations than the Recognition first group.
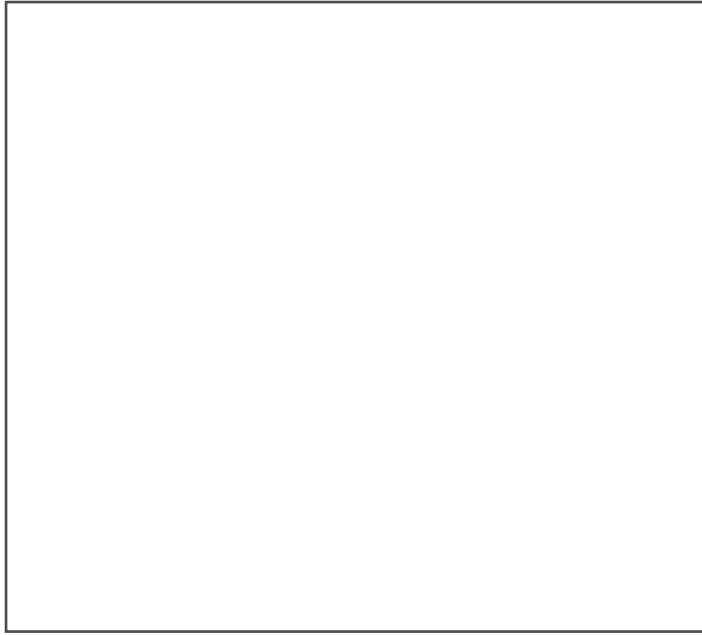
Appendix 5e (Exp 1): The statistics for the post-hoc paired samples comparisons
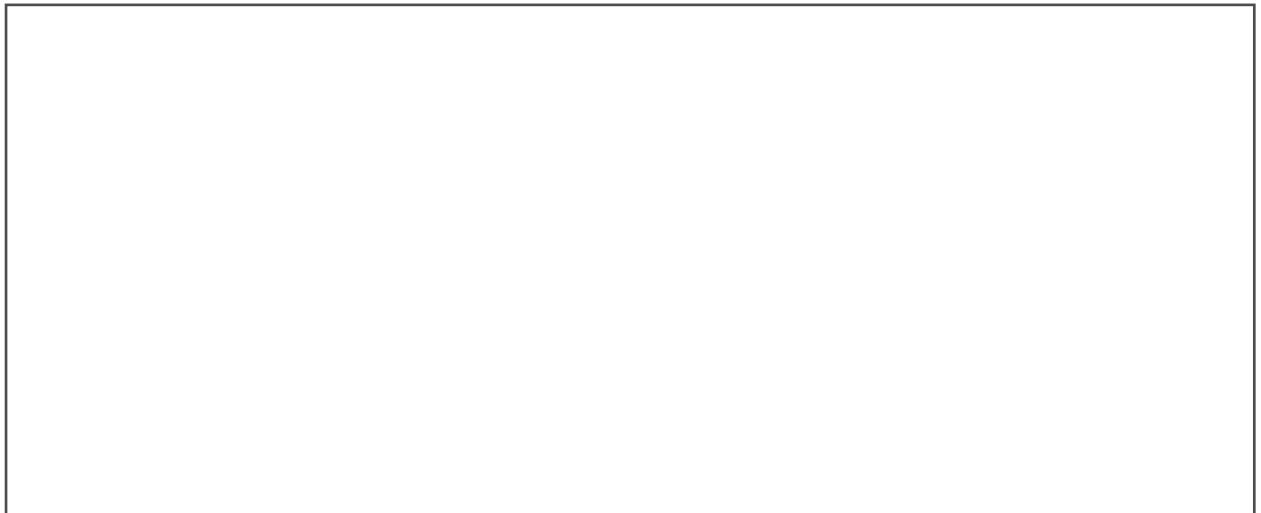
Appendix 5f (Exp 1): Post-hoc paired samples T-tests to further investigate the main effect of condition (test phase) on the average saccadic amplitude per picture.

Appendix 6a (Exp 1): Descriptive statistics for the average string similarities for each comparison, for each group.
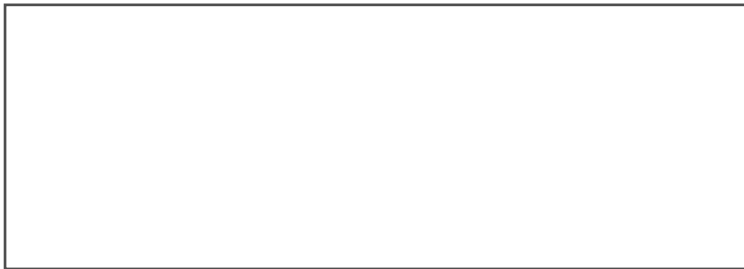
Appendix 6b (Exp 1): The results from a mixed-design ANOVA, revealing no reliable main effect of group.

Appendix 6c (Exp 1): A mixed-design ANOVA, revealing a reliable main effect of condition (test phase) and a reliable interaction between condition and group.

Appendix 6d (Exp 1): The statistics for the post-hoc paired samples comparisons for the Imagery First group

Appendix 6e (Exp 1): Post-hoc paired samples T-tests for the Imagery First group.

Appendix 6f (Exp 1): The statistics for the post-hoc paired samples comparisons for the Recognition First group
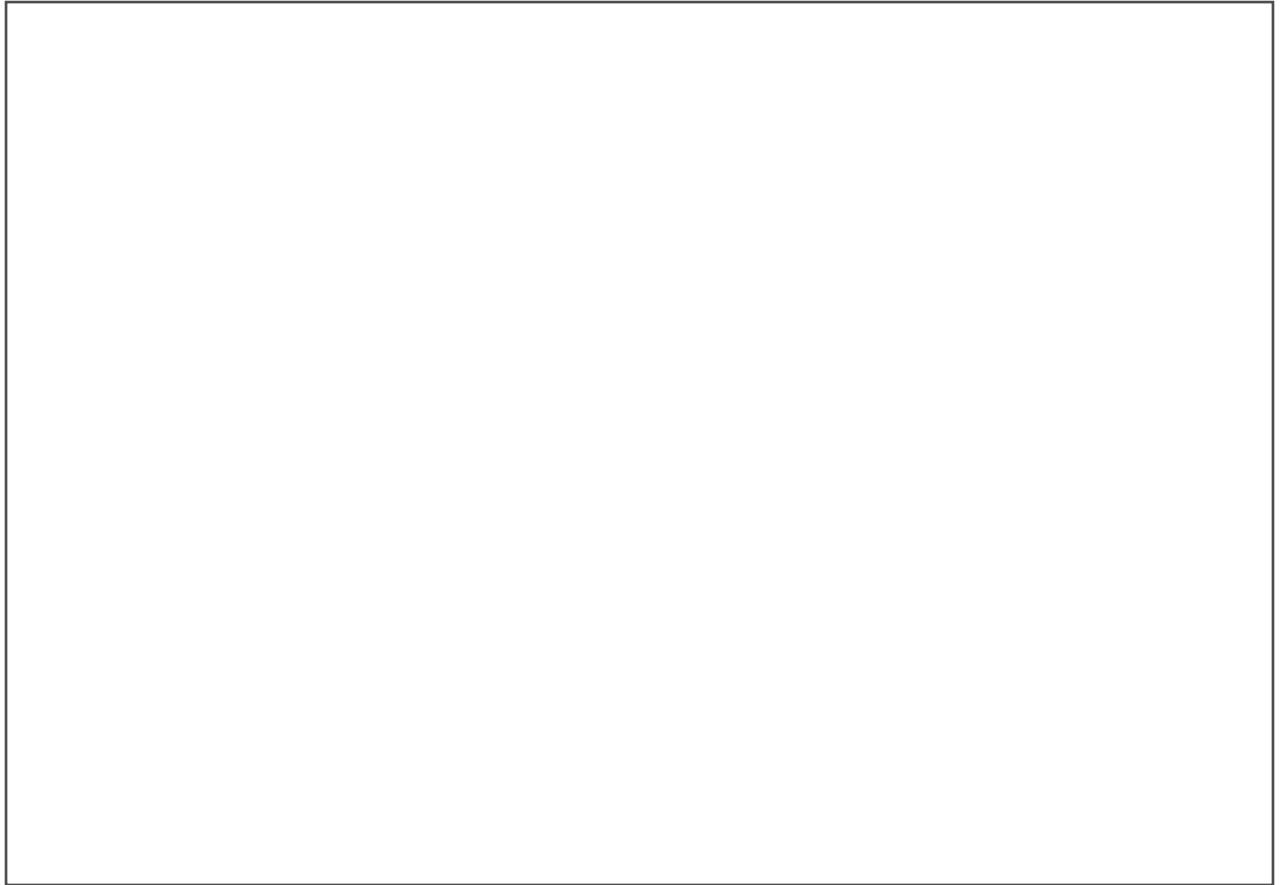
Appendix 6g (Exp 1): Post-hoc paired samples T-tests for the Recognition First group.

Appendix 7a (Exp 2): Statistics for the accuracy of old and new pictures at recognition test

Appendix 7b (Exp 2): Paired samples T-test for the accuracy of old and new pictures. There is a reliable difference – participants are more accurate at identifying new pictures than old ones.

Appendix 8a (Exp 2): Within group ANOVA for the average fixation duration per picture. There is a main effect of experimental condition.
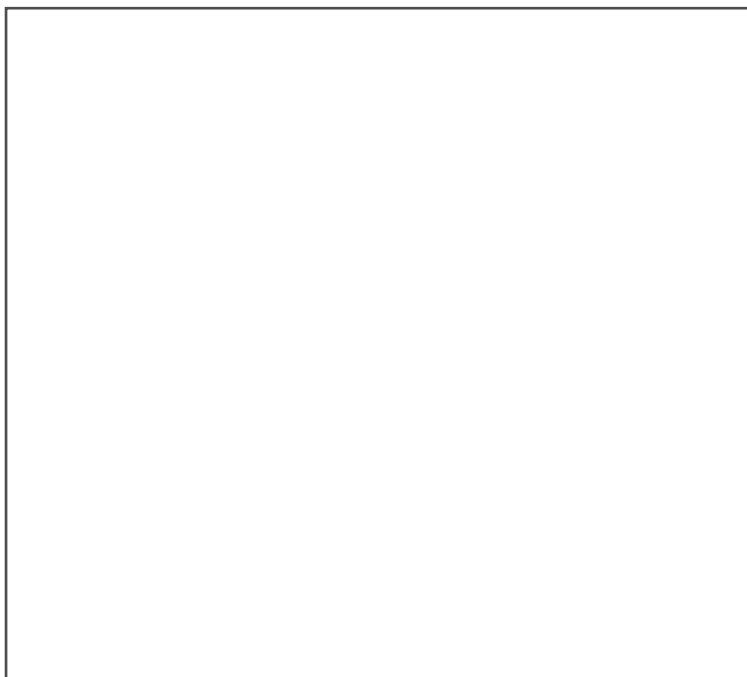
Appendix 8b (Exp 2): Statistics for the paired samples T-tests for average fixation duration per picture
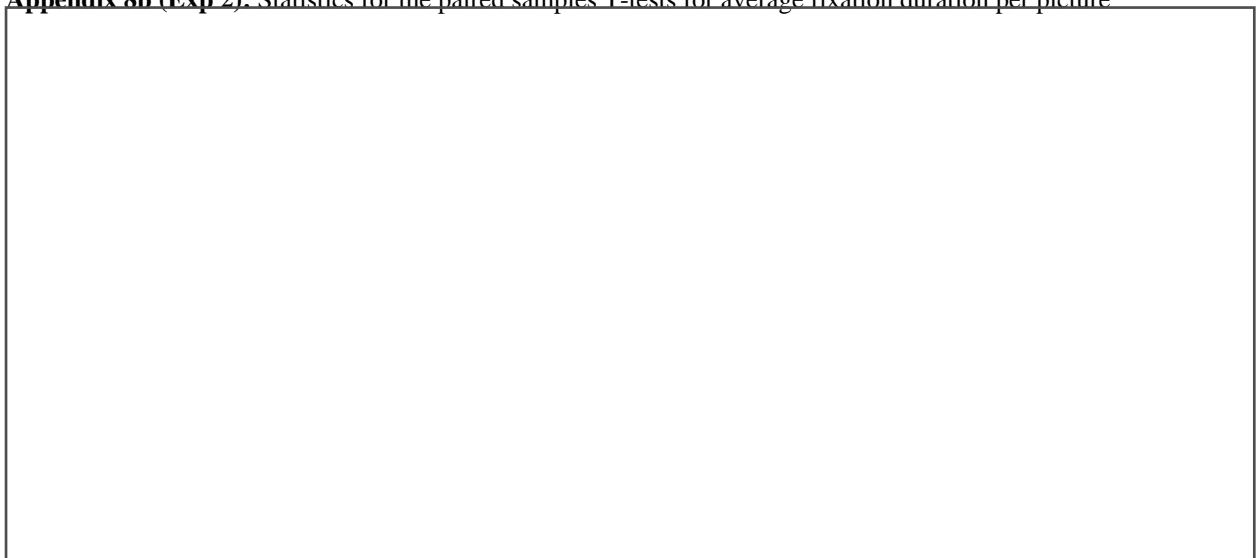
Appendix 8c (Exp 2): Paired samples T-test for average fixation duration per picture, for the different experimental conditions.
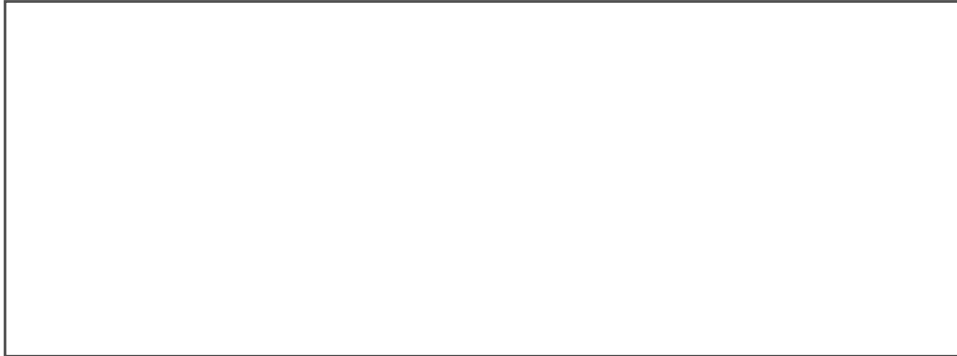
Appendix 9a (Exp 2): Within group ANOVA for the average number of fixations per picture. There is a main effect of experimental condition.
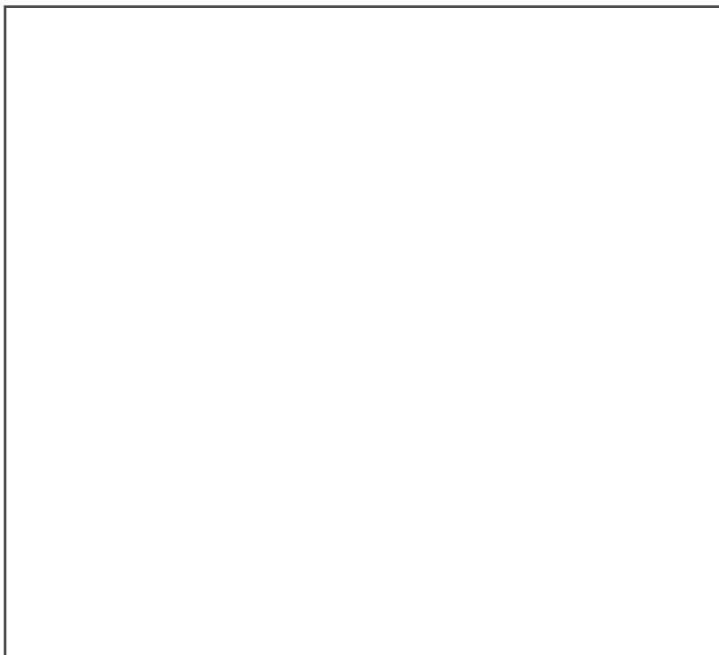
Appendix 9b (Exp 2): Statistics for the paired samples T-tests for average number of fixations per picture
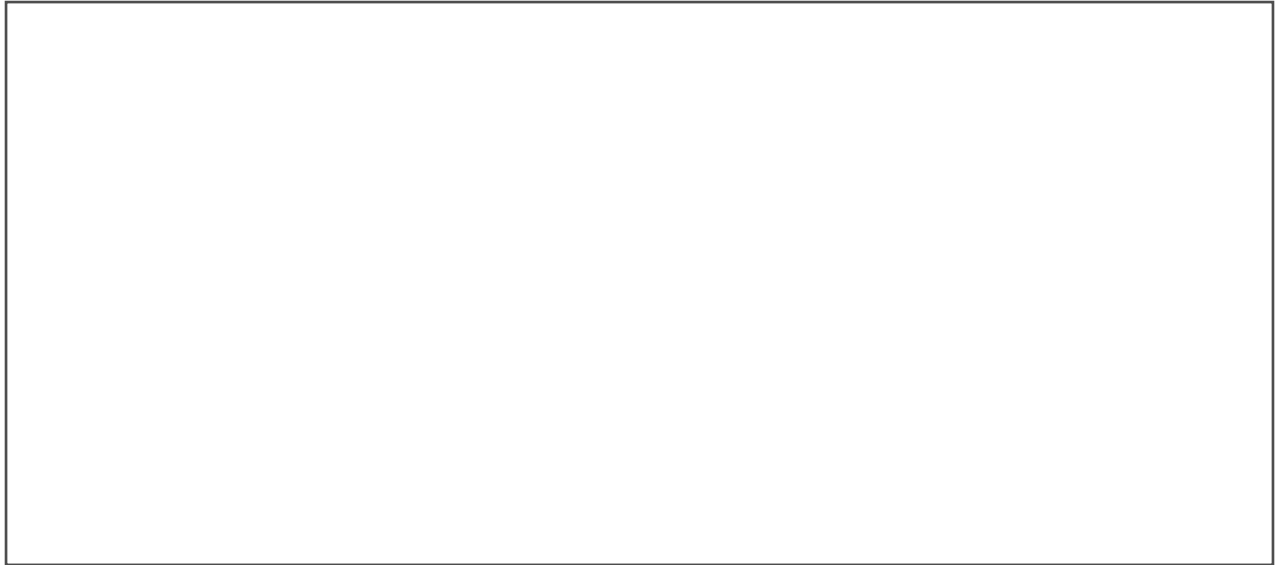
Appendix 9c (Exp 2): Paired samples T-test for average number of fixations per picture, for the different experimental conditions.

Appendix 10 (Exp 2): Within group ANOVA for the average saccadic amplitude per picture. There is no reliable main effect of experimental condition.

Appendix 11a (Exp 2): Within group ANOVA for the average saccadic amplitude per picture. There is a reliable main effect of experimental condition.
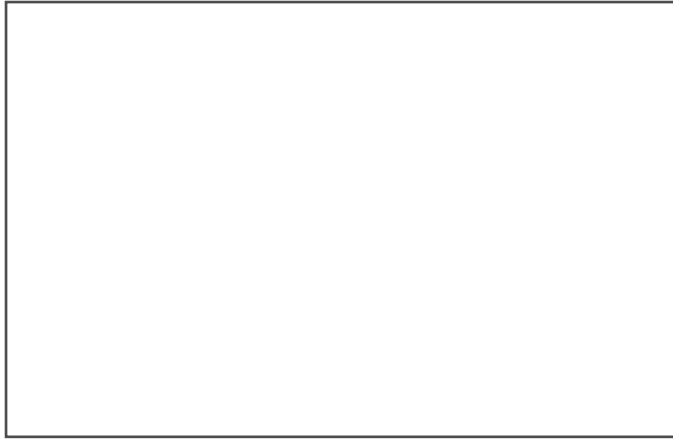
Appendix 11b (Exp 2): Statistics for the paired samples T-test for the average percentage of stimuli where participants fixated outside the picture boundary.

Appendix 11c (Exp 2): Paired samples T-test for the average percentage of stimuli where participants fixated outside the picture boundary.

Appendix 12a (Exp 2): Within group ANOVA for the average string similarity scores. There is a reliable main effect of experimental condition.
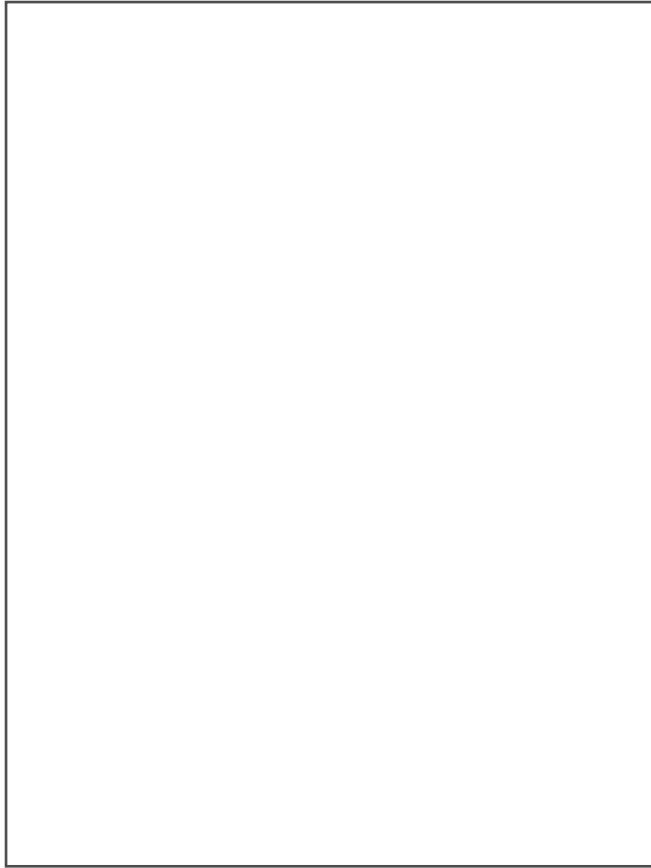
Appendix 12b (Exp 2): Statistics for the paired samples T-tests for the average string similarity score comparisons.

Appendix 12c (Exp 2): Paired samples T-tests for the average string similarity score comparisons.
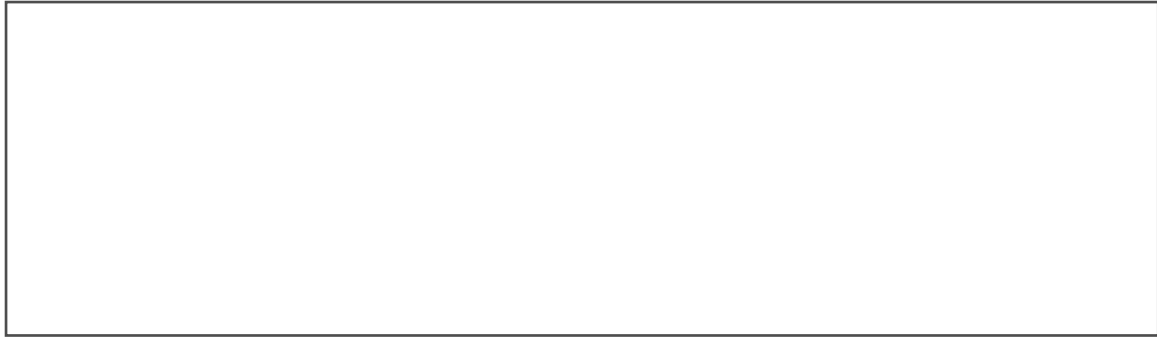
Appendix 12d (Exp 2): Paired samples T-tests comparing each string comparison against chance.

Appendix 13a (Exp 3): Paired samples T-tests comparing string similarities of encoding v imagery and encoding v verbalisation (describe). Strings were cropped to 5 letters.

Appendix 13b (Exp 3): Paired samples T-tests comparing each string comparison against chance. Strings were cropped to 5 letters.

Appendix 13c (Exp 3): Paired samples T-tests comparing string similarities of encoding v imagery and encoding v verbalisation (describe). Strings were cropped to 13 letters.

Appendix 13d (Exp 3): Paired samples T-tests comparing each string comparison against chance. Strings were cropped to 13 letters.

Appendix 13e (Exp 3): Paired samples T-tests comparing string similarities of encoding v imagery and encoding v verbalisation (describe). Strings were left unrestricted in length.

Appendix 13f (Exp 3): Paired samples T-tests comparing each string comparison against chance. Strings were left unrestricted in length.

Appendix 14a (Exp 4): Descriptive statistics for the mixed design ANOVA for accuracy between groups for the three types of stimuli.

Appendix 14b (Exp 4): ANOVA results showing a reliable main effect of group.

