The Fidelity of Visual Memory for Scenes

Rose Emily Wastling

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Psychology

June 2020

INTELLECTUAL PROPERTY

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

ACKNOWLEDGMENTS

I would like to say a huge thank you to my supervisors, Denis McKeown, David Bunce, Richard Allen and Charity Brown for all the support, encouragement, and guidance they have given me throughout my PhD. Special thanks must go to Charity who has really helped to support me through the final few months.

Thank you to the ESRC for funding my research and giving me this opportunity. I also want to thank all of the staff in the School of Psychology at the University of Leeds who have been so supportive over the years. I have also been very lucky to make many friends during my PhD, the experience wouldn’t have been the same without you all.

Thank you to my family for always supporting me, and reassuring me when I have been worried about the enormity of this task. Most of all, I would like to thank and dedicate this thesis to my husband, Diarmaid, and my daughter, Annabelle. You have both been so patient, looked after me, and of course, provided much-welcomed distraction throughout.

ABSTRACT

Previous research investigating visual memory for scenes has focused predominantly on memory for intentionally encoded static scenes, which may not be reflective of the way in which we view scenes every day. Across seven experiments, a range of experimental paradigms were used to test the fidelity of visual memory for natural, everyday scenes, over different periods of time. This included experiments which used incidental encoding and dynamic scene stimuli.

In two initial experiments (Chapter Two), participants first viewed static scene images incidentally, whereby they were unaware of the later memory test. An explicit recognition task was later given (a two-alternative forced choice task). In addition, an adapted version of the recent-probes task, an implicit memory task, where participants were again unaware their memory for earlier viewed scenes was being tested. Here, participants took part in a task in which previously viewed scenes were embedded in certain trials; participants' responses to these old stimuli compared to novel stimuli were assessed. Findings from these experiments support the notion of a detailed memory trace for scenes that endures over several days, even when scenes are incidentally encoded, and not explicitly probed.

The research that uses dynamic scene stimuli is then reported. To enable this, a stimuli set of dynamic scenes was created, and through two experiments, separate groups of participants rated these scenes on a number of attributes, including complexity, distinctiveness and attractiveness (Chapter Three). Using these novel stimuli, three experiments are then reported (Chapters Four and Five). Participants viewed short film episodes depicting everyday dynamic scenes, followed by a recognition test in which they were presented with selected static frames taken either from those film episodes, or highly similar foil images taken from the same scenes, but from a different time point to the extract participants viewed. Participants performed well and above chance when they were given a memory test immediately or after a longer delay period, but were reliably more accurate when tested immediately. Furthermore, to understand the detail in which the dynamic scenes were remembered across time, accuracy for static 'target' frames taken from the beginning, middle and end of scenes was compared, to test if certain parts of a scene are better remembered than others. Overall, accuracy was good for all target types, regardless of temporal position. When tested immediately, there was some evidence that accuracy was higher for target frames taken from the middle or end of the films compared to the beginning, which is consistent with outcomes for serial order recency effects in visual memory. The results are interpreted conceptually within the notion of a highly detailed memory ‘map’ or trace, preserving spatiotemporal detail from recent dynamic visual episodes.

The experiments reported in this thesis demonstrate the fidelity with which natural, everyday scenes are remembered. This is the case even under circumstances of incidental encoding and retrieval, varying periods of delay between encoding and test, and when scene stimuli are static or dynamic.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ACRONYMS

AFC	Alternative forced choice
M	Mean
MS	Milliseconds
NRN	Non-recent negative
RN	Recent negative
RT	Response time
SE	Standard error
VLTM	Visual long-term memory
VSTM	Visual short-term memory

CHAPTER ONE
GENERAL INTRODUCTION

1.1 Introduction

We live and interact within complex and dynamic visual environments, encountering many different scenes every day; for example, when travelling, working, shopping and exercising. All these events present new opportunities to be exposed to new scenes, or familiar scenes that have changed in some way since we last viewed them. Given the vast amount of scenes that we view regularly, and the level of detail contained within each scene, the challenge for researchers has been to understand how we structure and record our memory of scenes with apparently remarkable ease. Consistent with definitions used in the literature, a visual scene is defined here as depicting a real life place or setting, containing a number of different objects against a fixed background (Henderson & Hollingworth, 1999; Wolfe, Horowitz & Michod, 2007).

Within experimental psychology, research has been conducted to address the question of how many scenes we are capable of remembering and over what time period. In more recent years, much focus has also been given to addressing the question of what visual information from scenes is remembered, and in how much detail. In relation to these questions, the stimuli of choice within the majority of research undertaken has been static images of objects or scenes, that is, still photographs rather than filmed or dynamic scenes. Typically, participants are shown a number of static images – either real-life photographs or digitally created images – and then, following a variable retention interval, given a memory test. One attraction of such experimental methods is of course the control over the stimuli, their presented features, the timed exposure of those features to experimental participants and the delay between scene viewing and subsequent memory test. The focus has often been on ‘intentional’ memory, in which participants are aware when first viewing a scene that their memory will later be tested. Indeed, this research has shown that we are capable of remembering many previously viewed scenes (Standing, 1973).

These scene images are different to the dynamic, changing scenes we often view every day. Thinking about the numerous scenes we view daily, many will be highly familiar, such as the street that we live on. Although the fundamental, fixed background elements to these scenes will remain unchanged, the objects (such as people and cars) may change every time we view our road, place of work, or other familiar scene. Not only do scenes change across subsequent viewings, but many are dynamic and changing whilst we view them. Take the example of a street; multiple cars may be moving across the scene following different trajectories, whilst there might also be movement from objects such as people and animals.

The aim of this thesis is to explore the fidelity of visual memory for everyday scenes, with a focus on understanding incidental scene memory, and memory for both static and dynamic scenes. This is important as it allows greater understanding of memory for naturalistic scenes under encoding and testing conditions that are more likely to approximate scene memory in every day life. The following section considers relevant research into the fidelity of visual memory for scenes, including the typical experimental arrangements that have been used. Theoretical perspectives to explain visual memory for scenes are reviewed, and the research questions that will be addressed in this thesis will also be outlined.

1.2 The fidelity of visual memory for static scenes

1.2.1 Detailed visual memory for objects and scenes following scene viewing

Research into the longer-term capacity of visual memory has shown it to be very high indeed, when tested over different time periods of up to a year. A seminal experiment by Standing (1973) demonstrated that participants were capable of successfully recognising very many previously viewed pictures. Across different experiments, Standing tested the accuracy of memory for pictures and words. In one condition, a small group of participants were shown up to 10,000 pictures for five seconds each, over a period of five days. The pictures were varied, and collated from a range of sources. Participants were instructed to try and learn the items and were informed that their memory for such items would later be tested. After viewing all pictures, participants' memory for a subset of 160 items was tested through a two-alternative forced-choice (2-AFC) recognition task. In this 2-AFC task, participants were presented with an image previously seen, and a new, novel image (the ‘foil’) simultaneously; it was their task to identify which of the two images they had previously viewed. Recognition performance was found to be over 80% (Standing, 1973).

This high recognition memory also appears to persist following much longer retention periods of up to a year (Nickerson, 1968). These studies and others (for example, Shepard, 1967) suggest that we have a high capacity visual long-term memory (VLTM) for visual stimuli that endures across time periods of up to a year. In these studies however, the foil items that were paired with previously viewed images during the memory test may be very different in terms of their characteristics and details to those previously viewed images. Therefore, it was unclear whether this impressive recognition accuracy relies on the participants simply remembering the general gist of the previously viewed items, or whether they do indeed acquire a more fine-detailed discriminative ability.

It has been demonstrated that people are remarkably fast in appreciating the gist (that is, the general meaning or concept) of a presented image of a visual scene, and can do so in less than 500ms (Oliva, 2005; Potter, 1976; Rousselet, Joubert & Fabre-Thorpe, 2005). Although observers may grasp this gist information quickly, research into the so-called ‘change blindness’ effect could be taken to suggest that relatively little detailed information appears to be encoded into memory during brief viewing episodes. The change blindness effect is the phenomenon in which observers frequently fail to detect quite large changes in a visual scene, when made during a brief interruption such as a flicker or saccade (Rensink, 2000; Rensink, O'Regan & Clark, 1997; Simons & Levin, 1997; Simons & Rensink, 2005). In change detection studies using a ‘flicker task’, participants are presented with an image of a scene and a modified version of that scene alternately, until they detect the change. Observers often take a surprising amount of time to notice and identify the change; for example, in one experiment reported by Rensink et al. (1997), it took participants an average of 17 alternations to detect the change.

The outcome from studies on change blindness may be taken to imply that little detail is encoded into memory across successive saccades (Rensink, 2000). In other words, gist information which informs broad conceptual understanding may not be accompanied by a detailed memory trace holding very much more than this broad concept. Theoretical explanations have been proposed to account for the change blindness effect, in which people appear unable to detect large changes in a scene. Coherence theory proposed that a detailed object representation is only limited to the one object that is being attended to (Rensink, 2000, 2002; Rensink, O'Regan & Clark, 2000). In this view, change blindness therefore occurs when the changed object is not the one currently being attended to, as a detailed object representation is not available (Rensink, 2000).

Other theoretical viewpoints suggest that detailed scene memories may not be retained across longer time periods. The object file theory of transsaccadic memory suggests that during scene viewing, as the observer visually attends to different objects within a scene, ‘object files’ are created in memory (Irwin, 1992b; Irwin & Andrews, 1996). These temporary ‘object files’ contain key information about an object, such as its identity, colour and spatial location within a scene, and can be updated upon during successive views if the object has changed (Gordon & Irwin, 1996; Irwin, 1992b; Kahneman, Treisman & Gibbs, 1992). However, in this view, the number of object files in memory is thought to be limited to a the most recently viewed three to five objects (Irwin, 1992a; Irwin & Andrews, 1996; Irwin & Zelinsky, 2002; also see Hollingworth & Henderson, 2002, for discussion).

The findings relating to change blindness suggest that specific detail may not be retained following scene viewing. In addition, the theoretical positions of coherence theory and object file theory indicate a key role for attention in creating visual memories for scenes, and that when created, these visual memories are temporary and limited to the most recently viewed item or items. However, a series of experimental investigations by Hollingworth and Henderson (2002), in addition to other research reviewed below, indicates that quite detailed representations of visual scenes may actually be retained for some time after scene viewing.

In the context of the viewpoints suggesting only a temporary memory for objects and scenes as described above (Irwin & Zelinsky, 2002; Rensink, 2000), Hollingworth and Henderson (2002) aimed to test if detailed visual memories can be retained after scene viewing. Across three experiments, they tested memory for objects within scenes during scene viewing, and after a period of delay. To do this they showed participants computer-generated images of everyday scenes which contained pre-determined ‘critical’ objects. Changes were made to these critical objects to test participants’ ability to detect a change, therefore testing the detail of this visual memory. These consisted of ‘type-change’ trials (where the object was changed with one from a different object-category), ‘token-change’ trials (where the object was changed with one from the same object category), or ‘rotation-change’ trials, (where the object itself was the same, but was rotated compared to its original orientation). In general, across the experiments, the results suggest detailed memory for objects within a scene, both when tested during scene viewing and after a delay.

For example, in an initial experiment, participants first took part in a change detection task, in which they viewed scenes for 20 seconds each. Eye tracking was used to monitor if and when participants had attended to the critical object before the change to it was made. During scene viewing, a type-change or token-change was made to the critical object in the scene during a saccade, when participants’ gaze was not fixated on the critical object. Participants were asked to press a button if they detected a change. On 12 out of the 36 trials, a control condition was used, in which the scene was not changed during the 20 second viewing. This allowed study of participants’ tendency to report ‘false alarms’ (inaccurately reporting they detect a change), and therefore gave an indication of bias in responding. The results from the change detection task showed that if the critical object had been attended to before the change had occurred (monitored through eye tracking), participants’ change detection performance during viewing of the scene was 51.1% for type-changes, but significantly lower at 28.4% for token-change trials. In other words, participants were better at detecting changes when the critical object was replaced with an object from a different object category (for example, a piece of paper with a CD), rather than an object from the same object category. However, accuracy on both tests was significantly higher than the false alarm rate in the control condition (no change trials). In other words, participants reported to detect a change significantly more on the token-and type-change trials compared to trials in which there was no change, demonstrating that participants were able to detect some of the changes in these trials. A subsequent AFC task was then given, after viewing all scenes, and an additional five minute break. Here, the 12 scenes used in the control condition were shown to participants, in which no change had been made to the critical object in the initial change detection test. Participants were presented with two scene images sequentially. One was the same scene image participants had viewed in the change detection test, and the other was the same scene, but with a change made to the critical object. Participants had to report which of the two scenes they had previously viewed. Participants performed significantly above chance; with accuracy of 93.1% and 80.6% in the type-change and token-change conditions, respectively. This again marked a significant difference between the two conditions, but both were significantly above chance performance.

As discussed by Hollingworth and Henderson (2002), coherence theory (Rensink, 2000) cannot account for these results. According to coherence theory, representation of an object should not be held in memory if it is no longer being attended to, after attention for that object has been withdrawn. Instead, Hollingworth and Henderson (2002) proposed "a descriptive model of scene perception and memory" (p. 132). This conception of scene memory aimed to address the disparity in findings between the change blindness research and their findings of high fidelity VLTM, to offer an explanation as to how visual information is retained in memory after scene viewing. The theory offers a conception arising out of the standard memory model. The distinction here is between a capacity-limited short-term memory (or working memory) versus a long-term memory store, which has much greater capacity to hold detailed scene information over longer periods of retention (Hollingworth, 2004; see Schurgin, 2018 for discussion). Short term memory is generally considered as temporary, and limited in capacity to around 3-5 'items' (Baddeley, 2003; Baddeley & Hitch, 1974; Cowan, 2001; 2010).[1]

In Hollingworth and Henderson's account, VLTM is involved in the process of 'online' scene viewing. When viewing a scene we attend to different objects within that

---

[1] Visual short term memory (VSTM) is referred to throughout this thesis as a broad term to encompass what is referred to in the literature as short-term memory, working memory and immediate memory. Our focus here is on memory in the initial (short) time frame after viewing a visual item or scene, which can be contrasted to visual long-term memory (VLTM). Where authors specifically refer to working memory, this phrasing is used.

scene and as we do so, object files are created in long-term memory. As described above, object files contain specific information about an object, including its identity and spatial position within a scene (Gordon & Irwin, 1996; Irwin, 1992b; Kahneman et al., 1992). However, in Hollingworth and Henderson's formulation these object files are created in long-term memory, in which the spatial location of the object is positioned within a spatial map of the scene. Through the course of viewing a scene, a scene image, or map, is then accumulated as attention is directed to different objects within a scene. Conversely, visual short term memory (VSTM) is limited in its capacity to the most recently attended visual items, or objects. Withdrawal of attention therefore results in the VSTM trace decaying quickly, whereas the detailed scene map remains within VLTM.

Hollingworth and Henderson's (2002) conception of scene memory therefore proposes a role for VSTM and VLTM during scene viewing. In the context of these proposals and to investigate the contribution of a limited VSTM and VLTM, Hollingworth (2004) tested the capacity of VSTM to hold such object representations. Across six different experiments, the serial order in which observers viewed objects within a scene was controlled using a dot to cue observer's attention from one object to the next, in a set order. The rationale for this attention-directing manipulation was that it allowed memory to be subsequently tested for objects that had been most recently viewed – and therefore likely to be held within VSTM – or for objects that were viewed a few objects previously. During scene viewing, the cue remained on an object for 300ms, with an 800ms gap between each object; participants were instructed to maintain their gaze on the location of the last cue during this time until the next cue appeared. Participants' memory for an object ranging from 0-back and 10-back within each scene was then tested immediately. For example, in a 'one-back' condition, memory for the penultimate object that had been cued was tested. During this process, an object in the scene was covered by a mask, participants were then given either a 2-AFC or change detection test. In the change detection task, participants were immediately shown the same scene image, in which the masked object was unmasked and shown in its original form, or a change had been made to the object; participants had to indicate if they thought there had been a change through a keypress. Participants demonstrated reliably higher accuracy for the last two objects that were previously viewed, demonstrating a recency effect in scene memory. Although there was a drop in performance when tested for items that had been viewed further back than two objects, performance remained comparable across trials in which items had been viewed four-back or 10-back. The authors present that this outcome of stable performance across sequence items suggests a role for long-term memory. An additional condition in one experiment (experiment five) also used a delayed test to compare accuracy at 10-back to accuracy over much longer delay periods. On some trials in this experiment, testing was delayed until participants had viewed all of the scenes; on average this meant that participants had viewed 402 objects between viewing a target object and the memory test. Although performance on these trials was significantly lower when compared to immediate testing, performance was still above chance, suggesting maintenance of object information from scenes in VLTM.

As outlined by Hollingworth (2005), the conception described by Hollingworth and Henderson (2002) would predict that after scene viewing, the majority of objects from that scene, apart from the most recently attended two items, should be held within VLTM. Therefore performance when tested within the short time frame after viewing, compared to longer periods of hours, should yield comparable, or only slight decreases in levels of change detection performance with increasing delay periods, as the majority of scene information is held within VLTM. To test this, Hollingworth (2005) tested visual memory through change detection tasks, over different periods of delay, across several experiments. Participants were shown computer generated scene images, consisting of a number of different objects. Then, in the change detection task, participants were shown a previously viewed scene, and a critical object within that scene was either changed or unchanged. A small arrow on the screen marked the critical object to participants, and it was their task to report if there had been a change made to that object. The changes to the critical object were either a token-change (object replaced with another from the same object category), or rotation change (same object, rotated 90 degrees). Across experiments, participants were shown the scene images, and for half the scenes they were tested immediately (200ms) after viewing. The test for the other half of the scenes was delayed, and the amount of delay varied across experiments, either until after one trial, until the end of the study (once participants had viewed all the scenes) or 24 hours later. In the one trial condition, participants took part in a separate block of trials in which they were shown a scene image for 20 seconds, followed by a test image for the scene viewed on the previous trial, which they had finished viewing around 30 seconds earlier. The author suggests that according to Hollingworth and Henderson’s conception of scene memory, even in the immediate test condition, memory would most likely rely on VLTM, unless the change happened to be made to the two objects most recently fixated, and should therefore be comparable with the delayed tests, particularly when tested after one trial only. However, eye movements were not measured in this experiment, so it is difficult to ascertain how many objects had been fixated, and which ones had been most recently fixated. Consistent with the prediction laid out by Hollingworth, it was found that there was no significant difference in performance between immediate test and after a one-trial delay, across both conditions (token change and rotation change). When the delayed test occurred at the end of the session (an average delay of just under ten minutes between viewing a scene and test), there was no significant main effect of test time (immediate or delayed test). However, an interaction demonstrated that participants were significantly better at the rotation change detection task when tested immediately compared to one trial later. There was no significant difference for the performance in the token change tasks. Therefore some object-specific detail may be forgotten over this longer period of delay. After a delay of 24 hours, change detection performance was significantly lower than at immediate test, however participants still performed above chance, suggesting some specific detail for scenes is retained over this time period.

Hollingworth (2005) interpreted this evidence in relation to Hollingworth and Henderson’s (2002) conception of visual memory for scenes, suggesting that VLTM supports detailed scene representations during scene viewing. Detailed visual memories appear to be maintained over longer delay periods of up to 24 hours, demonstrated by above chance performance. However, there was some forgetting of visual information across the longer periods of delay; especially for object-specific detail, as demonstrated by lower performance in the rotation change condition after a delay of around 10 minutes (and many intervening scenes).

It has not escaped researchers’ attention that there is therefore an apparent disparity between this research demonstrating impressive memory performance for longer-term visual memory, and the change blindness phenomenon, pointing to imprecise encoding of experienced pictures of objects or scenes. A number of suggestions have been made to account for these differences in findings. For example, they may be at least partially attributed to procedural differences between the types of tasks used. Hollingworth and Henderson (2002) suggest the differences may be due to the fact that past change blindness research has often not tracked eye movements during scene viewing or change detection test; and therefore it is unclear whether the critical object was fixated before (or even after) the change. Other procedural considerations, such as limited encoding time given in previous studies showing change blindness may also account for at least some of this poor performance (see Brady, Konkle, Oliva & Alvarez, 2009, for discussion). In addition it has also been suggested that the paradigms used may make the task of comparing between alternations difficult (see Simons & Ambinder, 2005).

The results from Hollingworth and Henderson (2002) and others demonstrates that detailed representations for objects within scenes may persist in VLTM for some time after attention is withdrawn. Research by Brady and colleagues, amongst others, has specifically focused on the fidelity of this long-term visual memory. For example, Brady, Konkle, Alvarez and Oliva (2008) tested the fidelity of visual memory for individual objects. They used a methodological approach in which they carefully manipulated the similarity of previously viewed target objects to the foil objects used in recognition tasks. Participants viewed thousands of pictures of everyday objects (such as a clock, food item or a piece of furniture) for three seconds each, followed by a subsequent test. They were given 300 2-AFC trials, in which the foil image was either novel (from a different object category), an object from the same object category, or the same object but in a ‘different-state’ (for example, if participants had previously viewed an apple, in the recognition task this image might be presented next to a foil image of a half-eaten apple). The authors report both accuracy and response times (RTs) to make 2-AFC judgements in these three conditions. Participants were more accurate on trials in which the foil object was novel (92.5%), compared to both same category and different-state trials (where the foil and target image were more similar). There was no significant difference between these two latter conditions. Despite this, accuracy was still around 87% in the conditions in which the foil object was from the same object-category or was the same object in a different state, demonstrating participants were still very good at distinguishing between the previously viewed object and the similar foil. In terms of RTs, the more similar the foil object was to the previously viewed image, the slower participants were to make their recognition judgement; that is, they were slowest in the condition where the foil was a same object in a different state, and fastest when the foil was novel. The authors interpret these results to infer that responding to the different-state condition requires more detail, and accessing this additional detail from memory may take more time. Despite this, the results from Brady et al. suggest that this level of detail, about specific object properties and their current 'state', is retained in memory after only brief viewing times of a few seconds, up to the retention period of hours later.

However, there do appear to be limitations to the fidelity for memory of scenes within VLTM. As reviewed above, Hollingworth (2005) tested visual memory for objects within scenes after different periods of delay using a change detection task. After a delay of 24 hours, performance on the task was significantly worse than when tested immediately. Research testing the fidelity of memory for objects has also demonstrated forgetting of object information over longer periods of delay, days after initial viewing, despite performance being above chance (Andermane & Bowers, 2015; Brady, Konkle, Alvarez & Oliva, 2013a).

Thus, despite the findings relating to change blindness, there is extensive evidence from studies using both object and scene stimuli, that detailed visual memory persists over longer periods of retention. Multiple studies testing memory for object stimuli have found that performance in later recognition tasks is still above chance when tested after a delay of minutes, hours or days. However, some forgetting of detailed object information could occur over time periods up to a week (Andermane & Bowers, 2015). The theory of visual memory for scenes proposed by Hollingworth and Henderson (2002) provides a useful framework for interpreting the results outlined in this section, with the notion that VLTM is recruited during scene viewing, with a scene map containing object files indexed to their location being created in VLTM.

1.2.2 Incidental visual memory for objects and scenes

One notable feature of the recognition tests described in the sections above (Brady et al., 2008; Hollingworth & Henderson, 2002) is that participants are perfectly aware during the scene viewing phase that their memory will be tested later, and explicit instructions to participants often involve mention of a later test (for example, Andermane & Bowers, 2015). Presumably these participants therefore make every attempt to consign the pictures to memory. Yet arguably, this ‘trying to remember a scene’ is unnatural and does not represent the manner in which people view scenes in everyday experience. The nature of what might be termed ‘everyday memory’ for scenes presents a particular challenge for researchers, in that it is difficult to develop methods that avoid participants actively attempting to encode and rehearse visual information. In some studies, researchers have tested memory performance in viewing situations where participants were unaware that their memory would later be tested, referred to throughout this thesis an ‘incidental’ viewing (Williams, 2010; Williams, Henderson & Zacks, 2005). Developing methodologies to test incidental visual memory is important, as this is core to understanding our memory for the scenes we view outside of experimental settings and when the observer is not intentionally trying to remember images they are viewing. It is a natural question to ask whether the apparently high recognition performance reported in the intended viewing situation is achieved under conditions of incidental viewing.

One way in which this has been addressed, is through research that has asked participants to take part in a search task when originally viewing a scene or object, followed by later, surprise recognition tasks (Olejarczyk, Luke & Henderson, 2014; Williams et al., 2005). In one study, Castelhano and Henderson (2005) aimed to test whether longer-term retention of detailed object information from scenes was still maintained in memory when scenes were observed under incidental viewing conditions. In this study, memory for objects from previously viewed scenes was tested following intentional and incidental viewing conditions. In the intentional viewing condition, participants viewed scenes knowing that their memory for objects within those scenes would later be tested; this is similar to previous research in which participants are made aware of a later memory test before viewing images (Andermane & Bowers, 2015). In the incidental viewing condition however, participants were unaware they would be given a later memory test. Instead, they were given a visual search task, and asked to search the scene for certain objects. Both groups (incidental and intentional encoding) then took part in a subsequent memory test minutes later, testing memory for objects within the scene through different tests. For those that had previously viewed scenes incidentally, memory for specified objects that they were asked to search for was not tested, and instead memory for other objects within the scene was tested. In an initial experiment, a token discrimination forced-choice test was used, where previously viewed target objects were presented next to foil objects from the same conceptual category; the participant's task was to identify which object had been taken from a scene they had previously viewed. In a second experiment, a mirror-image discrimination test was used, in which a previously viewed object was displayed next to a mirror-reversal image of itself. The rationale the authors give for this was that previous research has found that, following intentional viewing, participants can remember object orientation information. The aim therefore was to see if this type of detailed information is still encoded following incidental viewing of a scene. It was found that performance in these two different tests did not statistically differ following incidental compared to intentional viewing conditions, suggesting detailed visual memory regardless of intent to memorise during scene viewing. Reports from participants confirmed that they were unaware or did not guess that their memory would be tested in the incidental condition. The authors interpreted the results in relation to Hollingworth and Henderson's (2002) framework of visual memory for scenes, which suggests that detailed visual information is accumulated in VLTM, without intentionally trying to remember the scene information.

1.3 Visual memory for dynamic scenes

1.3.1 Defining a dynamic scene

The static pictorial stimuli used in the research described above, often does not account for the dynamic nature of scene memory in everyday life. A visual scene was earlier defined as a picture depicting a real life place or setting, containing a number of different objects against a fixed background (Henderson & Hollingworth, 1999; Wolfe et al., 2007). This definition can be expanded to acknowledge the dynamic, changing nature of 'real life' scenes. Dynamic scenes contain not only spatial information (that is, objects in certain locations within a scene), but also temporal information (objects in different locations within a scene, across time). Therefore a dynamic scene, also often referred to as a 'moving' scene, is one depicting a real life place or setting, containing a number of static and moving objects, against a fixed background. Visual memory for dynamic scenes is relatively under-researched compared to memory for static stimuli, such as objects and scenes. An overview of the research testing visual memory for dynamic scenes is given in this section.

1.3.2 The fidelity of visual memory for dynamic scenes

Experiments have been conducted to test memory for dynamic scenes. In memory experiments comparing recognition accuracy for static and dynamic scene stimuli, a ‘dynamic superiority effect’ has been found, in which recognition accuracy is higher for previously viewed dynamic scenes compared to static (Buratto, Matthews & Lamberts, 2009; Matthews, Benjamin & Osborne, 2007; Matthews, Buratto & Lamberts, 2010). For example, Matthews et al. (2007) compared memory for static and dynamic scenes over time periods of one and four weeks. The mode at encoding (static or dynamic) matched the mode in a later recognition task. An old/new recognition task was given, in which participants are presented with an old, previously viewed item, and a new item. It is their task to identify which item is the old one. Recognition performance in the old/new recognition task was found to be superior for dynamic scenes. Furthermore, there was a main effect of delay time, suggesting some loss of fidelity in long-term memory over the one to four week period. There was no interaction, suggesting this decline in fidelity did not affect memory for dynamic and static scenes differently. Intentionality at encoding was also considered. In one experiment, Buratto et al. (2009) compared the within-subjects factors of encoding and test mode (static, multi-static, dynamic) and the between-subjects factor of intentionality at encoding (intentional or incidental). Participants first viewed a scene for three seconds each that was either a static scene image (static condition) or a short dynamic scene film (dynamic condition). In the multi-static condition, six images were taken from across a three second film clip, and presented in succession for 0.5 seconds each. Across all encoding conditions, participants were asked to search for a woman and report her presence during scene viewing. This was all that happened for participants in the incidental condition, whereas those in the intentional condition were also told about the later memory test. Participants were then given an old/new recognition task three days later. A congruency effect was observed, in that accuracy was highest when scene stimuli were dynamic in presentation and test mode. Furthermore, there was no main effect or interaction dependant on encoding condition (incidental or intentional) suggesting that, as appears to be the case for static scenes, information from dynamic scenes may also be accumulated in memory during natural scene viewing, without the need for intentional encoding.