Appendix 14c (Exp 4): Mixed design ANOVA for accuracy between groups for the three types of stimuli. There is no reliable main effect of stimulus type, but there is a reliable interaction between stimulus type and group.
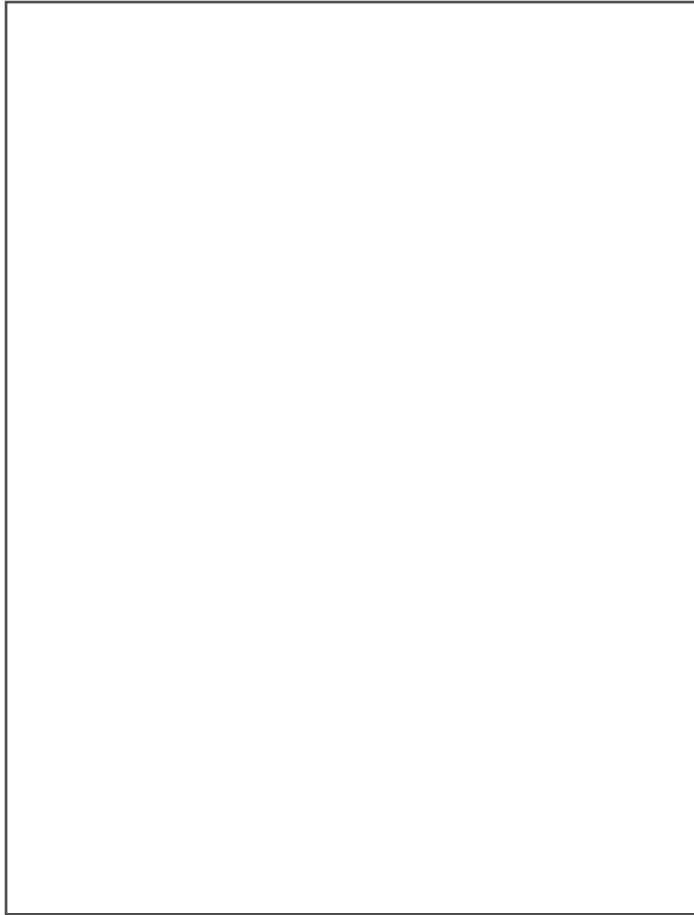
Appendix 14d (Exp 4): Statistics for the paired samples T-tests for accuracy at recognition test
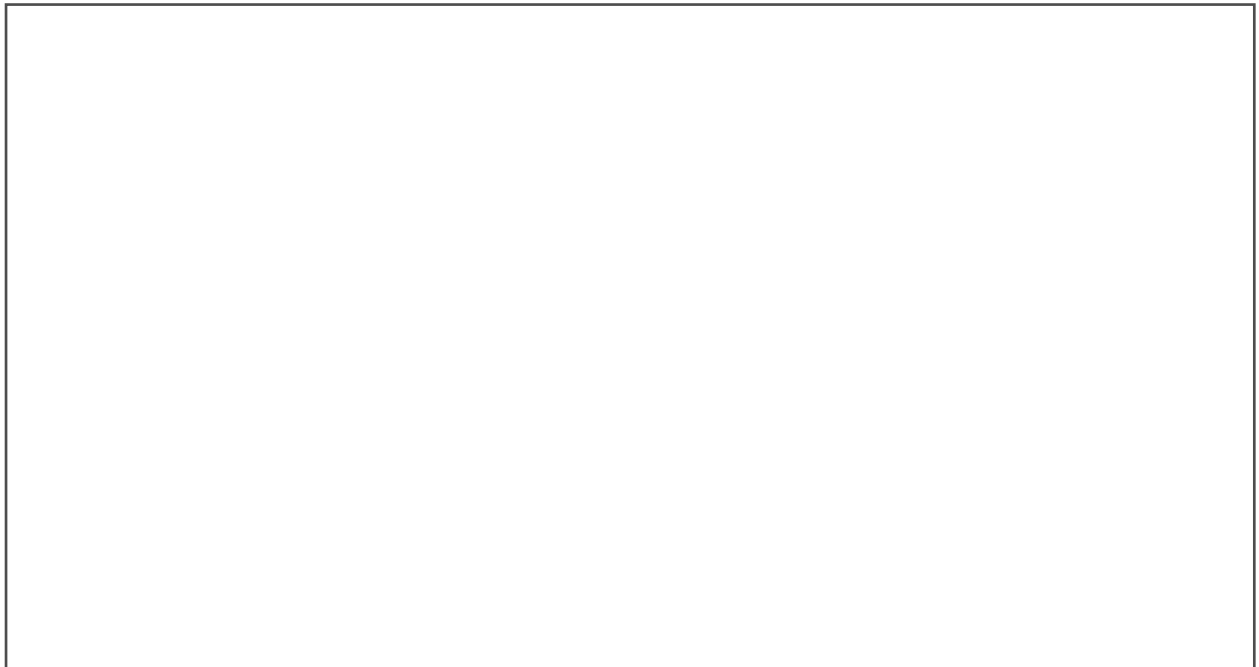
Appendix 14e (Exp 4): Paired samples T-tests for accuracy at recognition test
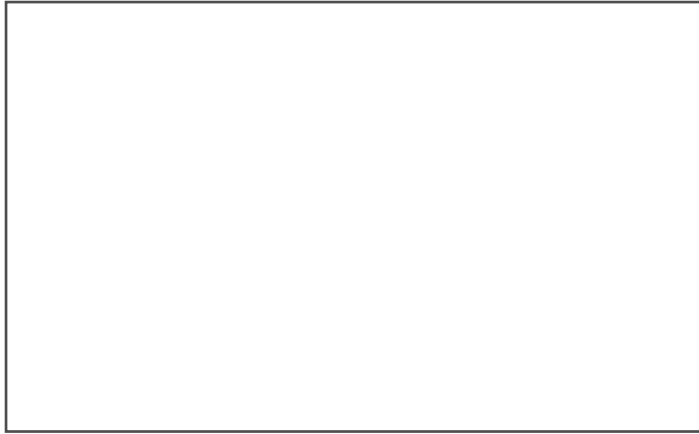
Appendix 15a (Exp 4): Descriptive statistics for the mixed design ANOVA for percentage of all fixations falling in the 5 most salient regions.

Appendix 15b (Exp 4): ANOVA results for the percentage of all fixations falling in the 5 most salient regions. There is no reliable main effect of group.

Appendix 15c (Exp 4): Mixed design ANOVA for percentage of all fixations falling in the 5 most salient regions. There is a reliable main effect of stimulus type and a reliable interaction between group and stimulus type.
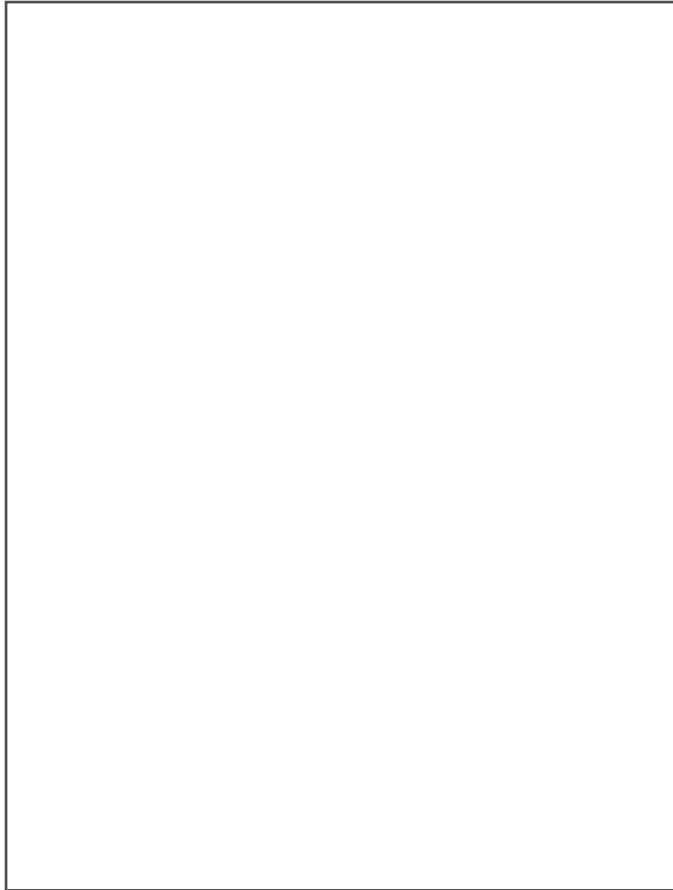
Appendix 15d (Exp 4): Statistics for the paired samples T-tests for percentage of all fixations falling in the 5 most salient regions. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
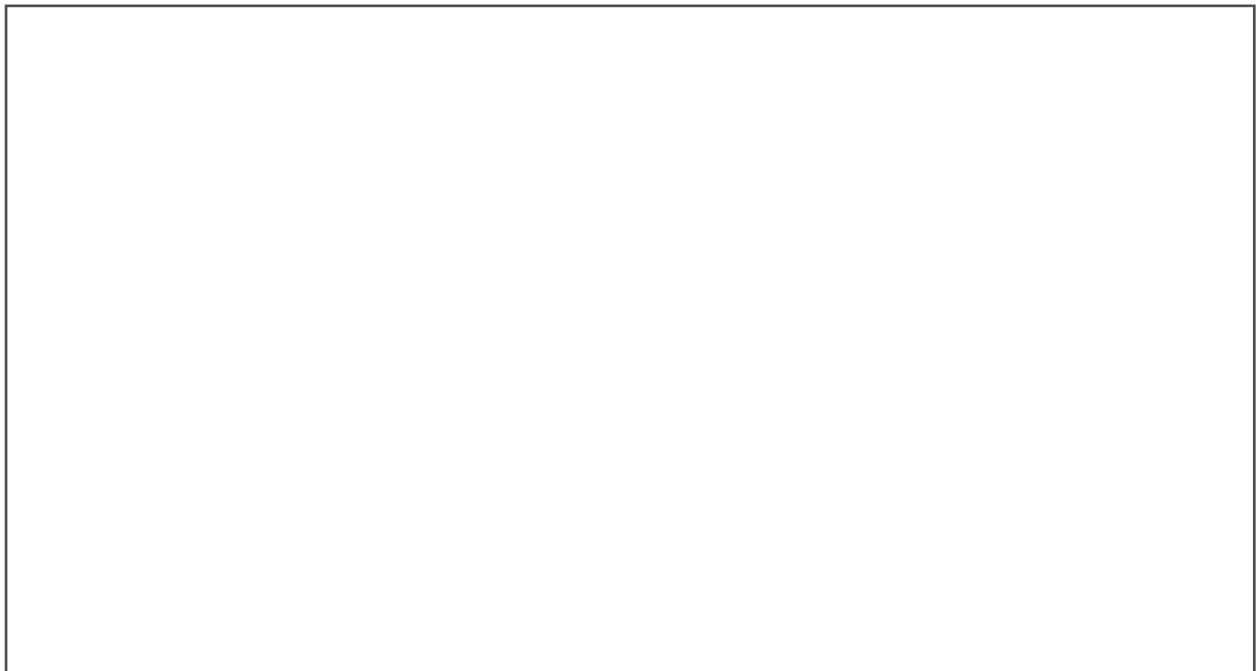
Appendix 15e (Exp 4): Paired samples T-tests for percentage of all fixations falling in the 5 most salient regions.

'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
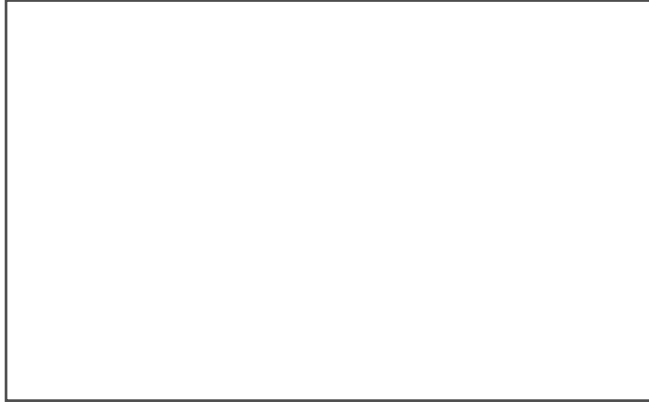
Appendix 16a (Exp 4): Descriptive statistics for the mixed design ANOVA for percentage of the first 5 fixations falling in the 5 most salient regions.

Appendix 16b (Exp 4): ANOVA results for the percentage of the first 5 fixations falling in the 5 most salient regions. There is no reliable main effect of group.

Appendix 16c (Exp 4): Mixed design ANOVA for percentage of the first 5 fixations falling in the 5 most salient regions. There is a reliable main effect of stimulus type and a reliable interaction between group and stimulus type.
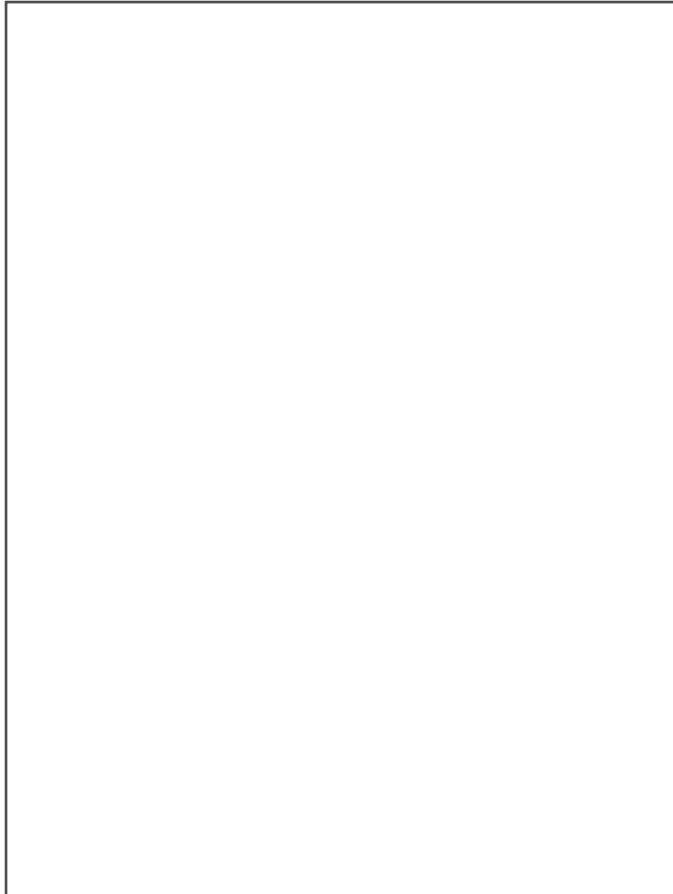
Appendix 16d (Exp 4): Statistics for the paired samples T-tests for percentage of the first 5 fixations falling in the 5 most salient regions. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
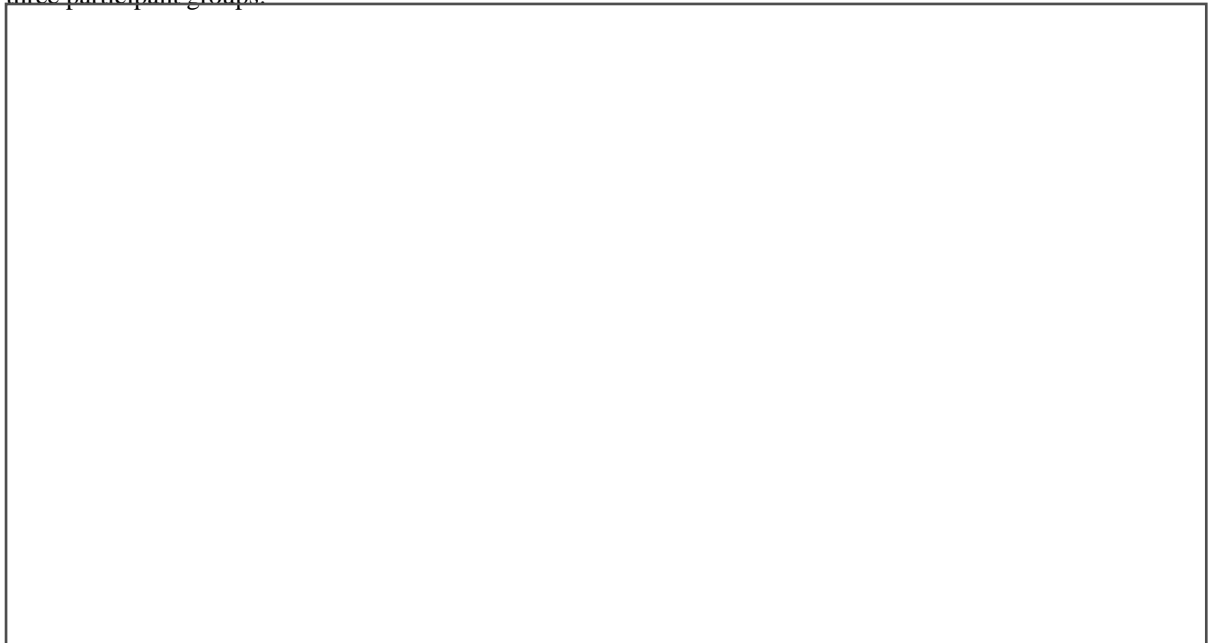
Appendix 16e (Exp 4): Paired samples T-tests for percentage of the first 5 fixations falling in the 5 most salient regions. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.

Appendix 17a (Exp 4): Descriptive statistics for the mixed design ANOVA for the average total number of fixations on each stimulus type, by each participant group.

Appendix 17b (Exp 4): ANOVA results for the average total number of fixations on each stimulus type, by each participant group. There is no reliable main effect of group.

Appendix 17c (Exp 4): Mixed design ANOVA for the average total number of fixations on each stimulus type, by each participant group. There is a reliable main effect of stimulus type and a reliable interaction between group and stimulus type.

Appendix 17d (Exp 4): Statistics for the average total number of fixations on each stimulus type, by each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
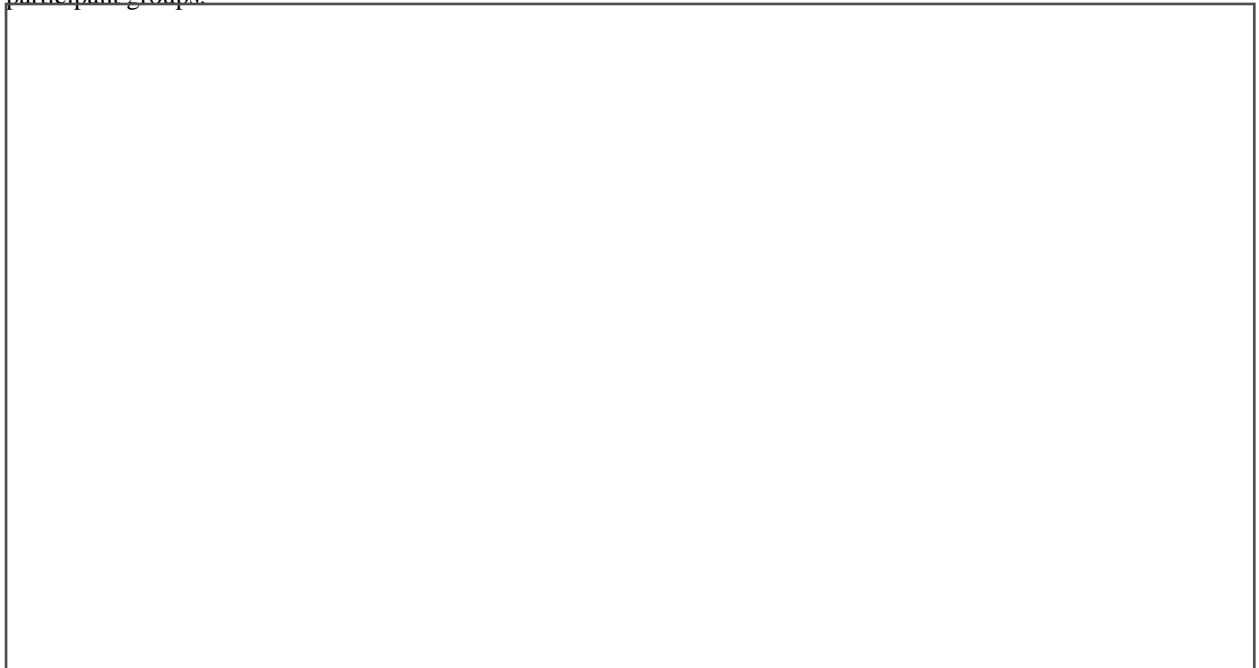
Appendix 17e (Exp 4): Paired samples T-tests for the average total number of fixations on each stimulus type, by each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.

Appendix 18a (Exp 4): Descriptive statistics for the mixed design ANOVA for the average saccadic amplitude on each stimulus type, by each participant group.

Appendix 18b (Exp 4): ANOVA results for the average saccadic amplitude on each stimulus type, by each participant group. There is no reliable main effect of group.

Appendix 18c (Exp 4): Mixed design ANOVA for the average saccadic amplitude on each stimulus type, by each participant group. There is a reliable main effect of stimulus type and a reliable interaction between group and stimulus type.

Appendix 18 (Exp 4): Statistics for the average saccadic amplitude on each stimulus type, by each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
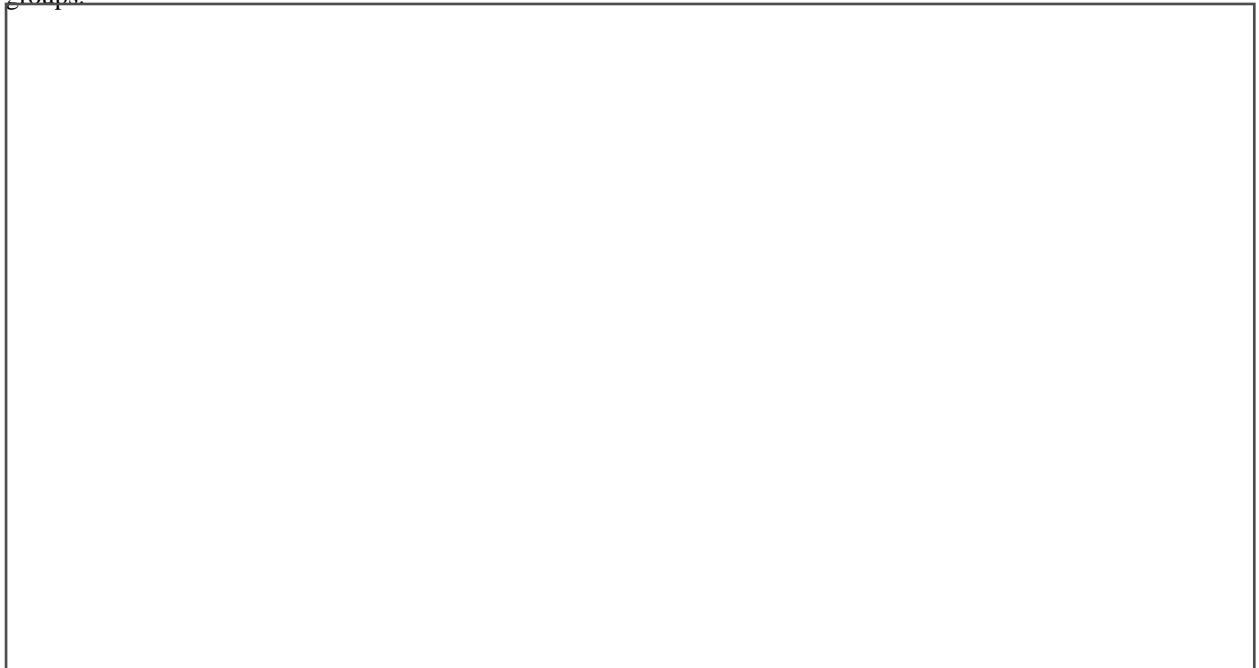
Appendix 18e (Exp 4): Paired samples T-tests for the average saccadic amplitude on each stimulus type, by each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.

Appendix 19a (Exp 4): Descriptive statistics for the average string similarity scores between encoding and model-predicted saliency scanpaths for each stimulus type and each participant group.

Appendix 19b (Exp 4): ANOVA results for the average string similarity scores between encoding and model-predicted saliency scanpaths for each stimulus type and each participant group. There is no reliable main effect of group.

Appendix 19c (Exp 4): Mixed design ANOVA for the average string similarity scores between encoding and model-predicted saliency scanpaths for each stimulus type and each participant group. There is a reliable main effect of stimulus type and a reliable interaction between group and stimulus type.

Appendix 19d (Exp 4): Statistics for the average string similarity scores between encoding and model-predicted saliency scanpaths for each stimulus type and each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
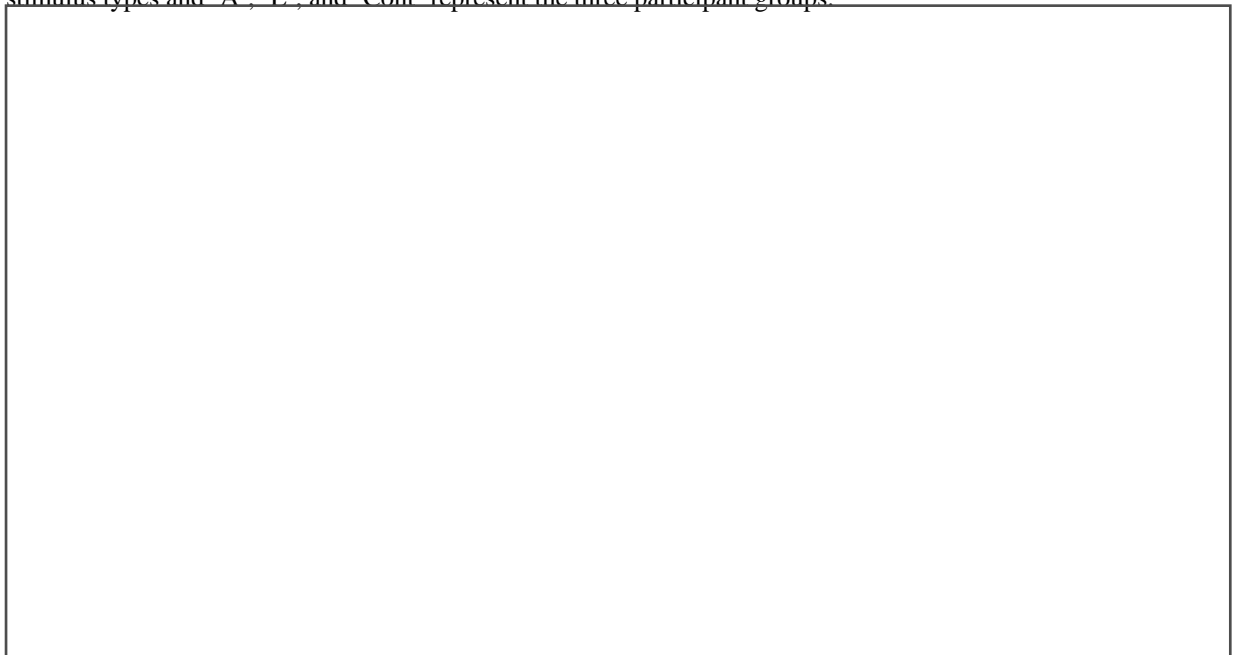
Appendix 19e (Exp 4): Paired samples T-tests for the average string similarity scores between encoding and model-predicted saliency scanpaths for each stimulus type and each participant group. 'Civil War', 'Eng' and 'Neut'

represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
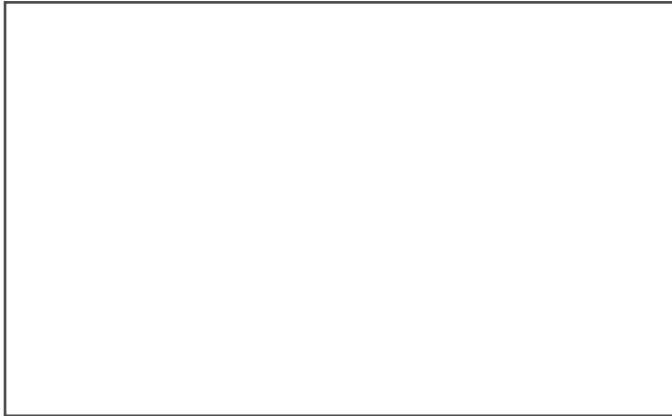
Appendix 20a (Exp 4): Descriptive statistics for the average string similarity scores between recognition and model-predicted saliency scanpaths for each stimulus type and each participant group.

Appendix 20b (Exp 4): ANOVA results for the average string similarity scores between recognition and model-predicted saliency scanpaths for each stimulus type and each participant group. There is no reliable main effect of group.

Appendix 20c (Exp 4): Mixed design ANOVA for the average string similarity scores between recognition and model-predicted saliency scanpaths for each stimulus type and each participant group. There is no reliable main effect of stimulus type but there is a reliable interaction between group and stimulus type.