Research with static scenes and object stimuli has demonstrated the fidelity of long-term memory by showing that participants are still able to distinguish previously viewed scenes, compared to highly similar foils in recognition tests (e.g. Brady et al., 2008; Konkle, Brady, Alvarez & Oliva, 2010a). Research into dynamic scene memory has also used foil images that are similar to the previously viewed scene. Matthews et al. (2010) compared memory for dynamic and static scenes across delay periods ranging from 90 minutes to 14 days. During the later old/new recognition task, participants were presented with foils that differed in their similarity to previously viewed scenes. A higher number of false alarms were made towards moving scenes foils, but only when those foils were most similar to the scenes participants had previously viewed. This research has demonstrated that highly detailed memory is stored in long-term memory for dynamic scenes, and recognition performance on later memory tests is higher than for static. However, this high fidelity memory trace may also lead to errors, as participants seem more likely to incorrectly report recognition of a scene when it is similar to a dynamic scene they have viewed before.

Similarly, Ferguson, Homa and Ellis (2017) tested visual recognition memory for dynamic scenes, using foil images in an old/new recognition test that were similar to the target images taken from the films participant had seen. Participants first viewed a film clip showing a dynamic scene, and then their memory for static frames taken from the clip was tested, alongside similar foils that were taken from unseen parts of the same film (experiment one), or cuts in the film clip (experiment two). In this second experiment, cuts between 0.5 and 30 seconds were taken from film clips, before the film was edited back together and shown to participants. The foil images were then taken from these unseen gaps. As discussed by Ferguson et al., the duration of the gap may be used as a proxy for similarity (that is, it may be assumed that frames taken from the shorter gaps will be more similar to the parts of the film participants actually viewed). It was found that when foil images were taken from the shorter gaps, the false alarm rate (incorrectly labelling foil images as ‘old’) was higher than the hit rate (correctly labelling target images as ‘old’).

A key question related to this research and understanding of visual memory for everyday scenes is: how are dynamic scenes remembered? Addressing this question links directly to the issue of fidelity; how much detail is remembered from a previously viewed dynamic scene? There are different explanations and findings to consider when contemplating how dynamic scenes may be remembered.

There may be certain parts of dynamic scenes that are remembered better than others (for example, the beginning or the end), especially after shorter retention intervals. Research into serial position effects in visual memory often use discrete individual items (a series of images, for example), rather than a continuous dynamic scene. What is often found in subsequent memory tests is higher accuracy for items presented first, or last (most recently), termed primacy and recency effects, respectively (Allen, Baddeley & Hitch 2006; Broadbent & Broadbent, 1981; Hurlstone, Hitch & Baddeley, 2014). Furthermore, when testing memory for object items within scenes in a change detection task, Hollingworth (2004) found that accuracy in detecting change was higher for the most recently fixated two objects, compared to objects fixated less recently. In addition, memory performance was comparable for items viewed further back, up to 10 items. It is not clear how these findings translate to memory for dynamic visual scenes. For example, it might be that the end of the scene, with objects in their final position, might be remembered in higher fidelity than parts of the scene first viewed.

Likewise, Hollingworth and Henderson’s (2002) visual scene theory describes object files as spatially indexed to a scene map, that is retained in good detail in VLTM. It may be that the spatial arrangement of dynamic scenes at a certain time point is represented in memory (for example, the initial or final view of the scene). The notion of ‘object files’ suggests that an object’s position within a scene can be updated in memory if the position of that object changes (Hollingworth & Henderson, 2002; Kahneman et al., 1992). Therefore it is possible that visual memory for a dynamic scene may be remembered with highest fidelity with objects in their most recent spatial position within a scene. This is also in line with recency effects in visual memory described above (*see* Hurlstone et al., 2014).

In a similar vein, representational momentum is the finding that individuals tend to report to ‘remember’ moving objects further along in their direction of motion than was actually viewed in the film, when tested after very short time periods (after delays of around 250 milliseconds) (Blättler, Ferrari, Didierjean & Marmèche, 2012; Freyd & Finke, 1984; Hayes & Freyd, 2002; Hubbard, 2005; Hubbard & Bharucha, 1988). When participants watch a video that implies self-motion, a representational momentum effect is also observed (Blättler et al., 2012). The representational momentum effect may suggest that the later parts of a dynamic scene may be remembered better than earlier parts.

Rather than higher fidelity memory for specific time points within a dynamic scene, such as the end, memory could be comparable for all different parts of a viewed scene. Matthews et al. (2007) have suggested that Hollingworth and Henderson’s (2002) theory could be developed to allow for the explanation of dynamic scene memory. In this view, object files could hold spatiotemporal information, whereby “the viewer establishes a record not only of where a fixated object is, but also when it is there, and quite possibly where it is going” (Matthews et al., 2007, p. 992). Such a dynamic memory map in VLTM may hold highly detailed visual representations of previously viewed dynamic scenes, including information about objects’ trajectories within scenes.

1.4 Overview of thesis

We see many scenes every day, some static and some dynamic in nature. Often we make no special, intentional effort, to remember these scenes. The aim of this thesis is to the understand the fidelity in which we remember everyday scenes, providing a more comprehensive understanding of the representations underlying this. To do this, experimental paradigms are developed that bring together the key strands of scene memory research, and address some of the current gaps in understanding of visual scene memory. The experiments test the fidelity of visual scene memory over different periods of delay between scene viewing and test, with some comparison between performance in immediate and delayed tests.

The research outlined above has demonstrated the high capacity and fidelity of VLTM for objects and static and dynamic scenes. VLTM appears capable of retaining quite detailed information, above category gist information, for hundreds of images of both scenes and single objects. This is the case even when scenes have been viewed incidentally, with no effort to memorise scenes during scene viewing. In the case of objects at least, memory performance remains significantly above chance when tested after longer delay periods of up to a week. There has been a focus across different research studies to compare the fidelity of visual memory for objects and scenes over different periods of delay (Andermane & Bowers, 2015; Brady et al., 2013a; Hollingworth, 2004; 2005). Forgetting of visual detail is often observed when participants are tested after a delay compared to immediately after scene viewing, and there may be some further forgetting over longer time periods of days and weeks (Andermane & Bowers, 2015; Hollingworth, 2005).

The first part of this thesis includes a focus upon incidental memory, using static scene images. The majority of past studies focusing on the question of fidelity of visual memory for scenes have specifically tested memory for individual objects, rather than whole scenes. Konkle et al. (2010a) demonstrated that detailed visual memories for whole scene images were available when tested within minutes to hours after viewing. The current research aims to test if this high fidelity memory lasts for longer periods of many days, even after incidental viewing. Furthermore, previous research testing incidental scene memory has used testing conditions that are still intentional in that participants are actively trying to identify the previously viewed object during the memory test. Although this research speaks to how much detail may be memorised following natural, incidental encoding, it is interesting to understand if this scene information is still accessible during subsequent encounters with a scene, without intentionally trying to remember. An aim of this thesis will be to test for the fidelity of visual scene memory over different periods of delay, when participants are not aware that their memory for a previously viewed scene is being tested. To do this, a paradigm to test incidental memory at encoding and retrieval is used for static scene images, to test if high fidelity incidental memory for scenes in maintained over longer retention intervals (Chapter Two).

The next part of the thesis focuses on dynamic scenes, to try and understand how these might be remembered and in how much detail. Research that has compared recognition memory for dynamic compared to static scene stimuli has found higher accuracy for dynamic stimuli (e.g. Matthews et al., 2007). However, there has been comparatively little research conducted to understand the detail in which dynamic visual scenes are remembered, and which parts of dynamic scenes are remembered best (if any). It is unclear if we have a highly detailed memory trace that can account for an object's location across its trajectory within a scene. Comparison of memory for different temporal parts of a dynamic scene is made, to attempt to understand representations of dynamic visual scenes in memory, and how this relates to theories of visual memory developed through research with static scenes (e.g. Hollingworth & Henderson, 2002). The research reported in this thesis also specifically examines how memory for dynamic visual scenes changes over different retention periods, by comparing memory performance in immediate and delayed tests.

To fulfil the aim of understanding memory for everyday scenes, both strands of work will use images and films of real-life scenes, and use highly similar foil images to test memory fidelity. Varied delays between scene viewing and test are used throughout, to see how this fidelity may change across the time course after scene viewing.

CHAPTER TWO

INCIDENTAL VISUAL MEMORY FOR STATIC SCENES

2.1 Chapter Summary

The current set of experiments aimed to test incidental visual memory for natural, everyday scenes. Using novel methodology across the two experiments reported in this chapter, incidental scene memory was assessed, where participants were not told during the encoding, retention or test phase that their memory for certain scene stimuli was being tested.

To test incidental visual scene memory, the ‘recent-probes’ task used in previous research was adapted. In a first phase, participants viewed a set of 80 everyday highly detailed scene images and made simple aesthetic judgements about them. Following a delay (from minutes to days) they then took part in a memory test, in which on each trial they viewed five new images in succession, followed by a ‘probe’ image which either matched one of the five just presented or did not. On a third of the trials, the probe image was a novel, previously unseen image (termed a ‘non-recent negative’), which was not one of the five just presented or one of the original 80. Crucially, on a third of the trials, a ‘recent-negative’ image was presented, which was not one of the five images just presented but was one of the original 80 images that participants had previously viewed in phase one. To the extent that the prior image is still within visual memory, we expected the response to be affected; previous research using the recent-probe paradigm over shorter durations, has found lower accuracy and/or slower RTs when rejecting a recent-negative image compared to a non-recent negative image (Craig, Berman, Jonides & Lustig, 2013; McKeown, Holt & Delvenne, 2014).

When tested within minutes of initial scene viewing, there was no significant difference between RTs to correctly reject a novel probe image compared to a recent-negative probe image (Experiment One). However, supporting the notion of an enduring memory trace for scenes it was found that recent-negative probe trial responding was faster than when responding to novel probes five days later in both experiments, and ten days later in Experiment One. Therefore, the expectation of inhibited responding observed in previous research was not replicated here.

Participants also took part in a two-alternative forced choice (2-AFC) recognition task after the recent-probes task, at delays of five or ten days. Here, previously viewed scenes were presented alongside scene images taken from the same ‘scene category’. Accuracy was found to be good and significantly above chance, highlighting that incidentally encoded visual memories can be maintained in good detail, above categorical gist information, over a period of days. Furthermore, remember-know-guess response judgements were made by participants alongside these 2-AFC tests, to help understand performance.

2.2 Introduction

We appear to have a large capacity for long-term visual memory for scenes. Studies of visual recognition have shown that participants are capable of remembering many different images that have been viewed days earlier (Standing, 1973). In studies such as Standing’s (1973), the foil images used in each recognition trial were selected from a large stimulus set of different pictures, and therefore may be quite distinct from the target image they are paired with at test. It has been suggested that participants may do rather well in such tests, as their ‘accurate’ response may be based on them simply recognising the general gist of previously viewed items, and essentially, simply having a feeling or sense of familiarity towards a previously viewed image compared to a completely novel one (e.g Konkle et al., 2010a).

However, more recent research has aimed to test the fidelity of visual memory for scenes, by manipulating the similarity of previously viewed images and the foils used in subsequent recognition tasks. Konkle et al. (2010a) used scene image stimuli to demonstrate the fidelity in which participants were capable of distinguishing between previously viewed scenes, and scenes from the same scene category (for example, an image of a previously viewed kitchen from an image of another kitchen). They presented participants with 2912 scene items from different ‘scene categories’ (for example, bedroom or street scenes). Participants viewed a different number of items from each scene category, ranging from four to 64. They then took part in a 2-AFC recognition test, 20 minutes later. On some trials, participants were presented with a previously viewed scene, alongside a foil scene taken from a different category of scenes; for example, a bedroom scene next to a kitchen scene. On other trials, the foil image was from the same scene category; for example, a previously viewed bedroom next to a new, previously unseen bedroom. The rationale for this arrangement was that successful recognition in this latter condition, using same scene category foil images, would mean participants being forced to rely on more than just gist or category information, as they would need to discriminate two scenes images from the same category. The outcome was that participants did indeed perform very well in the task. Performance was higher when the foil item used in the test was from a novel scene category (96%), compared to when it was from the same scene category. Performance decreased as participants viewed a larger number of items from the same scene category (up to 64 items). However, even on trials using same scene category foils and when participants had viewed 64 items, participants were still accurate on 76% of the trials and performed above chance. This supports the view that our VLTM may hold more information than just gist information for whole scenes, when tested within twenty minutes after scene viewing.

Similar research using object stimuli has also demonstrated that visual representations can maintain high detail over a longer period of many days. Andermane and Bowers (2015) extended the interval between the exposure and test phase, to test whether extended delays affected the level of detail that is remembered. In this study, participants viewed over a thousand objects for three seconds each. Their memory for 300 items was then tested following a delay of either 10 mins or seven days, using two types of recognition tasks: a 2-AFC and an old/new task. Using similar methodology to Brady et al. (2008), the foil object in each 2-AFC trial was either from a different object category, the same object category, or the same object but in a different state (for example, the half-eaten compared to the full apple). In the 2-AFC test, participants were more accurate overall when the foil image used was of an item from a different object category, compared to items from the same category, or the same object in a different state. Accuracy was higher when tested minutes later compared to a delay of seven days, suggesting some forgetting of the visual information occurs in long-term memory over this time frame. However, accuracy seven days later was still significantly above chance performance. Interestingly, there was no significant interaction; the effect of varying the delay between each of the three foil types of 2-AFC trial did not affect accuracy. The same pattern of results was shown for participants in the old/new recognition test condition. Therefore, as discussed by the authors, it appears that increased retention time does not affect the forgetting of highly detailed memories differently compared to categorical gist memory representations.

The studies described above have demonstrated that our memory can hold many items in detail, above and beyond basic gist information, over extended time periods. However, an issue with this work is that it focuses upon ‘intentional’ memory, whereby participants attempt to memorise the material during the initial viewing phase. In these paradigms, participants were aware during the original viewing of items that their memory would later be tested. Also, this research used recognition tests, such as 2-AFC or old/new tests, where participants intentionally attempted to recall visual information from memory during the later recognition task (Brady et al. 2008). However, memory for scenes in everyday settings is often incidentally encoded and retrieved.

The task demands of trying to remember a scene is unnatural compared to how we typically view scenes every day, whereby we often do not try and intentionally remember scene information. A key aim of this current research was to test whether this high-fidelity memory for scenes as reported in previous research is observed after incidental viewing.

In recent years, researchers have tested memory performance on tasks in which participants are unaware of a later memory test, referred to here as ‘incidental’ memory (Williams, 2010; Williams et al. 2005). Generally, these experiments employing an incidental encoding condition, in which participants are unaware when viewing the scene of a later memory test, have still found evidence of detailed VLTM. In one experiment, Castelhano and Henderson (2005) showed participants scene images, followed by 2-AFC recognition tasks testing memory for objects from within the scenes. Performance was comparable for those in the incidental and intentional memory encoding conditions.

Research has also focused on incidental visual memory only, without an intentional condition for comparison. Kuhbandner, Rosas-Corona and Spachtholz (2017) focused on incidental memory for objects over different retention periods, where no intentional encoding condition was used. Participants were first shown 120 object images for 500ms, twice each, with words displayed over the top. The participants were given a task related to the words, and were not asked to attend to the object images, and importantly were not told that there would be a later memory test. A 2-AFC task was given immediately after viewing the objects, and 24hrs later. The foil images used in the 2-AFC were either from a different object category, the same object category, or the same object in a different state. Accuracy was found to be lower after 24hrs or when the foils were more similar, but remained significantly above chance at both time points, and regardless of foil type. Interestingly, participants were asked if they had guessed after each recognition choice, reporting that they had on the majority of trials (77% in the immediate test and 95% after 24hr). An additional analysis was run using only trials in which participants reported to have guessed, and accuracy was found to still be above chance. Therefore even when participants are unaware during encoding that their memory would later be tested, they are still able to distinguish previously viewed objects from similar foils. This appears to be the case even when participants do not confidently remember previously viewing the object at test.

Whilst previous experimental paradigms have explored visual memory for items incidentally encoded when first viewed, the subsequent recognition task still requires participants to actively try to recall or recognise items they have previously seen. Instead, the experiments reported in this chapter aimed to develop a novel methodology in which participants are also unaware during the memory test that their memory for previously viewed items was being tested. Therefore, the aim was to more generally approximate how we may remember visual scenes in everyday situations. The recent-probes paradigm (described below) may lend itself to examining the nature of scene memory when encoding is incidental and the retrieval of scene information from long-term memory from scenes viewed in a previous task is not explicitly demanded. These are conditions that model many of the ways in which scene information may be encountered and utilised within everyday naturalistic settings.

2.2.1 The current study

The current study includes two experiments, and aimed to test incidental visual memory for natural, everyday scenes. VLTM memory for static scenes was tested across different periods of delay (up to 10 days), in different tasks using same scene category foils. The ‘recent-probes’ paradigm measures the influence of previously encoded information upon memory performance in a current trial or task (Berman, Jonides &

Lewis, 2009; McKeown et al., 2014; Monsell, 1978). Therefore, it is a useful experimental arrangement to test visual memory without participants having to actively attempt to remember scenes from a previous viewing episode during the memory test. It is also highly unlikely that participants would attempt to rehearse between trials, as they are not aware their memory for past trials is being tested.

In the recent-probes test, within a single trial, participants are presented with a small set of visual (or verbal) stimuli. They are then presented with one visual item, the ‘probe’, to which they must respond whether or not it was one of the items just presented in that trial (yes/no response). On ‘recent-negative’ (RN) trials, the probe has not been presented to participants within the current trial, but has been presented on a prior trial. On ‘non-recent-negative’ (NRN) trials, the probe has not previously been presented and is completely new to the participant. The usual reported outcome is that participants are less accurate, and/or take slightly longer to reject RN probes compared to when rejecting NRN probes. The usual explanation for the slowed RN probe response is that the lingering memory trace of the item from the prior trial proactively interferes with performance on the current trial (Craig et al., 2013; McKeown et al., 2014; McKeown et al., 2020). Indeed, the task has been popular within the proactive interference literature (see Jonides & Nee, 2006). Research has been conducted with both visual and verbal stimuli (Berman, Jonides & Lewis, 2009). In visual research, the items typically differ from complex scenes, for example, McKeown et al. (2014) used abstract visual items. The timings used in these studies are also very short, in the order of seconds; for example, McKeown et al. used inter-trial intervals of between one and six seconds, where a RN probe image always came from the preceding trial.

In this study the recent-probes paradigm is adapted in a novel way to explore whether incidental scene memory in terms of highly detailed representations is maintained following retention intervals of a few minutes to several days, rather than over delays of a few seconds. The factor of delay will be manipulated in this experiment to allow for the longevity of scene memory to be assessed. In an initial experiment, the delays of five minutes, five days and 10 days are tested, and a second experiment used only the five day delay. These are similar delay periods to those used in similar studies of visual memory for scenes, for example, Andermane and Bowers (2015) used delay periods of 10 minutes and seven days. In the current study, participants take part in a typical recent-probes experiment, however on RN trials, the probes are not targets from earlier trials within the same testing session, but instead they have been viewed either minutes or days previously. To the extent that a prior image is still within visual memory, it was expected that RTs and/or accuracy would be affected; previous research using the recent-probes procedure over much shorter delays, as described above, has demonstrated slowing of RTs and/or decreased accuracy when responding to RN compared to NRN trials (McKeown et al., 2014).

This study also examined explicit memory retrieval of scene information in a more traditional 2-AFC test. This task took place after participants had completed the recent-probes task, and were then made aware that their memory for the previously viewed scenes was being tested. In this task, participants were also given 2-AFC trials which displayed scene images viewed previously, next to same scene category foil images. This allows a test for detailed scene memory days after incidental viewing.

To understand the decision making underlying performance in this test, participants were also required to make subjective meta-memory judgments about each recognition decision, through making an associated remember, know or guess response. Remember-know judgments and the processes underlying them are thought to be distinct (Tulving, 1985). Although there is ongoing debate over the processes underlying these different judgements; typically, a report of ‘knowing’ is associated with a sense of familiarity that you have encountered a previously viewed stimuli before, whereas recollection is understood to be the more vivid recall of the particular memory (Gardiner, Ramponi & Richardson-Klavehn, 1998; Tulving, 1985; Yonelinas, 2002). Therefore, in this study, participants were asked to make remember-know-guess judgements after each 2-AFC trial, to provide additional information about the processes underlying the retrieval of this highly detailed scene information from memory days after incidental viewing. Subjective meta-memory judgments have been used in different studies of visual recognition to help researchers understand what processes underlie the retrieval of visual information (e.g. Horry, Wright & Tredoux, 2010; Rimmele, Davachi, Petroy, Dougal & Phelps 2011). More recently, Kuhbandner et al. (2017) asked participants to make judgements on a 2-AFC task following incidental object viewing, and found that despite good memory performance, participants reported that they were guessing on the majority of trials. It will be interesting to see if this pattern of results is replicated in the case of scenes in this study.

Across both tasks (recent-probes and 2-AFC), same scene category foil items were used. In other words, within every trial, all scene images used were from the same scene category. Therefore, differences in performance are likely to be due to a detailed memory representation that preserves enough detail to discriminate one scene image from another image from the same scene category.

2.3 Experiment One

2.3.1 Method

2.3.1.1 Participants

Twenty-nine (26 female, 3 male) participants took part in the study, with an average age of 22.34 years (SD = 3.4; range = 18 - 30). All reported having normal or corrected-to-normal vision. Participants were recruited using the Psychology department’s participant pool scheme at the University of Leeds and were eligible to earn participant pool credits for their participation. Participants were not made aware that the study was investigating memory performance, and the study was advertised as one interested in ‘aesthetic judgments of visual scene images over time’. Participants were required to attend the laboratory for two separate sessions five or ten days apart. Seventeen attended the second session five days later and 12 attended the second session 10 days later. The study received ethical approval from the School of Psychology’s Research Ethics committee at the University of Leeds (reference number: 15-0046).

2.3.1.2 Materials and Stimuli

The visual scene stimuli used in Experiments One and Two were taken from the same stimuli set used in Konkle et al. (2010a) (Stimuli accessed and downloaded from: http://cvcl.mit.edu/MM/ in January 2015, with permission). The scenes consisted of static colour images of highly detailed, natural, everyday scenes from different scene-categories. Twenty scene categories were chosen representing both indoor and outdoor, and natural and man-made scenes: country road; bar; garden; bedroom; bridge; castle; classroom; church; amusement park; dining room; bathroom; gym; hair salon; house; swimming pool; library; street; restaurant; lobby; and kitchen. There were 38 images in each category (760 in total). From this pool of images, four discrete sets of stimuli were created, each set assigned for use in one of four different experimental tasks (as described below). The 38 images from one category were randomly distributed across the four discrete stimulus sets in the following manner: four in the exposure phase, 16 images in each of two recent-probe tests, and two images as foils in a 2-AFC test. This procedure was repeated for the remaining 19 scene categories. An additional scene category of ‘doors’ was used for the practice trials. The experiment was presented in E-prime 2.0 (Psychology Software Tools, Inc.; www.pstnet.com/eprime) on a computer, and participants were sat approximately 70cm from the screen. All scenes were presented individually and centrally on the computer screen, with the exception of the 2-AFC task, in which two scene images were presented side-by-side.

2.3.1.3 Design and Procedure.

The basic procedure employed in this study was an exposure phase, followed by two test phases, one immediately (following a five-minute delay, during session one) and one that was delayed (either five or ten days later, in session two). Participants completed a recent-probes test in both their first and second session, and then took part in one 2-AFC test at the end of session two.

The recent-probe test used a 2 x 2 x 2 mixed-design, with the repeated measures variables of trial type (RN and NRN trials were compared) and test time (immediate or delayed test). The additional independent measures variable of delay interval (5 days or 10 days) allowed comparison of participants’ performance when session two occurred five or ten days later. To measure performance in the recent-probes task, participants’ accuracy and RT to respond to the probe image in each trial was measured, and responses when responding to NRN probe and RN probe images was compared.

A 2-AFC task was then undertaken at the end of session two. Recognition performance for the two independent groups of participants undertaking the 2-AFC five days versus ten days following the exposure phase was then compared. An old, previously viewed image was shown alongside a new, ‘foil’ image. Participants’ accuracy in recognising the previously viewed image was recorded. For each trial in the 2-AFC task, participants were also asked for an associated remember-know-guess response, and the pattern of these responses was then compared on trials eliciting correct compared to incorrect responses.

Of the 80 scenes that the participants viewed in session one, distinct images were used in each of the subsequent memory tests as follows: 20 were used in the first recent-probes test (in session one), 20 were used in the second recent-probe test (in session two), and 40 were used in the final 2-AFC task at the end of session two. Detailed instructions were given verbally to participants before each task and summarised written instructions were also given to participants at the beginning of each task (Appendix A). Participants had the opportunity to ask questions and took part in practice trials for each task.

2.3.1.3.1 Session 1

Initial exposure to scene images. After taking part in four practice trials, participants viewed 80 different scene images in a randomised order (four images from each of the 20 scene categories). On each trial, a fixation cross appeared on the screen for one second, followed by a scene image, which was displayed for four seconds. After the four seconds, a subsequent screen asked participants to 'please rate the attractiveness of this scene' (in other words, how aesthetically pleasing the scene was) on a scale of one to six (where six is most attractive). Participants made their response using the keyboard, and a response ended that trial, starting the onset of a new trial (Figure 2.1). Following viewing of all 80 scenes, participants engaged in a verbal distracter task for five minutes. Here, participants were randomly given the number 50 or 100 by the experimenter and asked to count back to zero in multiples of three. The aim of this was to minimize the likelihood of any verbal rehearsal of the scene images from the participant. However, this was unlikely to have occurred, as participants were not told that their memory was going to later be tested. Once the five minutes had finished, the experimenter gave the participants the instructions for the recent-probes test.

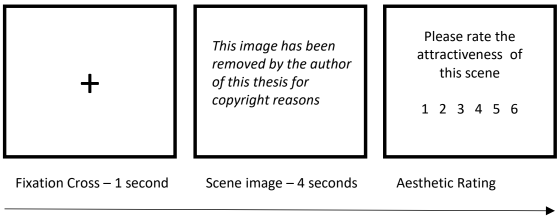


Figure 2.1. One trial in the initial exposure session; repeated for 80 trials during session 1.

Recent-Probes test. After the five-minute delay, participants were given two practice trials and then completed the recent-probes test, which consisted of 60 trials presented in a random order across three blocks. On each trial, participants were first shown five target scenes from one category for four seconds each, separated by a one second black fixation cross. After the five target scenes had been displayed, a red fixation cross was displayed to indicate that the next image would be the critical image to which participants had to respond (the ‘probe’). The probe was taken from the same scene category as the previous five images (for example, all bedrooms). Upon presentation of the probe, the participant’s task was to indicate if this probe was one of the previously viewed five images, and provide their answer (yes or no) using the keyboard (using the letters ‘z’ and ‘m' for yes and no, respectively) as quickly as possible. Participants’ RT and accuracy to make the yes/no judgment was recorded, and the importance of responding as quickly as possible was emphasised to participants through the written and verbal instructions. On a third of the trials (20/60), the probe was a ‘positive probe’; one of the five previously viewed target scenes (‘yes’ response required), on another third it was a completely novel (NRN) image that the participants had not seen before (‘no’ response required) and on a third it was one of the scenes previously viewed in the initial exposure phase – a RN probe (‘no’ response required) (see Figure 2.2). These 60 trials were intermixed across three blocks, with an equal number of trial types contained within each block (six of one trial type, seven of another, and seven of another). Participants were not made aware that there would be recent-probe trials (that is, that images from the attractiveness rating task they had already completed would also be presented). The time to complete the recent-probes test was around 30 minutes.