Appendix 20d (Exp 4): Statistics for the average string similarity scores between recognition and model-predicted saliency scanpaths for each stimulus type and each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.
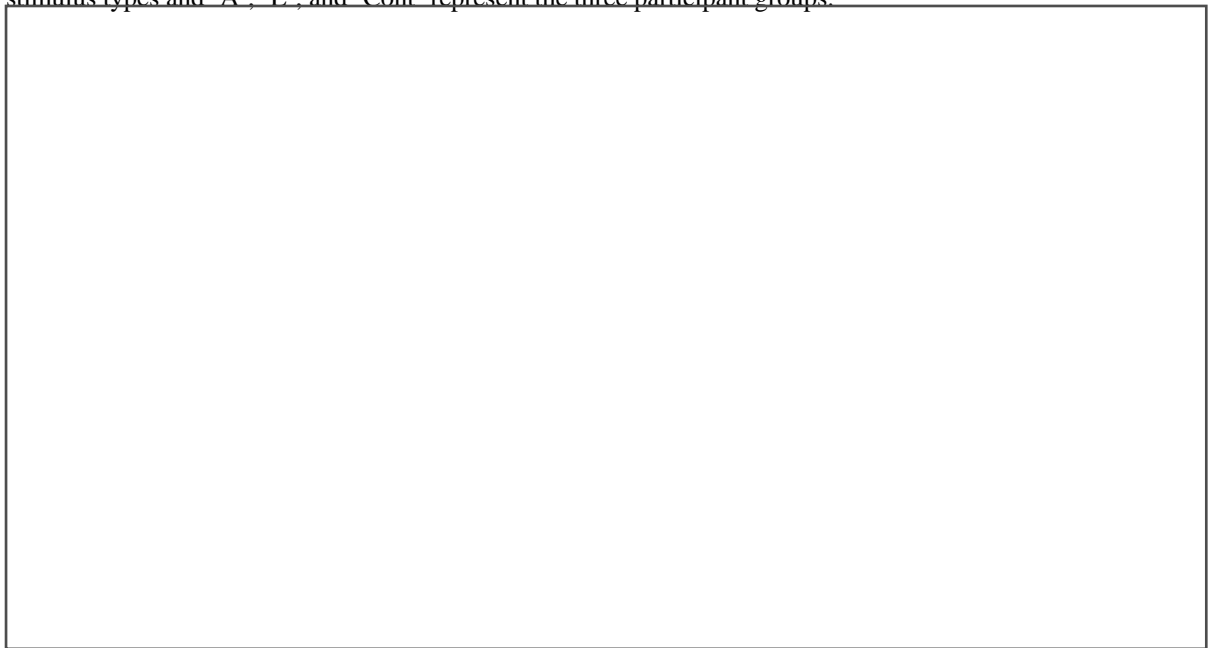
Appendix 20e (Exp 4): Paired samples T-tests for the average string similarity scores between recognition and model-predicted saliency scanpaths for each stimulus type and each participant group. 'Civil War', 'Eng' and 'Neut' represent the stimulus types and 'A', 'E', and 'Cont' represent the three participant groups.

Appendix 21a (Exp 4): Descriptive statistics for a repeated measures ANOVA for the average string similarity scores between scanpaths at encoding and recognition.
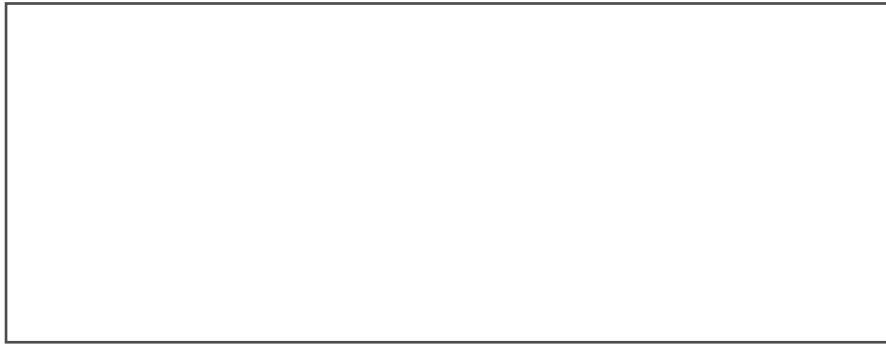
Appendix 21b (Exp 4): repeated measures ANOVA for the average string similarity scores between scanpaths at encoding and recognition. There is no reliable main effect of group.

Appendix 21c (Exp 4): Comparing the average string similarities for each participant group to chance. All comparisons are reliably greater than chance.
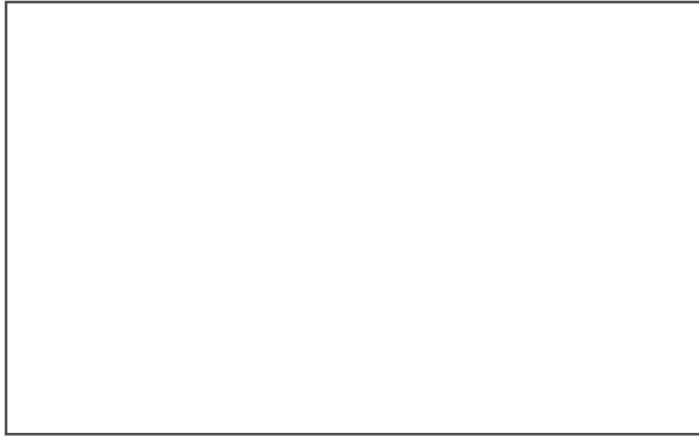
Appendix 21d (Exp 4): Statistics for the average string similarity scores between scanpaths at encoding and recognition for each stimulus type and each participant group.
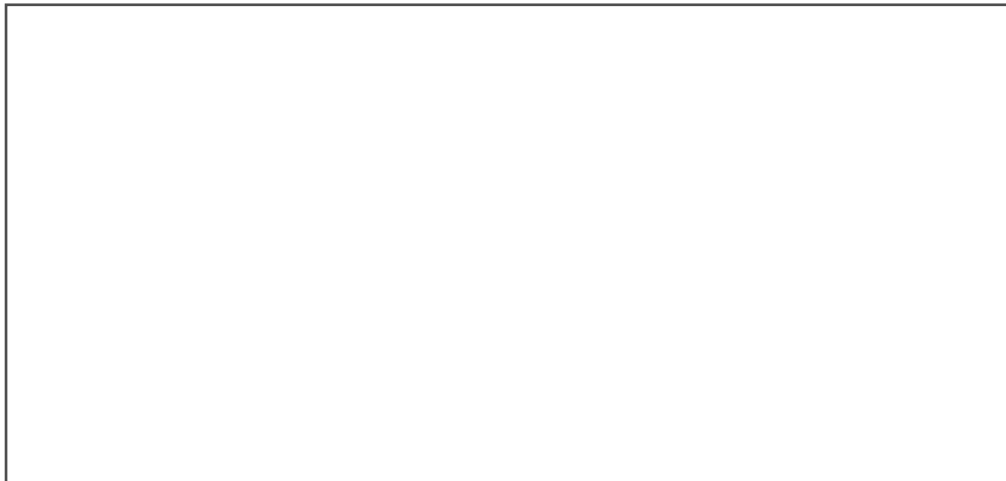
Appendix 21e (Exp 4): Mixed design ANOVA for the average string similarity scores between scanpaths at encoding and recognition for each stimulus type and each participant group. There is no reliable main effect of stimulus type and no reliable interaction between stimulus type and group.
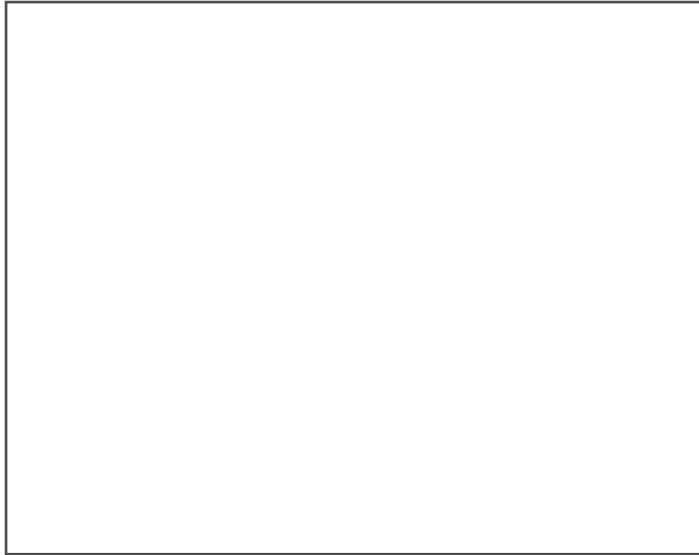
Appendix 22a (Exp 5): Descriptive statistics for accuracy at immediate and delayed recognition test, for each type of picture, for each participant group.

Appendix 22b (Exp 5): ANOVA between-subjects results for accuracy at immediate and delayed recognition test, for each type of picture, for each participant group. There is no main effect of group.

Appendix 22c (Exp 5): Mixed design ANOVA for accuracy at immediate and delayed recognition test, for each type of picture, for each participant group. There is a main effect of picture type and an interaction between picture type and group.

Appendix 22d (Exp 5): Statistics for post-hoc independent samples T-tests for accuracy at immediate and delayed recognition test, for each type of picture and each participant group.
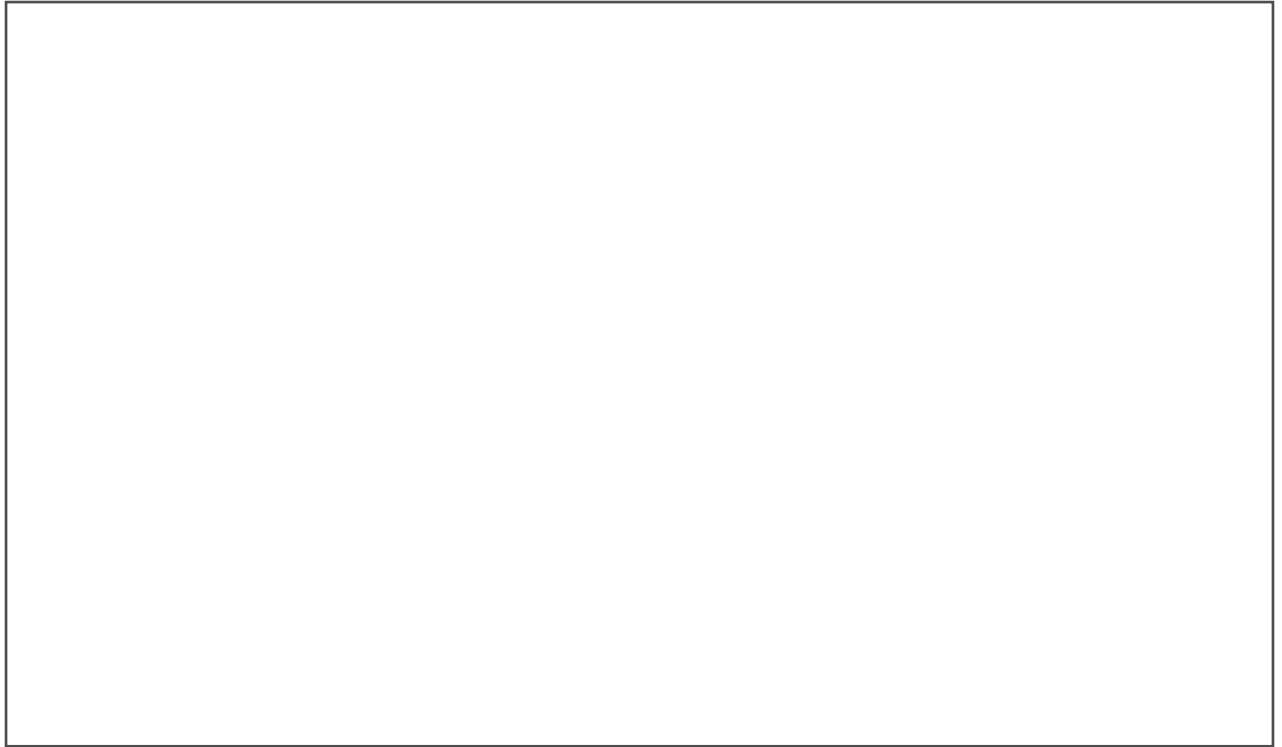
Appendix 22e (Exp 5): Post-hoc independent samples T-tests for accuracy at immediate and delayed recognition test, for each type of picture and each participant group. There are reliable differences between the participant groups for both Civil War and Engineering stimuli at immediate test and for Engineering stimuli at delayed test.
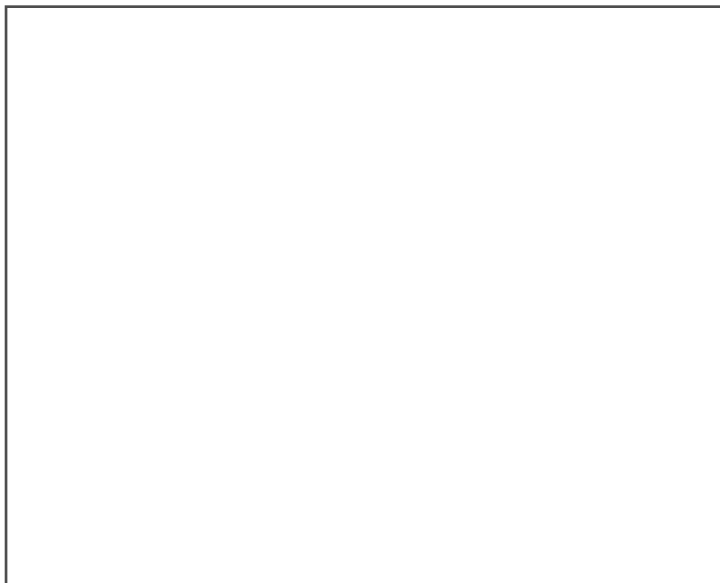
Appendix 22f (Exp 5): Statistics for paired samples T tests for American Studies students. Comparing accuracy at immediate and delayed recognition tests, for each type of picture.

Appendix 22g (Exp 5): Paired samples T tests for American Studies students. Comparing accuracy at immediate and delayed recognition tests, for each type of picture.
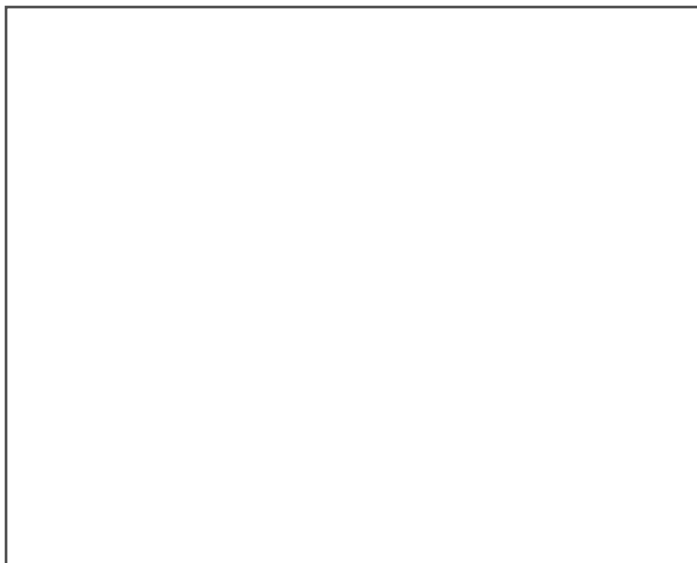
Appendix 22h (Exp 5): Statistics for paired samples T tests for Engineers. Comparing accuracy at immediate and delayed recognition tests, for each type of picture.

Appendix 22i (Exp 5): Paired samples T tests for Engineers. Comparing accuracy at immediate and delayed recognition tests, for each type of picture.

Appendix 23a (Exp 5): Descriptive statistics for the average number of fixations that fell into the first 5 most salient regions, for each type of picture, for each participant group.

Appendix 23b (Exp 5): ANOVA between-subjects results for the average number of fixations that fell into the first 5 most salient regions, for each type of picture, for each participant group. There is no main effect of group.
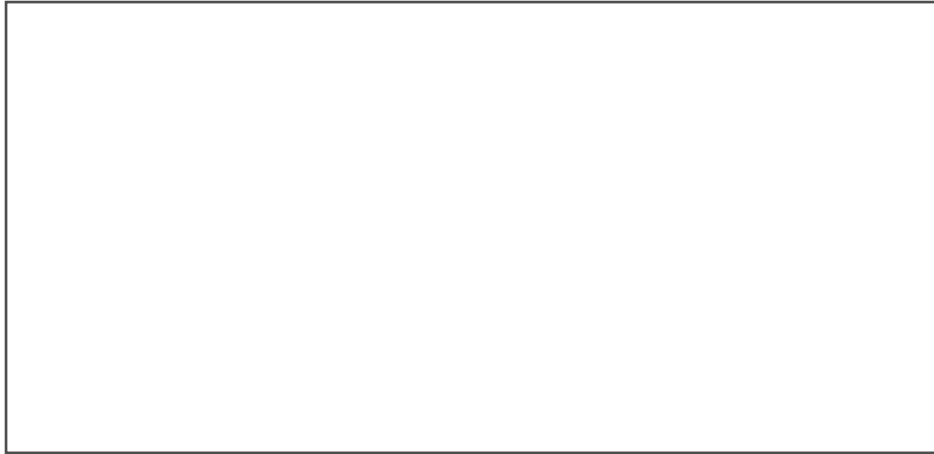
Appendix 23c (Exp 5): Mixed design ANOVA for the average number of fixations that fell into the first 5 most salient regions, for each participant group. There is a main effect of picture type and an interaction between picture type and group.
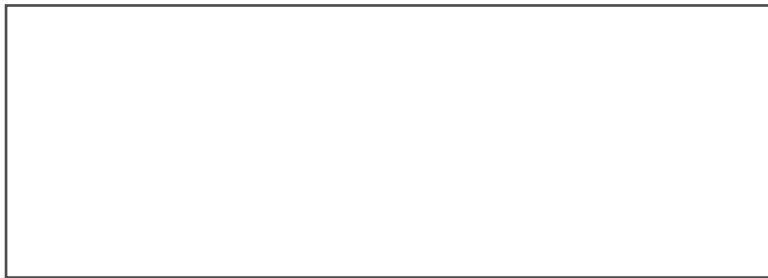
Appendix 23d (Exp 5): Statistics for post-hoc independent samples T- for the average number of fixations that fell into the first 5 most salient regions, for each type of picture and each participant group.

Appendix 23e (Exp 5): Post-hoc independent samples T-tests for the average number of fixations that fell into the first 5 most salient regions, for each type of picture and each participant group. There are reliable differences between the participant groups for both Civil War and Engineering stimuli.
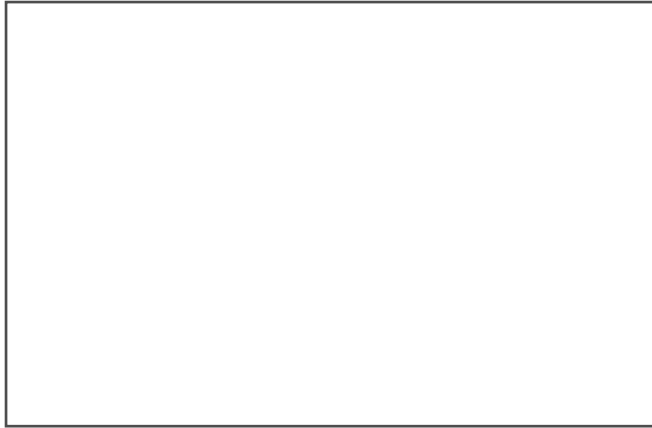
Appendix 23f (Exp 5): Statistics for the average number of fixations that fell into the first 5 most salient regions, for each type of picture.

Appendix 23g (Exp 5): Paired samples T tests for the average number of fixations that fell into the first 5 most salient regions, for each type of picture.

Appendix 24a (Exp 5): Descriptive statistics for the average string similarities between encoding and model-predicted saliency, for each stimulus type, for each participant group.

Appendix 24b (Exp 5): Between-groups results for the average string similarities between encoding and model-predicted saliency, for each stimulus type, for each participant group. There is no reliable main effect of group.
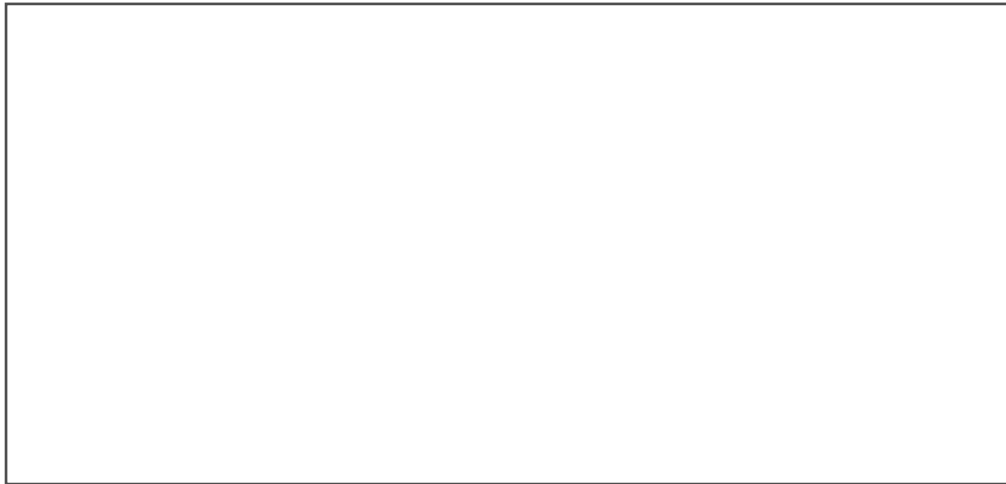
Appendix 24c (Exp 5): Mixed design ANOVA for the average string similarities between encoding and model-predicted saliency, for each stimulus type, for each participant group. There is a main effect of stimulus type and an interaction between stimulus type and group.

Appendix 24d (Exp 5): Statistics for post-hoc independent samples T-tests. Average string similarities between encoding and model-predicted saliency, for each stimulus type, for each participant group.
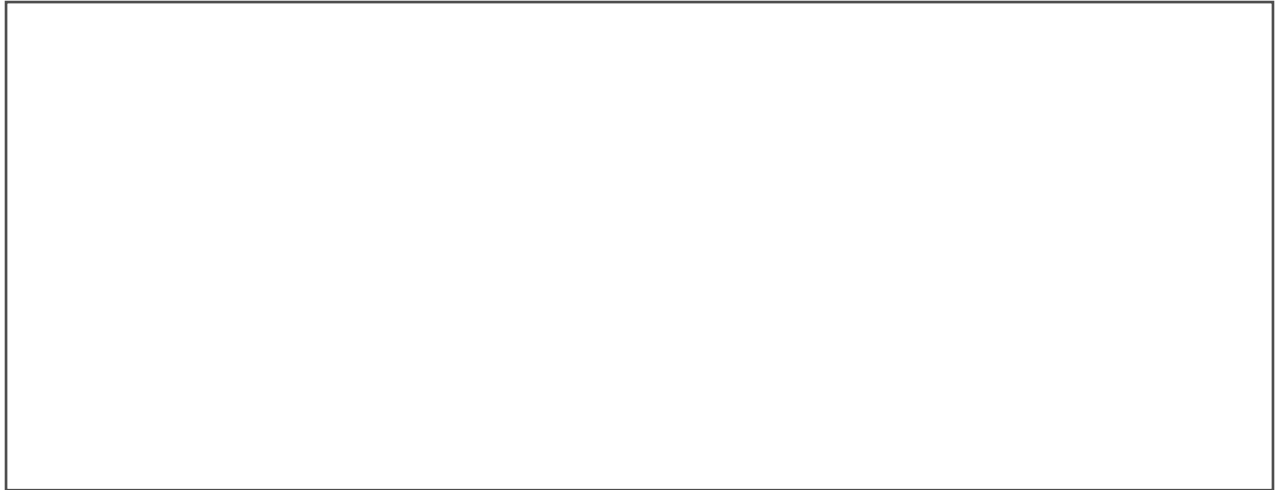
Appendix 24e (Exp 5): Post-hoc independent samples T-tests for the average string similarities between encoding and model-predicted saliency, for each stimulus type, for each participant group. There is a reliable difference between the groups for Engineering stimuli and Civil War but not for neutral stimuli.

Appendix 24f (Exp 5): Statistics for post-hoc paired samples T-tests. To further investigate the main effect of stimulus type, average string similarities between encoding and model-predicted saliency (for each stimulus type) are compared.

Appendix 24g (Exp 5): Post-hoc paired samples T-tests. Overall, string similarity scores are lower for neutral stimuli than for Civil war or Engineering stimuli.
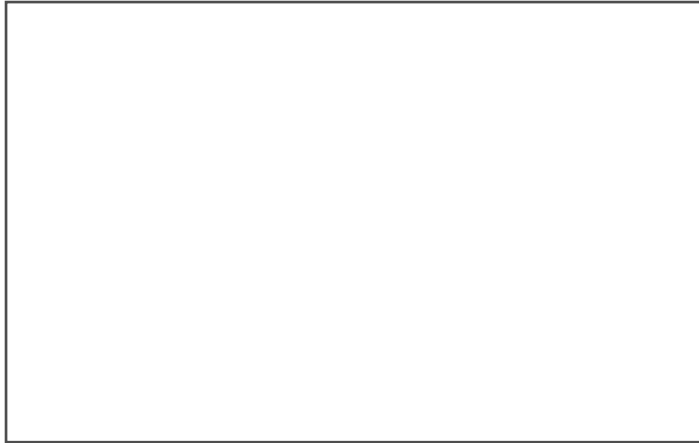
Appendix 24h (Exp 5): Statistics for post-hoc paired samples T-tests. Comparing actual string similarities to those expected by chance.
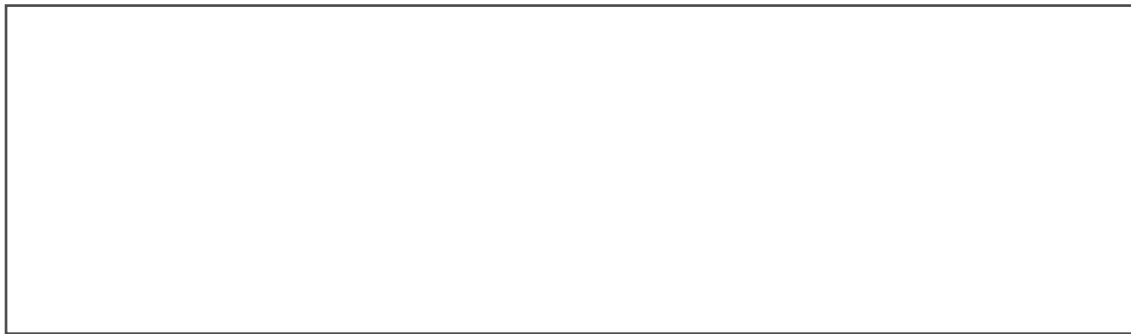
Appendix 24i (Exp 5): Post-hoc paired samples T-tests. All string similarities are reliably greater than expected by chance.

Appendix 25 (Exp 5)a: Descriptive statistics for the average string similarities between encoding and immediate recognition test, for each stimulus type, for each participant group.

Appendix 25 (Exp 5)b: Between-groups results for the average string similarities between encoding and immediate recognition test, for each stimulus type, for each participant group. There is a reliable main effect of group.

Appendix 25 (Exp 5)c: Average string similarities for each group. Overall, Engineers have higher similarity scores than American Studies students.
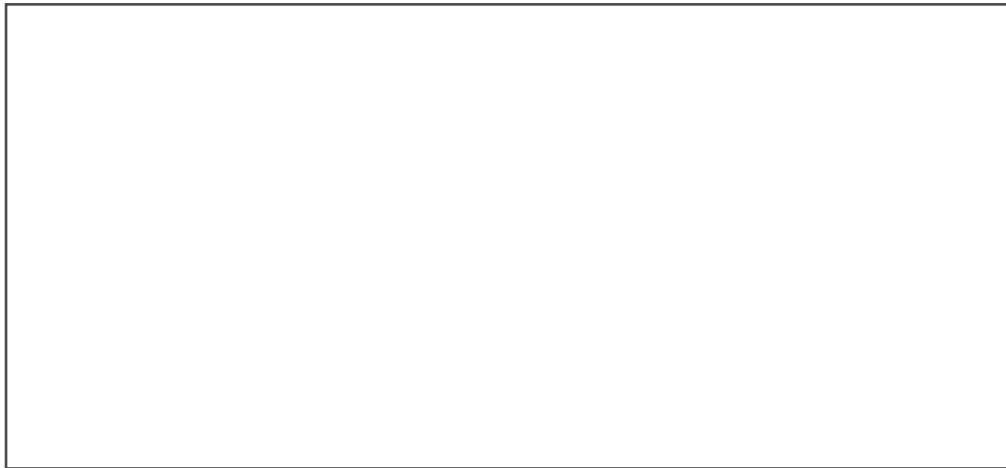
Appendix 25 (Exp 5)d: Mixed design ANOVA for the average string similarities between encoding and immediate recognition test, for each stimulus type, for each participant group. There is no main effect of stimulus type, but there is an interaction between stimulus type and group.

Appendix 25 (Exp 5)e: Statistics for post-hoc independent samples T-tests. Average string similarities between encoding and immediate recognition test, for each stimulus type, for each participant group.
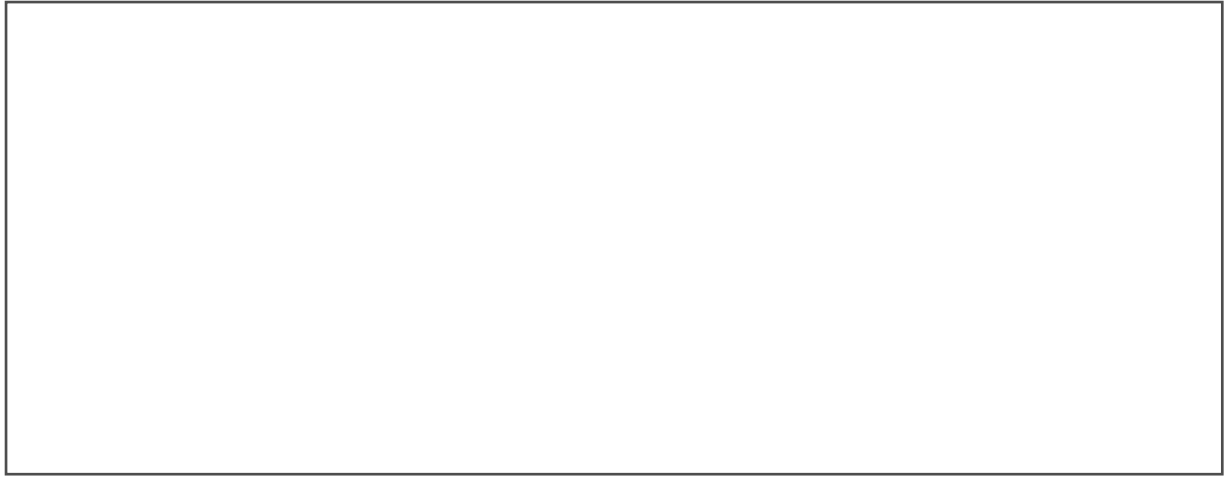
Appendix 25 (Exp 5)f: Post-hoc independent samples T-tests for the average string similarities between encoding and immediate recognition test, for each stimulus type, for each participant group. There is a reliable difference between the groups for Engineering stimuli, but not for Civil War or neutral stimuli.

Appendix 25 (Exp 5)g: Statistics for post-hoc paired samples T-tests. Comparing actual string similarities to those expected by chance.

Appendix 25 (Exp 5)h: Post-hoc paired samples T-tests. All string similarities are reliably greater than expected by chance.
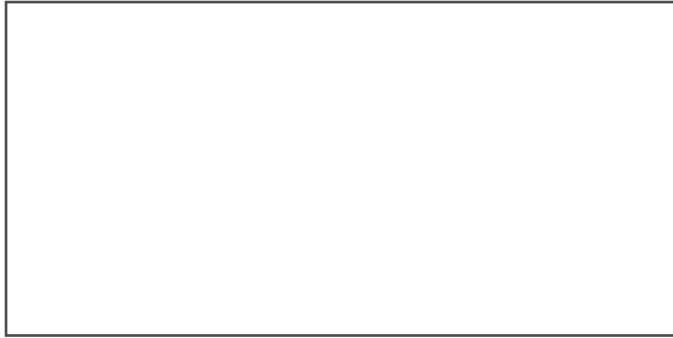
Appendix 26 (Exp 5)a: Descriptive statistics for the average string similarities between encoding and delayed recognition test, for each stimulus type, for each participant group.
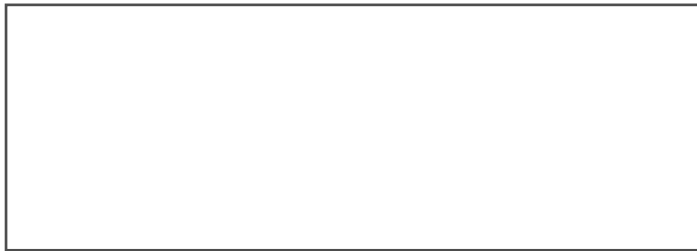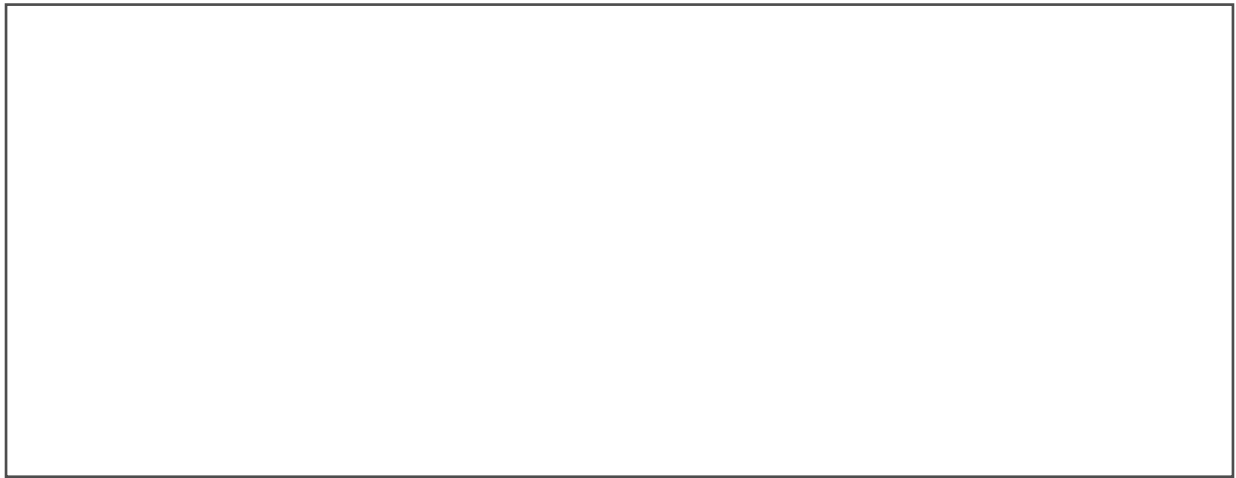
Appendix 26b (Exp 5): Between-groups results for the average string similarities between encoding and delayed recognition test, for each stimulus type, for each participant group. There is a reliable main effect of group.
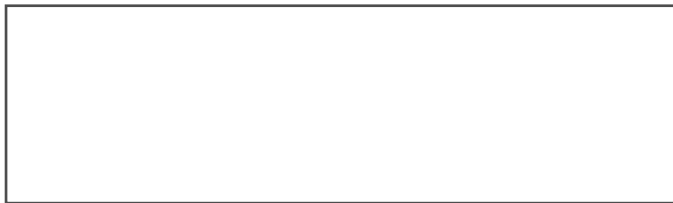
Appendix 26c (Exp 5): Average string similarities for each group. Overall, Engineers have higher similarity scores than American Studies students.

Appendix 26d (Exp 5): Mixed design ANOVA for the average string similarities between encoding and delayed recognition test, for each stimulus type, for each participant group. There is a main effect of stimulus type and an interaction between stimulus type and group.

Appendix 26e (Exp 5): Statistics for post-hoc independent samples T-tests. Average string similarities between encoding and delayed recognition test, for each stimulus type, for each participant group.

Appendix 26f (Exp 5): Post-hoc independent samples T-tests for the average string similarities between encoding and delayed recognition test, for each stimulus type, for each participant group. There is a reliable difference between the groups for Engineering stimuli and Civil War stimuli, but not for neutral stimuli.

Appendix 26g (Exp 5): Statistics for post-hoc paired samples T-tests. To further investigate the main effect of

stimulus type, average string similarities between encoding and delayed recognition test (for each stimulus type) are compared.

Appendix 26h (Exp 5): Post-hoc paired samples T-tests. Overall, string similarity scores are lower for neutral stimuli than for Civil war or Engineering stimuli.

Appendix 26i (Exp 5): Statistics for post-hoc paired samples T-tests. Comparing actual string similarities to those expected by chance.

Appendix 26j (Exp 5): Post-hoc paired samples T-tests. All string similarities are reliably greater than expected by chance.
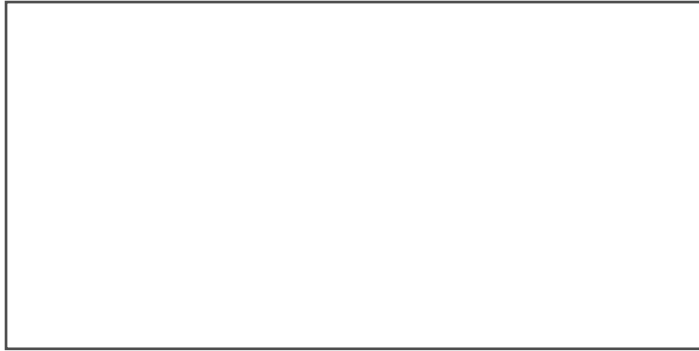
Appendix 27a (Exp 5): Descriptive statistics for the average string similarities between immediate and delayed recognition tests, for each stimulus type, for each participant group.

Appendix 27b (Exp 5): Between-groups results for the average string similarities between immediate and delayed recognition tests, for each stimulus type, for each participant group. There is no reliable main effect of group.

Appendix 27c (Exp 5): Mixed design ANOVA for the average string similarities between immediate and delayed recognition tests, for each stimulus type, for each participant group. There is a main effect of stimulus type and an interaction between stimulus type and group.

Appendix 27d (Exp 5): Statistics for post-hoc independent samples T-tests. Average string similarities between immediate and delayed recognition tests, for each stimulus type, for each participant group.

Appendix 27e (Exp 5): Post-hoc independent samples T-tests for the average string similarities between immediate and delayed recognition tests, for each stimulus type, for each participant group. There is a reliable difference between the groups for Engineering stimuli and Civil War stimuli, but not for neutral stimuli.

Appendix 27f (Exp 5): Statistics for post-hoc paired samples T-tests. To further investigate the main effect of stimulus type, average string similarities between immediate and delayed recognition tests (for each stimulus type) are compared.

Appendix 27g (Exp 5): Post-hoc paired samples T-tests. Overall, string similarity scores are lower for neutral stimuli than for Civil war or Engineering stimuli.
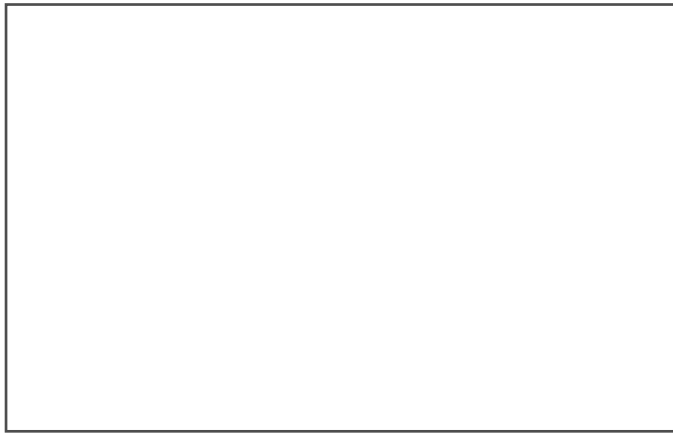
Appendix 27h (Exp 5): Statistics for post-hoc paired samples T-tests. Comparing actual string similarities to those expected by chance.

Appendix 27i (Exp 5): Post-hoc paired samples T-tests. All string similarities are reliably greater than expected by chance.

Appendix 28a (Exp 6): The group statistics for the independent samples T-test comparing overall accuracy of

specialists and control participants.

Appendix 28b (Exp 6): The group statistics for the independent samples T-test comparing overall accuracy of specialists and control participants.

Appendix 28c (Exp 6): Descriptive statistics for a mixed-design 2x2x2 ANOVA comparing accuracy of old and new computer and non computer stimuli, for each participant group.
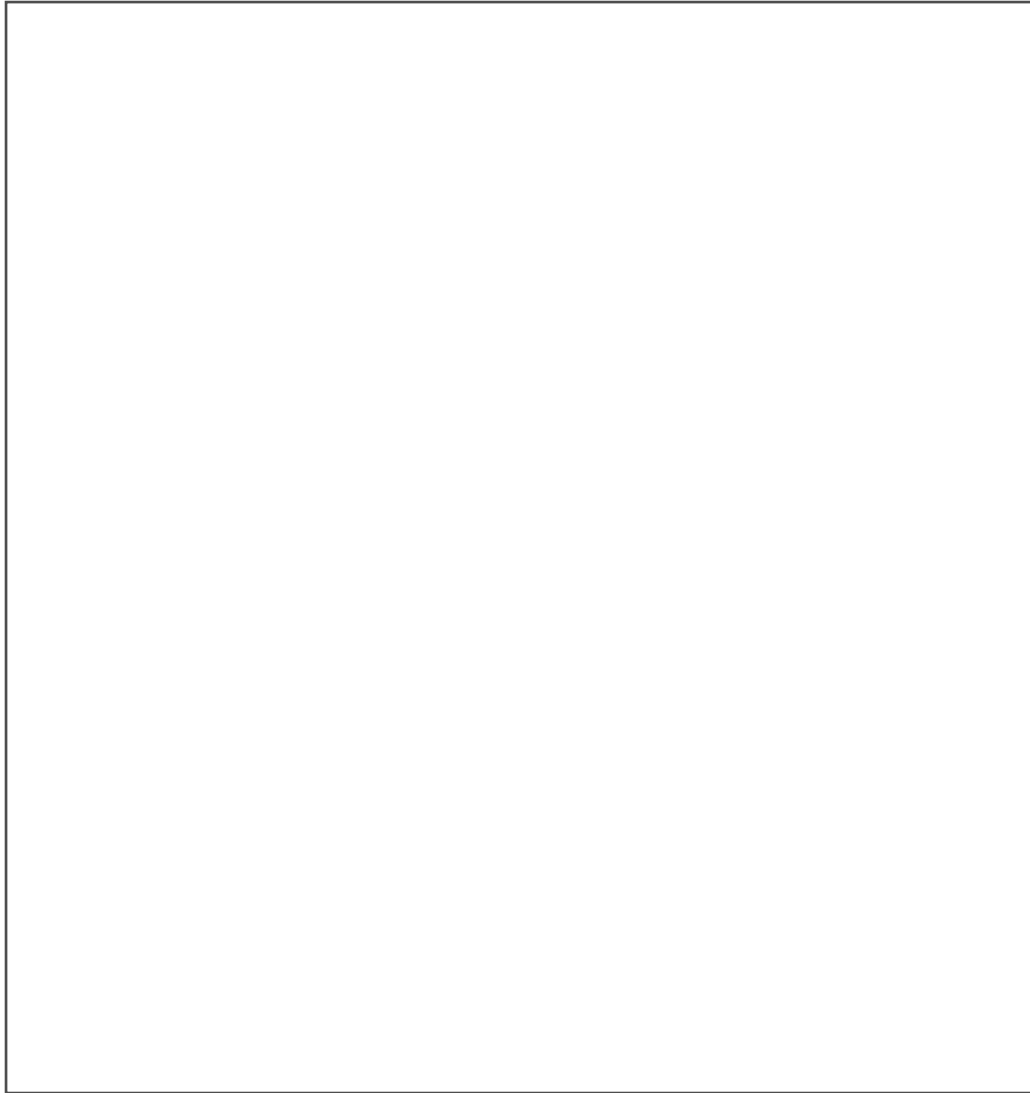
Appendix 28d (Exp 6): Mixed-design 2x2x2 ANOVA comparing accuracy of old and new computer and non computer stimuli, for each participant group. There is a 2-way interaction between new/old and comp/non comp stimuli, and a 3-way interaction between group, old/new and comp/non comp stimuli.


Appendix 28e (Exp 6): Descriptive statistics for post-hoc paired-samples T-tests to further investigate the 2-way interaction between new/old and comp/non comp stimuli.
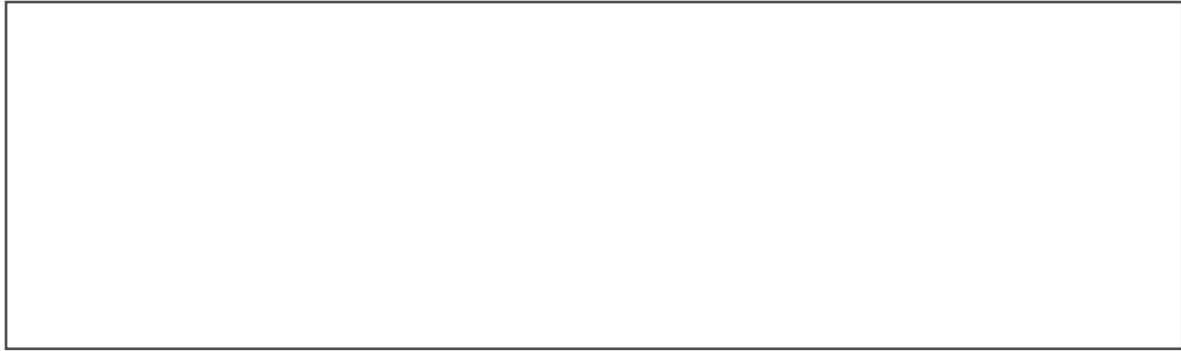
Appendix 28f (Exp 6): Post-hoc paired-samples T-tests to further investigate the 2-way interaction between new/old and comp/non comp stimuli. There is a statistically reliable difference between old computer stimuli and old non computer stimuli.

Appendix 28g (Exp 6): Descriptive statistics for post-hoc independent-samples T-tests to further investigate the 3-way interaction between group, old/new and comp/non comp stimuli.

Appendix 28h (Exp 6): Post-hoc independent-samples T-tests to further investigate the 3-way interaction between group, old/new and comp/non comp stimuli. There are reliable differences between the participant groups for new non computer stimuli and for old computer stimuli.
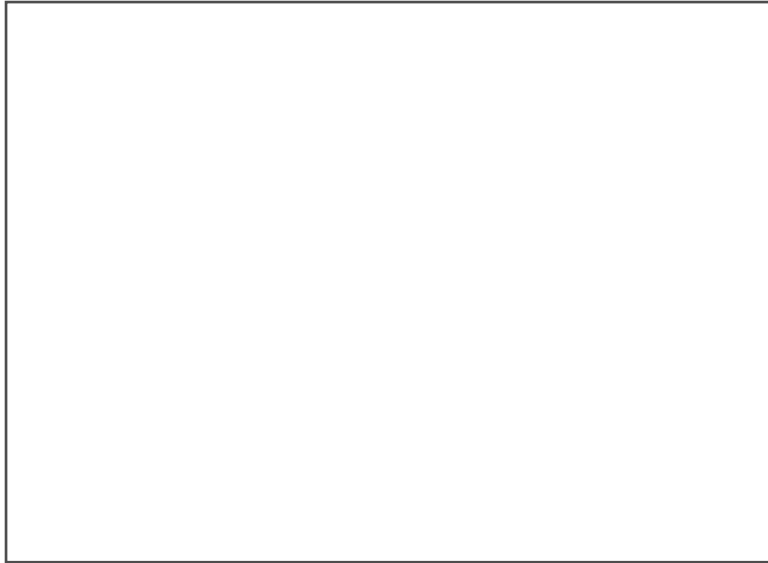
Appendix 28i (Exp 6): Descriptive statistics for post-hoc paired samples T-tests to further investigate the 3-way interaction between group, old/new and comp/non comp stimuli.
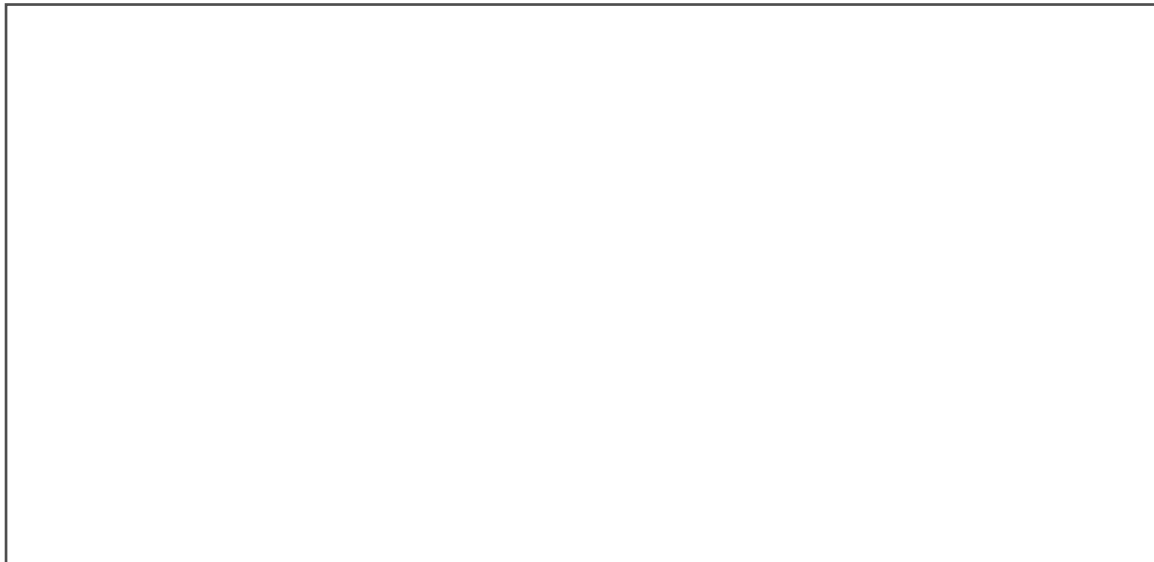
Appendix 28j (Exp 6): Post-hoc paired samples T-tests to further investigate the 3-way interaction between group, old/new and comp/non comp stimuli. There are reliable differences between old computer pictures and old non computer pictures for control participants and between new non computer pictures and old non computer pictures for control participants
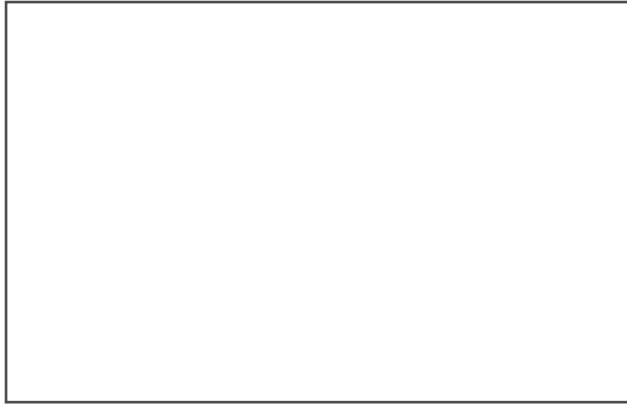
Appendix 29a (Exp 6): Descriptive statistics for a 2x2x2 mixed-design ANOVA comparing accuracy scores for computer and non computer pictures that were imagined or described, by each participant group.
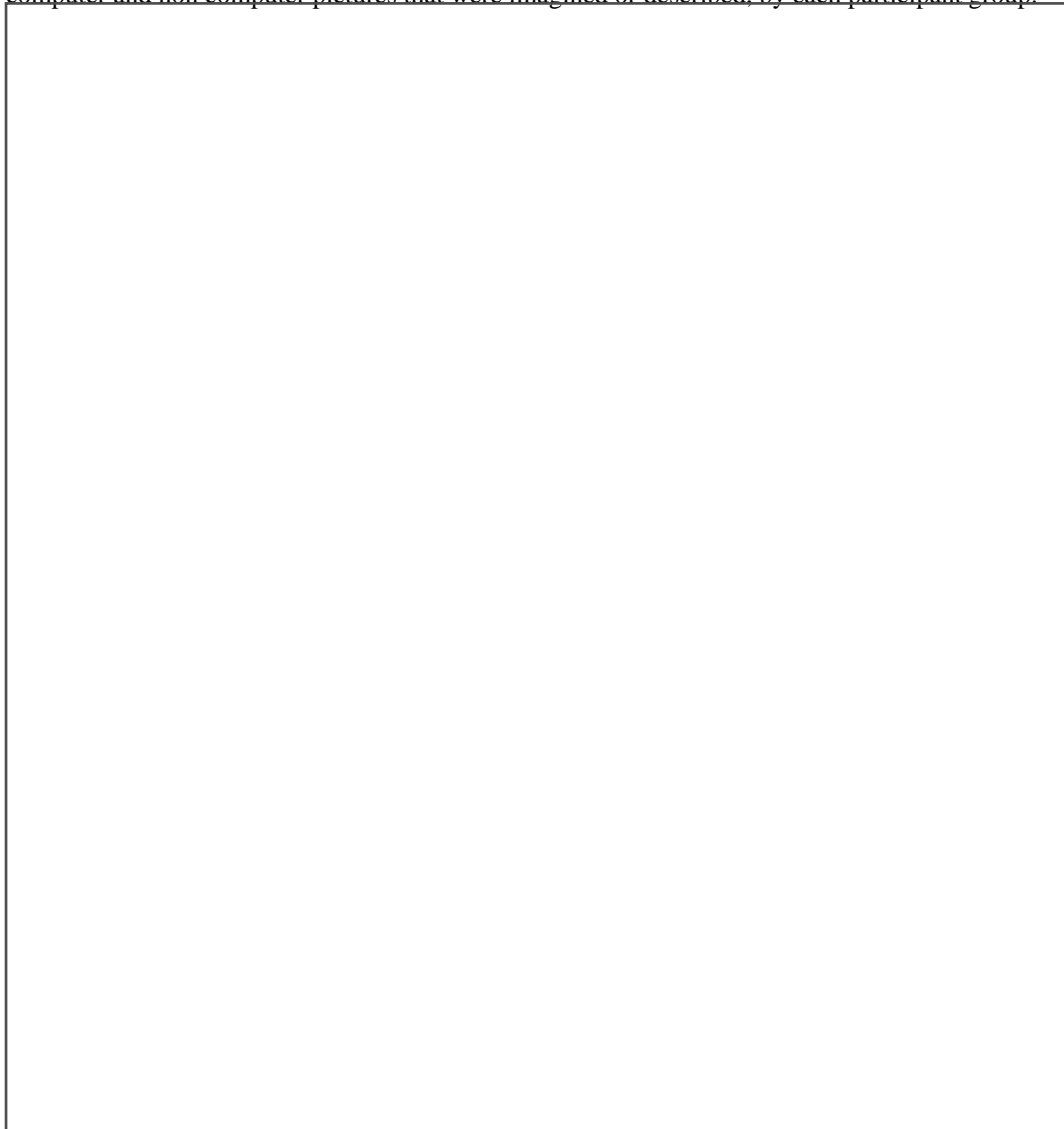
Appendix 29b (Exp 6): A 2x2x2 mixed-design ANOVA comparing accuracy scores for computer and non computer pictures that were imagined or described, by each participant group. There is a main effect of condition (imagery or describe) and a main effect of stimulus type (computer or non computer picture). There is also a reliable interaction between stimulus type and group.