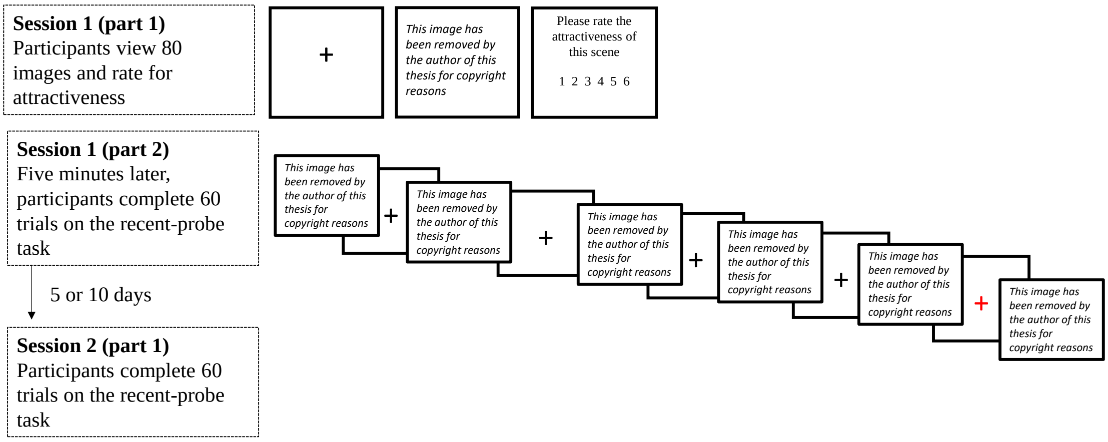


Figure 2.2. Schematic of the paradigm used for the recent-probes task, following earlier exposure to the probe scene during part one of session one (top row). Participants took part in the recent probes task after five minutes, and then either five or ten days later.

2.3.1.3.2 Session Two.

Recent-Probes test. Five or ten days later, participants completed a second recent-probes memory test, which followed the same procedure as in session one, with a new set of scene images. They were first given a practice trial to re-familiarise themselves with the task.

2-AFC test. Participants then completed a 2-AFC recognition test to assess visual recognition memory for the remaining 40 scene images that they had viewed during the initial exposure phase in session one but had not been used in the two recent-probe tests. For each trial, participants were shown an old image (previously viewed) and a novel image (the foil) side-by-side, taken from the same scene category. Their task was to indicate if the image on the left or the right was the one they had previously viewed, and communicate their answer using the keypad (key 1 or 3). The task was self-paced, and once participants had made their response, it initiated the next 2-AFC. Trial order, and foil image location (left or right) were randomised between participants. Following a selection of one image, participants were then asked to make a remember-know-guess judgment about their decision, using the keypad (keys 1-3). Participants were instructed to select 'remember' if they had a vivid memory of seeing the image during session one, to select 'know' if they did not have this vivid memory, but instead a feeling of familiarity when viewing the picture, and to select 'guess' if they did not remember either image and their answer was a random, forced guess. These choices were explained clearly to participants in the instructions prior to starting the 2-AFC test.

Following completion of the 40 trials, participants were debriefed regarding the true aims of the study.

2.3.2 Results

Unless otherwise stated, data in this experiment and throughout the thesis is handled as outlined here. Throughout, for analyses by ANOVA, where appropriate, Greenhouse-Geisser correction is reported for those cases where Mauchley's test of sphericity was significant. Where an ANOVA is found to be significant ($p < 0.05$) and follow-up multiple comparisons are made, a Bonferroni correction was applied, to reduce the risk of incorrectly rejecting the null hypothesis (Field, 2009). In this thesis this was only required when making up to three comparisons. For ANOVA, partial eta squared ($\eta_p{}^2$) is reported as a measure of effect size. Cohen's d is reported as a measure of effect size for t-test calculations, where a Cohen's d value of 0.2, 0.5, and 0.8 relate to a small, medium and large effect, respectively (Cohen, 1988).

2.3.2.1 Recent-Probes test

Only data from RN and NRN trials were included in the analysis (and not positive probe trials) as these are the two conditions of interest to this study. This is consistent with past research (McKeown et al., 2014; Mercer & Duffy, 2015). Preliminary analyses focused upon the delayed recent-probes task (in session two) and compared performance for participant groups completing their second recent-probes task following a five or 10 day delay. No significant differences emerged between these two groups for accuracy or RT measures. For mean accuracy (as a proportion) on RN probe trials there was no difference between the five day delay (M = .93, SE = .02) or 10 day delay (M = .96, SE = .01), $t(27) = 1.28$, $p = .21$. This was also the case when comparing accuracy on NRN trials at a five day delay (M= .95, SE = .01) and ten day delay (M = .93, SE = .02), $t(27) = 1.13$, $p = .27$. For mean RTs, on RN probe trials there was no difference between the five day delay (M = 868.68ms, SE = 47.95) and 10 day delay (M = 779.00ms, SE = 38.35) conditions, $t(27) = 1.37$, $p = .18$. This was also the case when comparing RTs on NRN trials, where there was no difference between the five day delay (M = 916.32ms, SE = 44.03) and 10 day delay (M = 804.33, SE = 48.95) conditions, $t(27) = 1.68$, $p = .10$. Therefore, in the following analyses, scores were collapsed across these two between-participant groups.

2.3.2.1.1 Accuracy

Accuracy scores were converted to a proportion (out of 20 trials for each trial type)[2]. Average accuracy in each condition is displayed in Table 2.1. Accuracy scores were high across RN and NRN conditions. A 2 x 2 mixed measures ANOVA was conducted on accuracy scores with the variables of test time (immediate, delayed) x trial type (NRN, RN). This showed there were no main effects or significant interactions (test time: $F(1, 28) = .058$, $p = .81$, $\eta_p^2 = .002$; trial type: $F(1, 28) = .68$, $p = .42$, $\eta_p^2 = .02$; interaction: $F(1, 28) = .74$, $p = .40$, $\eta_p^2 = .026$).

Table 2.1. Mean proportion accuracy (SE) for NRN and RN trials at immediate and delayed test. Note: Participants in five and 10 day delay interval conditions at session two are collapsed here.

Test Time	NRN Trials	RN Trials
Immediate	.95 (.01)	.93 (.02)
Delayed	.94 (.01)	.94 (.01)

2.3.2.1.2 Response times (RTs)

RTs were included in the analysis for correct trials only. Median RTs were calculated for each participant for each condition. Median RTs were used as they are less susceptible to extreme RTs biasing the data. This approach is consistent with other research using the recent-probe procedure (Atkins, Berman, Reuter-Lorenz, Lewis & Jonides, 2011; Craig et al., 2013). Average median RTs across all participants for each condition are shown in Figure 2.3. A 2 x 2 repeated-measures ANOVA was conducted with the variables of test time (immediate, delayed) x trial type (NRN, RN). There was

---

[2] The E-prime programme ‘froze’ for one participant during the recent-probe test in session 1, meaning they did not complete 5/60 trials (1 RN, 3 NRN, 1 positive probe). Their data is included in analysis and accuracy is still presented as a proportion.

not a significant main effect of test time, *F*(1,28) = 1.56, *p* = 0.22, η_p^2 = .05, or trial type, *F*(1,28) = 1.47, *p* = .24, η_p^2 = .05.

The interaction between test time and trial type was significant, *F*(1, 28) = 7.42, *p* = .01, η_p^2 = .21. Follow-up analyses revealed that NRN and RN probe RTs did not differ at immediate test, *t*(28) = 1.12, *p* = .27, d = .21, but did differ after a delay, as RTs were faster on RN trials (M = 831.57, SE = 32.85) compared to NRN probe trials (M = 869.92, SE = 33.88) at the delayed test, *t*(28) = 2.76, *p* = 0.01, *d* = .51.

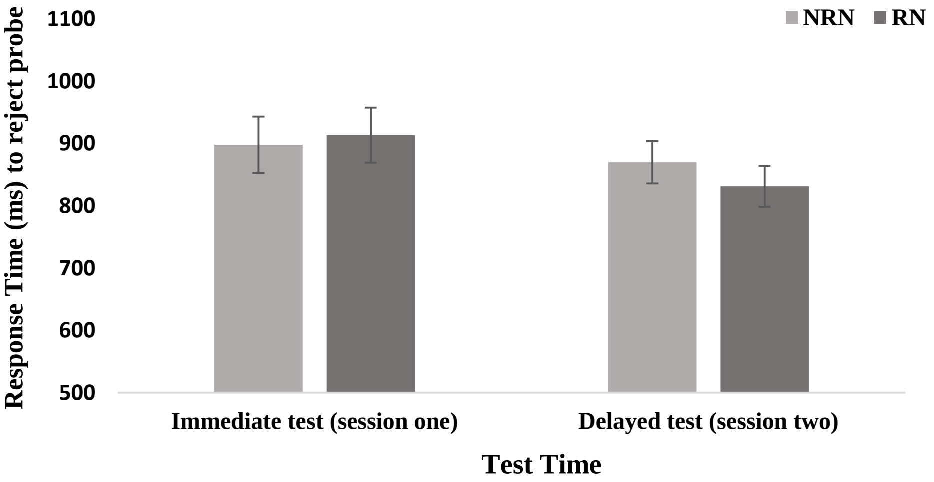


Figure 2.3. Average median RTs for NRN probe and RN probe correct trials at immediate and delayed test, in ms (error bars show standard error). Note: Participants in five and 10 day delay interval conditions at session two are collapsed here.

2.3.2.2 Recognition

An independent t-test found that there was no significant difference between recognition scores after five days (M = .77, SE = .02) and ten days (M = .76, SE = .02), t(27) = 0.22, *p* = .82, *d* = 0.09. The data was therefore collapsed across all conditions for the following analysis. The average recognition score across all participants, as a proportion, was .77 (SE = .01, range = .55 - .88). A one sampled t-test compared this to chance performance, finding that accuracy was significantly above chance (.5), *t*(28) = 18.67, *p* < .001.

The proportion of remember, know and guess judgments associated with the 2-AFC trials was calculated separately for correct and incorrect trials and are displayed in Table 2.2.

Table 2.2. Mean proportion (SE) of remember-know-guess responses averaged for correct and incorrect trials, separately.

Response	Correct trials	Incorrect trials
Remember	.38 (.03)	.07 (.02)
Know	.35 (.02)	.32 (.04)
Guess	.27 (.02)	.61 (.04)

Two one-way repeated measures ANOVAs with the factor of response (remember, know, guess) were conducted for correct and incorrect trials separately. For correct trials, there was no significant main effect of response, F(2,56) = 2.99, p = .06, η_p^2 = .10.

For incorrect trials, there was a significant main effect of response, F(1.37, 36.91) = 39.52, p < .001, η_p^2 = .59. Three follow up repeated measures t-tests to compare proportions for each response were all significant. Participants were more likely to select ‘know’ and ‘guess’ compared to ‘remember’, t(28) = 4.97, p < .001, d = 0.92 and t(28) = 11.00, p < .001, d = 2.04, respectively, and select the response ‘guess’ more compared to ‘know’, t(28) = 3.72, p = .001, d = 0.69.

2.3.3 Discussion

Using novel methodology based on the principles underlying the recent-probes paradigm, Experiment One sought to investigate incidental visual memory for scenes over a period of up to 10 days. The outcome was surprising, revealing a facilitation effect when responding to RN probe images, in that participants were faster to reject RN probes compared to NRN probe images. This is the opposite effect to previous research, in which proactive interference has been demonstrated by decreased accuracy (higher false alarms) and slower RTs when responding to RN probes.

It is intriguing that changing key premises of the methodology in this way led to these opposing findings. The length of delays used in previous research employing the recent-probes procedure have been much shorter, with retention intervals in the order of seconds (Berman, Jonides & Lewis, 2009), as opposed to minutes or days, as was used here. This successful adaptation of the recent-probes methodology to investigate longer-term memories allowed for incidental long-term memory for scenes to be tested. Here, in contrast to previous research, encoding and retrieval of the recent-probe image did not occur within the same context. In this experiment the recent-probe images were taken from a different task earlier in the session, or from a previous session. For example, in McKeown et al. (2014), participants were shown two abstract images side by side, and then following a delay of one to six seconds they were shown an individual image and had to respond as to whether the image was one of the previous two images displayed during that trial. On RN probe trials, the individually presented image was not one of the two presented on that trial, but instead was presented in the preceding trial. The time between trials (inter-trial interval) also varied between one and six seconds. Therefore, the interference effects observed in past research using the recent-probe methodology are likely to be due to the very short delays used, and therefore a possible result of residual memory traces still present in immediate memory, interfering with performance on the current trial. This is quite different to the conditions used in the current experiment.

Interestingly, there were no differences in RTs in the immediate test in Session One, which occurred at least five minutes following the initial viewing of the images, and spanned across approximately 30 minutes, demonstrating that after a delay of only minutes (as was the case in session one), the interference effect observed in previous research diminishes. After a period of days, participants were faster to reject the previously viewed RN probe compared to a NRN probe, despite only previously viewing this scene image once under incidental viewing conditions. This apparent facilitatory effect is intriguing. Indeed, recent research has demonstrated a facilitatory role of longer-term memory in a visual working memory task (Oberauer, Awh & Sutterer, 2017).

It is also possible that familiarity or practice with the recent-probes test in the immediate test session facilitated the faster RTs on recent-probe trials in the later test, and this possibility is explored further in Experiment Two, by removing the immediate test.

The recognition accuracy rates observed after five and ten days demonstrated good recognition memory for visual scenes, with accuracy over 75% and significantly above chance, even after viewing the scenes incidentally. This is similar to previous research where same scene category foils have been used, over shorter delay periods, using intentional encoding conditions (Konkle et al., 2010a). The lack of significant difference in recognition accuracy between the five and ten days suggest this scene detail is not forgotten over this time frame, even though the time of delay doubles. Previous research using object stimuli has found forgetting over a seven day period, when tested after a seven day delay compared to ten minutes. It may be the case that forgetting occurs more quickly in the initial timeframe after scene viewing, but memory fidelity remains more consistent over longer periods of delay.

The results from the remember-know-guess measure allowed further analysis and understanding of the processes underpinning recognition judgements in the 2-AFC. For correct trials, the pattern of results was that participants reported to remember the scene most frequently, and guess least frequently, although these differences were not significant. The results for incorrect trials followed the opposite pattern, in that participants were more likely to select ‘guess’, indicating that they were unsure on which of the two images they had previously seen.

2.4 Experiment Two

2.4.1 Overview

Experiment One found an effect on RTs when participants were given the recent-probes test after a number of days, but not when tested immediately. Experiment Two therefore represents an attempt to replicate the findings from Experiment One, through asking participants to complete a recent-probe test days after first viewing the scenes. In addition, the immediate test after scene viewing was omitted, to remove any practice effects that may have resulted in the findings in the subsequent session.

Therefore, in Experient Two, instead of two recent-probe tests (one in session one and one in session two), there was only one test given in session two. Participants attended a first session in which they viewed the scene images. They then left with no test, and came back five days later to complete the recent-probes test. Only one delay length of five days was used in this study as no differences between five and ten day delays emerged in the previous experiment, for responses on the recent-probe or 2-AFC test. This simplification of the design allows for a focus on using the recent-probes test to study incidental scene memory, days after scene viewing. As with Experiment One, a 2-AFC was also included here, following the recent-probes test.

2.4.2 Method

2.4.2.1 Participants.

Twenty-five participants (all female) took part in the study, with an average age of 19.67 years (SD = 1.27; range = 18-24). All reported having normal or corrected-to-normal vision. Participants were recruited using the Psychology department's participant pool scheme or paid participant database at the University of Leeds and were eligible to earn participant pool credits or a small fee (£10) for their participation. As in Experiment One, participants were not made aware that the study was investigating memory performance. Participants who had taken part in Experiment One were not eligible to take part in Experiment Two. The study consisted of two sessions, five days apart. The study received ethical approval from the School of Psychology's Research Ethics committee at the University of Leeds (ref no: 15-0260).

2.4.2.2 Materials, Design and Procedure.

Participants attended two sessions, the first in which they viewed scene images, and in the second they completed a recent-probes test and 2-AFC test. The materials used in Experiment Two were the same as those used in Experiment One. Participants were given written and verbal instructions for each task and took part in practice trials before each task. An overview of Experiment Two is given in Figure 2.4 below.

In session one, participants viewed the 80 scenes and rated them for their attractiveness on a scale of one to six. The order of the image presentation was counterbalanced between participants, with each participant being presented with one of a possible four order combinations. Session one lasted for approximately 20 minutes, and no test was given; instead all participants came back five days later to complete session two.

In session two, participants were first given the recent-probes test. The methodology here was the same as Experiment One, however participants completed 120 trials (40 RN probe, 40 positive probe, 40 NRN probe). This represents double the amount of trials that were used in Experiment One. All trials were allocated randomly to six blocks of 20 trials each. An equal number of trial types were contained within each block (six of one trial type, seven of another, and seven of another). The order of block and trial presentation within each block was counterbalanced between participants. Participants took breaks in-between each of the six blocks, and pressed the space bar to initiate the start of the next block. The independent variable of interest was therefore trial type (NRN vs RN), and these were compared for accuracy, and RTs on accurate trials.

Participants also completed the 2-AFC test. This was identical to Experiment One, and consisted of 40 2-AFC trials, presented in a counterbalanced order. Session two took approximately one hour and 20 minutes for participants to complete. Following completion of both tasks, participants were debriefed regarding the true aims of the study.

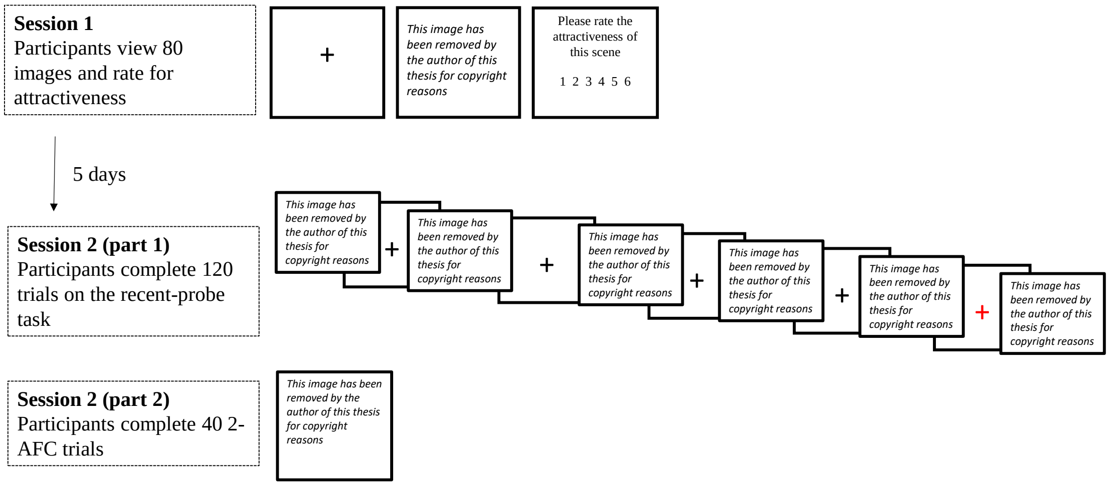


Figure 2.4. Schematic showing each task in Experiment Two. The image in each row depicts one trial within each task.

2.4.3 Results

2.4.3.1 Recent-Probes test

Accuracy and RTs are reported separately below. Only data from NRN and RN probe trials were included in the analysis.[3]

2.4.3.1.1 Accuracy

Accuracy scores were converted to a proportion (out of 40 trials for each trial type). Accuracy on the recent-probe test was high; mean proportion accuracy was .94 (SE = .01) and .92 (SE = .01), for RN probe and NRN probe trials, respectively. There was no significant difference in accuracy scores on RN and NRN probe trials, as demonstrated by a repeated-measures t-test, *t*(24) = 1.63, *p* = .12, *d* = .33. This is comparable to accuracy of participants in Experiment One, in which average accuracy in the delayed test was .94 (collapsing across the five and ten day delay conditions).

2.4.3.1.2 Response times (RTs)

Median RTs were calculated for each participant in each condition, for correct trials only. A repeated-measures t-test found that RTs were shorter for RN probe (M =

---

[3] The E-prime programme ‘froze’ for two participants during the recent-probe test, meaning they did not complete all trials. Their data is included in analysis and accuracy is still presented as a proportion of completed trials.

925.48, SE = 78.23) compared to NRN probe (M = 947.06, SE = 76.95) trials, *t(*24) = 2.84, *p* = .009, *d* = .57 (Figure 2.5).

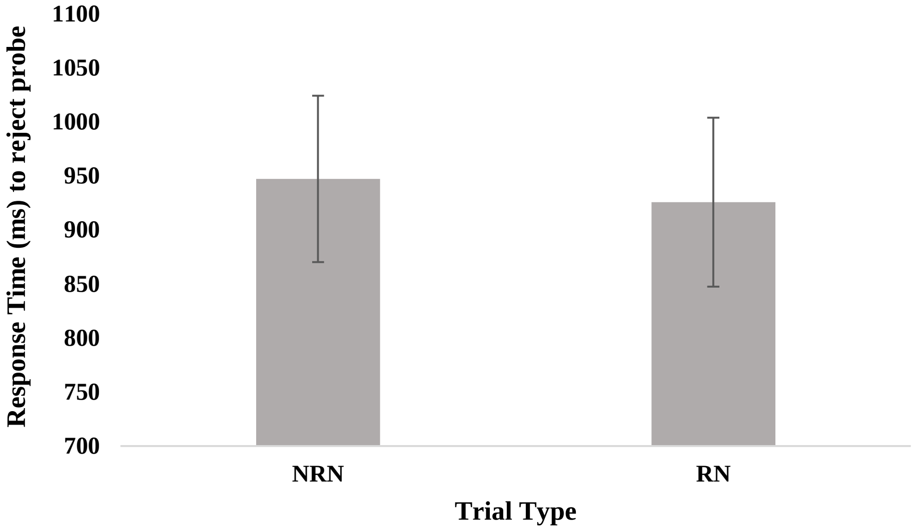


Figure 2.5. Median RTs in ms to reject NRN and RN probes (correct trials only). Error bars show standard error.

2.4.3.2 Recognition

The average recognition score across all participants, as a proportion, was .86 (SE = .02; range = .70 - .95). A one-sampled t-test showed that this was significantly above chance, *t*(24) = 23.42, *p* < .001.

The proportion of remember-know-guess judgments associated with each of the 2-AFC trials was calculated separately for correct and incorrect trials and displayed in Table 2.3.

Table 2.3. Mean proportion (SE) of remember-know-guess responses averaged for correct and incorrect trials, separately.

Response	Correct trials	Incorrect trials
Remember	.47 (.04)	.09 (.03)
Know	.32 (.02)	.33 (.06)
Guess	.21 (.03)	.59 (.07)

Two one-way repeated measures ANOVAs with the factor of response type (remember, know, guess) were conducted for correct and incorrect trials. For correct trials, there was a significant effect of response type, *F*(1.52, 36.3) = 14.01, *p* < .001, η_p^2 = .37. Three follow up repeated measures t-tests to compare proportions for each response were all significant. Participants were more likely to select 'remember' compared to 'know' *t(*24) = 2.85, *p* =.009, *d* = .57, and 'guess', *t*(24) = 4.49, *p* < .001, *d* = .90, and also select the response 'know' more compared to 'guess', *t*(24) = 3.26, *p* = .003, *d* = .65.

For incorrect trials, there was also a significant effect of response, *F*(1.29, 30.96) = 14.60, *p* < .001, η_p^2 = .38. Three follow-up repeated measures t-tests found that on incorrect trials, participants were less likely to select remember compared to know, *t*(24) = 3.96, *p* = .001, *d* = .79, and guess, *t*(24) = 5.90, *p* < .001, *d* = 1.18. Participants also selected guess more than know, but this difference was not significant when adjusting for multiple comparisons, *t*(24) = 2.10, *p* = .047, *d* = .42.

2.4.4 Discussion

In the recent-probes task, a similar pattern of results was observed when compared to Experiment One. Participants were faster when responding to RN probes compared to NRN probes after a delay of five days. In Experiment One, similar results were found in the delayed test in session two after also taking part in the recent-probes task in session one. These results from Experiment Two suggest the results in Experiment One were not a result of a type of practice effect or from familiarity with the recent-probes procedure, as the initial recent-probe task in session one was removed.

Recognition performance was again good in the 2-AFC test; on average 86% and significantly above chance. Remember-know-guess judgements were descriptively similar to Experiment One; when participants were correct, they were more likely to report a vivid memory of the image from session one (remember judgement), compared to know or guess responses. These differences were significant in this experiment. Incorrect responses were again linked to participants reporting guess judgements.

2.5 General Discussion

The current set of studies aimed to investigate detailed memory for visual scenes over the period of five minutes, five days and ten days, for scenes which had been viewed under passive/incidental viewing conditions.

Despite learning under incidental viewing conditions and not being explicitly aware that memory for scenes viewed earlier was being tested, RTs were affected in the recent-probes test, but only when tested days later (Experiments One and Two), and not when tested within minutes after exposure (Experiment One). Previous research using the recent-probes procedure in short-term memory studies, has found an inhibitory effect of RN probes on responding (McKeown et al., 2014). Contrary to predictions, RTs were significantly faster when rejecting these RN probes compared to rejection of novel, NRN probes. Thus, the RT patterns here may be taken as evidence of a memory trace facilitating performance. Although this was a small effect, it was replicated across two experiments.

Recent research has also found a faciliatory role of longer-term memories in a visual working memory task. Oberauer et al. (2017) conducted a series of studies to investigate the impact of long-term memories on working memory performance under different conditions, testing for any effects of proactive interference, but also proactive facilitation. They initially presented participants with a series of objects in different colours. Following this, a task was given to participants in which they viewed sets of three objects, and then had to recall the colour of the three objects that had just been displayed. Crucially, two objects had been previously learnt in the earlier phase; this time they were presented in a colour consistent with how they were originally shown, or in a different colour. The third object was a previously unseen object. They found that accuracy was higher when responding to an object that had been displayed in the same colour in which it had been previously viewed, reflecting a facilitatory effect from long-term memory. In the current study, a facilitatory effect was also observed, but this was displayed through quicker RTs. It is interesting that no effect on RTs was found after a delay of five minutes. It appears that at this length of time any proactive interference effect that has been observed in previous research may have diminished, but it may take more time for the facilitatory response to emerge.

There were no differences in accuracy scores in RN probe compared to NRN conditions. Accuracy was high across all conditions; even when visually similar (same category) images were used, participants could still accurately reject RN probes and novel (NRN) probe images that had not been present in the current trial, demonstrating the high fidelity of visual scene memory. It is possible that a ceiling effect may have occurred in accuracy scores on the recent-probes test; participants scored very highly in NRN and RN probe trials in both sessions. The high accuracy scores may have meant that any small differences in accuracy between conditions could not have been observed. However, high levels of memory accuracy may increase the sensitivity of RT measures to detecting differences between conditions (Vergauwe & Cowan, 2015), and this may have been the case here. In future research, the recent-probes test could be adapted to make it harder, which may allow for any differences in accuracy to be seen between conditions. This could be done, for example, by making image presentation briefer, or increasing the similarity of the foil images used in the trial.

In line with previous research, recognition accuracy also remained high even ten days after initial viewing, demonstrating good memory for the previously viewed visual scenes, even when same scene category foils were used. Past research has demonstrated that visual scene memory appears to maintain high fidelity records even after periods of days. The task here also differs to the previous research using same scene or object category foils (e.g. Andermane & Bowers, 2015; Konkle et al., 2010a), as the scene images here were first viewed under incidental viewing conditions. The current data adds to previous research by demonstrating that highly detailed incidental memory for scenes is maintained over a five to ten day period. Here, participants are unlikely to have been attempting to remember or indeed rehearse the level of detail needed to later differentiate from a same scene category foil, and therefore this seems to be encoded and retained incidentally.

Remember-know-guess observations were also collected following each recognition trial. Remember/know judgements (Tulving, 1985) are a useful tool to help to better understand the processes underlying participants' recognition judgements. There has been much debate over the processes underlying remember/know judgments, and whether or not they are qualitatively distinct (*see* Heathcote, Raymond & Dunn, 2006; Yonelinas, 2002, for discussion); dual process models position 'remember' judgements as representing more vivid recollection, whereas 'know' judgments are thought to be driven by a sense of familiarity (see Schurgin, 2018). In the current experiments, participants were asked to select 'remember' if they had a clear memory for seeing the image in the first session, whereas to select 'know' if they didn't have this vivid memory, but instead had a feeling of familiarity or knowing that they had seen the image before. For correct trials, participants were significantly more likely to select 'remember' compared to 'guess' and 'know' in Experiment Two. A similar, albeit non-significant, pattern of responses was observed in Experiment One. It is interesting that when accurately selecting the previously viewed image in the 2-AFC, participants were more likely to report having a clear memory for seeing this image previously. This builds upon previous research using these types of recognition tests with same scene foils, and shows that participants still feel that they have a clear memory for the image that they saw five days earlier, despite the similarity of foils to the previously seen images. The findings also tell us more about incidental scene memory. Where participants were correct, this was likely to be driven by increased confidence that they remembered previously viewing the image.

Across both experiments, participants were more likely to select 'guess' compared to 'remember' or 'know' on incorrect trials. Including a 'guess' option in addition to remember and know, allows participants to indicate that they have no memory for either image in the 2-AFC test (Eldridge, Sarfatti & Knowlton, 2002; Gardiner, Ramponi & Richardson-Klavehn, 2002). Again, this is useful as it suggests that when participants were inaccurate, it was not because they were necessarily picking the same scene category foil by mistake (perhaps due to its conceptual likeness), but instead that they did not have a memory for either image and due to the nature of the paradigm they had to make a forced guess.

A general limitation of incidental memory research is the uncertainty over whether participants are aware during the viewing of scenes that their memory will later be tested, given that they are taking part in a Psychology experiment. In addition, it is not certain that participants taking part in the recent-probe test were not aware that their memory for scene images viewed earlier was being tested. Indeed, the findings from the 2-AFC task and the associated remember-know-guess judgements are interesting, as where participants do accurately recognise a scene, this was more likely to be associated with a remember or know judgement, suggesting awareness of previously viewing that scene. In this study, steps were taken during the advertisement and the instructions to remove the likelihood of participants guessing during scene viewing or the recent-probe test that their memory for the scenes was being tested.

Taken together, the current results from the recent-probes and 2-AFC tests suggest that we passively maintain a highly-detailed memory trace of the recent visual past for days after viewing, even when intentional encoding has not taken place. Furthermore, this accuracy appears to largely be driven by participants reporting to have a vivid memory of originally seeing the scene, rather than a sense of familiarity with the scene, or lucky guesses.

The different tasks used to assess memory in the current study varied in that they required either intentional attempts to remember previously viewed scenes (2-AFC test) or measured the - presumably incidental - effects of previously viewed images on a current task (the recent-probes test). Despite the differences in these two methods, significant results were still found in both tests, suggesting that even after incidental viewing of many scene images, participants still remembered details about these scenes up to 10 days later. This good incidental memory for many scenes, despite incidental encoding, supports past research in which similar recognition rates have been obtained following intentional compared to incidental encoding (Castelhano & Henderson, 2005). Furthermore, this good incidental memory for scenes is for detail too - all the foils used in these two experiments across both memory tests were from the same scene category, therefore suggesting detail above basic gist information was encoded during incidental viewing, and this detail was retained in VLTM.

The current experiments have shown that incidental methods of testing long-term visual memory can be utilised. These types of methods may also more closely reflect how natural everyday scenes we have viewed previously, under incidental conditions, may be recalled if we re-encounter that scene at a later date.

CHAPTER THREE

DYNAMIC SCENE STIMULUS SET: DEVELOPMENT AND RATING STUDY

3.1 Introduction to Stimulus Set and Rating study

3.1.1 Rationale

Understanding the fidelity of memory for everyday naturalistic scenes demands attending to dynamic stimuli, and therefore the aim of the work described in this current chapter was to develop a stimulus set that more closely mimics the types of stimuli previously used within static scene research (including Experiments One and Two). Here, the aim was to create a novel stimulus set of dynamic natural scenes suitable for use in experimental work testing the fidelity of memory for dynamic scenes over differing retention periods. In particular, this experimental work will place emphasis upon assessing whether memory for information within a dynamic visual scene is affected by where that information is temporally placed within the scene (beginning, middle, or end). The rationale for the work in this chapter was so that future experiments in subsequent chapters can use the dynamic scenes, with assurance that there aren't specific features of the films or the static images taken from different temporal parts of the films that are particularly relevant or memorable, or affect the validity of the results.

Therefore the stimuli set created here will consist of dynamic scenes, each with a set of corresponding static target images taken from different temporal points across the scene, to be used as memory probes.

A dynamic, or moving, scene is considered here as a scene in which elements (such as people and objects) are moving within the scene, against other stationary objects and a stationary background. Although a range of stimulus sets have been used by researchers investigating natural scene memory for static scenes (for example, in: Konkle et al., 2010a; Wolfe, Alvarezm Rosenholtz, Kuzmova & Sherman, 2011), less research has used dynamic scene stimulus sets. Research that has investigated dynamic scene memory using larger stimulus sets have tended to use clips from films, documentaries and cartoons (Buratto et al., 2009; Ferguson et al., 2017), or filmed an actor moving within an indoor scene (Hirose, Kennedy & Tatler, 2010). However, in these films there is often a central actor/s, and they are very different to the scene stimuli used in static scene research (e.g. Hollingworth, 2005; Konkle et al., 2010a).

Research that has focused upon studying eye movements when viewing static compared to dynamic scenes have instead used films of natural everyday scenes. These have been taken from small stimuli sets of around 20 scenes (Dorr, Martinetz, Gegenfurtner & Barth, 2010; Smith & Mital, 2013). In these examples, short (around 20 second) films were taken on a digital camera in streets, car parks, shopping centres, parks and near rivers using a static camera position. For the current research, a similar approach is taken, with the aim to develop a larger collection of dynamic, natural everyday scenes. This required short films to be shot from a static viewpoint in which there is motion from different objects, without a focus on one particular actor. Filming was done outside in public places where movement came from cars, people, animals, wind, and other changes (such as traffic light changes, etc.).

Further, a key aim of the work is to develop a set of images that are perceived by participants as being similar on factors known to influence the accuracy of visual memory. Using such stimuli in experimental work that examines the role dynamic information plays in scene memory (as planned in Experiments Five, Six and Seven) would allow conclusions concerning the independent contribution of dynamic information (and how it unfolds temporally) to be better assessed. In two experiments reported in this chapter, participants were asked to rate the dynamic scene stimuli for the features of complexity, distinctiveness, and attractiveness. The complexity of visual scenes may affect memory and subsequent test performance (Eng, Chen & Jiang, 2005). Here, this was defined to participants as scenes that contain more objects and movement and are more complicated, whereas scenes with lower complexity may have less objects and movement within them. Scenes that are more distinctive compared to others in the stimuli set may also affect visual scene memory (Konkle et al., 2010a; Konkle, Brady, Alvarez & Oliva, 2010b). A highly distinctive scene was defined to participants as having characteristics that would distinguish it from other scenes of its type. In addition, participants were asked to make aesthetic/attractiveness judgements of scenes, which may be related to the memorability of images (*see* Isola, Xiao, Parikh, Torralba & Oliva, 2014, for discussion). It was important to test these characteristics for each scene, as it may be necessary to control for these characteristics across conditions in later studies. We control for these features by asking two different groups of independent participants to rate the dynamic scene stimuli on features of complexity, attractiveness and distinctiveness on a scale of one (low) to seven (high). One group of participants rated each dynamic scene and another the static target images that were taken from a corresponding dynamic scene.

Ethical approval for the development and use of the scene database in the thesis research, including the rating study described below was given by the School of Psychology's ethical review board at the University of Leeds (ref no: 16-0068).

3.1.2 Filming

The films recorded showed 'everyday' outdoor films. Films were recorded on a Sony HDR-PJ810E HD 1080p camcorder, in MPEG4-AVC/H.264 format. The camera was handheld in a stationary position and filmed public outdoor locations in UK cities. Around 100 films, 30 seconds in length were filmed. All films included multiple vehicles as the main source of movements. Other movement came from animals, people, objects being blown in the wind, and other dynamic changes within the scenes. To give an illustration of the type of dynamic scenes filmed, snapshots taken from some of the scenes are given in Figure 3.1.

Figure 3.1. 'Snapshots' taken from some of the dynamic scenes in the stimulus set.

Relevant guidelines and frameworks set out for visual researchers were followed (for example, Wiles et al., 2008). Careful attention was also given to not filming any people or vehicles close enough to be identifiable. All films were retrospectively checked to ensure no identifiable information was inadvertently captured. All films were taken in public places (such as towns and streets) and not in places which might be considered more private or ‘semi-public’, such as indoor shopping centres and restaurants. As familiarity to a scene may affect subsequent perception of and memory for that scene (Brockmole & Henderson, 2006), no filming was done in places likely to be known to potential participants (including Leeds City Centre and the University of Leeds campus).

3.1.3 Editing

Films were edited using Windows Movie Maker (http://windows.microsoft.com/en-us/windows/movie-maker), and film clips were trimmed to 22 seconds in length. Where filming inadvertently captured identifying information (such as registration plates), films were cropped to remove this where possible, and if not possible, those scenes were deleted. This left 80 scenes of 22 seconds in length. From these, a ten second film was taken from the middle, creating the dynamic film clips to be used in subsequent experiments. For approximately half (46) of the scenes, the 10 seconds was taken from eight seconds into the longer 22 second clip, the remaining 10 second clips were taken from four seconds in (illustrated in Figure 3.2). The rationale for this was to take a 10 second clip which began near (but not at) the beginning of the 22 second scene (four second start) and one which ended near (but not at) the end of the scene (four seconds from the end). Static ‘snapshots’ were then taken from three different time points throughout the 10 second clips, from one, five and nine seconds, representing the beginning, middle and end of the dynamic scene, respectively. These static images that are taken directly from the 10 second dynamic scene are referred to as ‘target’ images throughout[4]. The dynamic scenes were rated in Experiment Three,

---

[4] The remaining 12 seconds of the 10 second scene allowed static ‘foil’ images to be taken and used in subsequent recognition tasks, to create highly similar same-scene foils. The selection of foils is described in subsequent chapters.

and the static target images taken from these dynamic scenes were rated in Experiment Four.

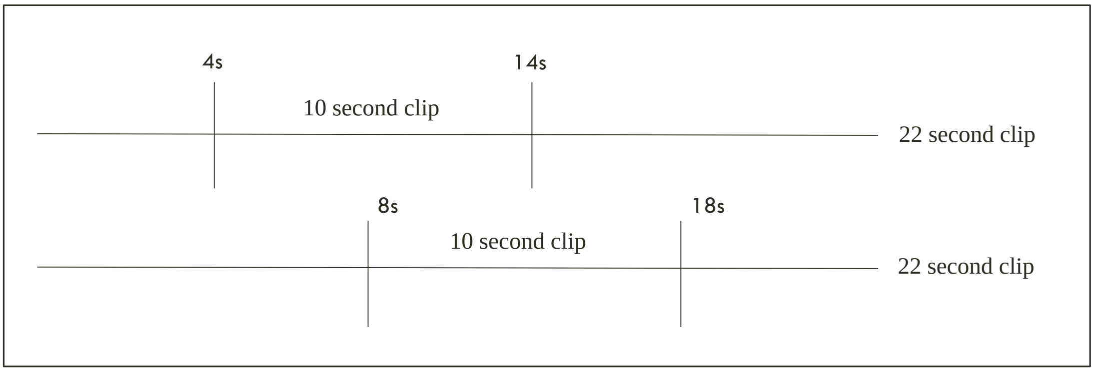


Figure 3.2 Illustration of where the short 10 second film clips were taken from relative to the 22 second clips. Note: Approximately half were taken from four seconds onwards, and half taken from eight seconds onwards.

3.2 Experiment Three: Ratings of Dynamic Scenes

3.2.1 Overview

In Experiment Three, participants rated the 22 second dynamic scene clips, on attractiveness, complexity and distinctiveness. The 22 second clips were used (rather than 10 seconds) to collect overall ratings of the dynamic scenes, as foils images were later taken from the remaining part of the dynamic scene. The aim was to produce a stimulus set that is suitable for testing the fidelity of memory for dynamic scenes over time, and across the unfolding temporal sequence of the scene, whilst controlling for other attributes about the dynamic scenes known to affect memory. Thus, for each scene participants' ratings were collected with a view to identifying (and potentially excluding) dynamic scenes that appear as outliers on any of these attributes.

In this experiment, participants also completed an event segmentation task for each dynamic scene. Event Segmentation Theory suggests that visual events are perceived as smaller units based on event boundaries, which are places where there is a natural boundary within an ongoing film or event (Speer, Zacks & Reynolds, 2007; Zacks, Speer, Swallow, Braver & Reynolds, 2007; Zacks & Swallow, 2007). Therefore a visual event or scene will be segmented into smaller events during perception. Such a 'boundary' could include an actor moving from one task to another, or an object or actor leaving or entering the scene (Zacks et al., 2001; Zacks & Tversky, 2001). Of most relevance to this research, later memory accuracy on recognition or recall tasks is found to be higher around these event boundaries (Newtson, 1973; Newtson & Engquist, 1976; Sargent et al., 2013; Schwan, Garsoffky & Hesse, 2000). The event segmentation task was included here to check that there were no systematic differences in the frequency with which event boundaries are perceived by participants during the beginning, middle and end of the dynamic scenes. This is important as static target images that were taken from these different time points within the scene will be used in later recognition tasks.

In this experiment, participants first watched each 22 second dynamic scene one time, to familiarise themselves with the scene, before watching it for a second time. On the second viewing, participants completed an event boundary task in which they were required to press a space bar on the key pad when they believed a perceptual ‘event boundary’ had occurred in the scene. After completing the event boundary task, participants then viewed the scene for a third time, and reported if they recognised the location of the scene, and rated the scene for the attributes of attractiveness, complexity and distinctiveness.

3.2.2 Method

3.2.2.1 Participants

Twenty two participants were recruited for the study (Female: 20, Male: 2), with a mean age of 20.36 (SE = .07). All participants were students at the University of Leeds and were recruited via the School of Psychology’s participant recruitment site. Participants were reimbursed with course credit for participation. Before starting the study, participants were given an information sheet and signed a consent form.

3.2.2.2 Materials and Procedure

To make the task more manageable for participants, the 80 scenes were separated into two sets of 40 for presentation to the participants, and participants rated one of the two sets (11 participants rated each set). The instructions for the task were explained to participants verbally and a summary of these instructions were also given on paper (Appendix B). Participants viewed 40 dynamic scenes. Each scene film was 22 seconds in length, and the order of scene presentation was counterbalanced.

Participants viewed each dynamic scene three times. On the first view, participants were instructed to simply watch the video. On the second viewing, they were required to respond using the space bar whenever they believed an event boundary occurred within the scene. It was explained to participants that the task was completely subjective and the number of times they pressed the space bar was up to them. Participants then watched the film for the third time and were then required to rate the scene on a paper form. They were asked to report whether they were familiar with the location of the scene (yes or no answer), and then rate the scene for attractiveness, complexity and distinctiveness, all on a scale of 1-7 (unattractive-attractive; not complex-complex; not distinctive- distinctive).

In this and future experiments using the dynamic films, the computer used had an additional NVidia graphics card inserted, and the experiment was presented through PsychoPy (Peirce, 2007; 2009).

3.2.3 Results and Discussion

3.2.3.1 Ratings of dynamic scenes

3.2.3.1.1 Familiarity with scene location

For each scene, participants were asked if they recognised the location in which the scene was filmed, to test the assumption that participants would not be familiar with the majority of the locations in which the dynamic scenes were filmed. To see if any scene was more familiar than the others to the participants, the frequency with which each scene location was reported to be familiar to participants was calculated. This ranged between 0 and 5 participants reporting to be familiar with the scene location, with an average of 1.05 participants per scene.

3.2.3.1.2 Attractiveness, Distinctiveness and Complexity Ratings

For each scene, an average rating was calculated from the ratings given by the 11 participants. Across all scenes, the average attractiveness ($M = 3.49$, $SE = 0.08$, $2 < M < 4.91$), distinctiveness ($M = 3.33$, $SE = 0.07$, $1.91 < M < 4.91$) and complexity ($M = 3.41$, $SE = 0.07$, $2.09 < M < 4.82$) ratings were calculated. The averages for each scene on the three different rating scales (attractiveness, distinctiveness and complexity) were screened for outliers, to identify scenes that had average ratings that were more than three SDs away from the mean on any of the scales. There were no outlying scenes for any of the three rating scales.

As the analysis found that no scenes were classed as an outlier based on the average rating they received for attractiveness, complexity or distinctiveness, this provides reassurance, from an independent group of participants, that it is unlikely that any of the dynamic scenes will be particularly memorable due to complexity, distinctiveness or attractiveness compared to others.

3.2.3.2 Event Segmentation Task

The focus of the analysis was on the 10 second part of the dynamic scene, taken from within the 22 second films. It was possible to analyse the number of event boundaries perceived for each scene across the beginning, middle and end of the 10 seconds of dynamic film clip that would be used in later experiments. To do this, the average number of times event boundaries occurred in three time segments of 3.33 seconds (beginning, middle, end) were calculated for each scene. On average, scenes received a similar number of event boundaries across the three time segments, as shown in Table 3.1. A one-way repeated-measures ANOVA was conducted to compare the number of event boundaries across each time segment, which was found to be non-significant, $F(2,158) = .003$, $p = .997$, $\eta_p^2 = .00$.

Table 3.1. The average (SE) number of times each scene received a report of an event boundary, by time segment (beginning, middle, end).

Time segment	Mean (SE)	Range
Beginning	0.37 (.03)	0 < M < 1.18
Middle	0.37 (.02)	0.09 < M < 1.09
End	0.37 (.03)	0.00 < M < 1.35

These results demonstrate that there are no inherent biases across the dynamic scene in terms of the temporal positioning of event boundaries. Overall for the stimulus set of dynamic scenes, the average frequency with which event boundaries are reported does not systematically differ between the beginning, middle and end of the 10 second dynamic films clips.

3.3 Experiment Four: Ratings of Static Target Images

3.3.1 Overview

As described earlier, static ‘snapshots’ (referred to as target images) were taken from across the ten second dynamic film clips, to be used in later recognition tasks. A selection of these target snapshots taken from one, five and nine seconds were rated in this experiment, representing the beginning, middle and end of the dynamic scenes. Ratings of complexity, attractiveness and distinctiveness were obtained for each of three target images that corresponded to each of 80 dynamic scenes.

3.3.2 Method

3.3.2.1 Participants

Twenty six participants were recruited for the study (Female: 26), with a mean age of 18.85 (SE = 0.13). All participants were students at the University of Leeds and were recruited via the School of Psychology’s participant recruitment site. Participants were reimbursed with course credit for participation. No participant had previously taken part in rating the dynamic scenes (Experiment Three).

3.3.2.2 Materials and Procedure

Participants rated one of the two sets of the 40 dynamic scenes (13 rated each set). For each dynamic scene, participants rated the three corresponding target images that had been taken from across the 10 second dynamic scene clips (at one, five and nine seconds). Therefore each participant rated 120 target images in total. The instructions and explanation of each rating scale for the task was explained to participants verbally and a summary of these instructions were also given on paper.

Each of the three target images was separated into three blocks (that is, one block of all images from one second, one block from five and one from nine). Order of block presentation was counterbalanced between participants, and the order of target image presentation within each block was randomised.

Participants viewed the target images on the screen one after the other in PsychoPy (Pierce, 2007; 2009), rating them on a sheet of paper provided. For each target image, participants first reported if they were familiar with the scene location (yes/no response; this was only measured for the first block) and then rated the target image for the three attributes on paper on a scale of 1-7 as described in Experiment Three. Participants pressed a space bar once they had finished rating an image, to initiate the presentation of the next image on the screen.

3.3.3 Results and Discussion

3.3.3.1 Familiarity with scene location

To test the assumption that participants would not be familiar with the majority of the locations in which the dynamic scenes were filmed, and to see if any scene was more familiar than the others to the participants, the frequency with which each scene location was reported to be familiar to participants was calculated. This ranged between 0 and 3 participants reporting to be familiar with each scene location, with an average of 0.61 participants per scene (out of the 13 who rated each scene).

3.3.3.2 Attractiveness, Distinctiveness and Complexity Ratings

The aim of the rating study was to compare responses to the three different target types (beginning, middle, end) on each of the three rating scales (attractiveness, complexity and distinctiveness). For each of the three target types (beginning, middle, end) pertaining to a specific dynamic scene, participants ratings were averaged to provide separate mean scores for the attributes of attractiveness, complexity and distinctiveness. Averages across all scenes, depending on target type and rating scale, are given in Table 3.2.

Table 3.2. Average (SE) scene ratings given to each target type on each of the three rating scales.

Target type	Complexity	Distinctiveness	Attractiveness
Beginning (1 second)	3.09 (.08)	3.13 (.08)	2.95 (.08)
Middle (5 seconds)	3.03 (.08)	3.08 (.08)	3.02 (.09)
End (9 seconds)	2.99 (.08)	3.17 (.08)	3.02 (.08)

Three separate repeated-measures ANOVAs were conducted for each of the rating scales. Target type (beginning, middle and end) was treated as a repeated-measures factor. This reflected the fact that three target types were taken from each of 40 dynamic scene stimuli. All were non-significant (*p*s > .10), showing no main effect of target type on any of the ratings across the three ratings scales.

Therefore, there were no significant differences observed in how participants rated the complexity, attractiveness or distinctiveness of the three static target images taken from different temporal locations within the dynamic scenes.

3.4 General Discussion

A stimuli set of dynamic visual scenes was required for use in subsequent experiments. Eighty scenes were filmed, edited and rated by an independent group of participants.

Overall reported familiarity with individual scenes used in this study was generally low across both experiments. This is relevant as the participants used in future research will be from the same participant pool, and therefore may have similar levels of recognition for the places depicted in the films. It is worth noting here that no participant that took part in Experiments Three or Four were eligible to take part in the subsequent experiments reported in this thesis that use the dynamic scenes. Likewise, participants couldn’t take part in more than one of the subsequent experiments.

A specific research aim that will be tested in the following experiments is to compare the fidelity of visual memory for different parts of a previously viewed dynamic scene (that is, to ask if certain temporal parts of dynamic scenes are remembered with higher fidelity). Across the two experiments reported here, there was no evidence of inherent differences within the dynamic scenes that would account for any differences in accuracy for target types taken across the temporal duration of the dynamic scenes. In Experiment Three, the number of event boundaries that were perceived for each scene across the beginning, middle and end was comparable. Furthermore, Experiment Three demonstrated that no particular scene was an outlier in terms of the ratings it received for the distinctiveness, attractiveness and complexity. Experiment Four demonstrated that comparable ratings of distinctiveness, attractiveness and complexity were given for targets taken from the beginning (one second), middle (five seconds), and end (nine seconds) of the 10 second dynamic film clips that participants in later experiments will be shown and tested on.

The dynamic scenes described here, and the corresponding target images, will be used in the subsequent experiments in this thesis (Experiments Five, Six and Seven). A key aim of these experiments is to test the fidelity of visual memory for dynamic scenes over different retention periods, and to explore if the pattern of responding varies for target images taken from different temporal locations across a previously viewed dynamic scene.

CHAPTER FOUR

THE FIDELITY OF VISUAL MEMORY FOR DYNAMIC SCENES – PART ONE

4.1 Chapter Summary

This chapter reports two experiments that explore the fidelity (that is, the detail) of visual memory for dynamic scenes. Here, participants viewed brief film clips depicting everyday dynamic scenes, followed by a 2-AFC recognition task which included static frames taken from the dynamic scene (targets), or highly similar foil frames taken from the same scene at a slightly different time point to the one participants viewed originally. Recognition accuracy for these frames was tested immediately following dynamic scene viewing (Experiment Five). In a subsequent Experiment (Experiment Six), a delayed test was also introduced, so that recognition accuracy was tested both immediately and after a delay.

Experiment Five tested recognition accuracy for static frames taken from five different temporal positions across the scene. Accuracy was high and above chance, suggesting the memory representation is detailed enough to recognise specific object details from dynamic scenes, as participants were able to distinguish previously viewed frames from same scene foils. There was no significant difference in accuracy in target recognition dependant on the temporal location of the dynamic scene that the target had been taken from. Two different lengths of intertrial interval (ITI) were also tested to allow comparison, however accuracy was high and comparable across both conditions.

In Experiment Six, participants performed above chance in both an immediate and delayed test. Recognition performance at immediate test was reliably higher, suggesting a more detailed initial scene representation that is forgotten to some extent across longer periods of time. Again there was no significant difference in accuracy when responding to the targets taken from different temporal positions within the scene in the immediate or delayed condition. These results help inform understanding of the precision of recognition memory for dynamic, changing visual scenes.

4.2 Introduction

Previous research has aimed to test VLTM memory for both static and dynamic scenes, with a focus on the fidelity of memory. Using static scene stimuli, Konkle et al. (2010a) presented participants with almost 3000 scene images across a time period of just under six hours. This was intentional encoding, as participants were aware of the upcoming memory test. Participants were then given a 2-AFC task 20 minutes after viewing all scenes. Accuracy was found to be lower when the foil image used in the 2-AFC task was taken from the same scene category compared to a different scene category. Furthermore, when participants had viewed more scenes from a scene category originally (between four and 64 were viewed originally), accuracy was lower. The authors suggest that the decline in performance when viewing more scenes from the same scene category is a result of interference in VLTM. Despite this, accuracy was still 76% on trials in which the foil image was from the same scene category, and participants had earlier viewed 64 scenes (the upper amount) from that category previously. Experiments One and Two in this thesis also tested visual recognition memory for static scenes using same scene category foils. Here, participants viewed scenes incidentally, and the memory test occurred five or 10 days later. It was found that participants were still able to accurately identify a previously viewed scene image in a 2-AFC test, when presented alongside a scene from the same scene category, up to 10 days later. The impressive recognition performance here and in Konkle et al.'s research suggests that the fidelity of memory for previously viewed scenes is detailed enough for participants to distinguish between the previously viewed image and a foil from the same scene category, even under conditions of viewing thousands of scenes at encoding, incidental encoding, or after longer delays of many days.

A series of studies conducted by Matthews and colleagues aimed to compare memory for static and dynamic scenes. Matthews et al. (2007) observed a 'dynamic superiority effect', in which recognition accuracy was higher on dynamic scene compared to static scene trials. Here, the mode of the scene stimuli was matched across encoding and test times (that is, a scene viewed dynamically during the encoding session, was subsequently also tested using a dynamic clip at test). Furthermore, a subsequent study varied the mode of the scene stimuli at encoding and test, and found a congruency effect, in that accuracy was highest when scenes were dynamic at both encoding and test (Buratto et al., 2009).

To understand the fidelity of memory for dynamic scenes, the similarity of foils to targets in the recognition test has also been manipulated. Matthews et al. (2010) used old/new recognition tasks to test memory for static and dynamic scenes using similar foils. In one experiment, participants viewed short dynamic scenes or static scene images taken from the films, for three seconds each, before taking part in the recognition test at some point subsequently (90 minutes up to 14 days). In this experiment, the mode of the scenes at encoding (static or dynamic) matched the mode at test. The similarity of the foils in the recognition test were manipulated so that they were either ‘similar’ to a previously viewed scene (the same people within the same scene), ‘related’ (the same people but in different locations) or ‘unrelated’. Hit rate when responding to targets was higher for moving compared to static scenes. In terms of false alarm rate, in which participants incorrectly report a new image as old, when the foils were related or similar, higher false alarm rates were observed in dynamic scene trials compared to static. This difference in false alarms between dynamic and static scenes was not present in the ‘unrelated’ foil condition, leading the authors to suggest the increased false alarm rate was not due to a change in participants’ response bias whereby they were just more likely to respond ‘old’ to dynamic scenes. Matthews et al. suggested memory for dynamic scenes (compared to static) may improve familiarity with particular people or objects within dynamic scene, even if they are later presented in a different context, hence increased false alarm rate when foils are similar. In terms of the differing durations of delay, there was no main effect of delay (90 minutes versus 14 days). However, more old responses (hits and false alarms) were made overall during the recognition test when tested at the longer delays of one or 14 days. Furthermore, with increased durations of delay, the hit rate when responding to previously viewed scenes decreased, and false alarms to unrelated scenes increased at test. Therefore although participants were able to do the test to some extent (hit rate was higher than false alarm rate at all delay conditions), this ability reduces as the delay between encoding and test increases (up to 14 days).

Ferguson et al. (2017) conducted a series of experiments in which participants were shown film extracts. This was followed by recognition tests using static target frames either taken from the film that participants saw, or foils that were taken from the same film clip, but not a part the participant had seen. In one experiment, participants watched three separate 26 minute film clips (with audio), containing small gaps of 0.5 to 30 seconds omitted from the film. Participants were then (around two minutes later) given an old/new recognition test in which they were presented with 100 static frames that had either been selected from the film clips (targets), or foil frames that were selected from the gaps that participants had not seen. In this way, the foil images were similar to the parts of the film participants had viewed, especially when taken from the shorter gaps. Accuracy, taking into account ‘hits’ (correctly identifying an old image) and ‘false alarms’ (incorrectly reporting to recognise a new, foil image), varied depending on the size of the gap the foil frames were taken from. On trials in which the foil frames were taken from the shorter gaps of 0.5 to 2.5 seconds - and therefore presumably more similar - the false alarm rate was actually higher than the hit rate; that is, participants reported to recognise the parts of the scenes they didn’t view more than the parts they did. However, overall accuracy reached 70% and was significantly above chance when foils were taken from the longer gap of 30 seconds.