Appendix 29c (Exp 6): Between-groups output from 2x2x2 ANOVA investigating accuracy scores for computer and non computer pictures that were imagined or described, by each participant group. There is no main effect of group.

Appendix 29d (Exp 6): Descriptive statistics for post-hoc paired samples T-tests to further investigate the 2-way interaction between stimulus type of group

Appendix 29e (Exp 6): Post-hoc paired samples T-tests to further investigate the 2-way interaction between stimulus type of group. There are no reliable differences.

Appendix 29f (Exp 6): Descriptive statistics for post-hoc independent samples T-tests to further investigate the 2-way interaction between stimulus type of group.
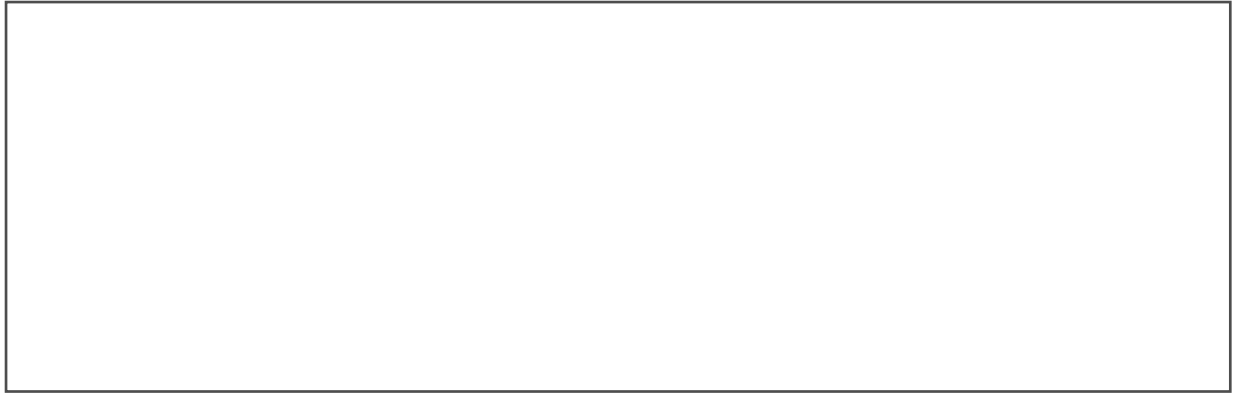
Appendix 29g (Exp 6): Post-hoc independent samples T-tests to further investigate the 2-way interaction between stimulus type of group. There is a reliable difference between the groups for computer pictures. Specialists are reliably more accurate at computer pictures than control participants.

Appendix 30a (Exp 6): Descriptive statistics for a 2x2 repeated measures ANOVA comparing the number of references to salient regions on computer and non computer pictures by specialists and control participants.
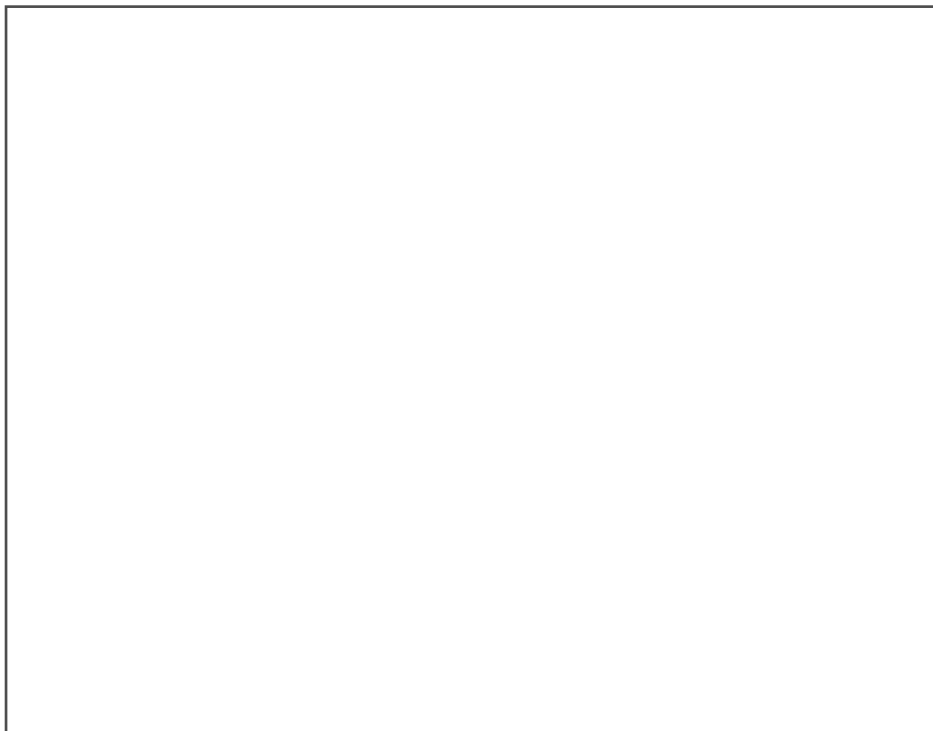
Appendix 30b (Exp 6): A 2x2 repeated measures ANOVA comparing the number of references to salient regions on

computer and non computer pictures by specialists and control participants. There is a reliable main effect of picture type and a reliable main effect of group. There is no reliable interaction between picture type and group.

Appendix 30c (Exp 6): Descriptive statistics for paired samples T-tests for the number of references to salient regions on computer and non computer pictures by specialists and non specialists

Appendix 30d (Exp 6): Paired samples T-tests for the number of references to salient regions on computer and non computer pictures by specialists and non specialists

Appendix 30e (Exp 6): Statistics for a paired-samples T-test comparing the number of references to specialist Regions of Interests when viewing computer stimuli.

Appendix 30f (Exp 6): Paired-samples T-Test comparing the number of references to specialist Regions of Interests when viewing computer stimuli. There is a reliable difference between specialist and control participants.

Group Statistics					
		Levene's Test for Equality of Variances	t-test for Equality of Means		
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1					
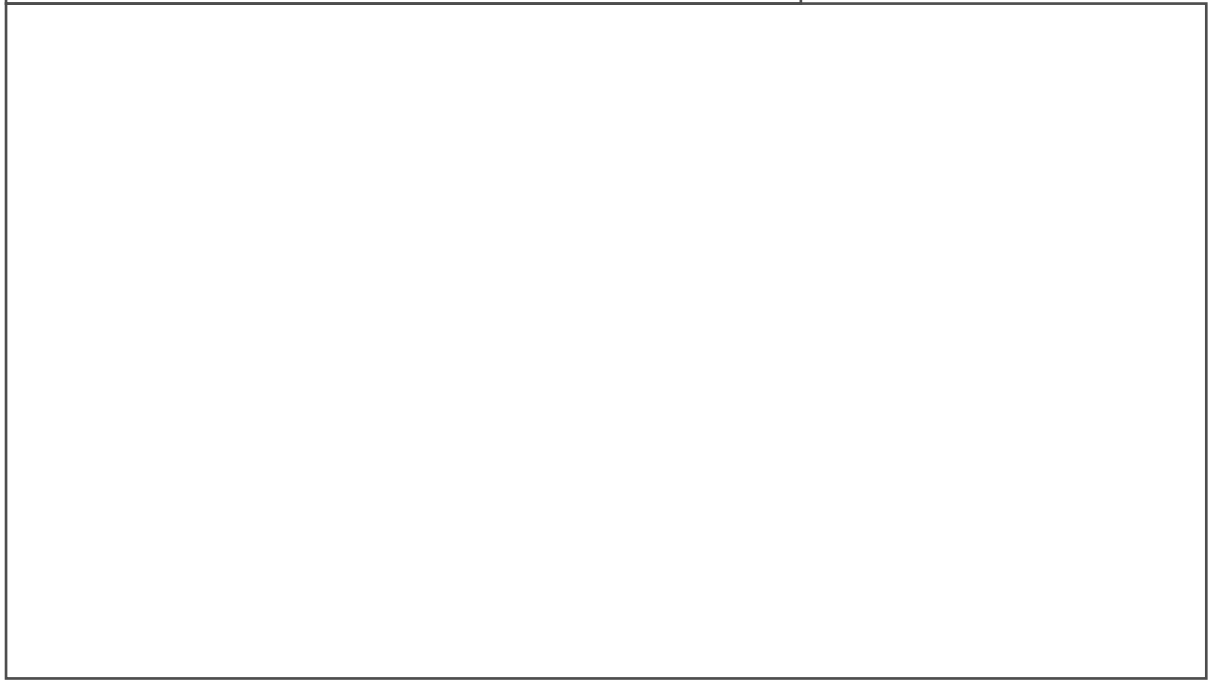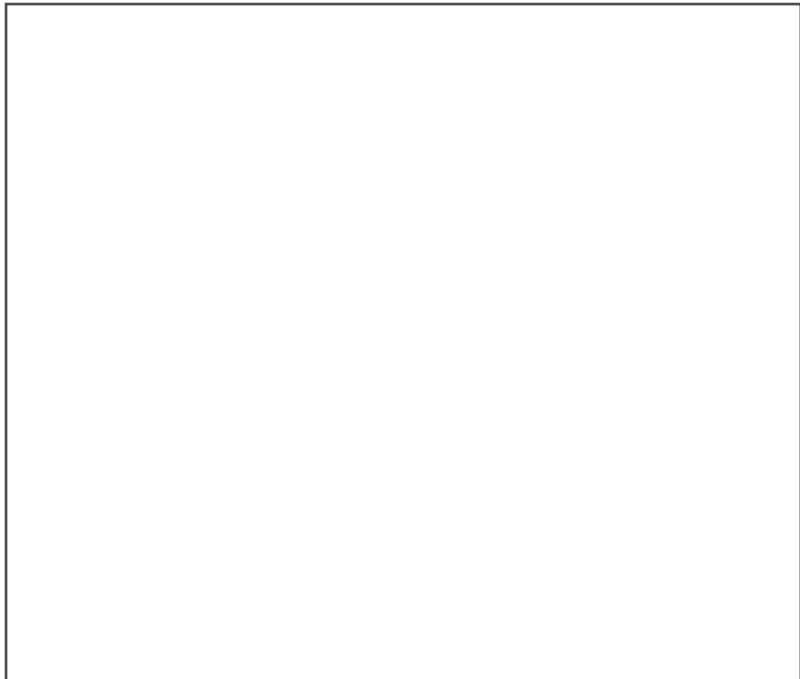
		Levene's Test	t-test for Equality of Means	
		for Equality		
		of Variances		
				95% Confidence Interval of the Difference
	Mean	N	Std. Deviation	Std. Error Mean
Pair 1				

		Levene's Test for Equality of Variances	t-test for Equality of Means		
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Paired Differences			t	df	Sig. (2-tailed)

|Mean |Std. Deviation |Std. Error Mean |95% Confidence Interval of the Difference | | | | | | | | |Lower |Upper | | | | |Pair 1
|Specialist_CompPic_Salient - Specialist_Non_Salient |-5.68598 |5.68260 |1.51874 |-8.96702 |-2.40494 |-3.744 |13 |.002 | |Pair 2
|Control_CompPic_Salient - Control_Non_Salient |-2.36686 |5.02285 |1.39309 |-5.40214 |.66842 |-1.699 |12 |.115 | |

Appendix 33c (Exp 6): The paired samples T-tests for the first fixation in RoIs on computer and non computer stimuli.

Appendix 34a (Exp 6): A mixed design ANOVA to compare the similarities of scanpaths at encoding compared to the different experimental conditions (imagery, verbalisation and recognition). There is a main effect of comparison and no reliable interaction between comparison and group.

Appendix 34b (Exp 6): ANOVA output for the scanpath similarities. There is no main effect of group.
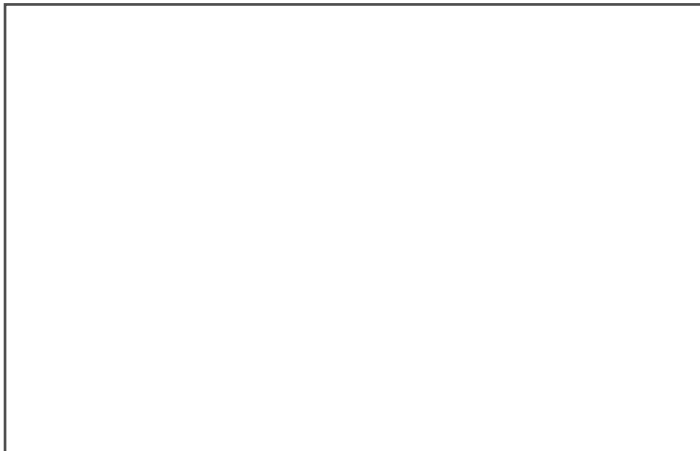
Appendix 34c (Exp 6): Statistics for post-hoc paired samples T-tests to further investigate the main effect of comparison. 'Enc' stands for Encoding, 'Vis' for visualisation and 'Desc' for describe or verbalise. Each comparison was also compared to chance.

Appendix 34d (Exp 6): Post-hoc paired samples T-tests of string similarity scores to further investigate the main effect of comparison.

Appendix 34e (Exp 6): Statistics for independent samples T-tests to compare string similarities at different stages of the experiment, for different stimulus types.
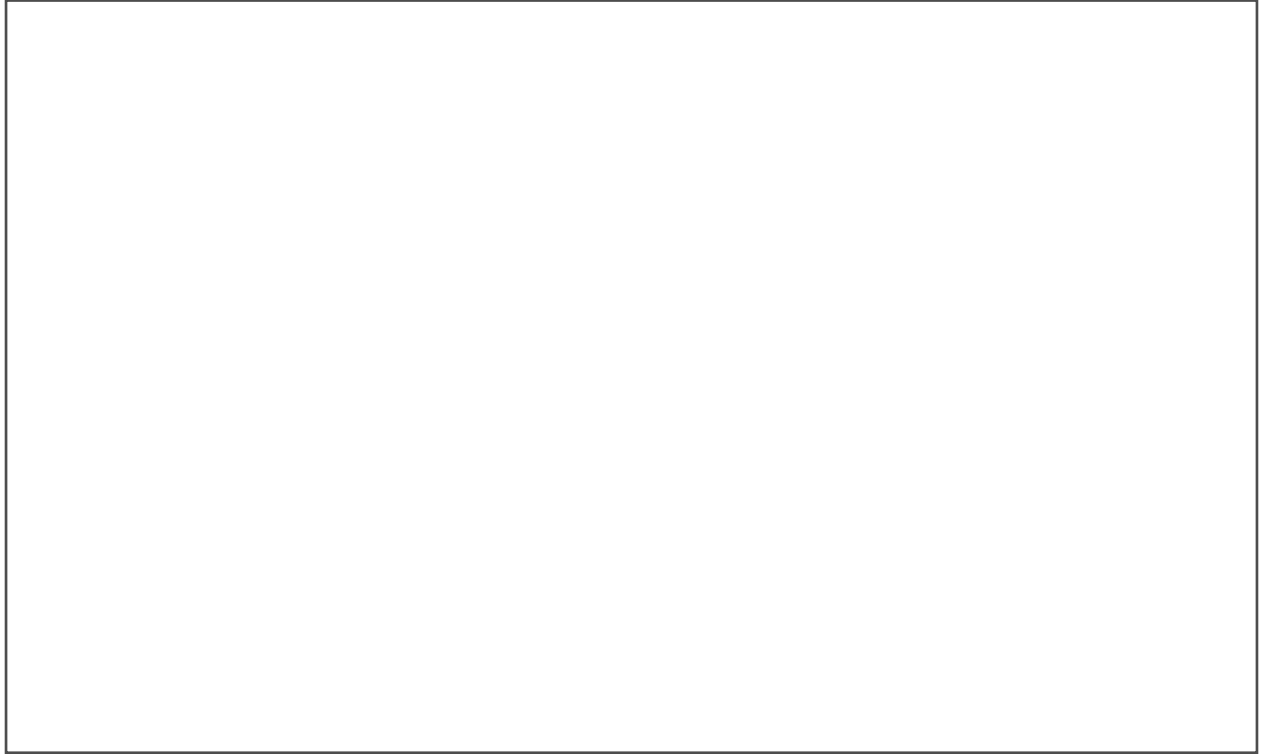
Appendix 34f (Exp 6): Independent samples T-tests to compare string similarities at different stages of the experiment, for different stimulus types.

Appendix 35a (Exp 7): Descriptive statistics for accuracy of chemists and control participants on congruent and incongruent chemistry and music pictures.
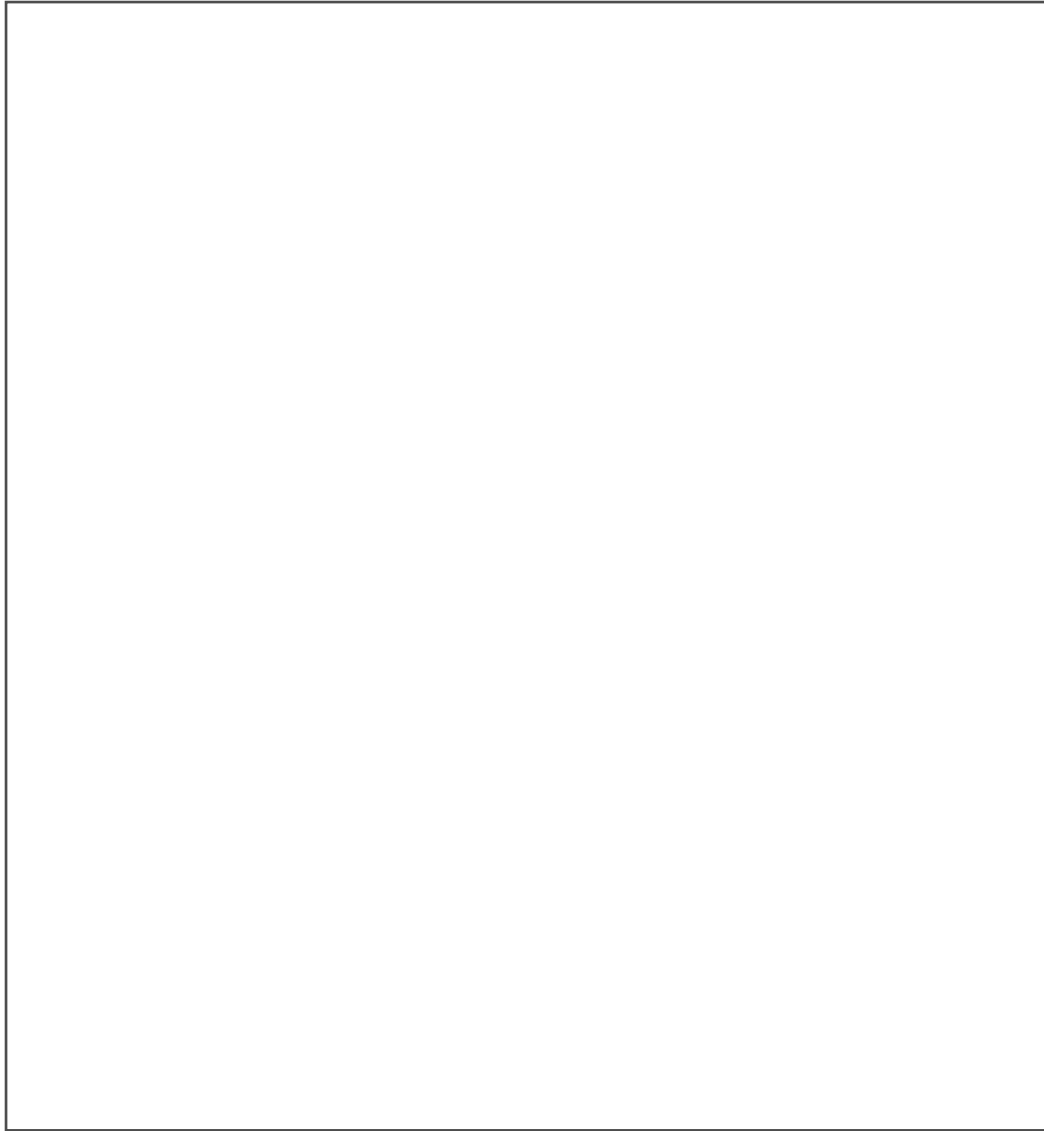
Appendix 35b (Exp 7): 2x2x2 mixed design ANOVA for accuracy of chemists and control participants on congruent and incongruent chemistry and music pictures. There is a main effect of picture type.

Appendix 35c (Exp 7): 2x2x2 mixed design ANOVA between-groups output for accuracy. There is a main effect of group.

Appendix 35d (Exp 7): Group statistics for the post-hoc independent sample T-test for accuracy.

Appendix 35e (Exp 7): Post-hoc independent sample T-test for accuracy

Appendix 35f (Exp 7): Paired samples statistics accuracy on each picture type for each participant group.

Appendix 35g (Exp 7): Paired samples t-test for accuracy on each picture type for each participant group.

Appendix 36a (Exp 7): Descriptive statistics for reaction time of chemists and control participants on chemistry and music pictures.
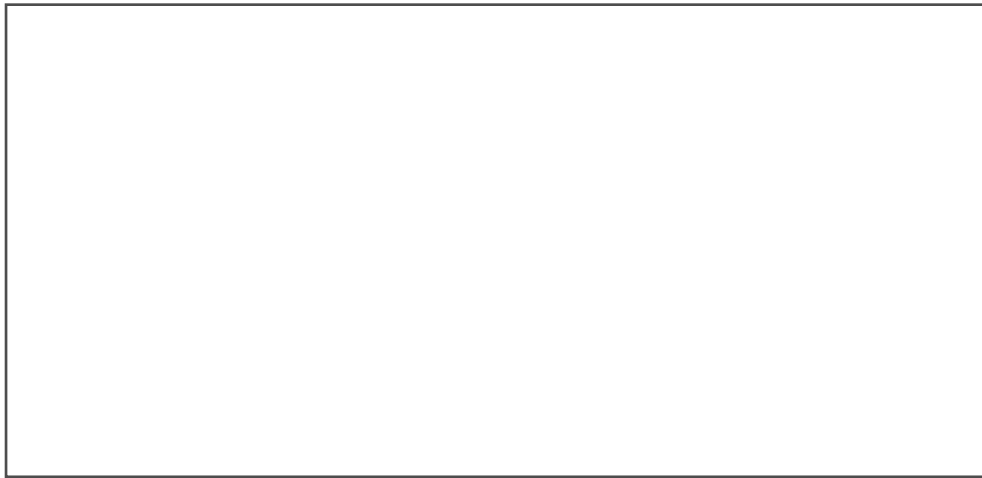
Appendix 36b (Exp 7): 2x2 mixed design ANOVA for reaction time of chemists and control participants on chemistry and music pictures. There is a reliable interaction between picture type and group.

Appendix 36c (Exp 7): Between-groups ANOVA output for reaction time of chemists and control participants on chemistry and music pictures. There is no reliable main effect of group.

Appendix 36d (Exp 7): Group statistics for reaction time on each picture type for each participant group.

Appendix 36e (Exp 7): Post-hoc independent samples T-tests for reaction time on each picture type for each participant group. There is a reliable difference between the groups for chemistry pictures.

Appendix 36f (Exp 7): Paired samples statistics for reaction time on each picture type for each participant group.

Appendix 36g (Exp 7): Paired samples T-tests for reaction time on each picture type for each participant group.
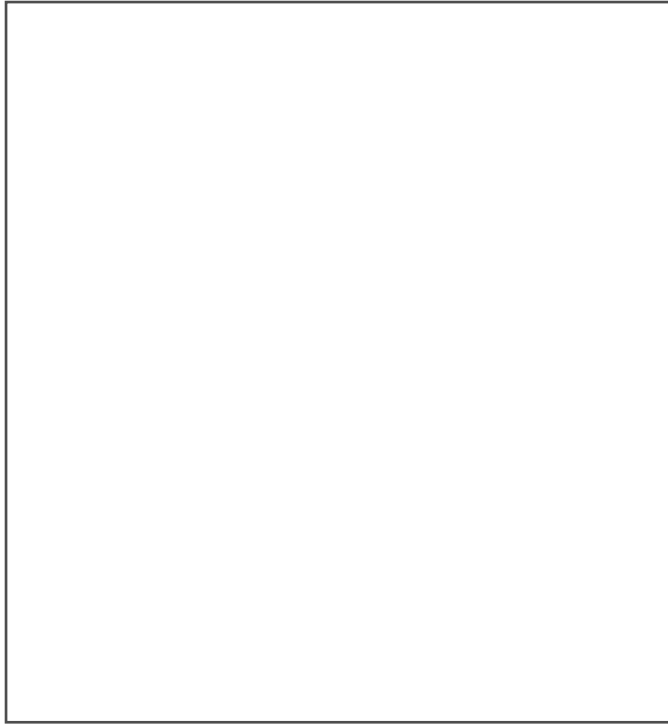
Appendix 36h (Exp 7): Descriptive statistics for reaction time of chemists and control participants on congruent and incongruent chemistry and music pictures.