Therefore, VLTM for dynamic scenes is reasonably detailed, and can persist after a delay of minutes, following viewing of many shorter scenes, or extended films containing different scenes and characters. However, there may be limitations to the fidelity of dynamic scene memory, in that discriminating highly similar foils from previously viewed parts of dynamic scenes may be difficult (Ferguson et al., 2017; Matthews et al., 2010).

Theories of scene memory such as Hollingworth and Henderson’s (2002; Hollingworth, 2007; 2009), offers a descriptive explanation of memory for scenes. In this account, VLTM is proposed to be involved in the process of scene viewing. As attention is directed to objects within a scene, detailed object representations, including the spatial layout, are held within object files in VLTM. VSTM is suggested to be limited in its capacity to the most recently attended visual items, or objects (Hollingworth, 2004). Therefore withdrawal of attention from an object results in the VSTM trace decaying quickly, whereas the detailed scene map remains within VLTM. Earlier theories suggest that object files can be updated with an object’s position within a scene, if the position of that object changes (Kahneman et al., 1992; also see Hollingworth, 2007). Although the fidelity of this long-term spatial map is lower than the current or most recently attended items in VSTM, the scene is still remembered in good detail, and only modest forgetting of visual detail may occur after the initial drop in accuracy (Hollingworth & Henderson, 2002; Hollingworth, 2005).

However, this does not readily account for dynamic scenes, where the presence and spatial position of objects within the scene changes across time. The question still remains of what part of a dynamic scene participants are most likely to remember, if any. This directly speaks to the question of fidelity. To test this, the current experiments specifically tested recognition accuracy for target frames taken from a previously viewed dynamic scene, dependant on their temporal location in the scene. It may be expected the latter parts of a dynamic scene are remembered when recalling or recognising a scene. Therefore a recency advantage in the current experiments, especially at extended retention intervals, may suggest a scene map that is ‘updated’, with the object files representing the object’s new locations within a scene (Hollingworth & Rasmussen, 2010; Kahneman et al., 1992).

In contrast, a high fidelity VLTM for dynamic scenes may contain the spatiotemporal detail of a scene, and all objects within a scene across time may be remembered (that is, what was where, and possibly when). In their research of memory for dynamic visual scenes, Matthews et al. (2007) suggested that a memory record for dynamic scenes may contain not only spatial, but spatiotemporal detail, building upon Hollingworth and Henderson's (2002) theory. This notion of a ‘dynamic memory map’, in which objects’ positions and trajectories across time are remembered, may expect comparable accuracy for all target images regardless of their temporal position.

4.2.1 The current study

The current study includes two experiments that aimed to test the fidelity of visual memory for dynamic scenes, to add to current understanding about how dynamic scenes might be remembered.

To do this, the purposefully-created dynamic scene stimuli described in Chapter Three were used here. This allowed for highly similar same scene foil images to be used in the recognition test. These foil images were taken from the same dynamic scene, but at a slightly different time point to the part of the dynamic scene that participants had viewed. Therefore, background scene elements and static objects were unchanged within the scene. The foil images here are more similar than same scene category foils used in the research with static scenes (e.g. Konkle et al., 2010a; also see Experiments

One and Two). They are also arguably more similar to the ‘similar’ foils (the same people in the same location) used in the research conducted by Matthews et al. (2010) on memory for dynamic scenes. To discriminate between the two images in the current recognition test, participants would have to recognise details of the moving objects within a scene, namely their trajectory or their various locations within the scene. 2-AFC tests were used in these experiments, with static foil and target images presented side by side, and participants had to discriminate between the two; an example of this similarity is given in Figure 4.1.

Figure 4.1. An example of two images (foil and target) used in the 2-AFC task. The foil is a same scene foil, meaning it is taken from the same scene as the target, and therefore highly similar.

Using this methodology of highly similar foils, the current research aims to test memory for dynamic scenes immediately, and after a delay. In the research described above (Ferguson et al., 2017; Konkle et al., 2010a; Matthews et al., 2010), there was a somewhat extended period of time between viewing and testing, as participants were tested after viewing many different scene images over hours, or viewed longer dynamic scene clips of up to 30 minutes. The memory representation of dynamic scenes may be more detailed if tested immediately, and therefore participants may be able to distinguish these highly similar foil images from previously viewed targets. In previous research utilising an immediate and a delayed memory test, typically lower performance is found at delayed test. This has been found to be the case for objects within static scenes (Hollingworth, 2005) and for stimuli depicting dynamic actions (Urgolites & Wood, 2013). The current experiments aim to test this finding with complex dynamic scene stimuli.

Furthermore, as described earlier, previous research does not address the question about what part of a dynamic scene participants are most likely to remember, if any.

To test memory based on temporal position, static frames were taken from previously viewed dynamic scene clips at various temporal locations, as described in Chapter Three. There may be a recognition accuracy pattern exhibited for parts of a dynamic scene dependant on temporal position. This is guided by a branch of research with static objects or items which has tested serial-position effects (Hollingworth, 2004; Hurlstone et al., 2014). In research testing serial order for visual memory, items are presented to participants in a specific order, and then memory for items is tested based on the order in which they were presented. Often what is observed is higher accuracy for items presented first (a ‘primacy’ effect), or items presented last, or most recently (a ‘recency’ effect) (Allen, Baddeley & Hitch 2006; Broadbent & Broadbent, 1981; Hollingworth, 2004; Hurslstone et al., 2014). Different mechanisms may account for such effects. For example, observed recency effects in visual memory may be indicative of a reduction in retroactive interference for that item, meaning that it is held in a more accessible state in memory (Allen et al., 2006). The research into serial position effects typically uses distinct, arbitrary individual items displayed to participants one at a time, in isolation. This is different to a more holistic dynamic scene, which contains rich dynamic visual detail, typically over a longer time period. It is an interesting question as to whether different mechanisms may apply when a dynamically changing scene is encountered.

An observed recency effect in these experiments may also be consistent with Hollingworth and Henderson’s (2002) theory accounting for visual memory for scenes. In this conception, object files are integrated into a ‘scene map’, and these object files may be updated upon subsequent scene viewing if the object has altered position (Hollingworth & Rasmussen, 2010; Kahneman et al., 1992). Therefore a recency advantage in the current experiments, especially at extended retention intervals, may suggest a spatial scene map that is ‘updated’.

On the other hand, an alternative proposition is a visual representation for dynamic scenes that contains detail about objects’ positions and trajectories across time; that is, what was where, and perhaps also when. This is in line with Matthews et al.'s (2007) suggestion that a memory record for dynamic scenes may contain not only spatial, but spatiotemporal detail, building upon Hollingworth and Henderson's (2002) theory. In this view, comparable accuracy for all target images regardless of their temporal position may be expected.

In summary, the current experiments therefore aimed to test visual memory for dynamic scenes over the initial time course after viewing. The focus of previous research using natural, dynamic scenes has been on testing VLTM after longer retention periods, rather than testing this question of fidelity in the immediate time frame after viewing a scene. In addition, the question of which part of a dynamic scene is remembered better, if any, needs to be further explored. The current set of experiments therefore tested two novel concepts. Firstly, testing for serial position effects of dynamic scene memory, and secondly, doing this for both the immediate timeframe after viewing and after a delay. In Experiment Five, an immediate test was used, and Experiment Six included an immediate and delayed test, to allow comparison. The overarching aim of the research is to understand in how much detail everyday visual scenes are retained in memory, and how much of this detail is forgotten across time.

Here, participants were first shown a short (10 second) film clip of a dynamic scene, and were then given a 2-AFC recognition test which included static target frames taken from different time points across the 10 second film clip, allowing serial position effects of dynamic scene memory to be explored across both experiments. Foil frames were taken from the same scene at a time point not shown as part of the film (described further below). Accuracy and RTs to make these judgements were recorded. Performance on this task was assessed when tested immediately only (Experiment Five) and at immediate compared to delayed test (Experiment Six).

4.3 Experiment Five

4.3.1 Overview

The aim of Experiment Five was to test the fidelity of visual memory for dynamic scenes in the immediate time frame after viewing (seconds later). Participants viewed short dynamic scene clips one at a time, and were then given a memory test for that scene immediately (seconds later). The research with static and dynamic stimuli described above, using longer periods of delay, has found memory performance in later recognition tasks is good (Konkle et al., 2010a; Matthews et al., 2010). This suggests that detailed visual memories would also be retained at shorter delay periods, when tested immediately. Here, we test the fidelity of immediate memory when using highly similar foil images. Furthermore, a specific aim was to compare recognition accuracy for targets taken from five different temporal locations across the ten second film clip. Therefore, participants’ ability to distinguish previously seen static scenes taken from differing time-points from within a dynamic scene from conceptually similar looking foils is tested.

To increase power, a multi-trial approach was adopted. For example, participants engaged in multiple trials, whereby each trial consisted of viewing a dynamic scene followed immediately by presentation of a 2-AFC test involving two static scene images. However, within this context other known memory effects associated with memory for multiple, conceptually similar stimuli may apply. In particular, as all of the scenes in this experiment are from conceptually similar scene categories (that is, outdoor scenes including streets, roads, etc), it is likely that there will be more interference among encountered stimuli present in the experiment, which can affect memory performance (Endress & Potter, 2014; Hartshorne, 2008; Konkle et al., 2010a, 2010b; Öztekin & McElree, 2007). Proactive interference from earlier visual memory trials in VSTM tasks can affect performance on subsequent trials (Jonides & Nee, 2006; Makovski & Jiang, 2008). In other words, interference from the scene in the most recently viewed trial may affect the subsequent trial. As memory accuracy for parts of the dynamic scene depending on their temporal position was being tested here, it is also possible that any interference effects from previous trials could disproportionately impact upon memory accuracy for beginning of the dynamic scene encountered within the following trial. To try and reduce any effects of this, two different lengths of intertrial interval (ITI), which is the amount of time between participants finishing a trial and starting the next one, were used in this experiment (two and ten seconds). The rationale for this was to inform experimental design, and to test if there were any transient carry-over effects from previous trials.

In different experimental designs, such as the recent-probe test (as described in Chapter Two), the ITI has been manipulated to assess the effects of longer periods of delay on the effects of proactive interference. Such research has found that the effects of proactive interference are reduced at longer ITIs (e.g. Mercer & Duffy, 2015). This suggests that interference effects may have a temporal gradient (in other words, reduce over time) (also see Souza & Oberauer, 2014). However it is worth noting that there have been mixed effects with longer ITI durations, and reductions in proactive interference effects have not always been observed (e.g. McKeown et al., 2014).

Different explanations may account for any observed differences in the effect of varying ITI (see Mercer & Duffy, 2015, for discussion). Decay based accounts of forgetting propose that forgetting occurs as a result of absolute time passing, in the absence of any rehearsal (Burgess & Hitch, 2006). It is unlikely participants would try to rehearse during the ITI in this experiment, as the memory test for the past scene has already been given. Under this view, longer ITIs should lead to increased forgetting of information on previous trials, thereby reducing the effects of proactive interference on subsequent trials. On the other hand, temporal distinctiveness accounts suggest that the likelihood of remembering something is related to its psychological distinctiveness in time (Brown, Neath & Chater, 2007; Ecker, Brown & Lewandowsky, 2015). In this view, the longer ITI may reduce the effects of proactive interference by making the current scene on any trial psychologically more distinct from the scene in the previous trial. Therefore, different accounts would predict longer ITIs to be associated with a reduction in proactive interference, meaning increased performance (higher accuracy) and/or faster responding compared to the short ITI condition. This may be particularly apparent for targets encountered earlier within the dynamic scenes viewed during each trial.

To summarise, this initial experiment aimed to test the fidelity of memory for dynamic visual scenes, and specifically test if accuracy differed depending on the temporal location within the scene that the target image had come from. Given the similarity in the scene stimuli used, two different ITI durations were used in this initial experiment as an additional condition of interest, to test if there were any transient carry-over effects from previous trials.

4.3.2 Method

4.3.2.1 Participants

Seventy participants were recruited for the study through the School of Psychology's online participant pool scheme and were offered a fee of £4 or participation credits for their time. A technical issue meant the data of one participant was lost from the analysis; resulting in 69 participants overall (female = 56, male = 13). The mean age of these participants was 21.86 years (SD = 4.58, range = 18 - 40). All participants reported normal or corrected-to-normal vision. Participants were randomly allocated to one of two ITI conditions at the beginning of the experiment (they were unaware of this), and the study took around 45 minutes to complete. For this and other experiments using the dynamic scene stimuli set, participants were not eligible to take part if they had previously taken part in other experiments that used the dynamic scene stimuli. That is, no participants took part in more than one experiment across Experiments Three to Seven. The study was granted ethical approval from the School of Psychology's ethics committee (ref: 16-0340).

4.3.2.2 Materials and Stimuli

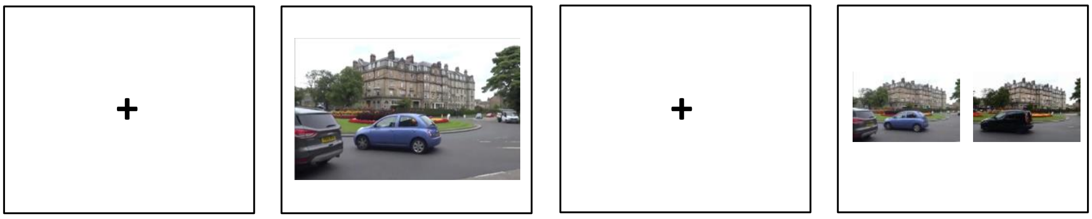
This experiment used the dynamic scene stimulus set described in Chapter Three. For Experiment Five, seventy scene stimuli were used, each comprised a dynamic 10 second film clip with seven associated test images. Five target images were taken from each film clip at one, three, five, seven and nine seconds. The earlier rating studies outlined in Chapter Three collected ratings on three of these targets at one, five and nine seconds, demonstrating that they didn't differ in ratings of attractiveness, complexity, distinctiveness or number of event boundaries across the ten seconds. In addition, two foil images were also taken from each scene: one from the three seconds before the 10 second clip was filmed, and one taken three seconds after. For each trial, one of the five target images and one of the foils were used in each 2-AFC. This is described further in the design and procedure section, below. Additional scenes were also used for the practice trials. The experiment was presented in PsychoPy (Peirce, 2007) on a 17 inch. monitor, and participants sat approximately 70cm from the screen.

4.3.2.3 Design and Procedure

A 5 x 2 mixed design had the within-groups factor of temporal position (five different targets taken from one, three, five, seven and nine seconds) and the between-groups factor of ITI (two or ten seconds). The dependent variables were recognition accuracy (measured as a proportion) and RTs (in ms). At the start of each session, the participant was randomly allocated to either the short or long ITI condition, with 35 participants taking part in each.

Participants first filled out an informed consent sheet, reporting their demographics. As part of this, participants were also asked to report if they were a driver. The scenes used in these experiments often contained moving vehicles. Research has shown that driving expertise can affect memory for visual road scenes (Blättler Ferrari, Didieriean, Van Elslande & Marmeche, 2010), and therefore this data was collected to test if this had any effect. A subsequent analysis showed that overall accuracy across all temporal positions did not significantly differ between participants who reported to drive compared to those that didn’t.

After completing the consent sheet, verbal and written instructions were then given to participants, and they had the opportunity to ask questions and take part in two practice trials before continuing with the main experiment. They were instructed that on each trial they were going to be shown a 10 second film clip followed by a fixation cross that was briefly presented for three seconds. As participants were aware that their memory for each scene was going to be tested immediately, this was intentional encoding. To reduce the likelihood of verbal encoding and rehearsal in visual memory research, an articulatory suppression task is often used, where participants are required to speak something out loud during the encoding and/or retention interval in an experiment (Hay, Smyth, Hitch & Horton, 2007; Hollingworth, 2005; Wheeler & Treisman, 2002). Here, participants were instructed to speak aloud the word ‘*la’* repeatedly whilst the film and fixation cross were on the screen (that is, during the retention interval). Asking participants to say a single word repeatedly is consistent with the approach taken in previous research (for example, in Saeki & Saito, 2004). After the fixation cross had left the screen, participants were presented with a 2-AFC task which included two images, one on the left and one on the right. It was the participant’s task to report which of the two images had been taken from the film they had seen by making a keypress using the keyboard (x = left, m = right). Their response and time to make the response was recorded, and participants were asked to respond as quickly and as accurately as possible. There was then a delay of either two or ten seconds (depending on the participant’s ITI condition) during which a fixation cross was displayed on the screen, before the next dynamic scene started. The procedure is illustrated in Figure 4.2. Participants were not given any feedback throughout the experiment.



ITI - Fixation Cross - 2 or 10 seconds

Dynamic Scene Clip – 10 seconds

Fixation Cross – 3 seconds

2-AFC

Figure 4.2. Schematic of one trial in the recognition test. Participants were either assigned to the two or 10 second ITI condition. The 2-AFC included a target image taken from the dynamic scene at a specific time point, alongside a same scene foil image.

There were 70 trials across three blocks (23 or 24 trials in each block), and participants were given the opportunity to take a break in between each block. The two independent variables of temporal position of the target and ITI created a 5 x 2 design. The ITI condition was a between-groups factor, however the temporal position of target was a within-groups factor, meaning each participant saw five temporal positions of the target. In addition, the allocation of foil images taken from either before or after the 10 second clip that participants viewed was also considered for counterbalancing, so that they were presented an equal number of times alongside each temporal position of the target. Therefore there were 10 trial types, each formed from one of these 10 different combinations (temporal position of target one to five, and foil image taken from before or after), with seven dynamic scenes presented in each of the ten condition combinations. The allocation of scenes to each of the ten combinations was fully counterbalanced, with each set of seven scenes occurring in each of the ten conditions. Therefore, multiples of ten participants created a counterbalanced design (70 participants were recruited for this study). The allocation of participants in each of the ITI conditions, and the position of each target image in the 2-AFC (left or right) was also counterbalanced (targets occurred equally in each position). The order of presentation of each scene within each block was randomised.

4.3.3 Results

4.3.3.1 Recognition Accuracy by ITI and temporal position of the target

Overall accuracy (as a proportion) across all trial types was 0.70. A one-sample t-test showed that this was significantly above chance performance of 0.5, $t(68) = 21.21$, $p < .001$. Accuracy scores for each temporal position of the target across the short and long ITI conditions are displayed in Figure 4.3. A 5 (temporal position of target) x 2 (ITI condition) mixed measures ANOVA was conducted. This revealed no significant main effect of temporal position, $F(4, 268) = .95$, $p = .43$, $\eta_p^2 = .01$, or ITI $F(1,67) = 3.19$, $p = .08$, $\eta_p^2 = 0.05$. There was also no significant interaction, $F(4, 268) = .07$, $p = .99$, $\eta_p^2 = .001$.

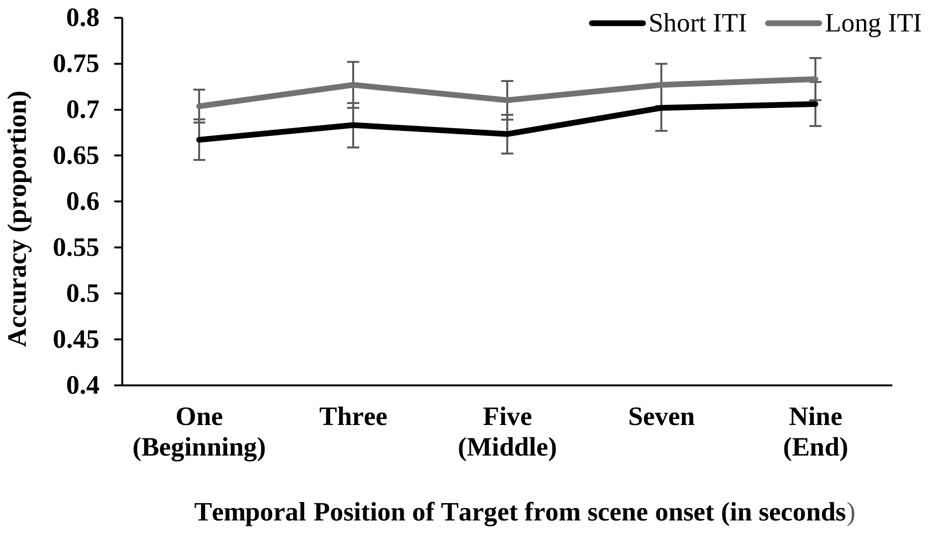


Figure 4.3. Accuracy for each target by temporal position from scene onset (in seconds) in the short and long ITI conditions.

RTs were first screened for outliers by identifying and removing RTs that were +/- 3SDs from the mean, for each participant within each condition (screening correct and incorrect trials separately)[5]. In this experiment, this resulted in 18 trials being removed across all participants and all trials. A preliminary analysis of RTs found that participants were faster on correct trials (M = 5212.34ms, SE = 353.41) compared to incorrect trials (M = 6343.44, SE = 492.37), $t(68) = 5.85$, $p < .001$, $d = .70$. RTs across each condition were calculated for correct trials only, and displayed in Table 4.1. A 5 x 2 repeated measures ANOVA was also conducted to compare these RTs. Again, there was no significant main effect of temporal position of target $F(3.31, 221.43) = 0.91$, $p = .44$, $\eta_p^2 = .01$ or ITI condition, $F(1,67) = 0.23$, $p = .63$, $\eta_p^2 = .003$. There was also no significant interaction, $F(3.31, 221.43) = 2.42$, $p = .06$, $\eta_p^2 = .04$. It is worth noting that these mean RTs are quite slow. This may be reflective of the difficulty of the task; the stimuli is complex and the participants are being asked to review and compare two highly similar images before making a decision.

---

[5] The same outlier screening procedure for RTs was used in Experiments Six and Seven.

Table 4.1. Mean RTs (SE) in ms for correct trials only, by temporal position of target and ITI condition.

	Temporal position from scene onset, in seconds					
	One	**Three**	**Five**	**Seven**	**Nine**	**Average**
Short ITI	5240.80 (611.34)	4847.67 (467.07)	5126.26 (535.11)	5232.90 (553.18)	4837.83 (522.39)	5057.09 (500.86)
Long ITI	5666.67 (620.27)	5534.95 (473.89)	5015.70 (542.92)	5009.19 (561.25)	5766.14 (530.02)	5398.53 (508.17)
Average	5453.74 (435.45)	5191.31 (332.69)	5070.98 (381.15)	5121.04 (394.02)	5301.99 (372..09)	

4.3.4 Discussion

Performance on the recognition test was high (70% on average), showing participants were able to distinguish between previously viewed frames and highly similar foil frames from the same scene on the majority of trials. In order to distinguish between foils and targets, participants must have attended to, and remembered the details of the moving objects within the scene, as the background elements and stationary objects remained unchanged between foils and targets.

There was no significant difference in accuracy or RTs dependent on the temporal position of the target, with no clear recency or primacy effects. This is inconsistent with what might be expected given the findings relating to primacy and recency effects in visual memory research, when using discrete individual items (Allen et al., 2006; Broadbent & Broadbent, 1981; Hollingworth, 2004, Hurslstone et al., 2014). In addition, Hollingworth (2004) found that accuracy in change blindness tests was higher for the most recently viewed objects within a scene. Therefore, we may have expected to observe a recency effect in this experiment.

However, when using much more detailed, full scene dynamic clips in this experiment, there was no difference in accuracy for the different temporal sections of the dynamic scene. These results however may be consistent with the notion of a highly detailed dynamic memory map, in which participants can accurately remember different specific parts of a previously viewed dynamic scene. This is a novel and important finding, and therefore subsequent experiments will still test memory for a range of temporal positions of the target image (taken from different positions across the scene).

As noted above, there have been mixed effects of varying ITI length on proactive interference effects in recent-probes experiments (McKeown et al., 2020; Mercer & Duffy, 2015). However, it was anticipated that performance may be superior at the longer ITI condition due to forgetting of information from the previous trial. The cause of such forgetting in VSTM is the topic of extensive and ongoing debate, particularly in relation to the roles of time-based decay, interference and temporal distinctness (Oberauer & Lewandowsky, 2008; Ricker et al., 2016; Souza & Oberauer, 2014). Notably, there was no significant difference between performance across both ITI conditions on accuracy or RT measures. Even with the shorter ITI of two seconds, accuracy was still high and significantly above chance performance. It may be the case that participants can still do the task well regardless of proactive interference, but the presence of interference slows down performance. However, the analysis of RTs showed that this wasn’t the case, as participants’ speed did not differ significantly when responding in the short ITI and long ITI conditions. Also of importance, is the finding that the longer ITI did not affect the accuracy for the different temporal positions of the target differently, as demonstrated by the non-significant interaction and the pattern of effects. This is of particular relevance to the first target, which may be most affected by proactive interference from previous trials. Therefore the results here do not provide a strong case for effects of proactive interference emerging in the present multi-trial paradigm. Thus, for methodological ease, the shorter ITI is used in the subsequent experiment.

Overall in this study, the fidelity of visual memory was still precise enough to distinguish information that has been viewed from same scene foil images. The results support the notion of a highly detailed memory trace for dynamic scenes over short periods of retention, up to three seconds. However it is not clear over what time period this high fidelity memory for dynamic scenes persists, and testing this was a key aim of Experiment Six.

4.4 Experiment Six

4.4.1 Overview

Experiment Six aimed to compare the fidelity of dynamic scene memory when tested immediately, and after a longer delay, building on the methodology and results described in Experiment Five. The main difference here was to introduce a delayed memory test, to allow comparison to the immediate test. In the current experiment, participants viewed a number of scenes and were either tested immediately, or not. Participants were later tested for all the scenes they had not been tested for immediately, and therefore the delay was up to around 20 minutes, and included many intervening scenes. The results from Experiment Five demonstrated that after the initial time course (when tested three seconds later) participants were able to distinguish previously viewed parts of a moving scene compared to highly-similar same scene foils. In the current experiment, the immediate condition aims to replicate the findings from Experiment Five, and verify the robustness of findings regarding temporal position of target.

Previous research has shown that there are limitations to the fidelity of visual memory for scenes, and forgetting of some detail can occur, especially when comparing performance in immediate and delayed testing conditions. This is often framed as a comparison between VSTM/working memory that is capacity-limited to a small number of recently viewed items, and VLTM (see Schurgin, 2018). Typically, what has been found is that when tested immediately, with no intervening trials, accuracy is higher than when tested after a much longer delay of minutes, hours and days (Hollingworth, 2005; Urgolites & Wood, 2013). For example, the fidelity of memory for objects when tested immediately, has been found to be superior to delayed test (Biderman, Luria, Teodorescu, Hajaj & Goshen-Gottstein, 2019; Schurgin & Flombaum, 2015). However, research has shown that under some conditions, memory fidelity for specific object details may be comparable (Brady, Konkle, Gill, Oliva & Alvarez, 2013b).

This forgetting across longer periods of delay might be more pronounced when the delays cover a period of days. Using static scenes, Hollingworth (2005) compared visual memory when tested immediately or after a delay, testing memory for objects within the scenes. Rotation changes (a previously viewed object rotated in position) or token changes (different object) were made to objects within the scenes. Across a number of experiments, participants were shown a set of scenes, and for half the scenes they were tested immediately (200ms) after viewing; whereas the test for the other half of the scenes was delayed by differing amounts. In one experiment, Hollingworth used a delay of 24 hours, and compared this to the immediate test. Performance on the task was significantly worse after 24 hours than when tested immediately. In a similar vein, forgetting of detailed information has been found to occur for object information over time periods up to a week (Andermane & Bowers, 2015). Therefore, there do appear to be limitations to the fidelity for memory of scenes within VLTM, in that longer delays lead to increased forgetting.

However, there is evidence to suggest that for delays of a few minutes (following intervening stimuli) forgetting may be more modest. In Hollingworth (2005), there was no significant difference in performance between immediate test and after a one-trial delay, across both change-type conditions (token change and rotation change). When the delayed test occurred at the end of the session (an average delay of just under ten minutes between viewing a scene and test), there was also no significant main effect of the time of test (that is, performance when tested immediately compared to the end of the session was comparable, overall). However, an interaction demonstrated that participants were significantly better at the rotation change detection task when tested immediately compared to the end of the session. There was no significant difference for the performance in the token change tasks. Therefore some object-specific detail may be forgotten over this longer period of delay.