Tests of Within-Subjects Effects

Measure: MEASURE_1

| Source		Type III Sum of Squares	df	Mean Square	F	Sig.
PicType	Sphericity Assumed	42996.041	1	42996.041	.004	.953
	Greenhouse-Geisser	42996.041	1.000	42996.041	.004	.953
	Huynh-Feldt	42996.041	1.000	42996.041	.004	.953
	Lower-bound	42996.041	1.000	42996.041	.004	.953
PicType * Group	Sphericity Assumed	196581040	1	196581040.4	16.045	.000
	Greenhouse-Geisser	196581040	1.000	196581040.4	16.045	.000
	Huynh-Feldt	196581040	1.000	196581040.4	16.045	.000
	Lower-bound	196581040	1.000	196581040.4	16.045	.000
Error(PicType)	Sphericity Assumed	330803843	27	12251994.20		
	Greenhouse-Geisser	330803843	27.000	12251994.20		
	Huynh-Feldt	330803843	27.000	12251994.20		
	Lower-bound	330803843	27.000	12251994.20		
Congruency	Sphericity Assumed	45022815.3	1	45022815.32	4.593	.041
	Greenhouse-Geisser	45022815.3	1.000	45022815.32	4.593	.041
	Huynh-Feldt	45022815.3	1.000	45022815.32	4.593	.041
	Lower-bound	45022815.3	1.000	45022815.32	4.593	.041
Congruency * Group	Sphericity Assumed	13852274.1	1	13852274.05	1.413	.245
	Greenhouse-Geisser	13852274.1	1.000	13852274.05	1.413	.245
	Huynh-Feldt	13852274.1	1.000	13852274.05	1.413	.245
	Lower-bound	13852274.1	1.000	13852274.05	1.413	.245
Error(Congruency)	Sphericity Assumed	264669755	27	9802583.511		
	Greenhouse-Geisser	264669755	27.000	9802583.511		
	Huynh-Feldt	264669755	27.000	9802583.511		
	Lower-bound	264669755	27.000	9802583.511		
Difference	Sphericity Assumed	104731199	1	104731199.3	4.290	.048
	Greenhouse-Geisser	104731199	1.000	104731199.3	4.290	.048
	Huynh-Feldt	104731199	1.000	104731199.3	4.290	.048
	Lower-bound	104731199	1.000	104731199.3	4.290	.048
Difference * Group	Sphericity Assumed	43017203.2	1	43017203.16	1.762	.195
	Greenhouse-Geisser	43017203.2	1.000	43017203.16	1.762	.195
	Huynh-Feldt	43017203.2	1.000	43017203.16	1.762	.195
	Lower-bound	43017203.2	1.000	43017203.16	1.762	.195
Error(Difference)	Sphericity Assumed	659191176	27	24414488.01		
	Greenhouse-Geisser	659191176	27.000	24414488.01		
	Huynh-Feldt	659191176	27.000	24414488.01		
	Lower-bound	659191176	27.000	24414488.01		
PicType * Congruency	Sphericity Assumed	6081280.547	1	6081280.547	.761	.391
	Greenhouse-Geisser	6081280.547	1.000	6081280.547	.761	.391
	Huynh-Feldt	6081280.547	1.000	6081280.547	.761	.391
	Lower-bound	6081280.547	1.000	6081280.547	.761	.391

Appendix 36i (Exp 7) *Part 1*: 2x2x2x2 mixed design ANOVA for reaction time of chemists and control participants on congruent and incongruent chemistry and music pictures with and without differences. Table continued on the next page.

|						
PicType * Congruency * Group	Sphericity Assumed	464929.021	1	464929.021	.058	.811
	Greenhouse-Geisser	464929.021	1.000	464929.021	.058	.811
	Huynh-Feldt	464929.021	1.000	464929.021	.058	.811
	Lower-bound	464929.021	1.000	464929.021	.058	.811
Error(PicType*Congruency)	Sphericity Assumed	215875191	27	7995377.453		
	Greenhouse-Geisser	215875191	27.000	7995377.453		
	Huynh-Feldt	215875191	27.000	7995377.453		
	Lower-bound	215875191	27.000	7995377.453		
PicType * Difference	Sphericity Assumed	6511049.638	1	6511049.638	.613	.441
	Greenhouse-Geisser	6511049.638	1.000	6511049.638	.613	.441
	Huynh-Feldt	6511049.638	1.000	6511049.638	.613	.441
	Lower-bound	6511049.638	1.000	6511049.638	.613	.441
PicType * Difference * Group	Sphericity Assumed	43948052.8	1	43948052.84	4.136	.052
	Greenhouse-Geisser	43948052.8	1.000	43948052.84	4.136	.052
	Huynh-Feldt	43948052.8	1.000	43948052.84	4.136	.052
	Lower-bound	43948052.8	1.000	43948052.84	4.136	.052
Error(PicType*Difference)	Sphericity Assumed	286881993	27	10625259.00		
	Greenhouse-Geisser	286881993	27.000	10625259.00		
	Huynh-Feldt	286881993	27.000	10625259.00		
	Lower-bound	286881993	27.000	10625259.00		
Congruency * Difference	Sphericity Assumed	1537612.828	1	1537612.828	.118	.734
	Greenhouse-Geisser	1537612.828	1.000	1537612.828	.118	.734
	Huynh-Feldt	1537612.828	1.000	1537612.828	.118	.734
	Lower-bound	1537612.828	1.000	1537612.828	.118	.734
Congruency * Difference * Group	Sphericity Assumed	24067803.5	1	24067803.51	1.851	.185
	Greenhouse-Geisser	24067803.5	1.000	24067803.51	1.851	.185
	Huynh-Feldt	24067803.5	1.000	24067803.51	1.851	.185
	Lower-bound	24067803.5	1.000	24067803.51	1.851	.185
Error(Congruency*Difference)	Sphericity Assumed	350994993	27	12999814.57		
	Greenhouse-Geisser	350994993	27.000	12999814.57		
	Huynh-Feldt	350994993	27.000	12999814.57		
	Lower-bound	350994993	27.000	12999814.57		
PicType * Congruency * Difference	Sphericity Assumed	73278087.3	1	73278087.27	5.723	.024
	Greenhouse-Geisser	73278087.3	1.000	73278087.27	5.723	.024
	Huynh-Feldt	73278087.3	1.000	73278087.27	5.723	.024
	Lower-bound	73278087.3	1.000	73278087.27	5.723	.024
PicType * Congruency * Difference * Group	Sphericity Assumed	74804209.1	1	74804209.13	5.843	.023
	Greenhouse-Geisser	74804209.1	1.000	74804209.13	5.843	.023
	Huynh-Feldt	74804209.1	1.000	74804209.13	5.843	.023
	Lower-bound	74804209.1	1.000	74804209.13	5.843	.023
Error(PicType*Congruency*Difference)	Sphericity Assumed	345686997	27	12803222.11		
	Greenhouse-Geisser	345686997	27.000	12803222.11		
	Huynh-Feldt	345686997	27.000	12803222.11		
	Lower-bound	345686997	27.000	12803222.11		

Appendix 36i (Exp 7) *Part 2*: Continuation of ANOVA output for reaction time of chemists and control participants on congruent and incongruent chemistry and music pictures with and without differences.
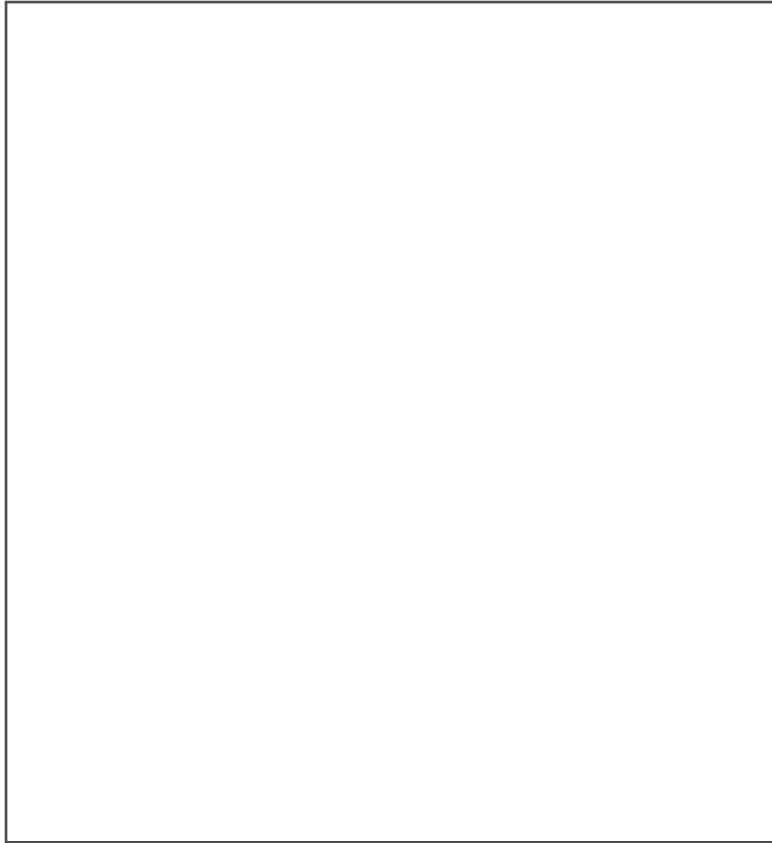
**Appendix 36j (Exp 7):** Between-groups 2x2x2x2 ANOVA output for reaction time. There is no main effect of group.
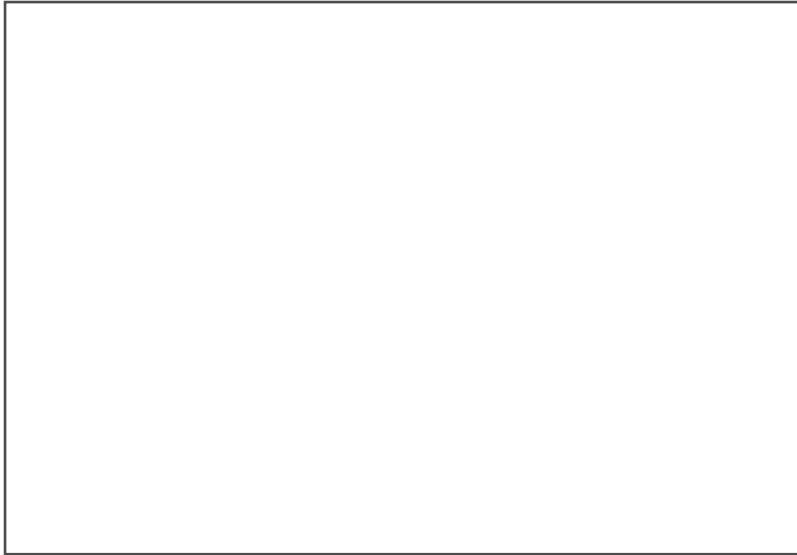
Appendix 36k (Exp 7): Group statistics for post-hoc independent sample T-tests for reaction time.
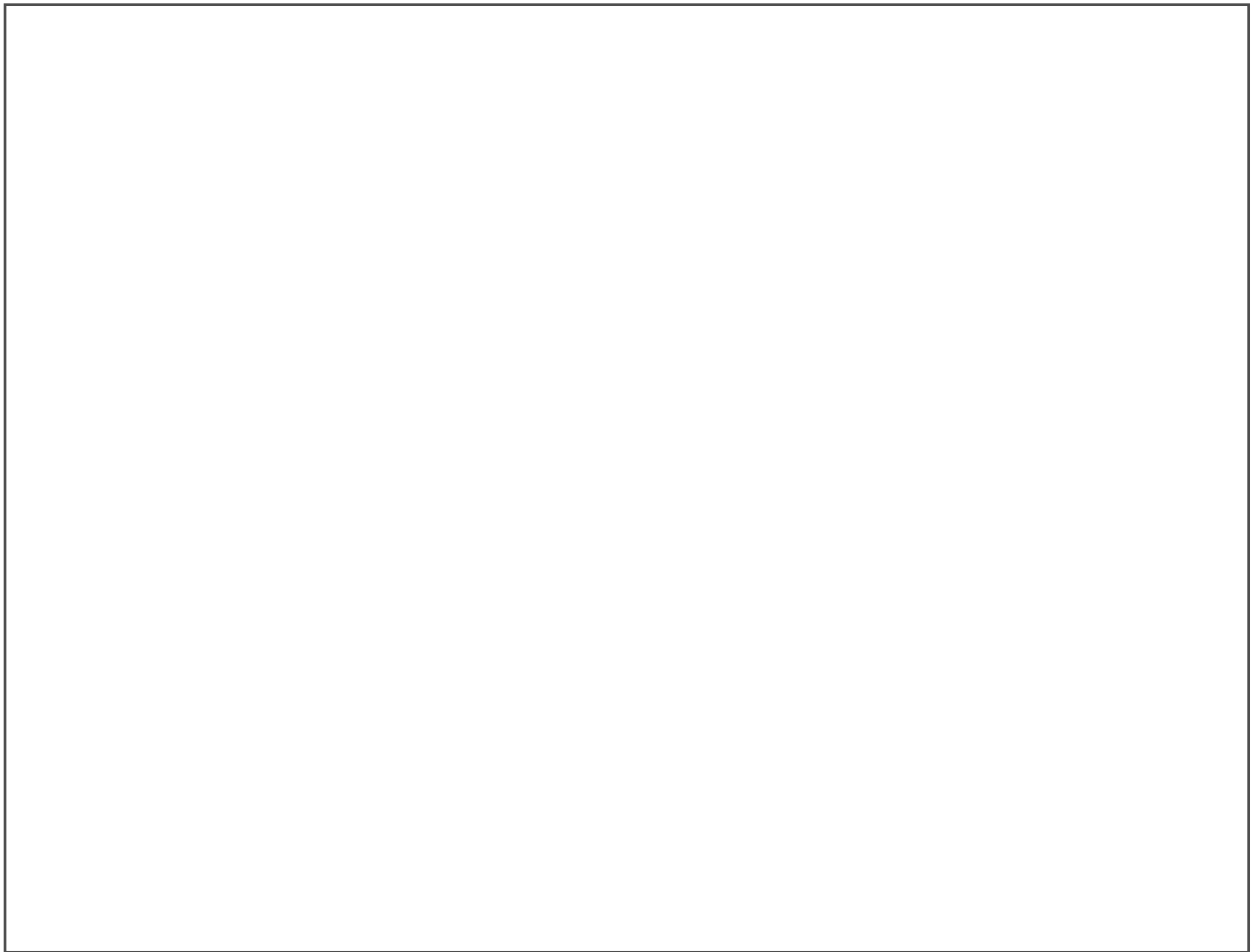
Appendix 36L (Exp 7): Post-hoc independent sample T-tests for reaction time.

Appendix 36m (Exp 7): Paired samples statistics for reaction time to help explain the 3-way interaction.
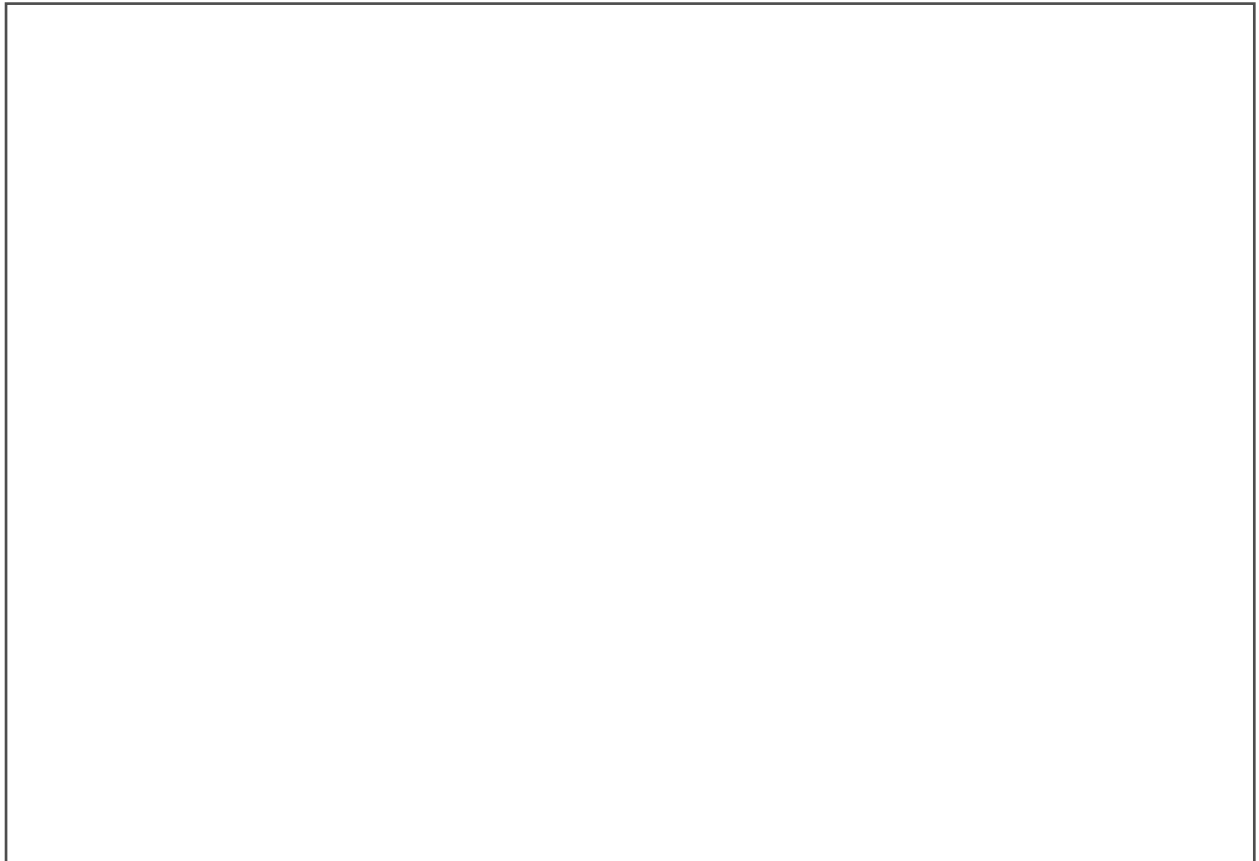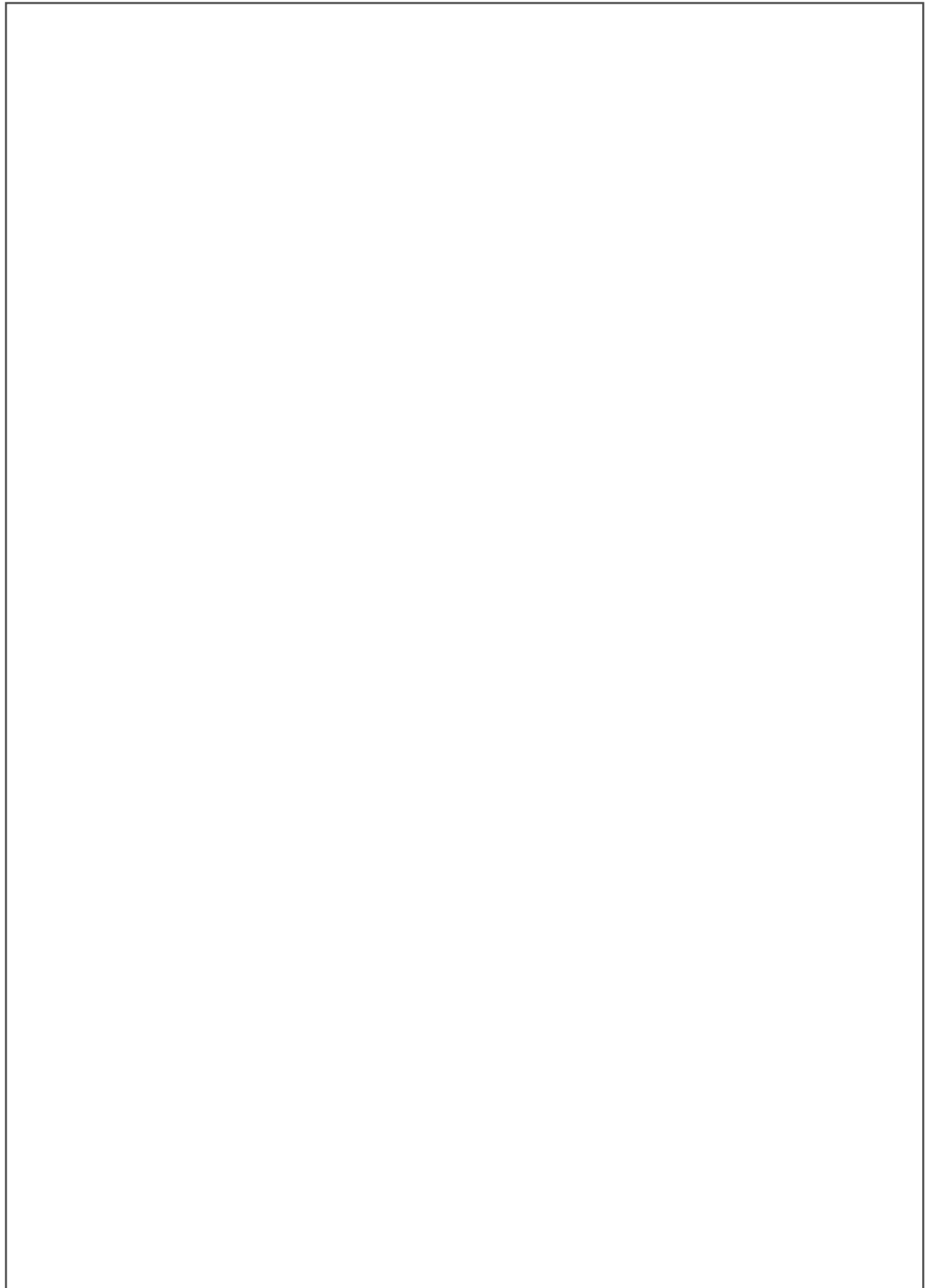
Appendix 36n (Exp 7): Post-hoc paired samples T-tests for reaction time to help explain the three-way interaction.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	C_Chem_Incon_Diff	5160.9151	13	2752.12216	763.30135
	C_Chem_Incon_NoDiff	5160.2734	13	2412.16361	669.01381
Pair 2	C_Chem_Incon_Diff	5160.9151	13	2752.12216	763.30135
	C_Chem_Cong_Diff	4017.7789	13	3151.28387	874.00889
Pair 3	C_Chem_Incon_Diff	5160.9151	13	2752.12216	763.30135
	C_Music_Incon_Diff	5541.2114	13	1766.12665	489.83540
Pair 4	C_Chem_Incon_NoDiff	5160.2734	13	2412.16361	669.01381
	C_Chem_Cong_NoDiff	9528.7764	13	4180.97600	1159.594
Pair 5	C_Chem_Incon_NoDiff	5160.2734	13	2412.16361	669.01381
	C_Music_Incon_NoDiff	9007.6847	13	4019.46066	1114.798
Pair 6	C_Chem_Cong_Diff	4017.7789	13	3151.28387	874.00889
	C_Chem_Cong_NoDiff	9528.7764	13	4180.97600	1159.594
Pair 7	C_Chem_Cong_Diff	4017.7789	13	3151.28387	874.00889
	C_Music_Cong_Diff	8470.7388	13	4155.57445	1152.549
Pair 8	C_Chem_Cong_NoDiff	9528.7764	13	4180.97600	1159.594
	C_Music_Incon_NoDiff	9007.6847	13	4019.46066	1114.798
Pair 9	C_Chem_Cong_NoDiff	9528.7764	13	4180.97600	1159.594
	C_Music_Cong_NoDiff	8361.3784	13	4518.43947	1253.190
Pair 10	C_Music_Incon_Diff	5541.2114	13	1766.12665	489.83540
	C_Music_Incon_NoDiff	9007.6847	13	4019.46066	1114.798
Pair 11	C_Music_Incon_Diff	5541.2114	13	1766.12665	489.83540
	C_Music_Cong_Diff	8470.7388	13	4155.57445	1152.549
Pair 12	C_Music_Incon_NoDiff	9007.6847	13	4019.46066	1114.798
	C_Music_Cong_NoDiff	8361.3784	13	4518.43947	1253.190
Pair 13	C_Music_Cong_Diff	8470.7388	13	4155.57445	1152.549
	C_Music_Cong_NoDiff	8361.3784	13	4518.43947	1253.190
Pair 14	Cont_Chem_Incon_Diff	7700.8390	16	4673.78594	1168.446
	Cont_Chem_Incon_NoDiff	7797.0313	16	5477.84685	1369.462
Pair 15	Cont_Chem_Incon_Diff	7700.8390	16	4673.78594	1168.446
	Cont_Chem_Cong_Diff	9333.9274	16	5957.98760	1489.497
Pair 16	Cont_Chem_Incon_Diff	7700.8390	16	4673.78594	1168.446
	Cont_Music_Incon_Diff	5092.5152	16	2046.49330	511.62333
Pair 17	Cont_Chem_Incon_NoDiff	7797.0313	16	5477.84685	1369.462
	Cont_Chem_Cong_NoDiff	7784.0069	16	3849.97638	962.49410
Pair 18	Cont_Chem_Incon_NoDiff	7797.0313	16	5477.84685	1369.462
	Cont_Music_Incon_NoDiff	7589.3496	16	3528.80587	882.20147
Pair 19	Cont_Chem_Cong_Diff	9333.9274	16	5957.98760	1489.497
	Cont_Chem_Cong_NoDiff	7784.0069	16	3849.97638	962.49410
Pair 20	Cont_Chem_Cong_Diff	9333.9274	16	5957.98760	1489.497
	Cont_Music_Incon_Diff	5092.5152	16	2046.49330	511.62333
Pair 21	Cont_Chem_Cong_Diff	9333.9274	16	5957.98760	1489.497
	Cont_Music_Cong_Diff	5871.0557	16	3134.72495	783.68124

Appendix 36o (Exp 7) *part 1*: Paired samples statistics for reaction time to help explain the 4-way interaction. Table continued on the next page

Pair 22	Cont_Chem_Cong_NoDiff	7784.0069	16	3849.97638	962.49410
	Cont_Music_Incon_NoDiff	7589.3496	16	3528.80587	882.20147
Pair 23	Cont_Chem_Cong_NoDiff	7784.0069	16	3849.97638	962.49410
	Cont_Music_Cong_NoDiff	6768.6060	16	2763.63895	690.90974
Pair 24	Cont_Music_Incon_Diff	5092.5152	16	2046.49330	511.62333
	Cont_Music_Incon_NoDiff	7589.3496	16	3528.80587	882.20147
Pair 25	Cont_Music_Incon_Diff	5092.5152	16	2046.49330	511.62333
	Cont_Music_Cong_Diff	5871.0557	16	3134.72495	783.68124
Pair 26	Cont_Music_Incon_NoDiff	7589.3496	16	3528.80587	882.20147
	Cont_Music_Cong_NoDiff	6768.6060	16	2763.63895	690.90974
Pair 27	Cont_Music_Cong_Diff	5871.0557	16	3134.72495	783.68124
	Cont_Music_Cong_NoDiff	6768.6060	16	2763.63895	690.90974

Appendix 36o (Exp 7) *part 2***:** Continuation of paired samples statistics for reaction time to help explain the 4-way interaction.

Appendix (Exp 7) 36p: Post-hoc paired samples statistics for reaction time to help explain the 4-way interaction.

Appendix 37a (Exp 7): Descriptive statistics for the total time in congruency RoIs
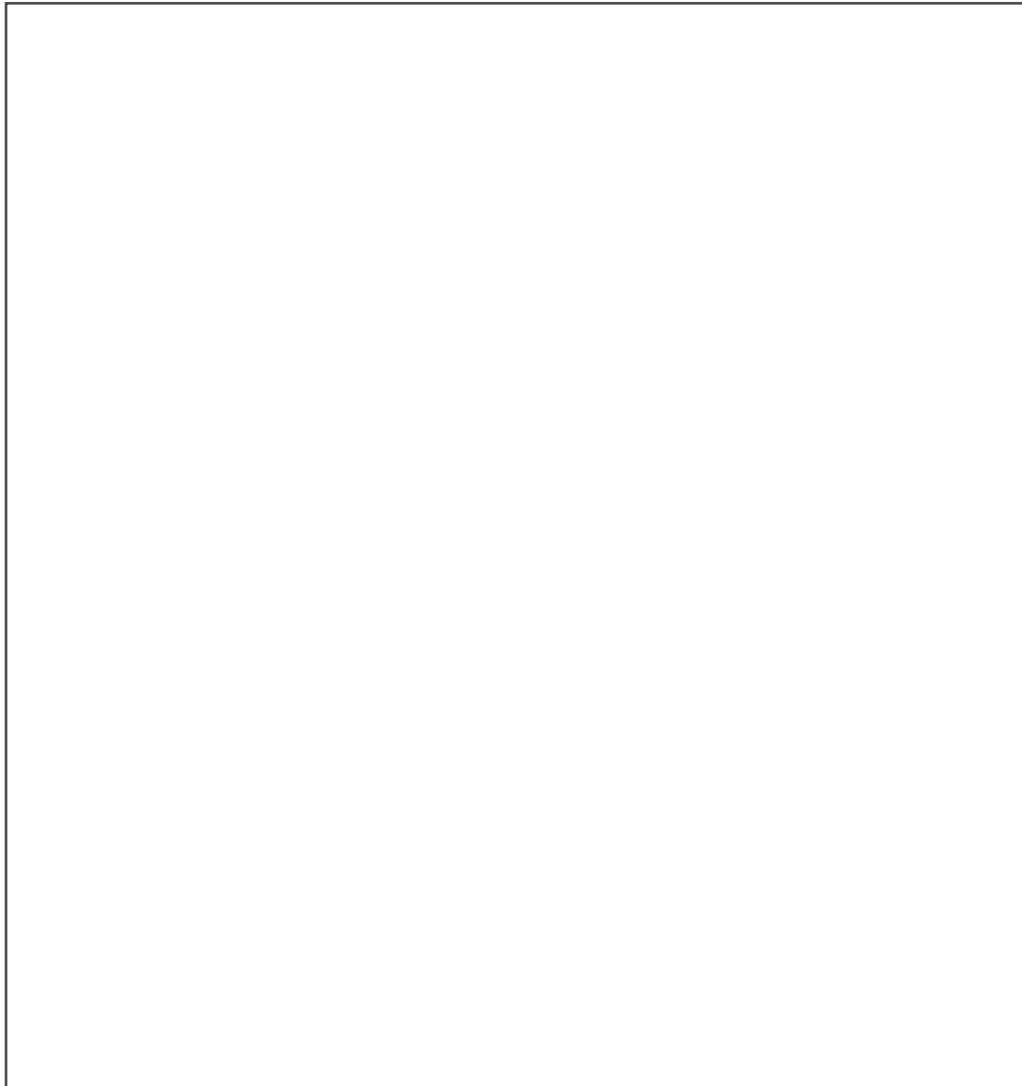
Appendix 37b (Exp 7): 2x2x2 mixed design ANOVA for the total time in congruency RoIs. There is a main effect of picture type, an interaction between picture type and group and an interaction between difference and group.