In terms of dynamic visual stimuli, Urgolites and Wood (2013) compared the fidelity of visual memory for dynamic actions in an immediate and delayed task. They showed participants short clips of computer animated dynamic actions (for example, a person raising an arm) followed by a 10-AFC test in which all images showed the same person completing the same action but with a different range of motion (varying from a little motion to a large range of motion). Therefore to successfully select the previously viewed action, participants would need to have a very specific memory for the action. Participants were assigned to either an immediate or a delayed test condition. In the immediate testing condition, participants completed the 10-AFC test straight after viewing each dynamic action. In the delayed memory test, participants viewed all 12 dynamic actions before being tested. They found significantly lower accuracy in the delayed test compared to when tested immediately, but participants performed above chance in both conditions. The authors interpreted these findings in relation to the fidelity of working memory and VLTM, suggesting that we store detailed dynamic representations of previously viewed actions with “only slightly less detail” (pg. 409) in VLTM compared to working memory. The dynamic actions used in Urgolites and Wood’s (2013) study were quite simple, artificial in nature, and did not involve the motion of multiple objects, which is often the case in complex dynamic scenes. Therefore although informative, it is hard to generalise results. It is therefore difficult to predict how such findings might apply to memory for complex, dynamic scenes in the current experiment. Based on the research reviewed, it was expected that performance would be significantly lower at delayed test (following many minutes and intervening scenes) compared to when tested immediately (seconds later).

In summary, the current experiment expands upon the research conducted in Experiment Five by introducing a delayed test to allow understand of the fidelity of VLTM for dynamic scene memory, whilst contributing to the wider discussion on the fidelity of visual memory when tested over different periods of retention. Here, participants watched short dynamic film clips, and were given a 2-AFC recognition test for each scene either immediately after viewing the scene, or in a delayed test after viewing all the scenes (in a similar way to: Hollingworth, 2005; Urgolites & Wood, 2013). Accuracy for targets taken from three different temporal positions within the scene was also compared, in both the immediate and delayed test conditions.

4.4.2 Method

4.4.2.1 Participants

Forty-eight participants (Female = 43, Male = 5) were recruited for the study through the School of Psychology’s paid participant database and were offered a fee of £3 or participation credits for their time. The mean age of participants was 24.98 years (SD = 5.78, range = 18 - 40). All participants took part in all conditions, and the study took around 45 minutes to complete. All participants reported normal or corrected-to-normal colour vision. The study was granted ethical approval from the School of Psychology’s ethics committee (approval number: 16-0163).

4.4.2.2 Materials and Stimuli

Seventy-two of the scene stimuli from the same scene set described in Chapter Three were used for this Experiment. Each scene comprised a dynamic 10 second film clip with five associated images used at the test phase: three target images (taken from one, five, and nine seconds, creating a beginning, middle and end target),[6] and two foil images (one taken from three seconds before and one taken three seconds after the 10 second clip, as per Experiment Five). Additional scenes were also used for the practice trials. The experiment was presented in PsychoPy (Peirce, 2007), with the same computer monitor and set up as Experiment Five.

4.4.2.3 Design and Procedure

The two independent variables were temporal position of target (beginning, middle or end) and test time (immediate or delayed). Three targets were tested based on temporal position, rather than five as used in Experiment Five, as no significant difference based on target type was found. Using three target types therefore allowed for more data points per target condition, whilst still enabling a comparison of memory across the whole scene duration, namely the beginning, middle and the end.

The 3 x 2 design resulted in six different combinations of trial types. The position the foil image had been taken from (before or after) was also considered, resulting in 12 different combinations of trials for counterbalancing, with six scenes in each of the twelve conditions. The allocation of scenes to each of the 12 conditions was fully counterbalanced, with each set of six scenes occurring in each of the 12 conditions. Therefore multiples of 12 participants created a completely counterbalanced design (48 participants were recruited for this study). The position of each target image in the 2-AFC (left or right) was also counterbalanced (target and foils occurred equally in each position). Participants first completed a consent form, verbal and written instructions were given to participants, and then they took part in two practice trials.

The procedure was similar to that for Experiment Five. In the current experiment, participants viewed 72 dynamic film clips, and their memory was tested using 2-AFC

---

[6] The static target snapshots taken from one, five and nine seconds will be referred to as the beginning, middle and end targets throughout this experiment.

tasks. Following each ten second film clip, a three second fixation cross was shown on the screen. Participants were instructed to speak aloud the word '*la'* repeatedly ('*la la la...')* whilst the film and fixation cross were on the screen. On half the trials, this was followed by a 2-AFC task in which a target was presented next to a foil image. It was the participant's task to select the image they had previously seen in the film clip. However, memory for the other half (36/72) of the trials was tested later, and therefore instead of two images, a message was displayed on the screen reading 'no test, press x or m to continue'. On these trials participants pressed any key and a two second fixation cross was displayed before the next 10 second scene started. This ITI is the same used in the short ITI condition in Experiment Five. An illustration of this is given in Figure 4.4.

Figure 4.4. Schematic of the immediate test. After each scene, participants were either given a 2-AFC (A) or were not tested for that scene (B).

Once participants had watched and responded to all 72 scenes, they were given a two minute break. This was followed by the delayed test for the 36 scenes that were not tested for immediately. Participants were presented with the 36 2-AFC tests sequentially and made a response to each one, in the same way in which they had done for the immediately tested trials. Across all trials, feedback was not given to participants. The order of presentation of each scene within both the immediate and the delayed block was randomised, yielding a range of different delays between viewing the scene and test in the delayed condition (ranging from 0 to 106 scenes).

4.4.3 Results

4.4.3.1 Overall recognition accuracy and response times

Overall accuracy (as a proportion) was 0.60 across all trial types, the breakdown of which is displayed in Figure 4.5. A one-sample t-test showed this was significantly above chance (0.5), $t(47) = 10.11$, $p < .001$. A 3 (temporal position of target) x 2 (test time) repeated measures ANOVA comparing accuracy scores revealed a main effect of test time, $F(1, 47) = 47.58$, $p < .001$, $\eta_p^2 = .50$, as participants were significantly more accurate at immediate (M = 0.66, SE = 0.02) compared to delayed (M = 0.54, SE = 0.01) test. The main effect of temporal position of target $F(2, 94) = 1.80$, $p = .17$, $\eta_p^2 = .04$ and interaction, $F(2, 94) = 2.44$, $p = .09$, $\eta_p^2 = .05$, were not significant. An additional one-sample t-test was conducted comparing performance in the delayed test to chance, and performance was significantly above chance, $t(47) = 3.42$, $p = .001$.

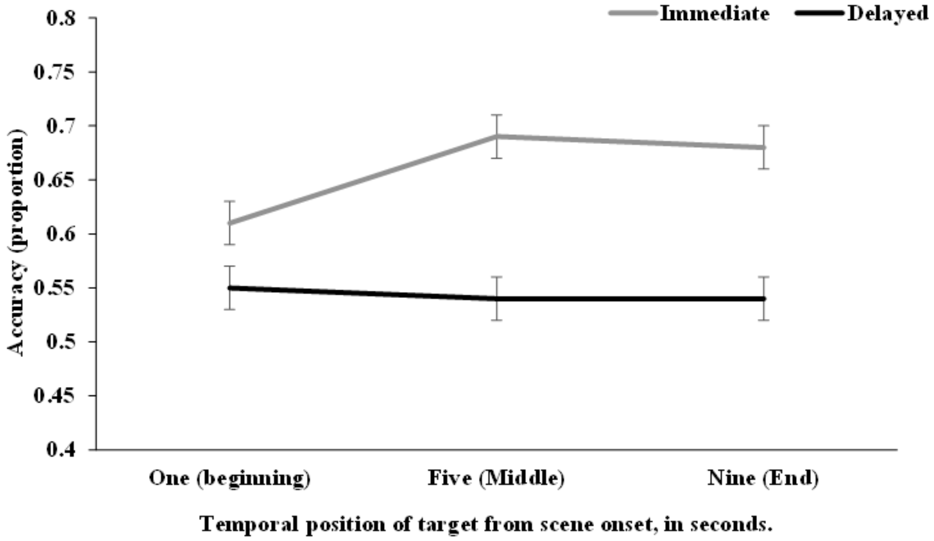


Figure 4.5. Accuracy for each target by temporal position from scene onset (in seconds) in the immediate and delayed test.

RTs were screened for outliers, but there were no outliers found. A repeated measures t-test compared average RTs across all correct and incorrect trials, demonstrating that participants were significantly faster to respond on correct (M = 4636.36ms, SE = 293.97) compared to incorrect (M = 5032.75ms, SE = 330.17) trials, $t(47) = 4.53$, $p < .001$, $d = .65$. This was also the case when comparing correct and incorrect trials at immediate test separately *t*(47) = 4.13, $p < .001$, $d = .60$, but not when comparing correct and incorrect trials at the delayed test *t*(47) = 1.88, $p = .07$, $d = .27$.

A 3 (temporal position of target) x 2 (test time) repeated measures ANOVA was also conducted on RTs for correct trials only (see Table 4.2). This also revealed a main effect of test time, *F*(1,47) = 9.93, $p = .003$, $\eta_p^2 = .17$, as participants were significantly faster at the immediate (M = 4226.96ms SE = 250.21) compared to the delayed (M = 5061.84ms, SE = 385.29) test. There was a significant interaction, *F*(1.57,73.72) = 3.78, $p = .04$, $\eta_p^2 = .07$. Follow-up t-tests showed that participants were significantly faster at immediate test compared to the delayed test when responding to the beginning target, t(47) = 2.47, $p = .017$, $d = .36$, and end target *t*(47) = 3.46, $p = .001$, $d = .50$. The difference for the middle target was not significant, *t*(47) = 2.01, $p = .05$, $d = .29$, when adjusting the *p* value used for multiple comparisons (.05/3 = .017). The main effect of temporal position of the target *F*(2, 94) = 0.62, $p = .54$, $\eta_p^2 = .01$ was not significant.

Table 4.2. Mean RTs (SE) in ms for correct trials only, by the temporal position of the target and test time condition.

	Temporal Position of Target			
	Beginning	**Middle**	**End**	**Average**
Immediate	4444.35 (295.66)	4271.27 (296.30)	3965.26 (219.72)	4226.96 (250.21)
Delayed	5056.03 (341.83)	4879.02 (432.25)	5250.49 (449.51)	5061.84 (385.29)
Average	4750.19 (294.52)	4575.14 (338.20)	4607.88 (301.22)	

4.4.3.2 Recognition accuracy by number of intervening scenes

The order of scene presentation in both the immediate and delayed block was randomised, meaning that there were a range of possible delays (from two minutes to approximately twenty minutes), and therefore intervening scenes (from 0 to 106). For example, if the last scene participants saw in the immediate testing block was the first scene they were then tested on as part of the delayed test, this would mean that there were 0 intervening scenes for this scene for that participant, and a delay period of only two minutes for the break. To test if there was a change in accuracy depending on the number of intervening scenes, accuracy in the delayed test was compared depending on the number of intervening scenes (similarly to Hollingworth, 2005). The number of intervening scenes were calculated into groups. Proportional accuracy was calculated for each participant for intervening scenes (including scenes from both the immediate and delayed test) of 0-20, 21-40, 41-60, 61-80 and 81+ (Table 4.3). A one-way repeated measures ANOVA to compare accuracy across these different number of intervening trials found no significant main effect, $F(3.18, 139.89) = 0.09$, $p = .99$, $\eta_p^2 = .002$.

Table 4.3. Average accuracy (as a proportion) in the delayed test, by number of intervening scenes.

Number of Intervening Scenes	Average Accuracy (SE)
0-20	0.52 (.05)
21-40	0.52 (.03)
41-60	0.53 (.03)
61-80	0.54 (.03)
81+	0.51 (.04)

4.4.4 Discussion

Overall, participants performed above chance at both time points, demonstrating a detailed visual memory trace for dynamic scenes. There was little difference in memory accuracy for the different temporal positions of the target image, across immediate and delayed tests in Experiment Six, which is consistent with the results found in Experiment Five.

Good fidelity for long-term dynamic scene memory was also observed in this study, with performance also significantly above chance at this longer period of delay. However, performance was significantly lower in the delayed test condition (when tested two to 20 minutes later) compared to the immediate test condition. The analysis of RTs was also in line with this, as participants were not only less accurate at delayed test, but also slower to make correct recognition judgements. However, an interaction demonstrated this immediate versus delayed difference in RTs was only significant when responding to the end target.

This pattern of results is similar to those of Urgolites and Wood (2013), in their study comparing memory for dynamic actions when tested immediately or after a delay. Here, they also found that accuracy was above chance in both the immediate and delayed test, but significantly lower when tested after a longer delay. Although the pattern of results are similar, the current experiment used much more complex stimuli, in the form of natural, everyday dynamic scenes. Participants also saw many more scenes (12 were viewed and tested for in Urgolites & Wood). Despite this, performance was still significantly above chance at delayed test, although it was not very high, at 54% on average.

Given the randomisation of scene presentation, a varying number of scenes occurred in the delayed condition. More intervening scenes would increase possible interference, in addition to the time delay between scene viewing and test. Analysis of the results showed that the number of intervening scenes (and therefore delay) did not impact overall accuracy. This is similar to previous research that has tested for the effect of intervening scenes. For example, in Hollingworth's (2005) study using static scene stimuli, memory for objects within the scene was tested immediately, and after a delay, following viewing of all scene stimuli (in a similar way to this experiment). Hollingworth analysed the results to test if there was a relationship between the number of intervening scenes in the delayed condition and accuracy on the change detection task, finding that there was a negative but not a significant correlation for either test condition (rotation or token changes).

4.5 General Discussion

Two reported experiments tested the fidelity of visual memory for dynamic scenes, using a novel dynamic scene stimulus set. Across both experiments, a high detailed memory for dynamic scenes was demonstrated, with participants able to discriminate previously viewed target frames from highly similar foil frames taken from the same scene just seconds before or after the dynamic scene clip was filmed.

Previous research into the fidelity of dynamic scene memory has demonstrated that participants are able to discriminate between previously viewed scene images and images from the same scene category, with accuracy above 75% when tested after a delay of at least 20 minutes (Konkle et al., 2010a). In the present research, however, foil frames were potentially more similar, and contained the same background elements and static objects; the foil images here were taken from three seconds before or after the 10 second clip the participant's viewed. Therefore, it is likely the location of moving objects within a scene needed to be represented in memory to ensure correct recognition of target frames.

Experiment Six also included a comparison of memory performance at an immediate and delayed task. The observed results suggest comparable fidelity of dynamic scene memory is not maintained from the seconds after viewing over longer periods of delay spanning many minutes, as performance was significantly higher when tested immediately compared to after a delay. The lower accuracy after a delayed test is very similar to previous research that has attempted to compare the fidelity of visual memory when tested immediately and after a delay using less complex stimuli (Schurgin & Flombaum, 2015; Urgolites & Wood, 2013).

Despite this, performance (accuracy) was still above chance when tested after a delay of many minutes and intervening scenes, and participants could still distinguish between target and foil images. Research has shown that VLTM recognition accuracy for objects and scenes was lower when more objects/scenes from the same object/scene category had been previously viewed (Konkle et al., 2010a; 2010b). However, the intervening scene analysis suggests that when tested after a delay of two minutes minimum (the break between viewing the last dynamic scene and the delayed test block), performance did not appear to drop with increasing time (here, up to 20 minutes) and intervening trials. These results suggest there may be forgetting of visual detail soon after viewing (sometime between the initial seconds and two minutes following the end of scene viewing). After this time period the fidelity of the memory trace is maintained to a greater extent. The time frame over which this initial forgetting occurs for dynamic scenes is investigated further in the following chapter.

Further, the results from the two experiments reported here suggest that the fidelity of visual memory for dynamic scenes is not negatively affected to a great extent by proactive interference from previous trials. Despite the similarity in the scene stimuli. accuracy in Experiment Five was not significantly different across different durations of ITI. There was also no interaction, suggesting the shorter, compared to the longer, ITI period between stimuli did not negatively affect memory for the beginning part of the subsequent scene.

In an attempt to understand any serial position effects in visual memory for dynamic scenes, both experiments here specifically aimed to compare accuracy for targets taken from varying temporal positions within the dynamic scene clip previously viewed. There were no significant differences in accuracy depending on temporal position of the target image, however accuracy was above chance for all targets, regardless of temporal position. The similar accuracy across the temporal position of the target in the current experiments, may be consistent with the concept of a dynamic memory map, in which the memory trace for dynamic visual scenes holds detail of what objects were where, across different time points (that is, spatiotemporal detail). The dynamic scene stimuli is very different to the static scene images or objects used in previous research of serial order, and may to some extent account for the difference in the pattern of results.

To summarise, the two experiments reported here used novel methodology to demonstrate detailed memory accuracy for highly detailed moving scene images. Accuracy remained above chance in all conditions. This was despite viewing at least 70 dynamic scenes from the same scene category. Performance was above chance, enduring after a delayed test. Accuracy was comparable and above chance for all target frames taken from across the time course of the dynamic scenes (beginning, middle and end); there was no clear evidence for serial position effects, when tested immediately or after a delay.

CHAPTER FIVE

THE FIDELITY OF VISUAL MEMORY FOR DYNAMIC SCENES – PART TWO

5.1 Chapter Summary

This chapter reports one experiment that further explores the fidelity of visual recognition memory for dynamic scenes over the initial, immediate time course after scene viewing. Here, participants viewed brief film clips depicting everyday dynamic scenes. Each scene was followed by an old/new recognition task which included a series of static frames that were either taken from different time points across the dynamic scene, or highly similar foil frames taken from the same scene at a slightly different time point to the one participants viewed originally. The recognition task occurred either two, four or eight seconds after initial scene viewing.

In line with the results from Experiments Five and Six, recognition accuracy was high and above chance, demonstrating a highly detailed visual memory for dynamic scenes. This was despite the use of a different recognition paradigm, the similarity of target and foil images used in the recognition task, and the similarity of scenes across trials. Recognition accuracy (taking into account hits and false alarms) decreased at longer delay intervals, but remained above chance when tested after eight seconds. This demonstrates forgetting of visual details occurs over this initial time frame.

The hit rate for each temporal position of the target image was also analysed, and it was found that participants were more accurate in identifying the targets taken from the middle and end compared to the beginning target. This finding that latter parts of a recently viewed scene appeared to be recognised with higher accuracy indicates there may be some recency effects in dynamic scene memory in the initial time frame after scene viewing.

5.2 Introduction

The two experiments reported in Chapter Four demonstrated the high fidelity of visual memory for dynamic scenes. Participants were able to discriminate target images taken from previously viewed scenes compared to highly similar foil images from the same scene, at different time points. Although accuracy was still above chance at a delayed test.

The results from Experiment Six demonstrated that participants had significantly lower recognition accuracy after a delay (of at least two minutes up to 20 minutes) compared to when tested immediately (three seconds after viewing the scene). Accuracy did not decrease as the number of intervening scenes and time between viewing and test increased. Therefore, in the study of visual memory for dynamic scenes in Experiment Six, it appears that there may be some forgetting of visual detail soon after scene viewing (some point between the initial seconds up to two minutes), but after this time period the fidelity of the memory trace is maintained. The aim of Experiment Seven therefore was to further test the fidelity of immediate visual memory for dynamic scenes, over the retention interval of two to eight seconds after viewing.

Various retention intervals have been used to investigate visual memory for scenes. Static scene experiments have used durations of 200ms to one trial (30 seconds on average) (Hollingworth, 2005). Other experiments have tested the fidelity of longer-term visual scene memory, with delays of minutes (Ferguson et al., 2017) or days (Experiments One and Two, this thesis). Typically, what has been found is when tested immediately, with no intervening trials, accuracy is higher than when tested after a much longer delay of minutes, hours and days (Hollingworth, 2005; Urgolites & Wood, 2013). This was mirrored in results of Experiment Six, in which accuracy was higher after being tested immediately compared to after a delay. However, these are quite different comparisons in which there are multiple scenes that intervene between the encoding and test in the delayed test condition, but not when tested immediately.

In the current experiment, participants were shown a dynamic scene, followed by a series of six images (three target images and three foils), one at a time. They had to respond to each image to indicate if it was taken from the part of the dynamic scene that they had viewed, or not. Therefore the number of intervening scene images here was limited, and ranged from 0 to five images from the same scene. To allow for a test of memory with no intervening scenes, this was specifically considered as part of the analysis; that is, accuracy was calculated for the first image of the six on each trial, in addition to the overall average.

In addition, it is not clear at what rate forgetting occurred in the retention interval in Experiment Six. In the current experiment, accuracy was tested after delays of two, four and eight seconds, thereby testing the fidelity of memory in the immediate time period after viewing, across a range of delays.

An old/new recognition task was used in this experiment, to test if the high performance observed in the 2-AFC test in Experiments Five and Six is replicated across different experimental paradigms. This type of recognition test has been used in previous studies of memory for dynamic visual scenes over longer time periods of retention (Buratto et al., 2009; Ferguson et al., 2017). This differs to the 2-AFC test methodology used in Experiments Five and Six, in which participants were presented with two images at a time, and had to decide which of the two they had seen previously. There was also only one 2-AFC test per scene in Experiment Five and Six, but here participants were presented with all three temporal positions of the target (separately) as part of the recognition test after each scene.

There are a number of reasons why it is interesting to use this different recognition paradigm. Experiment Five and Six found minimal evidence for differences in the fidelity of memory for the different temporal positions. The method used here (in Experiment Seven) allows for assessment of temporal position patterns from within the same trial, for each scene, to further investigate any potential differences in positioning effects. In addition, it allows the study of false alarm rates to foils that stand alone, when not presented alongside a previously viewed target. In previous research using similar foil images when testing recognition memory for dynamic scenes, high false alarm rates have been found, especially when the foils were similar.

For example, Matthews et al. (2010) used an old/new recognition test. Across three experiments, participants viewed short (three second) dynamic scenes or static images taken from the films, before taking part in the recognition test. When the foils were most similar to the previously viewed scenes, higher false alarm rates, in which participants incorrectly report a new image as old, were observed in dynamic scene trials compared to static. Therefore motion may improve recognition of previously viewed scenes (hits), but also perhaps make it harder to distinguish similar foils to targets at test. Likewise, Ferguson et al. (2017) conducted a series of experiments in which they showed participants short film clips. This was followed by recognition tests using static frames either taken from the film that participants saw (targets), or foils that were taken from the same film clip, but not a part the participant had seen. When the foils were most similar, the false alarm rate was actually higher than the hit rate. Therefore, again, accuracy for dynamic scenes in recognition tests is impressive, but when foils are highly similar to previously viewed targets, the false alarm rate can also be high. The current methodology will allow exploration of responses to stand alone foil images, and how this affects overall performance and accuracy on the recognition test. Given that accuracy when tested immediately was found to be higher than delayed test in Experiment Six, we may expect that participants would have a high fidelity memory trace for scenes after short periods of delay, and possibly also have a high false alarm rate to foil images, reducing overall recognition accuracy.

It is also possible there may be different processes underlying different recognition judgements. For example, Cunningham, Yassa and Egeth (2015) suggested that different processes may underlie performance in old/new compared to AFC tasks, whereby AFC tasks may rely more on familiarity-based judgements, whereas old/new recognition tasks rely more on recollection. Cunningham et al. tested recognition memory for objects, comparing performance in both types of recognition test. Foil images in the recognition test varied in their similarity to the target object, and were either from a novel object category, the same object category, or the same object in a different state. It was found that accuracy was higher when responding to AFC compared to old/new trials. The authors suggest a reason for this is that when making an AFC judgement, participants can either reject the foil as they do not recognise it, or accept the target as they do recognise it, meaning more information is available to help make that judgement. Conversely, when responding to an image in an old/new recognition task, participants only have one of these pieces of information (also see Schurgin, 2018, for discussion). Experiments Five and Six established that participants were capable of performing in the 2-AFC test above chance. It is therefore interesting to understand if performance is still high and above chance when responding to dynamic scene images in an old/new recognition test, when the foils are highly similar (taken from the same scene), and presented separately.

The basic paradigm used in this experiment was as follows: on each trial, participants watched a 10 second film clip showing a dynamic everyday scene. After a brief delay - of two, four or eight seconds - they were given an old/new recognition task in which they were presented with six images sequentially, to which they had to respond via a key press to report if each image was taken from the film clip they just viewed or not. The six images comprised three target frames (from the beginning, middle and end, as was used in Experiment Six), and three foil frames. The three foil frames were taken from the same scene video, but at a different time point to the 10 seconds shown to participants. The main aim of the experiment was to understand the fidelity of memory for dynamic scenes over these different periods of retention. A subsidiary aim, given that the nature of the paradigm allows, was to compare performance when responding to targets taken from different temporal positions in a previously viewed dynamic scene. This will also allow a test of whether a similar pattern of results is found when using an old/new test compared to a 2-AFC task (as was used in Experiment Five and Six).

5.3 Method

5.3.1 Participants

Thirty-one participants were recruited for the study through the School of Psychology's online participant pool scheme and were offered a fee of £5 or participation credits for their time. The data of two participants was incomplete due to a failure of data collection (the output was not saved), and therefore they were excluded from the analysis. This left 29 participants (female = 24, male = 5), with a mean age of 22.1 years (SD = 3.82, range 18-31. All participants reported normal or corrected-to-normal vision. Participants took part in all conditions, and the study took around one hour to complete. The study was granted ethical approval from the School of Psychology's ethics committee (ref number: 15-0222).

5.3.2 Materials and Scene Stimuli

This experiment used the dynamic scene stimulus set described in Chapter Three. Sixty scene stimuli were used, each comprised a dynamic 10 second film clip with six associated images: three target images taken from each film clip at one, five and nine seconds, creating a beginning, middle and end target, respectively[7] (as used in Experiment Six), and three foil images. The foil images were taken from the remaining 12 seconds of the 22 second clips (as described in Chapter Three). For half the scenes, one foil image was taken from before and two after; for the other half, two foils were taken from the eight seconds before, and one from after.

Additional scenes were also used for the practice trials. The experiment was presented in PsychoPy (Peirce, 2007) on a 17 inch. monitor, and participants sat approximately 70cm from the screen.

5.3.3 Design and Procedure

A 3 x 3 repeated measures design had factors of temporal position of target (beginning, middle, end) and delay (two, four, eight seconds). The dependent variables were recognition accuracy, and RT. Accuracy was analysed as both corrected recognition accuracy (taking into account ‘hit’ and ‘false alarm’ rate for all trials), and also ‘hit’ rate, when responding to the different target images only. Each trial consisted of a 10 second film clip followed by a fixation cross lasting two, four or eight seconds. This was followed by six images shown in succession for 2.5 seconds each, separated by a 0.5 second fixation cross. Participants were asked to make an old/new recognition judgment for each image using the keypad, to report if they thought the image was taken from the specific clip of scene that they viewed. Their response and time to make the response was recorded, and participants were asked to respond as quickly and as accurately as possible, and no feedback was given. There were 60 trials across three blocks (20 trials in each block), and participants were given the opportunity to take a break in between each block. An example of one trial in the task is given in Figure 5.1.

---

[7] The static target images taken from one, five and nine seconds will be referred to as the beginning, middle and end targets throughout this chapter, to avoid confusion of the different periods of delay used in this experiment.

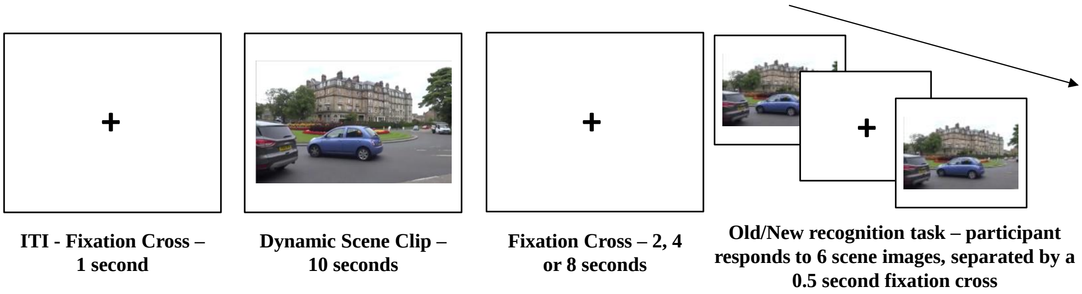


Figure 5.1. Schematic of one trial in the recognition task. Participants viewed and responded to six images for each dynamic scene clip.

All six image types (three foils and three temporal positions of target) occurred equally often within each of the six presentation positions during the test phase. The number of trials with each of the three different delay periods was separated equally between the three blocks (six, seven, seven, times in each block). The order of trials within each block and the order of block presentation were counterbalanced between participants. Before starting the experiment, all participants provided informed consent, received written and verbal instructions, and took part in two practice trials before beginning the test trials. They were then fully debriefed upon completion of the study.

5.4 Results

5.4.1 Overview of Analysis

For each trial, there was six images to which participants had to respond, three of these were ‘old’ target images (that participants had seen before), and three were ‘new’ foil images, taken from the same scene at a different time point. Responses to each of these images were taken, and classed as: ‘hits’ (correctly identifying an old image as old), ‘misses’ (incorrectly identifying an old image as new), ‘correct rejections’ (correctly identifying a new image as new) and ‘false alarms’ (incorrectly identifying a new image as old). Where there was no response made to a test image (within the 2.5 seconds opportunity), trials were excluded; this only happened on a small amount of trials (2.02%).

Corrected recognition was then calculated as a measure of accuracy on each trial. This was calculated by taking the proportion of hits minus the proportion of false alarms for each participant in each condition. Temporal position of the target images was not relevant to this particular analysis.

In addition to corrected recognition, the hit rate to respond to each temporal position of the target was then analysed. Although this no longer considers performance on the foil images, it does allow for a comparison between performance when responding to the beginning, middle, and end targets.

5.4.2 Accuracy and response times, by delay

5.4.2.1 Overall Accuracy

Accuracy for each image type was calculated as a proportion for each participant. Average hits and false alarms across all participants were calculated and displayed in Table 5.1.

Table 5.1. Average proportion of hits and false alarms during old/new recognition test, across all delay periods

Hits (target images)	0.71
False Alarms (foil images)	0.34

5.4.2.2 Corrected Recognition

Corrected recognition was 0.37 (SE = .02) on average across all participants and all conditions. A one-sampled t-tested found that this was significantly above chance performance (0), $t(28) = 15.29$, $p < .001$.

Mean corrected recognition scores were compared for each delay condition through a repeated measures ANOVA, and displayed in Figure 5.2[8]. There was a significant main effect of delay, $F(2,56) = 20.16$, $p < .001$, $\eta_p^2 = .42$. Follow up repeated measures t-tests were conducted to compare performance in each delay condition. For the post-hoc comparisons here and throughout, the p value used was adjusted for Bonferroni multiple comparisons (.05/3 = .017). There was no significant difference between the two (M = 0.41, SE = 0.03) and four (M = 0.41, SE = 0.03) second delay, $t(28) = 0.16$, $p = .88$, $d = .03$. However, performance after an eight second delay (M = 0.30, SE = 0.03) was lower than after both two seconds and four seconds; $t(28) = 5.76$, $p < .001$,

---

[8] An alternative measure used in the literature is d’. For completeness, this was also analysed and the same pattern of results was found in terms of significant differences across conditions. This data is reported in Appendix C.

d = 1.07, and *t*(28) = 6.06, *p* < .001, *d* = 1.13, respectively. Performance after eight seconds was still above chance, t(28) = 11.20, *p* < .001.

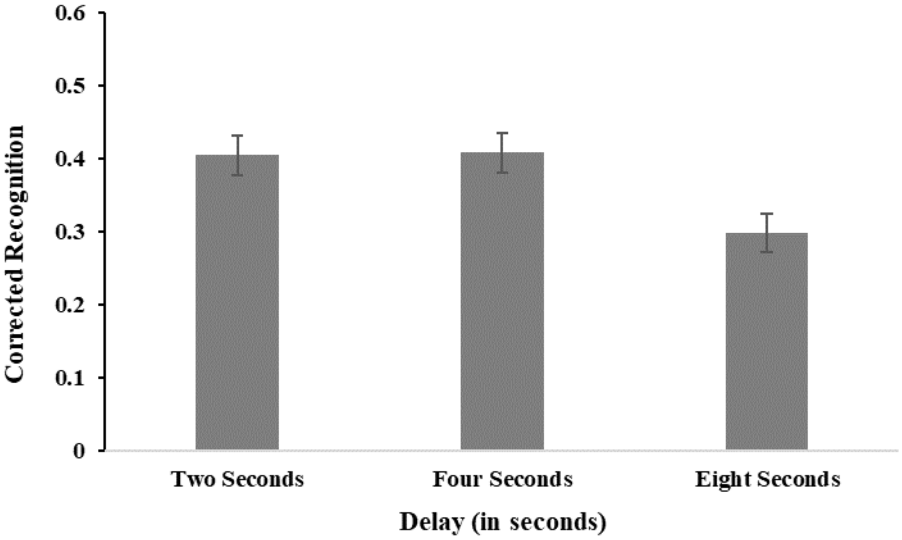


Figure 5.2. Average corrected recognition across all participants for delays of two, four and eight seconds between viewing a scene and test. Error bars show SE.

The above analysis takes into account all six images (three target and three foil images) on each trial, and therefore the delay between viewing the dynamic scene and the test was variable and longer on average than the two, four and eight second delay when responding to only the first image on each trial (as examined in this analysis). Therefore participants' corrected accuracy when responding to the first images on each trial was calculated as displayed in Figure 5.3. A one-way repeated-measures ANOVA with the three levels of delay was conducted. It was found that there was a significant main effect of delay, *F*(2,56) = 10.76, *p* < .001, η_p^2 = 0.28. Follow up repeated measures t-tests showed that accuracy at the two seconds delay (M = 0.48, SE = 0.05) was significantly better than both the four seconds (M = 0.27, SE = 0.05), *t*(28) = 3.83, *p* = .001, *d* = .71, and eight seconds delay (M = 0.27, SE = 0.04), *t*(28) = 3.96, *p* < .001, *d* = .74. The difference between four and eight seconds was not significant, *t*(28) = .07, *p* = .95, *d* = .01.

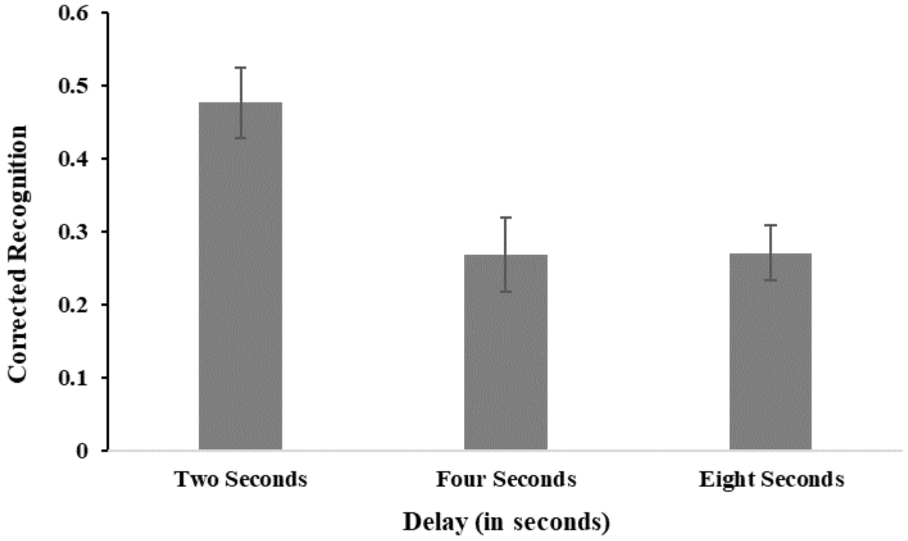


Figure 5.3. Average corrected recognition across all participants - for the first image only - for delays of two, four and eight seconds between viewing a scene and test. Error bars show SE.

5.4.2.3 Response times

RTs were screened for outliers; this resulted in the removal of 54 (0.53%) of all the trials. Average RTs to respond on incorrect (miss and false alarm responses) and correct (hits and correct rejections) trials were calculated for each participant and compared in a repeated-measures t-test. Participants were faster on trials in which they were correct (M = 1081.48ms, SE = 33.17) compared to incorrect (M = 1135.85ms, SE = 38.87), $t(28) = 5.23$, $p < .001$, $d = .97$.

Participants' average RTs to respond to images at each delay period was compared through a one-way ANOVA for correct trials only. This found that there was no significant main effect of delay period on RTs, $F(2,56) = .92$, $p = .40$, $\eta_p^2 = .03$. RTs were similar on trials that included a two second (M = 1080.27ms, SE = 36.50), four second (M = 1070.63ms, SE = 32.37) and eight second (M = 1088.59ms, SE = 34.45) delay.

5.4.3 'Hit' Rate and response times for target images only

The average hit rate (accurately identifying target images as previously seen) as a proportion across all participants was 0.71 (SE = 0.02).

A repeated-measures ANOVA comparing hit rate for each temporal position of the target (beginning, middle and end) found a significant main effect of temporal position of target, $F(2, 56) = 16.97$, $p < .001$, $\eta_p^2 = .38$ (see Figure 5.4). Follow up repeated measures t-tests were conducted to compare performance for each temporal position of the target. Participants were more accurate in identifying the middle (M = 0.75, SE = 0.02) and end targets (M = 0.72, SE = 0.02) compared to the beginning frame (M = 0.66, SE = 0.03); $t(28) = 6.02$, $p < .001$, $d = 1.12$ and $t(28) = 3.31$, $p = .003$, $d = .61$, respectively. The difference between the middle and end targets was not significant, $t(28) = 2.20$, $p = .036$, $d = .41$, when adjusting for multiple comparisons.

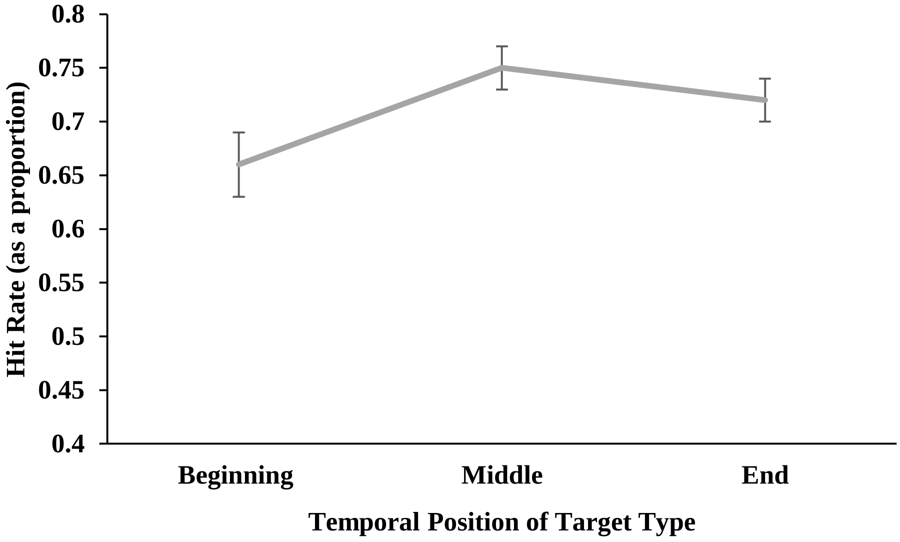


Figure 5.4 Average hit rate (as a proportion) by temporal position of the target. Error bars show standard error.

RTs to respond correctly to previously viewed targets was compared in a one-way repeated measures ANOVA. This found that there was no difference in RTs when responding to the beginning (M = 1063.09ms, SE = 33.73), middle (M = 1050.38ms, SE = 36.09) and end targets (M = 1046.99ms, SE = 33.04), $F(2,56) = 0.85$, $p = .43$, $\eta_p^2 = .03$.

5.5 Discussion

Using an old/new recognition task to test the fidelity of visual memory for dynamic scenes, the results of Experiment Seven demonstrated a high fidelity memory for dynamic visual scenes. Accuracy, when correcting for false alarms, was significantly above chance.

This high fidelity recognition memory for dynamic scenes was also found in Experiment Five and Six, in both immediate and delayed tests using a 2-AFC test. The old/new paradigm used in the current experiment differs in that foil images are not presented alongside target images, which may make the task more difficult (Cunningham et al., 2015). Considering the data in this and the previous experiments, it suggests that performance overall was more accurate in the 2-AFC test used in Experiments Five (average accuracy was 70% after three seconds) and Experiment Six (average accuracy was 66% at immediate test, after three seconds). In the current experiment, corrected recognition for the first probe only was 48% (at 2 seconds) and 27% (at four seconds). Previous research testing dynamic scene memory has also used old/new tests, finding higher false alarm rates when foil images are more similar to the target image (Ferguson et al., 2017). Here, despite foil images being very similar, participants performed well and above chance.

The delay between the end of the dynamic scene clip and the onset of the first test image ranged from two to eight seconds. Accuracy on the old/new recognition task was found to decrease over the longer delay periods, whereby accuracy was lowest after an eight second delay. When analysing the first encountered image only, there was evidence that performance may be detrimentally affected relatively early, between two and four seconds. Accuracy was highest after two seconds, but lower and comparable at four and eight seconds after viewing. Together, the results show a decrease in accuracy of visual memory over the immediate time course after viewing, within eight seconds. This is consistent with other research that has found forgetting on VSTM tests over increased retention intervals (e.g. Mercer & Barker, 2020), and the current experiment extends these findings to dynamic scenes. Future research may want to further investigate if, and at what rate, forgetting continues to occur at longer delay periods. Performance may stabilise after delays longer than eight seconds, or further forgetting may continue with increased time.

Accuracy when responding to target images from different temporal positions was also analysed. Performance for the middle and end targets was higher than for the target frame taken from the beginning position. This pattern is somewhat reminiscent of the familiar recency effects within the visual and verbal memory literature, without the primacy gradient present (Allen et al., 2006; Hurlstone et al., 2014). These results are therefore to some extent in line with the notion of an updating ‘scene map’, and the decay of attended items in memory when attention is directed towards new objects, when tested immediately after viewing (Henderson & Hollingworth, 2002; Hollingworth, 2004). However, a clear recency effect was not demonstrated, and actually accuracy was descriptively higher for the middle target, albeit not significantly.

These differences in target type position were not observed in the previous Experiments (Experiments Five and Six). It is not clear why this might be the case, however there are some possibilities. More data points were collected here on each trial, as for each scene participants completed a recognition judgement for all three temporal positions of the target. The recognition judgement required for an old/new test is also thought to be different, in that it relies more on recollective memory (Cunningham et al., 2015; Schurgin, 2018). Forgetting may occur for the beginning of the film clips, but more retrieval cues are available in the 2-AFC to compensate for this forgetting, whereas for the old/new recognition test this is not the case. This may have led to reduced likelihood of successful recognition for the beginning target when presented in isolation (in the old/new test), rather than in an 2-AFC.

To summarise, Experiment Seven tested visual recognition memory for dynamic scenes in the immediate time frame after viewing. Despite using an old/new task, participants accuracy remained high. However, accuracy decreased over the initial eight seconds, demonstrating some forgetting of scene detail over this timeframe. The pattern of effects suggest forgetting may occur more often for the visual details at the beginning of the scene (that is, the beginning target in this experiment).

CHAPTER SIX

GENERAL DISCUSSION

6.1 Overview

The aim of the current thesis was to explore the fidelity of visual memory for scenes. The novel emphasis here was on memory for everyday, naturalistic scenes, that are viewed under encoding and testing conditions that are more likely to approximate the use of scene memory in everyday life.

There has been a focus in the scene memory literature on understanding visual memory for simple objects and static scenes. Some of this research has attempted to address the question of the fidelity of visual scene memory, over different periods of retention. Studies have compared performance in immediate and delayed tests, and interpreted these findings in relation to fidelity of visual short-term (or working) memory, compared to VLTM. Generally, the fidelity of memory has been found to be superior when tested immediately compared to a delay of many minutes (or more) and intervening items. This has been found to be the case for simple static items, and some dynamic stimuli (Hollingworth, 2005; Urgolites & Wood, 2013). Despite any forgetting that may occur between immediate and delayed testing, detailed long-term memory for many objects and scenes has been observed after retention periods of minutes, hours or days (Andermane & Bowers, 2015; Konkle et al., 2010a).

Hollingworth and Henderson conducted a series of studies in which participants viewed static scene images, and then their memory for specific objects within each scene was tested (Hollingworth, 2004, 2005; Hollingworth & Henderson, 2002). They proposed a theoretical approach to account for their findings, which outlines a role of both VSTM and VLTM during scene viewing (Hollingworth & Henderson, 2002). In this conception, the most recently attended items within each scene are held in higher fidelity within VSTM. However, memory for objects (in ‘object files’) previously attended to, form part of a coherent scene map in VLTM, allowing for accurate performance on later recognition tests.

Past research has often been conducted under intentional encoding conditions (whereby participants attempt to remember a scene for a later memory test). Recent research has also tested memory under incidental encoding conditions. For example, Castelhano and Henderson (2005) compared object memory following intentional and incidental scene viewing and found comparable performance. However, the testing condition here was still intentional, whereby participants attempted to remember the previously viewed scene at test, which is less comparable to the way in which we re-encounter scenes in every day.

In order to understand memory for naturalistic scenes, research has also focused on testing long-term memory for dynamic scenes over the period of minutes. This research has demonstrated that following dynamic scene viewing, performance on later recognition tests is significantly above chance, and recognition performance is superior to that for static scenes. However, false alarm rates to new foils in the recognition test have been found to increase when the foils are more similar to the target items (Ferguson et al., 2017; Matthews et al., 2010). It is unclear how, and in what fidelity, dynamic natural scenes are represented in memory, and how this changes (if at all) over different periods of retention.

To address these questions, and to attempt to understand the fidelity of the representations underlying visual scene memory for everyday scenes, seven experiments were conducted and reported in this thesis. A number of different experimental paradigms were used and developed to test memory for static and dynamic scenes under different conditions. Across the multiple experiments, participants were able to distinguish previously viewed scenes from highly similar foil images. Recognition performance remained above chance at longer delays, despite viewing multiple similar scenes at encoding. This finding generalised across the use of implicit and explicit encoding conditions.

6.2 Summary of findings, implications and future research

Experiment One and Two of this thesis tested incidental scene memory, and developed a method to enable an incidental test in addition to incidental encoding. A recent-probes test, a test typically used within the study of VSTM, was adapted to study the effects of VLTM for previously encountered scenes on a later, separate task. In these experiments, participants first viewed images of static scenes from different scene categories and completed an irrelevant task, in which they rated the scenes for their attractiveness. Participants completed a subsequent test minutes later (Experiment One) or in a second session either five or 10 days later (Experiment One and Two), in which they took part in the adapted version of the recent-probes test. Here, participants were shown a series of static scene images before being shown a sixth image, to which they had to report if that sixth image was one of the five just viewed. On trials in which the sixth image was not one of the five, participants should respond no. On some trials (RN trials), a scene image participants had viewed and rated earlier was used; participants should still reject this image as it was not one of the five just viewed. Accuracy and RTs were compared when responding to the novel images (NRN trials) and recently viewed images (RN trials). The advantage here was that participants were not made aware that their memory for the scenes viewed earlier was being tested. Across both experiments, RTs to correctly reject a RN probe were faster than when correctly rejecting NRN probes. This is the opposite effect to what is normally observed in recent-probes experimental paradigms investigating VSTM; accuracy is typically lower and/or RTs slower when responding to RN compared to NRN probes (Craig et al., 2013; McKeown et al., 2014; 2020). The opposite effect in the current experiments may have been due to a facilitatory effect. Indeed, recent research has demonstrated that participants perform with higher accuracy on a current task if it features object images that they have viewed previously (Oberauer et al., 2017). These findings are also in line with research demonstrating that VLTM for a previously viewed scene can guide subsequent visual search within that scene (Hollingworth, 2009; Kruijne & Meeter, 2016; Wolfe & Horowitz, 2017; Woodman & Chun, 2006). Taken together, these results suggest that incidentally-acquired VLTM memories for scenes are present up to five, and even 10 days later, and may support current and ongoing perception of previously encountered scenes. A further implication is that incidental visual memory for scenes can be probed using incidental tests, thereby presenting an option for researchers interested in understanding incidental scene memory.

The recent-probes methodology was supplemented by a 2-AFC recognition task and associated remember/know/guess responses, which was taken after the recent-probes test, in both experiments. This builds upon past research in which participants have been able to distinguish previously viewed scenes from other scenes from the same scene category. For example, in research by Konkle et al. (2010a), participants viewed scenes from different scene categories (between four and 64 scenes from each scene category), followed by a later recognition task 20 minutes after viewing all scenes. They found that accuracy was at 76% for trials in which participants attempted to recognise scenes in which they had previously viewed 64 scenes (the upper amount) from that same scene category in the initial encoding phase. Although the number of scenes viewed from each scene category was at the lower end in the reported experiments in this thesis (four in the initial exposure phase), the retention interval was much longer (five or 10 days), and incidental encoding had occurred. Despite this, accuracy was still high and above chance across two experiments. Furthermore, Experiment One also demonstrated that accuracy didn't significantly differ when tested after a delay of 10 days compared to five days. This expands on previous research by Andermane and Bowers (2015), who tested memory for objects after delays of up to seven days, showing that object recognition memory was still above chance, even when placed next to an object from the same object category, or the same object in a different state (for example, a cup half full rather than full) in a later recognition task.

The results from the additional remember/know/guess judgement made after each 2-AFC trial showed that on the trials in which participants were correct, this was more likely to be associated with reporting to remember (that is, have a vivid memory of seeing the scene image before), rather than having familiarity towards one image, or a lucky guess (this pattern occurred across both experiments, but the difference was not significant in Experiment One). It is important to highlight this was the case despite participants previously viewing the scene images incidentally. Conversely, inaccurate trials were linked most often to guesses, suggesting that participants did not recognise either image at all. In other words, the inaccurate response was less likely to be driven by a false memory or feeling of remembering or knowing that they viewed the foil image in the previous session. These findings help to understand successful recognition judgements in tests in which very similar foil images are used.

Using similar recognition test methodology, visual memory for dynamic scenes was then tested. To enable this, a dynamic scene stimuli set was created comprising 80 dynamic scenes. The aim here was to create a stimulus set that resembled the stimuli used in studies of static scene memory, avoided use of a central actor, whilst being naturalistic and depicting real-life scenes (Experiments Three and Four). Through three experiments (Experiments Five to Seven), visual recognition memory for these dynamic scenes was then tested. Both 2-AFC (Experiments Five and Six) and old/new (Experiment Seven) recognition tests were used. In each of these, highly similar same scene foil images were used, in which the foil images were taken from the same scene seconds before or after the segment of dynamic scene that participants viewed. Therefore, to successfully discriminate target and foil images, participants must rely on memory for moving objects and their position within the scene. Despite the difficulty of this task, detailed memory was observed overall, as participants could distinguish between same scene foil images after various retention intervals, across the different tasks. However, forgetting of scene information occurred over short and longer periods of retention.

In Experiment Five, when participants were given the 2-AFC immediately (three seconds) after viewing a 10 second dynamic scene, accuracy was 70% on average, and significantly above chance. In Experiment Six, the immediate test finding of Experiment Five was replicated, and performance was compared at immediate test to performance in a delayed test. It was found that accuracy was significantly lower when tested after a delay (following viewing of all the scenes) compared to when tested immediately (three seconds later) (54% compared to 66%, respectively). This is consistent with other research with static scene stimuli comparing performance in an immediate and delayed test (Hollingworth, 2005; Urgolites & Wood, 2013). Thus, despite the finding of the dynamic superiority effect, in which accuracy is highest for dynamic compared to static scenes (Buratto et al., 2009; Matthews et al., 2007; Matthews et al., 2010), accuracy for dynamic scenes is still significantly reduced after longer delays of minutes (and many intervening scenes) compared to the immediate time frame after dynamic scene viewing.

RTs were also collected and analysed. Generally, across Experiments Five to Seven, participants were faster on correct compared to incorrect trials. Furthermore, there was a difference between retention interval conditions in Experiment Six, in that participants were faster to respond at the immediate compared to the delayed test, for the beginning and end target. Therefore the forgetting across this timeframe from immediate to delayed test, not only affected later recognition accuracy, but also the time to make accurate judgements.

These results also demonstrate that despite the overall findings that VLTM for scenes is detailed, it is not perfect, and subject to forgetting over extended delay periods. However, it was not clear what happens to the fidelity of dynamic scene memory in the timeframe in between immediate and delayed test (in Experiment Six). Furthermore, the issue here was that the tests are quite different. In the immediate testing condition, participants are given a blank retention interval before their memory for that scene is tested. In the delayed testing condition, participants view many scenes before being tested for the dynamic scene (therefore both time has passed, and the number of intervening scenes). An intervening scene analysis was conducted in Experiment Six, to see if the accuracy pattern at the delayed test differed depending on the number of intervening scenes participants had viewed. It was found that accuracy did not differ dependant on the number of intervening scenes (and as a result, time), suggesting that accuracy may decrease to a greater extent over the initial timeframe (of around two minutes), but less so at longer time frames.

Experiment Seven therefore explored this forgetting over the initial time frame after scene viewing, using an old/new recognition test. It was found that accuracy decreased over the initial few seconds, and was lowest after the longest retention interval of eight seconds. This data supports a pattern of forgetting that occurs very early on in the retention interval.

Across all the dynamic scene experiments (Experiment Five to Seven), accuracy was above chance in all tests on average and participants were able to do the task under different conditions. In addition, accuracy was still above chance at a delayed test (Experiment Six), showing that despite time passing and intervening scenes, accuracy was still high, and actually may be unaffected by the number of intervening scenes. The careful selection of very similar foils here (see Chapter Three) provided a strong test of the fidelity of dynamic scene memory, and the outcome was very impressive. The results across these experiments suggest that participants are able to distinguish frames from previously viewed dynamic scenes, when compared to same scene foils. This ability was demonstrated across two recognition tasks. It has been suggested that AFC tests may be easier than old/new recognition trials, as in AFC tests, participants have a foil image present to help discriminate between the foil and target (Cunningham et al., 2015). The recognition accuracy data across the experiments are in line with this view, as accuracy was higher in the 2-AFC tests in Experiment Five and Six compared to the correct recognition accuracy in the old/new test in Experiment Seven. Despite this, average accuracy in Experiment Seven which used an old/new recognition test was still above chance.

A seminal account of visual memory for scenes, put forward by Hollingworth and Henderson (2002), suggests that VLTM is involved during the process of scene viewing, and representations are held within VLTM. In this view, only the most recently fixated objects are held within VSTM; research by Hollingworth (2004) estimated that this was limited to two objects. Hollingworth (2005) conducted research comparing memory for object details, following static scene viewing, at immediate (200ms) and delayed test (varying lengths of delay, the earliest occurring after a one-trial delay). They predicted comparable, or only slight decreases in levels of change detection performance with increasing delay periods, resulting from the assumption that all but the most recently attended two items in a viewed scene should be held within VLTM. On this basis, the authors suggested that, even in the immediate test condition, memory would most likely rely on VLTM, unless the change happened to be made to the two objects most recently fixated, and should therefore be comparable with the delayed tests, particularly when tested after one-trial only (approximately 30 seconds later). It was found that there was no significant difference in performance between immediate test and after a one-trial delay, across both change conditions tested (a condition where the change involved rotating an object, or a condition where the change involved replacing the object with another from the same object category).

There are some differences between the immediate test conditions in the experiments reported in this thesis and that used by Hollingworth (2005). In Hollingworth's research, a delay period of 200ms was used for immediate test, whereas three seconds was used in Experiment Five and Six, and a minimum of two seconds in Experiment Seven. Furthermore, dynamic, rather than static scenes were used. To enable accuracy in the current experiments, participants had to remember the identity and location of moving objects within a scene, as that was the only way to distinguish previously viewed frames from unseen, same scene foils. In dynamic scenes, if moving objects are attended to more in scenes, it might be that VSTM is involved for longer, depending on the number of moving objects within the scene. Indeed, motion within a scene has been shown to be highly effective in capturing viewers' attention, as shown through research tracking eye movements (Carmi & Itti, 2006; Mital, Smith, Hill & Henderson, 2011). Therefore, it is possible that both VSTM and VLTM may be recruited and relied upon in the immediate test conditions in Experiments Five to Seven. Thus, within the current experiments, the initial drop in memory performance over the first few seconds of retention (in Experiment Seven) may be due to both the partial loss of information occurring from VSTM as time passes and attention is redirected to more recent aspects of the scene, as well as due to the high demands that are likely placed immediately on VLTM due to the complexity of the unfolding scene.