Appendix 37c (Exp 7): Between-groups ANOVA output for the total time in congruency RoIs. There is no main effect of group.

Appendix 37d (Exp 7): Paired samples statistics for the total time spent in congruency RoIs in chemistry and music pictures for the two participant groups.
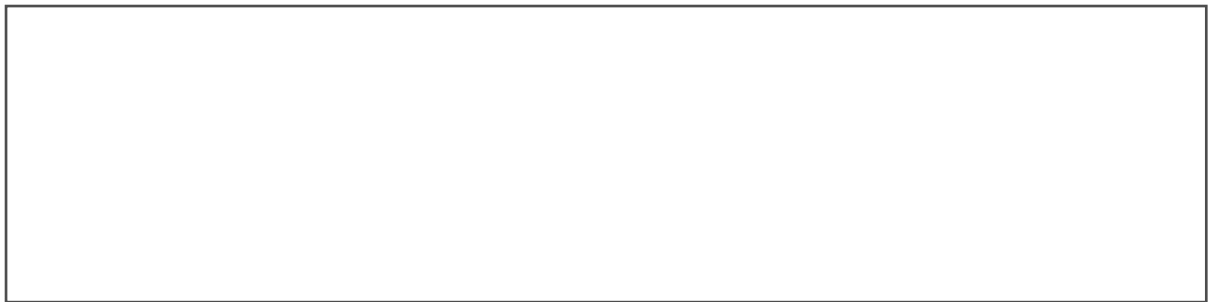
Appendix 37e (Exp 7): Post-hoc paired samples T-tests for the total time spent in congruency RoIs in chemistry and music pictures for the two participant groups.

Appendix 37f (Exp 7): Paired samples statistics for the total time spent in congruency RoIs in pictures with and without a difference for the two participant groups.
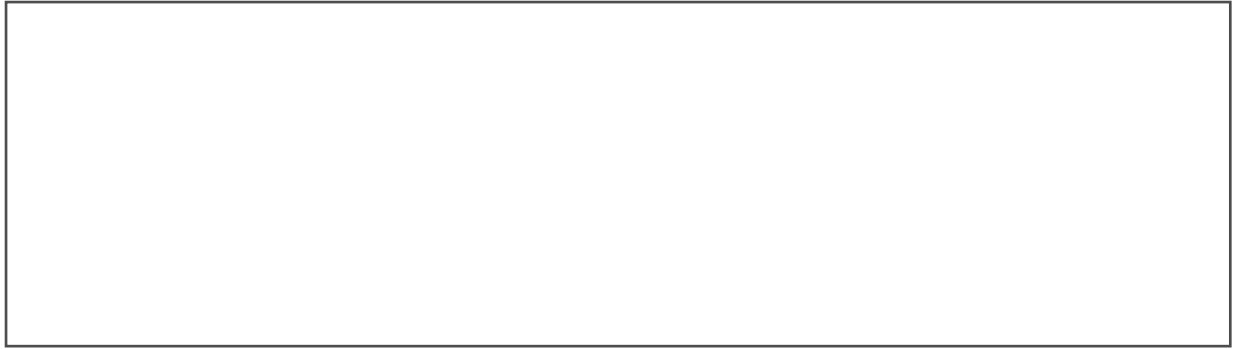
Appendix 37g (Exp 7): Paired samples T-tests for the total time spent in congruency RoIs in pictures with and without a difference for the two participant groups.
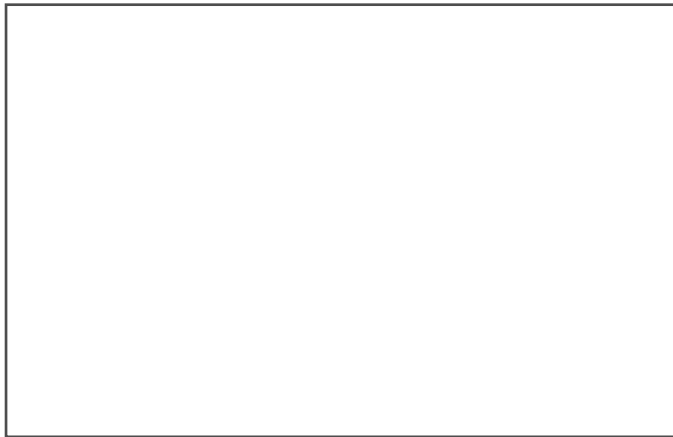
Appendix 38a (Exp 7): Descriptive statistics for the total time spent in difference RoIs in congruous and incongruous chemistry and music pictures.
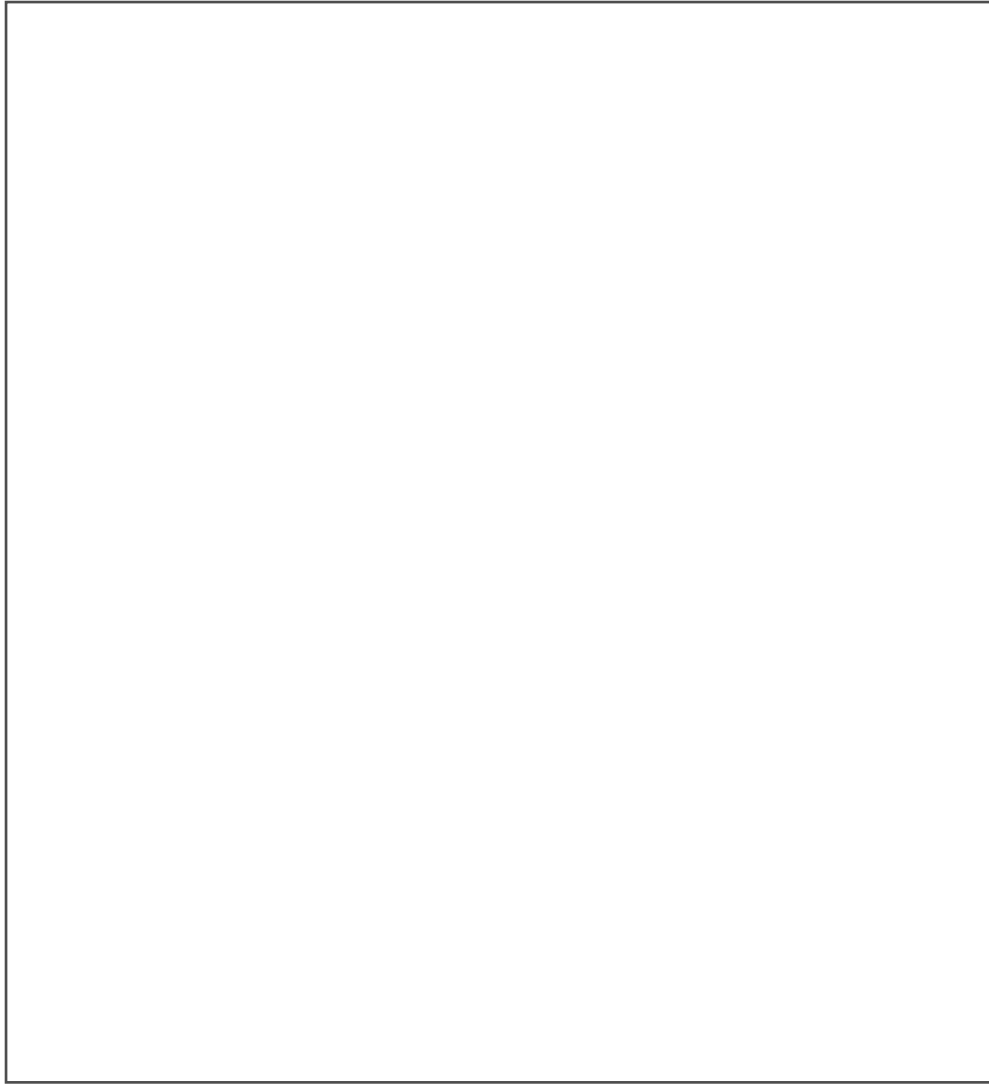
Appendix 38b (Exp 7): 2x2x2 mixed design ANOVA for the total time spent in difference RoIs in congruous and incongruous chemistry and music pictures. There is a reliable main effect of congruency.

Appendix 38c (Exp 7): Between-groups ANOVA output for the total time spent in difference RoIs in congruous and incongruous chemistry and music pictures. There is no main effect of group.
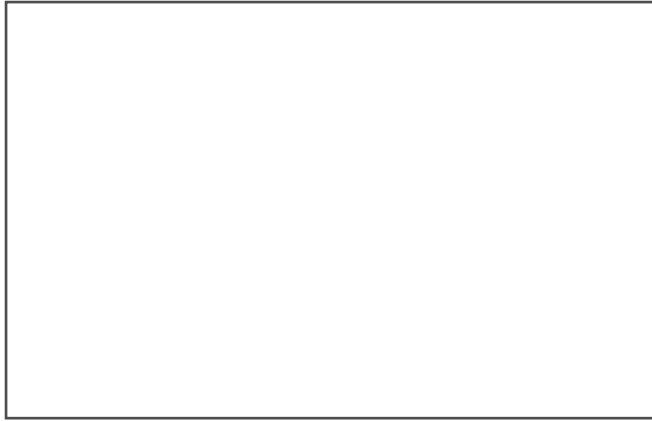
Appendix 39a (Exp 7): Descriptive statistics for the number of fixations in congruency RoIs for chemists and control participants when viewing chemistry and music pictures with and without differences.
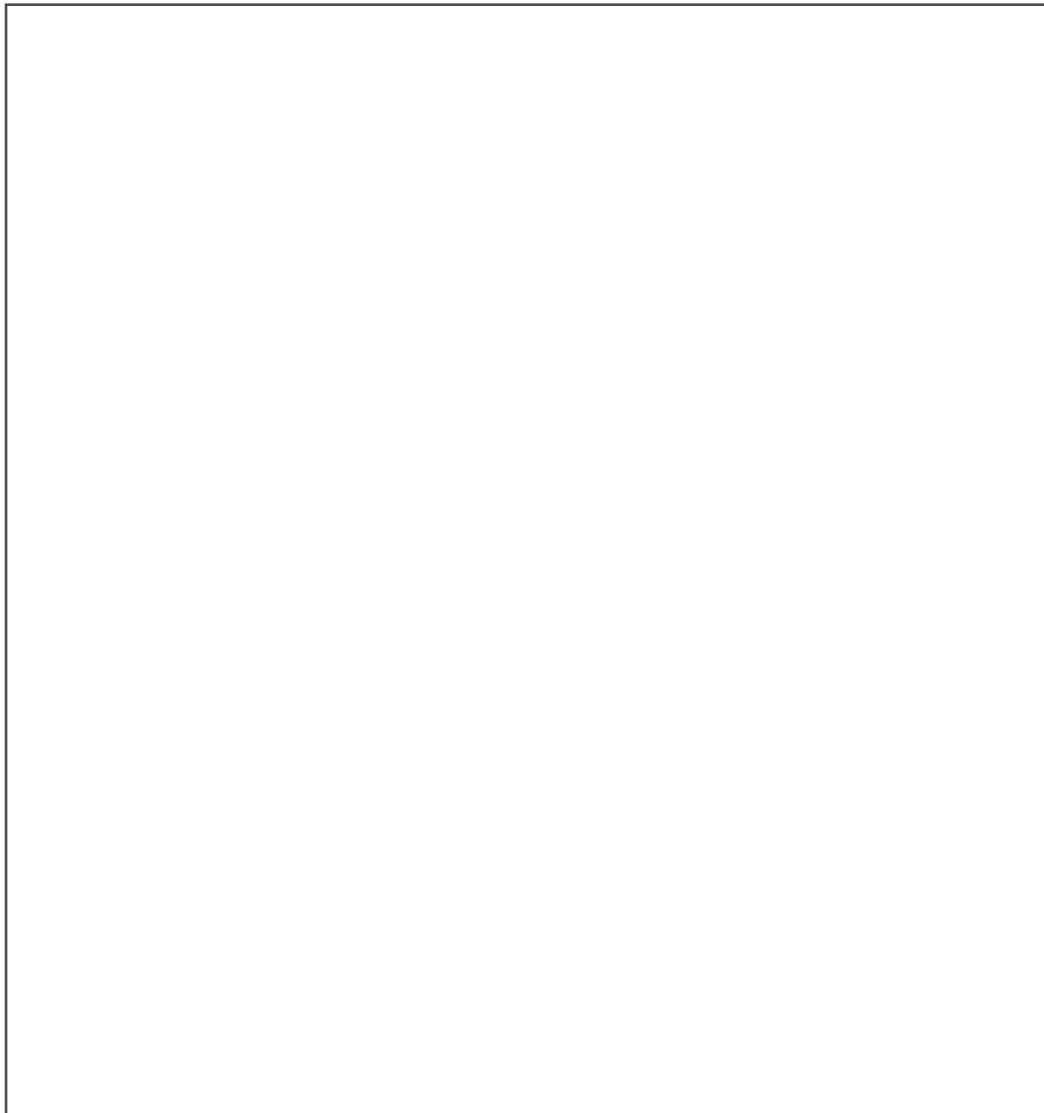
Appendix 39b (Exp 7): 2x2x2 mixed-design ANOVA for the number of fixations in congruency RoIs for chemists and control participants when viewing chemistry and music pictures with and without differences.

Appendix 39c (Exp 7): Between-groups ANOVA output for the number of fixations in congruency RoIs. There is no main effect of group.

Appendix 39d (Exp 7): Descriptive statistics for the number of fixations in congruency RoIs for pictures with and without a difference.

**Appendix 40a (Exp 7):** Descriptive statistics for the number of fixations in difference RoIs for chemists and control participants when viewing congruent and incongruent chemistry and music pictures.
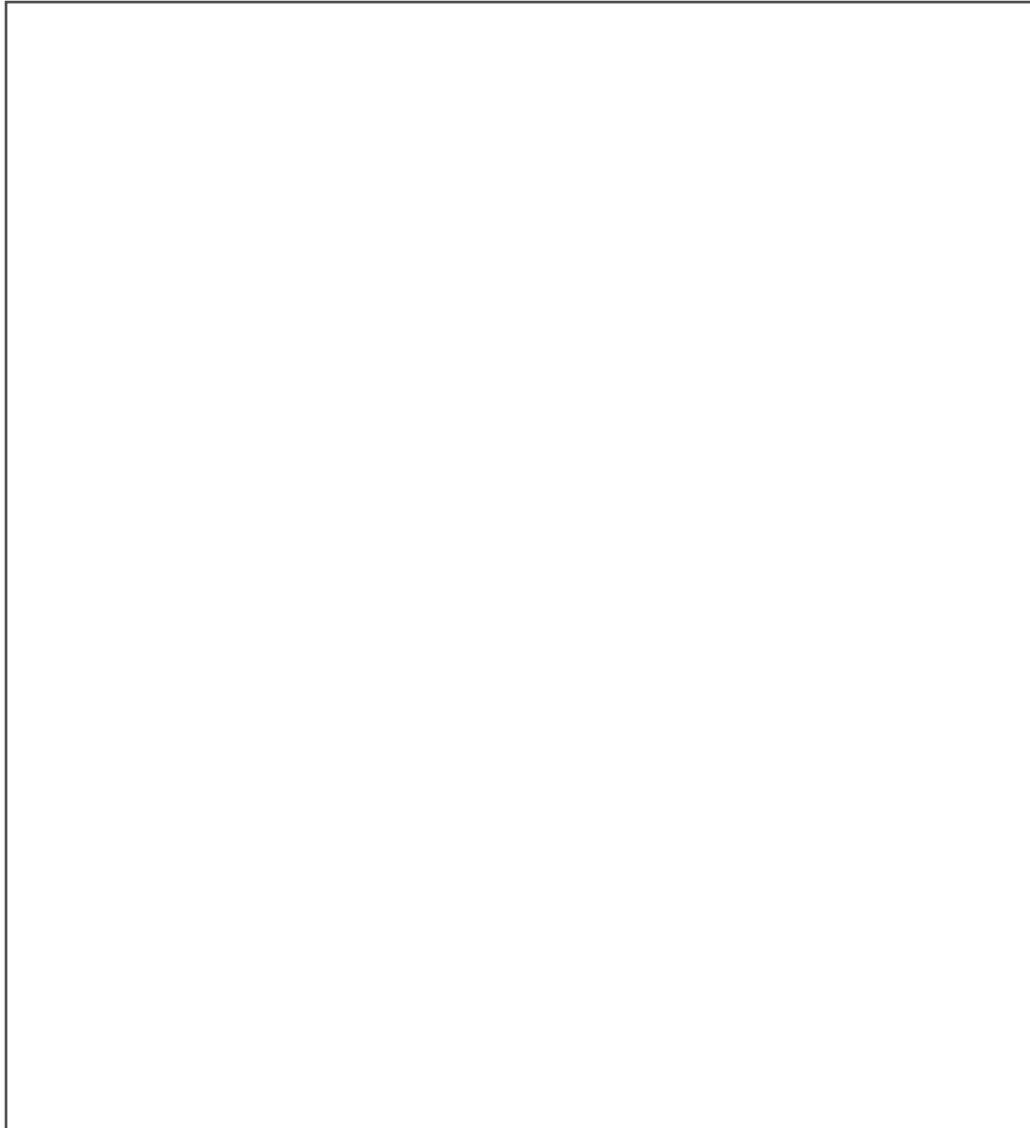
Appendix 40b (Exp 7): 2x2x2 mixed-design ANOVA for the number of fixations in congruency RoIs for chemists and control participants when viewing chemistry and music pictures with and without differences.

Appendix 40c (Exp 7): Descriptive statistics for the number of fixations in difference RoIs for pictures with and without a difference.

Appendix 40d (Exp 7): Descriptive statistics for the number of fixations in difference RoIs for congruent and incongruent pictures.

Appendix 40e (Exp 7): Paired samples statistics for the number of fixations in difference RoIs for congruent and incongruent pictures.

Appendix 40f (Exp 7): Paired samples T-tests for the number of fixations in difference RoIs for congruent and incongruent pictures.

Appendix 41a (Exp 7): Group statistics for the percentage of first fixations that fell in congruency RoIs for chemistry and music pictures.
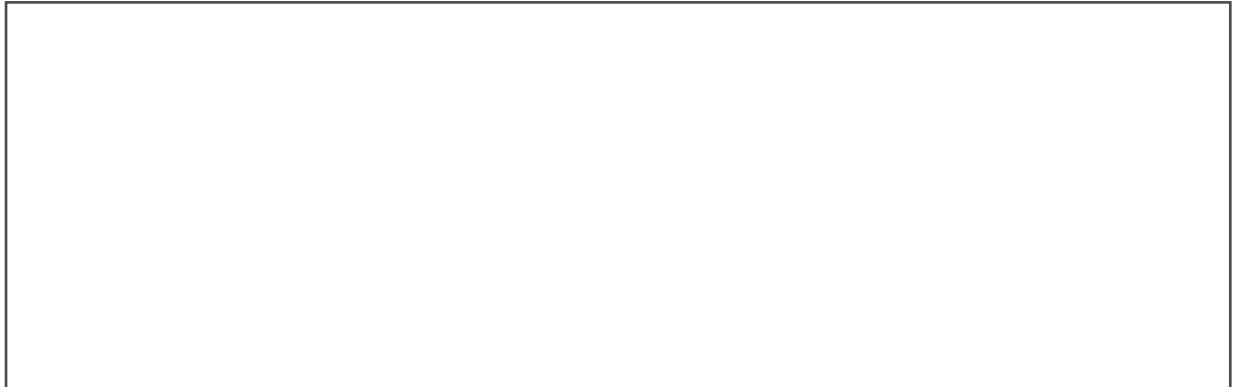
Appendix 41b (Exp 7): Independent samples T-test for the percentage of first fixations that fell in congruency RoIs for chemistry and music pictures. There is a reliable difference between the groups for chemistry pictures.

Appendix 41c (Exp 7): Paired samples statistics for the percentage of first fixations of chemists and control participants that fell in congruency RoIs for chemistry and music pictures.

Appendix 41d (Exp 7): Paired samples statistics for the percentage of first fixations of chemists and control participants that fell in congruency RoIs for chemistry and music pictures.

Appendix 42a (Exp 7): Group statistics for the percentage of first fixations that fell in difference RoIs for chemistry and music pictures.
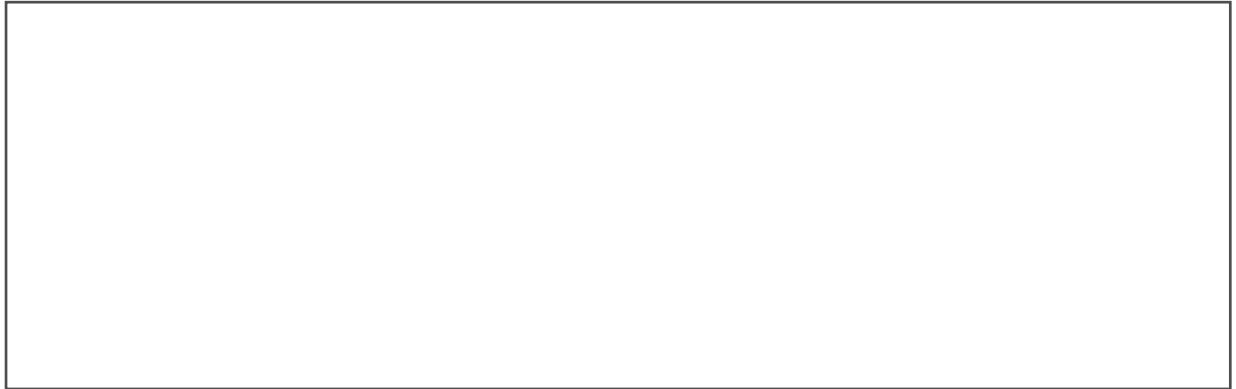
Appendix 42b (Exp 7): Independent samples T-test for the percentage of first fixations that fell in difference RoIs for chemistry and music pictures.

Appendix 42c (Exp 7): Paired samples statistics for the percentage of first fixations of chemists and control participants that fell in difference RoIs for chemistry and music pictures.

Appendix 42d (Exp 7): Paired samples statistics for the percentage of first fixations of chemists and control participants that fell in difference RoIs for chemistry and music pictures.
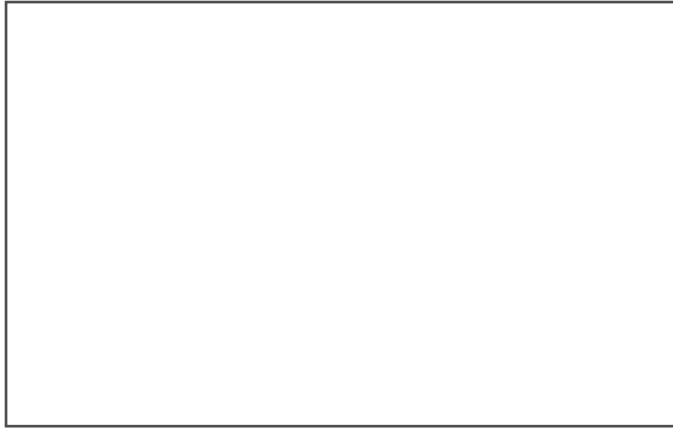
Appendix 42e (Exp 7): Descriptive statistics for time to first fixate the incongruous and difference regions.
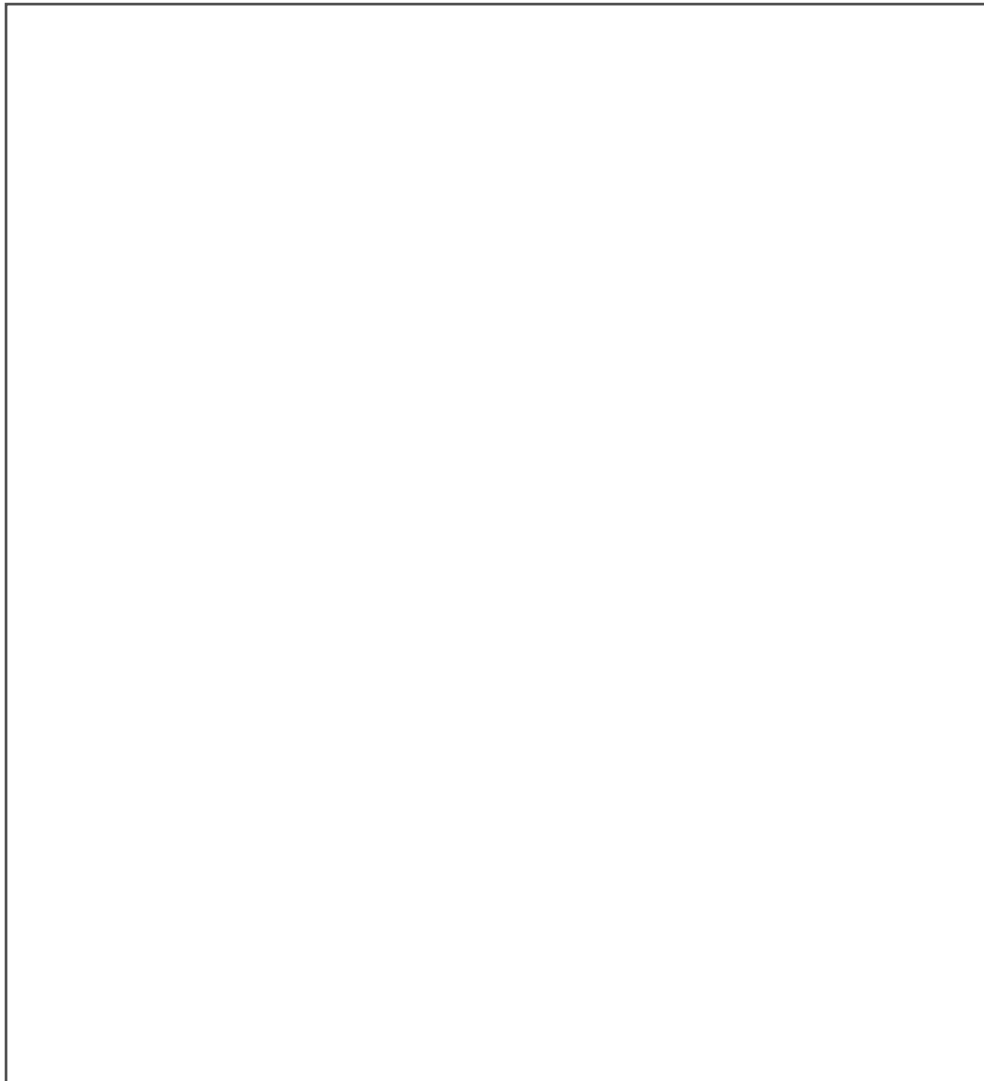

Appendix 42f (Exp 7): 2x2x2 mixed-design ANOVA for time to first fixate incongruous and difference RoIs. There are no main effects and no reliable interactions.

Appendix 42g (Exp 7): Between groups output from the 2x2x2 mixed-design ANOVA for time to first fixate incongruous and difference RoIs. There is a reliable difference between the participant groups.

Appendix 43a (Exp 8): Paired samples statistics for the accuracy of identifying 'people' and 'no people' pictures at recognition test.
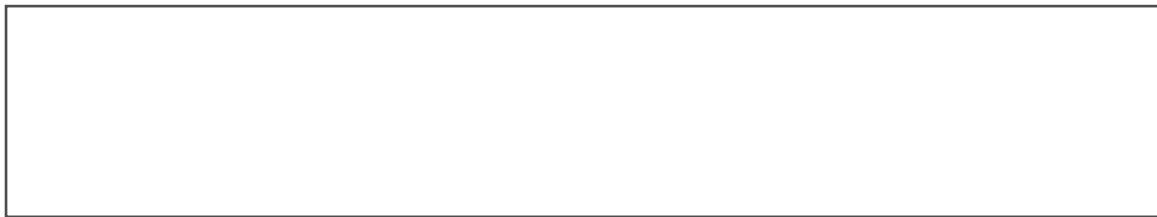
Appendix 43b (Exp 8): Paired samples T-test for the accuracy of identifying 'people' and 'no people' pictures at recognition test. There was a reliable difference between the stimulus types.

Appendix 43c (Exp 8): 2x2 repeated measures ANOVA for accuracy in identifying old and new 'people' and 'no people' pictures at recognition test. There was a reliable main effect of stimulus type and a reliable main effect of whether the picture was old or now. There was no reliable interaction.


Appendix 43d (Exp 8): A one-way ANOVA comparing accuracy on pictures containing one person compared to those containing many people. There is no reliable difference in accuracy.

Appendix 43e (Exp 8): A one-way ANOVA comparing accuracy on pictures with RoIs less than 20 pixels and those with RoIs of 20 pixels or more. There is no reliable difference in accuracy.

Appendix 43f (Exp 8): A one-way ANOVA comparing accuracy on pictures with RoIs less than 10 degrees of visual arc from the centre and those with RoIs of 10 degrees or more from the centre. There is no reliable difference in accuracy.
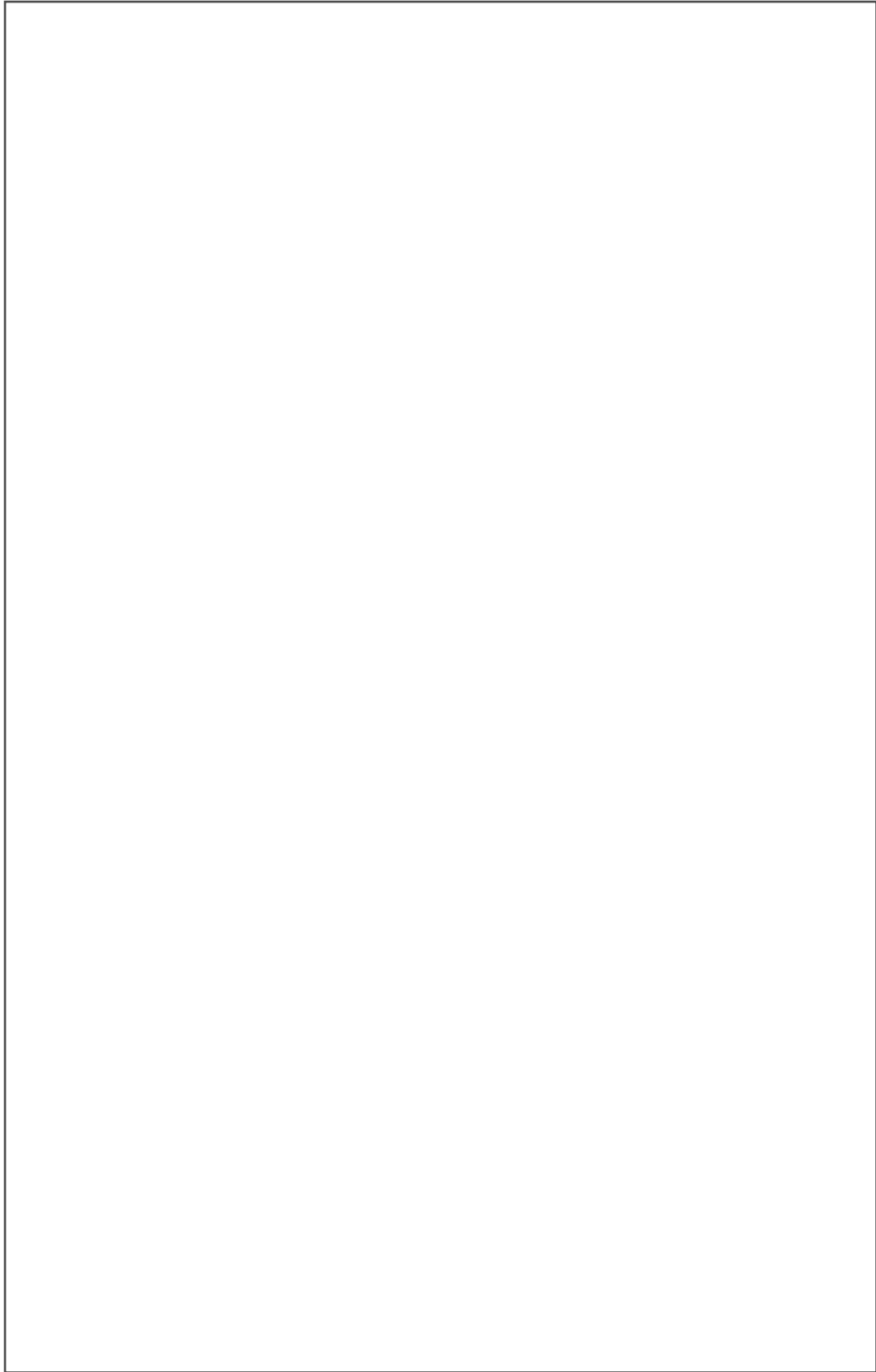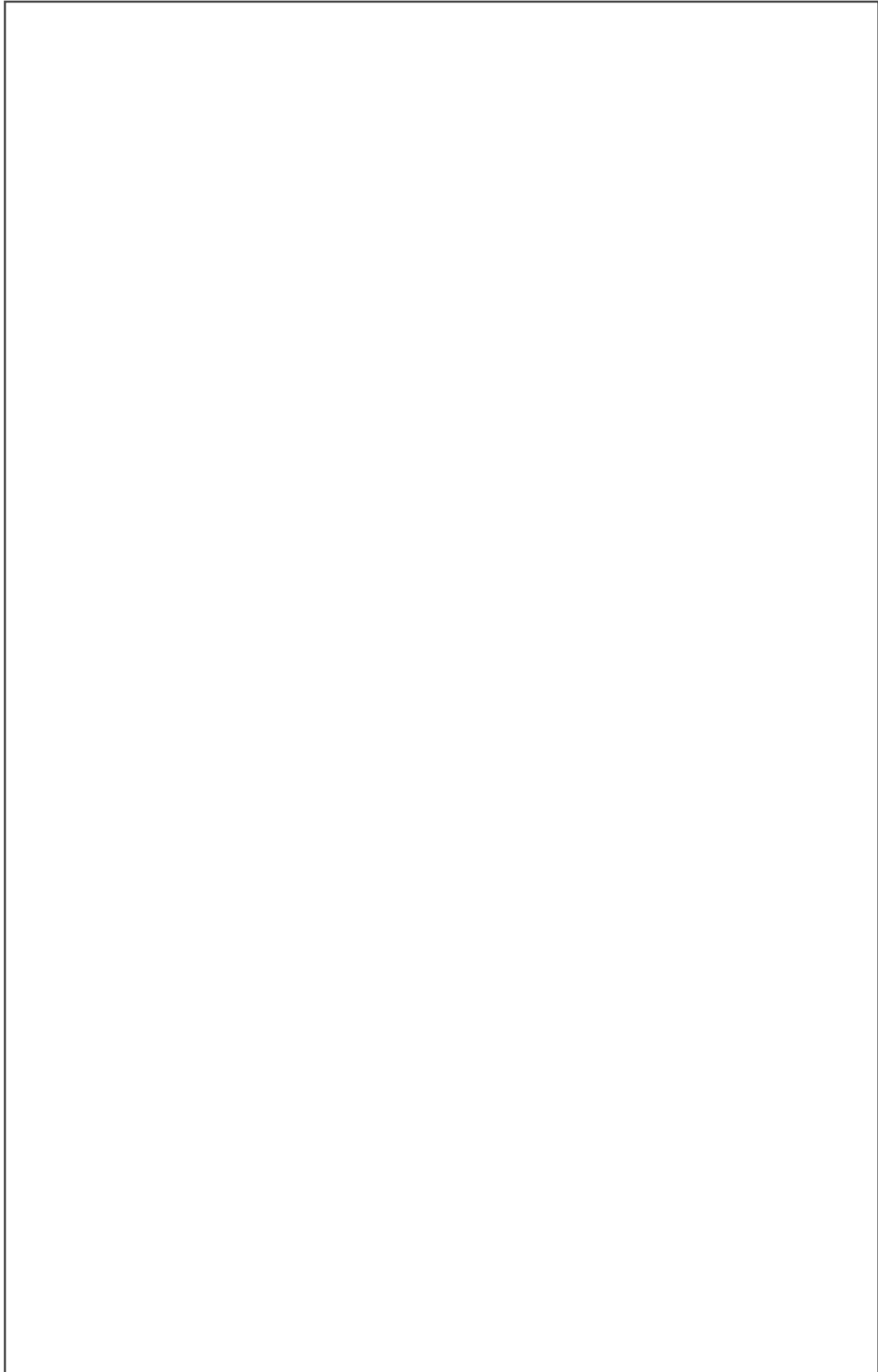
Appendix 44 (Exp 8): A 2x2x2 repeated measures ANOVA comparing fixation duration on old and new 'people' and 'no people' pictures that participants correctly and incorrectly identifies at recognition. There is a reliable main effect of whether the picture was old or new.

Appendix 45a (Exp 8): Paired samples statistics for the number of fixations and total average time in RoIs, compared to chance. 'No people' represents the chance baseline whereby the people RoI has been applied to a randomly selected no people picture.

Appendix 45b (Exp 8): Paired samples T-tests for the number of fixations and total average time in RoIs, compared to chance. 'No people' represents the chance baseline whereby the people RoI has been applied to a randomly selected no people picture. More fixations fell inside the RoIs of the people pictures than would be expected by chance and that more time was spent fixating within the RoIs of the people pictures than would be expected by chance.
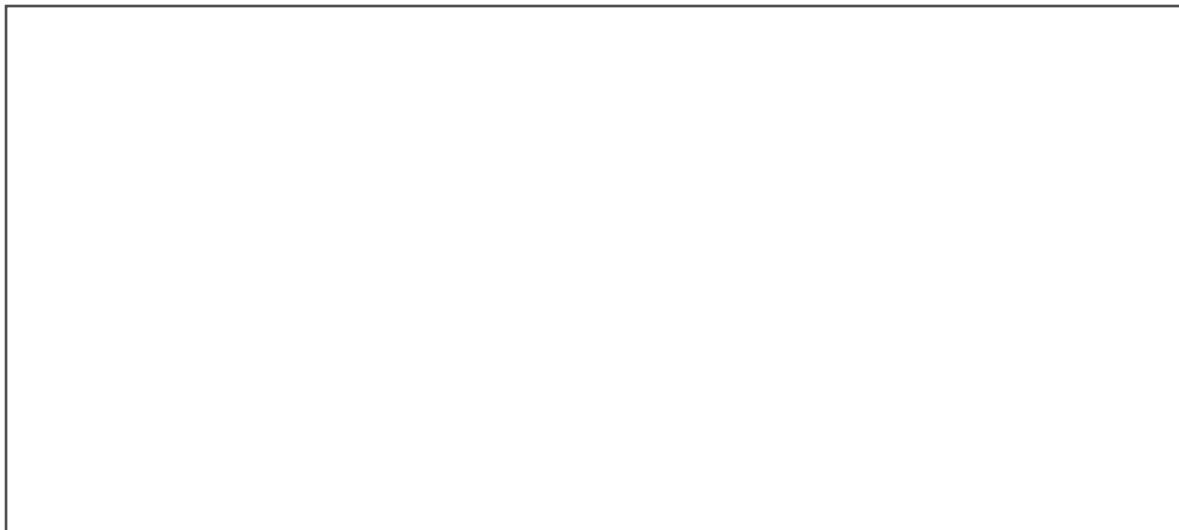
Appendix 46a (Exp 8): 2x2x2 repeated measures ANOVA for the average saccadic amplitude on old and new 'people' and 'no people' pictures that participants correctly and incorrectly identifies at recognition.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Incorrect/No Ppl/Old	3.7333	14	.61912	.16547
	Incorrect/Ppl/Old	4.6350	14	1.08791	.29076
Pair 2	Incorrect/No Ppl/New	4.2949	14	1.11761	.29870
	Incorrect/Ppl/New	4.3712	14	.76084	.20334
Pair 3	Correct/No Ppl/Old	4.3090	14	.84922	.22696
	Correct/Ppl/Old	4.1116	14	.71211	.19032
Pair 4	Correct/No Ppl/New	4.2949	14	1.11761	.29870
	Correct/Ppl/New	4.6477	14	.52483	.14027
Pair 5	Incorrect/No Ppl/Old	3.7333	14	.61912	.16547
	Correct/No Ppl/Old	4.3090	14	.84922	.22696
Pair 6	Incorrect/Ppl/Old	4.6350	14	1.08791	.29076
	Correct/Ppl/Old	4.1116	14	.71211	.19032
Pair 7	Incorrect/No Ppl/New	4.2949[a]	14	1.11761	.29870
	Correct/No Ppl/New	4.2949[a]	14	1.11761	.29870
Pair 8	Incorrect/Ppl/New	4.3712	14	.76084	.20334
	Correct/Ppl/New	4.6477	14	.52483	.14027

Appendix 46b (Exp 8): Paired samples statistics to further investigate the 3-way interaction in saccadic amplitude.

Appendix 46c (Exp 8): Paired samples T-tests to further investigate the 3-way interaction in saccadic amplitude between correct/incorrect, new/old, and people/no people.
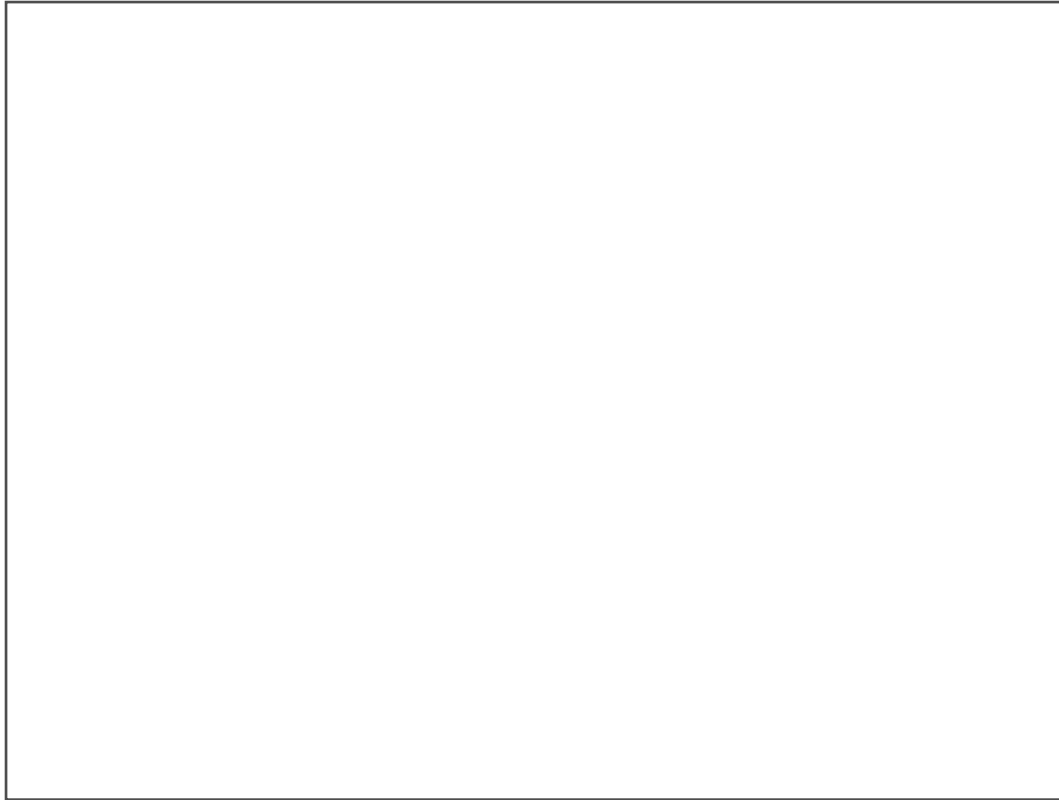
Appendix 47a (Exp 8): 2x2 repeated measures ANOVA for string similarities of scanpaths on 'people' and 'no people' pictures that participants identified correctly and incorrectly. There is a main effect of whether the picture contains people.

Appendix 47b (Exp 8): Mean string similarities scores for people and no people pictures.

Appendix 47c (Exp 8): Paired samples statistics comparing each string similarity comparison to chance

Appendix 47d (Exp 8): Paired samples T-tests. Three out of the four comparisons were reliable greater than chance.

Appendix 48a (Exp 9): Repeated measures ANOVA for recognition accuracy of negative, neutral and positive stimuli. There is a reliable effect of stimulus type.

Appendix 48b (Exp 9): Paired samples statistics for recognition accuracy of negative, neutral and positive stimuli.
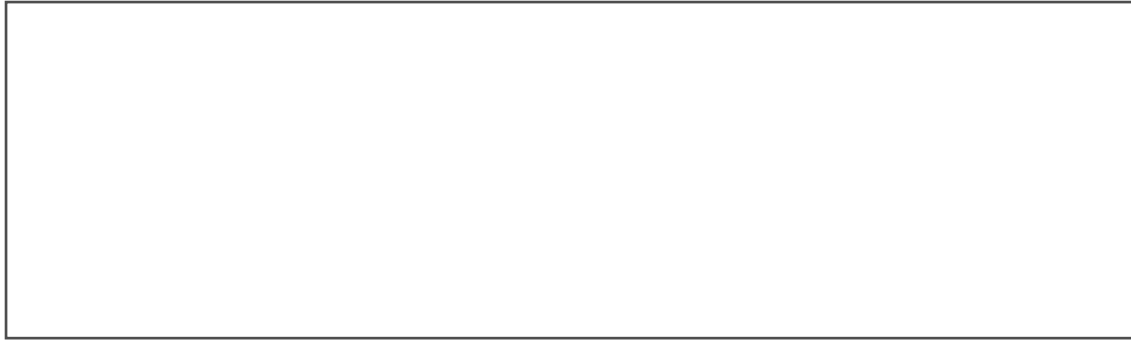
Appendix 48c (Exp 9): Paired samples T-tests for recognition accuracy of negative, neutral and positive stimuli. Negative pictures are recognised more accurately than neutral pictures.

Appendix 49a (Exp 9): Repeated measures ANOVA for the average number of references to visually salient RoIs of negative, neutral and positive pictures.

Appendix 49b (Exp 9): Paired samples statistics for the average number of references to visually salient RoIs of negative, neutral and positive pictures.

Appendix 49c (Exp 9): Paired samples T-tests for the average number of references to visually salient RoIs of negative, neutral and positive pictures.

Appendix 49d (Exp 9): Paired samples statistics for the average number of references to emotive RoIs of negative and positive pictures.

Appendix 49e (Exp 9): Paired samples T-test for the average number of references to emotive RoIs of negative and positive pictures.

Appendix 49f (Exp 9): Paired samples statistics for the average number of references to salient RoIs and emotive RoIs, for positive and negative stimuli.
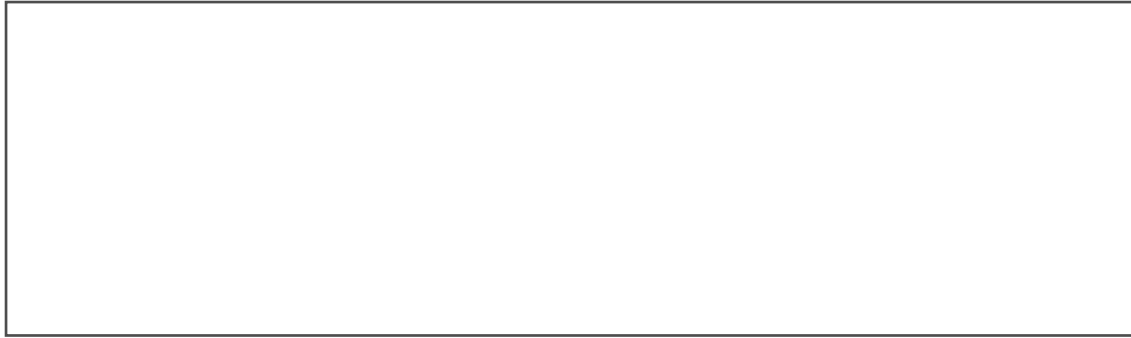
Appendix 49g (Exp 9): Paired samples T-tests for the average number of references to salient RoIs and emotive RoIs, for positive and negative stimuli.
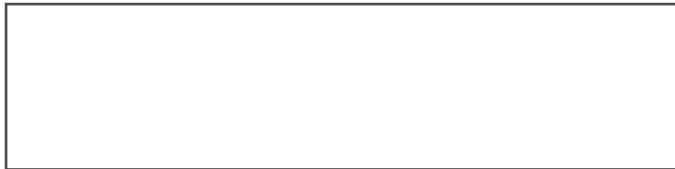
Appendix 50a (Exp 9): A repeated measures ANOVA comparing the percentage of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 50b (Exp 9): Paired samples statistics comparing the percentage of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 50c (Exp 9): Paired samples T-tests comparing the percentage of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 50d (Exp 9): Paired samples statistics for the percentage of fixations that fell in emotive RoIs on positive and negative pictures.

Appendix 50e (Exp 9): Paired samples T-test comparing the percentage of fixations that fell in emotive RoIs on positive and negative pictures. There is a reliable difference between the picture types.

Appendix 50f (Exp 9): Paired samples statistics for the percentage of fixations that fell in salient RoIs and emotive RoIs.

Appendix 50g (Exp 9): Paired samples T-test comparing the percentage of fixations that fell in salient RoIs and emotive RoIs.
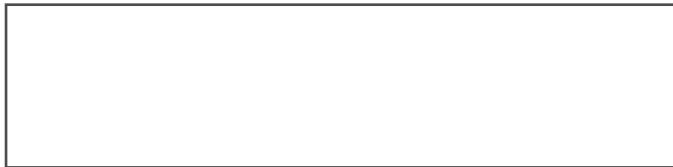
Appendix 51a (Exp 9): A repeated measures ANOVA comparing the average duration of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 51b (Exp 9): Paired samples statistics comparing the average duration of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 51c (Exp 9): Paired samples T-tests comparing the average duration of fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 51d (Exp 9): Paired samples statistics for the average duration of fixations that fell in emotive RoIs on positive and negative pictures.

Appendix 51e (Exp 9): Paired samples T-test comparing the average duration of fixations that fell in emotive RoIs on positive and negative pictures.

Appendix 51f (Exp 9): Paired samples statistics for the average duration of fixations that fell in salient RoIs and emotive RoIs.

Appendix 51g (Exp 9): Paired samples T-test comparing the average duration of fixations that fell in salient RoIs and emotive RoIs.

Appendix 52a (Exp 9): A repeated measures ANOVA comparing the percentage of first fixations that fell inside salient RoIs, for each picture type.

Appendix 52b (Exp 9): Paired samples statistics comparing the percentage of first fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 52c (Exp 9): Paired samples T-tests comparing the percentage of first fixations that fell inside the visually salient RoIs for each stimulus type.

Appendix 52d (Exp 9): Paired samples statistics for the percentage of first fixations that fell in emotive RoIs on positive and negative pictures.

Appendix 52e (Exp 9): Paired samples T-test comparing the percentage of first fixations that fell in emotive RoIs on positive and negative pictures.

Appendix 52f (Exp 9): Paired samples statistics for the percentage of first fixations that fell in salient RoIs and emotive RoIs.

Appendix 52g (Exp 9): Paired samples T-test comparing the average duration of fixations that fell in salient RoIs and emotive RoIs.

Appendix 53a (Exp 9): Paired samples statistics for the similarities of actual scanpaths compared to model-predicted saliency scanpaths, for each stimulus type.

Appendix 53b (Exp 9): Paired samples T-tests for the similarities of actual scanpaths compared to model-predicted saliency scanpaths, for each stimulus type.

Appendix 54a (Exp 10): Paired samples statistics for recognition accuracy of stimuli presented in the same, reverse or random order.

Appendix 54b (Exp 10): Paired samples statistics for recognition accuracy of stimuli presented in the same, reverse or random order.

Appendix 55a (Exp 10): Paired samples statistics for reaction time to recognise stimuli presented in the same, reverse or random order.

Appendix 55b (Exp 10): Paired samples T-tests comparing reaction time to recognise stimuli presented in the same, reverse or random order.
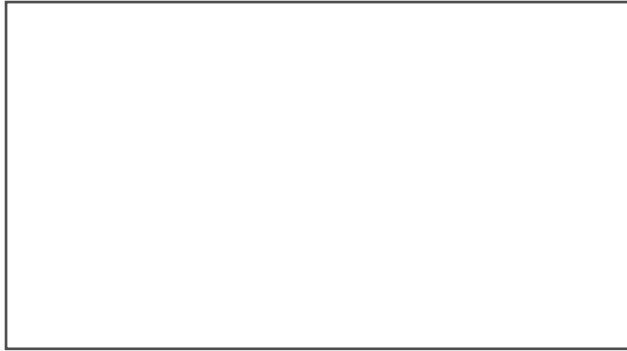
Appendix 56a (Exp 10): The position in the 'same' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 56b (Exp 10): The position in the 'reverse' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 56c (Exp 10): The position in the 'random' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 56d (Exp 10): The position in the 'same' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average reaction for that picture.

Appendix 56e (Exp 10): The position in the 'reverse' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average reaction time for that picture.

Appendix 56f (Exp 10): The position in the 'random' fixation sequence of the most visually salient feature was calculated for each picture and then correlated with the average reaction time for that picture.

Appendix 57a (Exp 10): The position in the 'same' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 57b (Exp 10): The position in the 'reverse' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 57c (Exp 10): The position in the 'random' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average accuracy for that picture.

Appendix 57d (Exp 10): The position in the 'same' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average reaction for that picture.

Appendix 57e (Exp 10): The position in the 'reverse' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average reaction time for that picture.

Appendix 57f (Exp 10): The position in the 'random' fixation sequence of the most semantically interesting feature was calculated for each picture and then correlated with the average reaction time for that picture.

%

correct

Chemistry