In previous studies, researchers have attempted to look at the contributions of VSTM through shorter delays and VLTM through longer delays, possibly with intervening items (Brady et al., 2013b; Urgolites & Wood, 2013). It may be interesting to look at the fidelity of dynamic scene memory over the shorter time frame of milliseconds, to understand forgetting over the more immediate timeframe. Another way to study VSTM is to look at the focus of attention (e.g. Hollingworth, 2004). The general aim of the experiments reported in this thesis was to increase understanding of the fidelity of dynamic scene memory across the specific time course, not by item. It cannot be concluded from these experiments in which order objects in a dynamic scene were fixated, and where in the sequence. To understand this with dynamic natural stimuli, monitoring of eye fixations during dynamic scene viewing could be used. Alternatively, an on-screen dot to cue attention to certain objects within a scene may also be appropriate here. For example, Hollingworth (2004) used this type of methodology is his study of scene memory, in which a dot was used over scene stimuli to direct observers’ attention. This allowed comparison of memory for objects that had been viewed most recently, compared to those viewed further back. Alternatively, using more tightly controlled dynamic scene stimuli in future may provide further insight into a more item-based understanding of what is attended to and when, and how this affects subsequent memory performance. These type of methods may also be useful in allowing for a comparison of memory for static and dynamic elements of a dynamic scene (also see Matthews et al., 2010, for discussion). In the current experiments, accurate performance must have relied on memory for the identity and location of moving objects within the dynamic scenes, but this may have been at the expense of memory for the static aspects of a scene.

To address the question of fidelity, in addition to assessing recognition performance over different retention intervals, there was also a focus in the dynamic scene experiments on the pattern of accuracy when responding to target frames dependant on their temporal positioning within the dynamic scene clip. This also links directly to the question of how a dynamic scene might be represented in memory. Targets were taken from different time points across the 10 second clips (there was a focus on the beginning, middle and end). A rating study was conducted, as reported in Chapter Three, to ensure that these different time points of the scene stimuli didn't vary on factors such as number of event boundaries, or scene attractiveness, distinctiveness or complexity. Across the three experiments (Five to Seven) using the dynamic scene stimuli, limited evidence was found of effects based on the temporal position that the target frames had been taken from.

Based on serial position research (Allen et al., 2006; Hurlstone et al., 2014), it may have been expected that recency effects would have been observed when memory was tested immediately. This would also be in line with Hollingworth's (2004) earlier research, in which the objects most recently attended to within a static scene were later remembered with higher fidelity than objects viewed four objects back or more. A clear recency effect was not consistently observed across these experiments at the immediate testing. In experiments Five and the immediate test in Experiment Six, there was no significant difference in accuracy for the different target types. However, in Experiment Seven, where a different paradigm was used – an old/new test, in addition to testing for all targets on each trial - participants were more accurate in identifying the middle and end targets compared to the beginning. It is possible that less retrieval support in the old/new (compared to the AFC) test, was at the detriment of recognition accuracy for the beginning target.

Furthermore, at longer retention intervals (of minutes), superior accuracy for a specific part of a previously viewed dynamic scene (such as the beginning or end) may be suggestive of a spatial-only map in VLTM. For example, high accuracy for the end target may have been indicative of an updated scene map, with objects in the final position that they were viewed (Hollingworth & Henderson, 2002; Kahneman et al., 1992). Over the longer duration used in the delayed test condition in Experiment Six (of up to 20 minutes), accuracy was comparable across all target types, regardless of temporal location. This suggests that the VLTM trace is detailed enough to hold information of which objects were where within the viewed scene. This gives some support to the conception of VLTM for scenes that is spatiotemporal, rather than spatial only, as here, participants could presumably remember the spatial location of various objects within a scene across the 10 second time frame. Matthews et al. (2007) suggested a possible expanding of Hollingworth and Henderson's (2002) conceptual model, of a memory trace that holds spatiotemporal information. Matthew et al.'s suggested that in this conception, "the viewer establishes a record not only of where a fixated object is, but also when it is there, and quite possibly where it is going" (Matthews et al., 2007, p. 992). The data collected in this thesis lends itself to a demonstration of memory with detail of 'what' objects were 'where', as accuracy in the task relied on this knowledge. The question of whether such a memory map includes detail of 'when' across a timeframe is less clear and could be tested in future research. Future research may aim to test specifically if the fidelity of memory for dynamic scenes is capable of remembering the order that visual events occurred within dynamic scenes. For example, Kwok and Macaluso (2015) showed participants short film clips, and then at test, participants were presented with two images from the previously viewed scene and asked to select the image that came first. Therefore different experimental paradigms, potentially similar to the one used by Kwok and Macaluso, may be developed to test memory for 'when' visual events occurred. This would allow further understanding of the fidelity of dynamic scene representations in memory.

Whether such a spatiotemporal map in VLTM is dynamic in nature cannot be concluded from the current data. However, there is some evidence in the literature that suggests it may be dynamic. For example, the finding of representational momentum, in which people tend to misremember a moving object as further along it its direction of travel than was observed, suggests a dynamic representation (Freyd & Finke, 1984; Hubbard, 2005; Hubbard & Bharucha, 1988; also see Matthews et al., 2007, for discussion). The dynamic superiority effect for scenes, and associated congruency effect, is also indicate that the representation of scenes in visual memory may be dynamic in nature (Buratto et al., 2009; Matthews et al., 2007; Matthews et al., 2010).

It is worth noting that static images were used in the recognition tasks testing dynamic scene memory. Previous research has demonstrated the so called dynamic superiority effect, and associated congruency effect, in which accuracy is higher when dynamic scenes are used in a subsequent recognition test, following dynamic scene viewing (Buratto et al., 2009; Matthews et al., 2007). Therefore, overall accuracy may have been higher across all conditions (in Experiments Five to Seven) if dynamic clips had been used in the recognition tests. Instead, static images were used in the current study to specifically test participants' memory for 'snapshots' of different time points throughout a previously viewed dynamic scene. This is in line with other research testing visual memory for dynamic scenes (Ferguson et al., 2017). A general uplift in accuracy, especially at the delayed test (in Experiment Six), may have also been observed if scenes from different scene categories had been used. For example, Konkle et al. (2010a) found an uplift in recognition performance if participants had previously viewed fewer scenes from the same scene category.

To summarise, across the experiments reported in this thesis, evidence for detailed visual memory representations was observed throughout, using different paradigms. After viewing complex dynamic scene stimuli, forgetting of visual detail may start to happen quickly (within seconds), but robust visual representations are still available minutes (and many intervening scenes) later. Generally, these detailed representations are available for all parts of a dynamic scene (across temporal locations), especially when tested at longer delays of minutes later. Therefore this doesn't suggest a VLTM spatial map which is fixed (for example updated in the final position), but instead a map which contains details of what objects were where. It may be possible to advance previous theoretical perspectives such as Hollingworth and Henderson's (2002), to not only describe a VLTM scene map for spatial locations of objects within scenes, but memory for spatiotemporal detail (Matthews et al., 2007). Future research could use the methodologies developed in this thesis, and others, to further understand this representation of memory for natural, everyday scenes.

6.3 Conclusion

We view many different scenes every day. The research described in this thesis focused on understanding the fidelity with which real life naturalistic scenes are remembered. A key contribution of this work is the development of methods to understand visual memory for natural scenes. The findings from the experiments reported here suggest that various parts of a short dynamic scene can be remembered from across the time viewed, indicating that VLTM for dynamic scenes may be spatiotemporal in nature. Overall, the experiments reported in this thesis demonstrate the fidelity with which natural, everyday scenes are remembered. This is the case even under circumstances of incidental encoding and retrieval, varying periods of delay between encoding and test, and when scene stimuli are static or dynamic.

REFERENCES

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135*(2), 298-313. https://doi.org/10.1037/0096-3445.135.2.298

Andermane, N., & Bowers, J. S. (2015). Detailed and gist-like visual memories are forgotten at similar rates over the course of a week. *Psychonomic Bulletin and Review, 22*(5), 1358-1363. https://doi.org/10.3758/s13423-015-0800-0

Atkins, A. S., Berman, M. G., Reuter-Lorenz, P. A., Lewis, R. L., & Jonides, J. (2011). Resolving semantic and proactive interference in memory over the short-term. *Memory and Cognition, 39*(5), 806-817. https://doi.org/10.3758/s13421-011-0072-5

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839. https://doi.org/10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation - Advances in Research and Theory, 8*, 47-89. https://doi.org/10.1016/S0079-7421(08)60452-1

Berman, M. G., Jonides, J., & Lewis, R. L. (2009). In search of decay in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(2), 317. https://doi.org/10.1037/a0014873

Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science, 30*(2), 223-237. https://doi.org/10.1177/0956797618813538

Blättler, C., Ferrari, V., Didierjean, A., & Marmèche, E. (2012). Role of expertise and action in motion extrapolation from real road scenes. *Visual Cognition, 20*(8), 988-1001. https://doi.org/10.1080/13506285.2012.716799

Blättler, C., Ferrari, V., Didierjean, A., van Elslande, P., & Marmèche, E. (2010). Can expertise modulate representational momentum? *Visual Cognition, 18*(9), 1253-1273. https://doi.org/10.1080/13506281003737119

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America, 105*(38), 14325-14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013a). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General*, *142*(3), 791-808. https://doi.org/10.1037/a0029649

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013b). Visual Long-Term Memory Has the Same Limit on Fidelity as Visual Working Memory. *Psychological Science, 24*(6), 981-990. https://doi.org/10.1177/0956797612465439

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects. *Communicative & Integrative Biology, 2*(1), 1-3. https://doi.org/10.1073/pnas.0803390105.www.landesbioscience.com

Broadbent, D. E., & Broadbent, M. H. P. (1981). Recency effects in visual memory. The *Quarterly Journal of Experimental Psychology Section A, 33*(1), 1-15. https://doi.org/10.1080/14640748108400762

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: evidence from eye movements. *Quarterly Journal of Experimental Psychology 59*(7), 1177–1187. https://doi.org/10.1080/17470210600665996

Brown, G. D. A., Neath, I., & Chater, N. (2007). A Temporal Ratio Model of Memory. *Psychological Review, 114*(3) 539-576. https://doi.org/10.1037/0033-295X.114.3.539

Buratto, L. G., Matthews, W. J., & Lamberts, K. (2009). When are moving images remembered better? Study-test congruence and the dynamic superiority effect. *Quarterly Journal of Experimental Psychology, 62*(10), 1896–1903. https://doi.org/10.1080/17470210902883263

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55(*4), 627-652. https://doi.org/10.1016/j.jml.2006.08.005

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*(26), 4333–4345. https://doi.org/10.1016/j.visres.2006.08.019

Castelhano, M., & Henderson, J. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017–1040. https://doi.org/10.1080/13506280444000634

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-114. https://doi.org/10.1017/S0140525X01003922

Cowan, Nelson. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science, 19*(1), 51-57. https://doi.org/10.1177/0963721409359277

Craig, K. S., Berman, M. G., Jonides, J., & Lustig, C. (2013). Escaping the recent past: Which stimulus dimensions influence proactive interference? *Memory and Cognition, 41*(5), 650-670. https://doi.org/10.3758/s13421-012-0287-0

Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning and Memory, 22*(11), 563-566. https://doi.org/10.1101/lm.039404.115

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*(10), 28. https://doi.org/10.1167/10.10.28

Ecker, U. K. H., Brown, G. D. A., & Lewandowsky, S. (2015). Memory Without Consolidation: Temporal Distinctiveness Explains Retroactive Interference. *Cognitive Science*, *39*(7), 1570-1593. https://doi.org/10.1111/cogs.12214

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548-565. https://doi.org/10.1037/a0033934

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review, 12*(6), 1127-1133. https://doi.org/10.3758/BF03206454

Ferguson, R., Homa, D., & Ellis, D. (2017). Memory for temporally dynamic scenes. *Quarterly Journal of Experimental Psychology, 70*(7), 1197-1210. https://doi.org/10.1080/17470218.2016.1174721

Field, A, (2009). Discovering statistics using SPSS (3rd ed.). Sage Publications.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 126-132. https://doi.org/10.1037/0278-7393.10.1.126

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of Remembering, Knowing, and Guessing. *Consciousness and Cognition, 7*(1), 1-26. https://doi.org/10.1006/ccog.1997.0321

Gordon, R. D., & Irwin, D. E. (1996). What’s in an object file? Evidence from priming studies. *Perception and Psychophysics, 58*(8), 1260-1277. https://doi.org/10.3758/BF03207558

Hartshorne, J. K. (2008). Visual Working Memory Capacity and Proactive Interference. *PLoS one, 3*(7). https://doi.org/10.1371/journal.pone.0002716

Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: A SIMPLE explanation? *Memory & Cognition, 35*(1), 176–190. https://doi.org/10.3758/BF03195953

Hayes, A. E., & Freyd, J. J. (2002). Representational momentum when attention is divided. *Visual Cognition, 9*(1-2), 8-27. https://doi.org/10.1080/13506280143000296

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language, 55*(4), 495-514.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*(1), 243-271. https://doi.org/10.1146/annurev.psych.50.1.243

Hirose, Y., Kennedy, A., & Tatler, B. W. (2010). Perception and memory across viewpoint changes in moving images. *Journal of Vision, 10 (4), 1-20.* https://doi.org/10.1167/10.4.2

Hollingworth, A. (2004). Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 519–537. https://doi.org/10.1037/0096-1523.30.3.519

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. https://doi.org/10.1037/0278-7393.31.3.396

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance, 33(*1), 31–47. https://doi.org/10.1037/0096-1523.33.1.31

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17*(1–2), 273–291. https://doi.org/10.1080/13506280802193367

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception & Performance, 28*(1), 113–136. https://doi.org/10.1037/0096-1523.28.1.113

Hollingworth, A., & Rasmussen, I. P. (2010). Binding objects to locations: the relationship between object files and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 543–564. https://doi.org/10.1037/a0017836

Horry, R., Wright, D. B., & Tredoux, C. G. (2010). Recognition and context memory for faces from own and other ethnic groups: A remember-know investigation. *Memory and Cognition, 38*(2), 134-141. https://doi.org/10.3758/MC.38.2.134

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin and Review, 12*(5), 822-851. https://doi.org/10.3758/BF03196775

Hubbard, T. L., & Bharucha, J. J. (1988). Judged displacement in apparent vertical and horizontal motion. *Perception & Psychophysics, 44*(3), 211-221. https://doi.org/10.3758/BF03206290

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339-373. https://doi.org/10.1037/a0034221

Irwin, D. E. (1992a). Memory for Position and Identity Across Eye Movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(2)*, 307-317. https://doi.org/10.1037/0278-7393.18.2.307

Irwin, D. E. (1992b). Visual Memory Within and Across Fixations. In Rayner, K (Ed.), *Eye Movements and Visual Cognition* (pp. 146-165). Springer, New York, NY.

Irwin, D. E., & Andrews, R. V. (1996). Integration and Accumulation of Information across Saccadic Eye Movements. In Inui, T & McClelland, J. L (Eds.), Attention and Performance XVI: Information Integration in Perception and Communication. Cambridge, MA: MIT Press.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: memory for things observed. *Perception & Psychophysics, 64*(6), 882–895. https://doi.org/10.3758/BF03196793

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(7), 1469-1482. https://doi.org/10.1109/TPAMI.2013.200

Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181-193. https://doi.org/10.1016/j.neuroscience.2005.06.042

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175-219. https://doi.org/10.1016/0010-0285(92)90007-O

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551-1556. https://doi.org/10.1177/0956797610385359

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558-578. https://doi.org/10.1037/a0019165

Kruijne, W., & Meeter, M. (2016). Long-term priming of visual search prevails against the passage of time and counteracting instructions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(8), 1293. https://doi.org/10.1037/xlm0000233

Kuhbandner, C., Rosas-Corona, E. A., & Spachtholz, P. (2017). High-fidelity visual long-term memory within an unattended blink of an eye. *Frontiers in Psychology*, *8*, 1859.

Kwok, S. C., & Macaluso, E. (2015). Immediate memory for “when, where and what”: short-delay retrieval using dynamic naturalistic material. *Human brain mapping*, *36*(7), 2495-2513. https://doi.org/10.1002/hbm.22787

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press. https://doi.org/10.4324/9781410611147

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition*, *36*(1), 43-52. https://doi.org/10.3758/MC.36.1.43

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review*, *14*(5), 989-993. https://doi.org/10.3758/BF03194133

Matthews, W. J., Buratto, L. G., & Lamberts, K. (2010). Exploring the memory advantage for moving scenes. *Visual Cognition*, *18*(10), 1393-1419. https://doi.org/10.1080/13506285.2010.492706

McKeown, D., Holt, J., Delvenne, J. F., Smith, A., & Griffiths, B. (2014). Active versus passive maintenance of visual nonverbal memory. *Psychonomic Bulletin & Review, 21*(4), 1041-1047. https://doi.org/10.3758/s13423-013-0574-1

McKeown, D., Mercer, T., Bugajska, K., Duffy, P., & Barker, E. (2019). The visual nonverbal memory trace is fragile when actively maintained, but endures passively for tens of seconds. *Memory & Cognition*, 48, 212-225. https://doi.org/10.3758/s13421-019-01003-6

Mercer, T., & Barker, E. (2020). Time-dependent forgetting in visual short-term memory. *Journal of Cognitive Psychology*, 32(4) 391-408. https://doi.org/10.1080/20445911.2020.1767627

Mercer, T., & Duffy, P. (2015). Rapid Communication: The loss of residual visual memories over the passage of time. *Quarterly Journal of Experimental Psychology, 68(2), 242-248. https://doi.org/10.1080/17470218.2014.975256*

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive computation*, *3*(1), 5-24. https://doi.org/10.1007/s12559-010-9074-z

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*(4), 465-501. https://doi.org/10.1016/0010-0285(78)90008-7

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28. https://doi.org/10.1037/h0035584

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, *12*(5), 436-450. https://doi.org/10.1016/0022-1031(76)90076-7

Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science*, *11*(2), 58-58. https://doi.org/10.3758/BF03330991

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 1. https://doi.org/10.1037/xlm0000302

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference?. *Psychological Review, 115*(3), 544-576. https://doi.org/10.1037/0033-295X.115.3.544

Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, *22*(7), 975-995. https://doi.org/10.1080/13506285.2014.941433

Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention,* 251-256. https://doi.org/10.1016/B978-012375731-9/50045-8

Öztekin, I., & McElree, B. (2007). Proactive interference slows recognition by eliminating fast assessments of familiarity. *Journal of Memory and Language*, *57*(1), 126-149. https://doi.org/10.1016/j.jml.2006.08.011

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics, 2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509-522. https://doi.org/10.1037/0278-7393.2.5.509

Rensink, R. A. (2000). The Dynamic Representation of Scenes. *Visual Cognition, 7*(1–3), 17–42. https://doi.org/10.1080/135062800394667

Rensink, R. A. (2002). Change Detection. *Annual Review of Psychology, 53*(1), 245–277. https://doi.org/10.1146/annurev.psych.53.100901.135125

Rensink, R. A., O’Regan, J. K., & Clark, J. J. (1997). To see or not to see: The Need for Attention to Perceive Changes in Scenes. *Psychological Science, 8*(5), 368-373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rensink, R. A., O’Regan, J. K., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition, 7*(1-3), 127-145. https://doi.org/10.1080/135062800394720

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology, 69*(10), 1969-1995. https://doi.org/10.1080/17470218.2014.914546

Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. A. (2011). Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. *Emotion, 11*(3), 553-562. https://doi.org/10.1037/a0024246

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition, 12*(6), 852–877. https://doi.org/10.1080/13506280444000553

Saeki, E., & Saito, S. (2004). Effect of articulatory suppression on task-switching performance: Implications for models of working memory. *Memory, 12*(3), 257-271. https://doi.org/10.1080/09658210244000649

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition, 129*(2), 241–255. https://doi.org/10.1016/j.cognition.2013.07.002

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition, 23*(7), 859-862. https://doi.org/10.1080/13506285.2015.1093243

Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, and Psychophysics, 80*(5), 1035-1056. https://doi.org/10.3758/s13414-018-1522-y

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory and Cognition, 28*(2), 214-223. https://doi.org/10.3758/BF03213801

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*(1), 156-163. https://doi.org/10.1016/S0022-5371(67)80067-7

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science, 14*(1), 44-48. https://doi.org/10.1111/j.0963-7214.2005.00332.x

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences, 1*(7), 261-267. https://doi.org/10.1016/S1364-6613(97)01080-2

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*(1), 16-20. https://doi.org/10.1016/j.tics.2004.11.006

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision, 13*(8), 1–24. https://doi.org/10.1167/13.8.16

Souza, A. S., & Oberauer, K. (2014). Time-based forgetting in visual working memory reflects temporal distinctiveness, not decay. *Psychonomic Bulletin and Review, 22*(1), 156-162. https://doi.org/10.3758/s13423-014-0652-z

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18*(5), 449-455. https://doi.org/10.1111/j.1467-9280.2007.01920.x

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology, 25*(2), 207-222. https://doi.org/10.1080/14640747308400340

Tulving, E. (1985). Ebbinghaus's Memory. What Did He Learn and Remember? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 485-490. https://doi.org/10.1037/0278-7393.11.3.485

Urgolites, Z. J., & Wood, J. N. (2013). Visual Long-Term Memory Stores High-Fidelity Representations of Observed Actions. *Psychological Science, 24*(4), 403-411 . https://doi.org/10.1177/0956797612457375

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131(*1) 48-68. https://doi.org/10.1037/0096-3445.131.1.48

Wiles, R., Prosser, J., Bagnoli, A., Clark, A., Davies, K., Holland, S., & Renold, E. (2008). Visual ethics: Ethical issues in visual research. Retrieved June 2020 from: http://eprints.ncrm.ac.uk/421/

Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding tasks? *Visual Cognition, 18*(9), 1348-1367. https://doi.org/10.1080/13506285.2010.486280

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception and Psychophysics, 67*(5), 816-827. https://doi.org/10.3758/BF03193535

Wixted, J. T., & Lee, K. (n.d.). Signal Detection Theory. Retrieved June, 2020, from: http://phonetics.linguistics.ucla.edu/facilities/statistics/dprime.htm

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics, 73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 1-8. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Horowitz, T. S., & Michod, K. O. (2007). Is visual attention required for robust picture memory? *Vision Research, 47*(7), 955–964. https://doi.org/10.1016/j.visres.2006.11.025

Woodman, G. F., & Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition, 14*(4-8), 808-830. https://doi.org/10.1080/13506280500197397

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517. https://doi.org/10.1006/jmla.2002.2864

Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin and Review, 9(*3), 575-582. https://doi.org/10.3758/BF03196315

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293. https://doi.org/10.1037/0033-2909.133.2.273

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science, 16*(2), 80–84. https://doi.org/10.1111/j.1467-8721.2007.00480.x

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21. https://doi.org/10.1037/0033-2909.127.1.3

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29-58. https://doi.org/10.1037/0096-3445.130.1.29

APPENDICES

Appendix A – Written Instructions for the recent-probes task in Experiment One

Note: The written instructions were intended to compliment the detailed verbal instructions provided by the researcher. Instructions for each task were given at the appropriate times, not all at the beginning. Instructions were adapted for Experiment Two.

Session 1 – Task 1

Thank you for agreeing to take part in this study. If at any time you want to stop the study, please just let your researcher know.

In this task, you will be expected to watch the computer screen and make judgments about the images using the computer keyboard. You will first see an image of a visual scene. Please look carefully at the image and think about the attractiveness of the scene. You will be asked to rate the attractiveness of that scene on a scale of 1-6, where 1 is very unattractive and 6 is very attractive. Please use the numbered keys on your computer keypad to make this judgement. You will be shown each image for a few seconds, therefore please make your response as quickly as possible. Following your answer, you will be briefly shown a blank screen with a '+' on it, before the next scene image is shown. Please carry on rating the different scene images until the experiment stops.

Session 1 – Task 2

Now you have viewed and rated the different images, we would like to ask you to take part in a different task. In this task, we will show you 5 different images, each separated by a small '+' on the screen. After the 5 images, you will be shown a red cross '+'. This symbolises that you are about to be shown the 6th image in that sequence. You will then be shown the 6th image and you must decide if that image was one of the 5 you have just seen by pressing z for 'Yes' or m for 'No'. Please make this judgment as fast, but as accurately as possible, as your reaction time is also being measured.

Session 2 – Task 1

Thank you for agreeing to take part in the second session of this study. If at any time you want to stop the study, please just let your researcher know.

You will now be asked to complete a similar task to the one you completed in the previous session. In this task, we will show you 5 different images, each separated by a small '+' on the screen. After the 5 images, you will be shown a red cross '+'. This symbolises that you are about to be shown the 6th image in that sequence. You will then be shown the 6th image and you must decide if that image was one of the 5 you have just seen by pressing z for 'Yes' or m for 'No'. Please make these judgments as fast, but as accurately as possible, as your reaction time is being measured.

Session 2 – Task 2

This is the final task of the study. For this task, you will be shown 2 scene images at a time. It is your task to identify which image you saw in the previous session (out of the scenes that you rated for visual attractiveness). To select the image on the left, please press 1. To select the image on the right, please press 3. You will then be asked to make a judgment about your answer and your memory for the image you have chosen by selecting 'remember' = 1, 'know' = 2 or 'guess' = 3. If you have a clear memory of seeing the image you have selected in the original session, please press 1 for 'remember'. If you do not have a vivid memory, but you have a feeling of familiarity or knowing, please press 2 for 'know'. If you do not remember either image and therefore your choice is a guess, please press 3 for 'guess'.

Appendix B – Written Instructions for the rating study task in Experiment Three

Note: The written instructions were intended to compliment the detailed verbal instructions provided by the researcher. Instructions were adapted for Experiment Four.

Thank you for agreeing to take part in this study. You may choose to withdraw from the study at any time up until the end of today's session and do not need to give a reason.

In this session, you will be shown a number of films displaying ‘everyday’ scenes. We would like you to rate these scenes for different attributes.

Instructions for task

You will first view a short film segment displaying an everyday scene – please watch this carefully.

You will then be shown the scene again, during which time you must complete a task in which you press the space bar if and when you think a boundary has occurred in the scene.

The scene will stop again, and you will then be asked to watch the scene for a final time, and rate it on the 3 attributes of visual attractiveness, complexity and distinctiveness. Before proceeding to rate each scene, please report if you recognise the location and the scene (i.e. is it somewhere you have been before).

Rating the visual images

You will be asked to rate each scene on the following attributes on a scale of 1-7.

Visual attractiveness – This attribute describes how attractive or ‘aesthetically pleasing’ this scene is compared to others like it. Please make your response on the scale 1-7, where 1 is unattractive and 7 is attractive.

Scene complexity – This attribute describes how complex a scene is. A simple scene with little complexity might contain less objects and movement, whereas in highly complex scenes many things may be happening and may be more complicated. Please make your response on the scale 1-7, where 1 is not complex, and 7 is complex.

Scene distinctiveness – This attribute links to how visually distinct you think the scene is. A scene may be perceived as more distinctive if you think a scene has characteristics which distinguish it from other scenes of its type. On the other hand, if a scene is very similar to other scenes of its type, it would be seen as undistinctive. Please rate distinctiveness on a scale of 1-7, where 1 is not distinctive and 7 is distinctive.

Appendix C – Additional Analysis for Experiment Seven

In addition to corrected accuracy to compare accuracy across delay conditions, d' was also calculated. d' was calculated by calculated by translating Hit and false alarm rates in each of the delay conditions to z scores. This was done through the 'NORMSINV' function in Microsoft Excel, which uses a mean of 0 and standard deviation of 1. Once z scores were calculated the scores were subject to the following equation to create d' data for each condition: z(hit rate)-z(false alarm rate) (Macmillan & Creelman, 2004; Wixted & Lee, n.d.).

These scores were subject to a one-way repeated measures ANOVA, which found a significant main effect of delay: F(2,56) = 19.44, $p < .001$, $\eta_p^2 = .41$. Follow up repeated measures t-tests were conducted to compare performance in each delay condition. The *p* value used was adjusted for Bonferroni multiple comparisons (.05/3 = .017). There was no significant difference between the two (M = 1.12, SE = 0.08) and four (M = 1.15, SE = 0.09) second delay, *t*(28) = .45, *p* = .66, *d* = .08. However, performance after an eight second delay (M = 0.81, SE = 0.08) was lower than after both two seconds and four seconds; *t*(28) = 5.46, $p < .001$, *d* = 1.01, and *t*(28) = 6.40, $p < .001$, *d* = 1.19, respectively. This was the same pattern of results found using the corrected recognition measure, reported in Chapter Five.