University of Southampton

Faculty of Social and Human Sciences

School of Psychology

The Relationship between Encoding and Recall for the Identities and Locations of Objects

David Corck-Adelman

Thesis for the degree of Doctor of Philosophy

November 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

SCHOOL OF PSYCHOLOGY

Doctor of Philosophy

THE RELATIONSHIP BETWEEN ENCODING AND RECALL FOR THE IDENTITIES AND LOCATIONS OF OBJECTS

By David Corck-Adelman

Five experiments are reported which investigated the relationship between the number of fixations that participants made on objects in a photograph of a visual scene and the memory that participants exhibited for the identities and locations of those objects. The results of Experiment 1 showed that there is a very close relationship between encoding and recall for both the identities and locations of objects, and provided evidence that information about identities and locations of objects might be encoded differently. Specifically, the data suggested that object identities were encoded across multiple fixations, thereby accumulating in memory across separate fixations. However, object locations were encoded accurately after the first fixation on an object and did not appear to improve significantly with subsequent fixations.

Experiment 2 demonstrated that the relationship between encoding and recall can be influenced using a lower-level attentional capture technique to draw attention to certain ‘boxed’ objects at the expense of ‘non-boxed’ objects, leading to superior memory for the ‘boxed’ objects.

Experiment 3 investigated whether a primacy or recency effect was found for object identity and location memory. No such evidence was obtained for either. Experiment 4 provided evidence that information about the overall configuration of objects was encoded without the need for each object to be directly fixated. Experiments 4 and 5 both found that objects must be fixated for their specific location to be encoded, but also provided some suggestive evidence that object location memory might also accumulate across fixations (though to a lesser extent than for object identity memory). Hollingworth and Henderson’s Visual Memory Model provided the theoretical framework for the experimental work. Modifications to this model were proposed to account for the current findings.

In conclusion, the five experiments provide significant insight into encoding processes associated with memory for the identities and locations of objects.

Contents

1 Literature Review 1

1.1 Research on picture recognition 1

1.2 Change blindness 9

1.3 Flicker paradigms 15

1.4 Why does change blindness occur? – Coherence Theory 18

1.5 A challenge to Coherence Theory 21

1.6 Basic characteristics of eye movements 24

1.7 Further challenges to Coherence Theory: Eye movements and change detection tasks 26

1.8 Why does change blindness occur? – Visual Memory Model 32

1.9 Why are change blindness results interesting? 34

1.10 Research into memory for object identities 36

1.11 Research into memory for object locations 46

1.12 Summary of previous literature 53

1.13 Proposed research and introduction to the novel paradigm 54

2 The role of visual encoding in memory for object identity and object location 57

2.1 Introduction 57

2.1.1 Explanation of key procedures used in the thesis: General methods 66

2.1.2 Calculation of placement error 67

2.1.3 Calculation of best fit error 70

2.1.4 Monte Carlo simulation to calculate random placement error 71

2.2 Method 73

2.3 Results 74

2.4 Discussion 85

3 Using attentional capture to modulate the relationship between encoding and recall for object identity and object location 95

3.1 Introduction 95

3.1.1 Higher-level influences on eye movements: Task instructions 96
3.1.2 Higher-level influences on eye movements: Prior knowledge 99
3.1.3 Higher-level Influences on eye movements: Linguistically-driven eye movements 100
3.1.4 Lower-level Influences on eye movements: Saliency maps 103
3.1.5 Lower-level influences on eye movements: Colour and visual search ... 106
3.1.6 The present experiment 110
3.2 Method 114
3.3 Results 117
3.4 Discussion 126
4 The influence of fixation order on memory for object identity and object location 133
4.1 Introduction 133
4.1.1 Serial order and memory for words in lists 134
4.1.2 Serial order and memory for the identity of objects in scenes 136
4.1.3 Serial order and memory for the locations of objects in scenes 141
4.1.4 Comparison of memory for word lists and memory for objects in scenes: Manner of encoding 142
4.1.5 Comparison of memory for word lists and memory for objects in scenes: Identification period 143
4.1.6 The present experiment 144
4.2 Method 147
4.3 Results 149
4.4 Discussion 171
5 The relationship between the number of fixations on an object and memory for the configuration of objects and specific object locations 187
5.1 Introduction 187
5.1.1 Hollingworth and Henderson's Visual Memory Model 187
5.1.2 The time course of encoding for the spatial map: The link between the spatial map and scene gist 190

5.1.3 Scene gist ... 190

5.1.4 The nature of the spatial map: The link between the spatial map and saliency maps ... 193

5.1.5 Are object identities and specific object locations encoded in the same way? What is the time course of encoding for object identities and object locations? ... 196

5.1.6 When does consolidation of information from short-term to long-term memory happen? ... 202

5.1.7 Summary of the Visual Memory Model and its relationship with the current data ... 203

5.1.8 Evidence for a distinction between the overall configuration of objects and the specific locations of objects ... 204

5.1.9 Evidence for the overall configuration of objects being encoded automatically ... 206

5.1.10 The present experiment ... 210

5.2 Method ... 212

5.3 Results ... 213

5.4 Discussion ... 223

6 The relationship between the number of fixations on an object and memory for both the specific location and correct surface location of an object ... 227

6.1 Introduction ... 227

6.1.1 Evidence for accumulation of specific object location information across multiple fixations ... 228

6.1.2 Comparison between Tatler et al. and the present experiments ... 228

6.1.3 Modification of the test phase of the present paradigm ... 230

6.1.4 The present experiment ... 232

6.2 Method ... 234

6.3 Results ... 235

6.4 Discussion ... 245

7 General discussion ... 257

7.1 Chapter outline ... 257

7.2 Motivation for the thesis 257
7.3 Key Findings 259
7.4 Modified Visual Memory Model 266
7.5 Closing Remarks 270
8 References 273

List of figures

Figure 1. The stimulus photograph viewed by participants in the encoding phase 62

Figure 2. Schematic diagram of calculation of the surface-specific co-ordinates for "home" locations 67

Figure 3. Schematic diagram of calculation of the generalised co-ordinates for "home" locations 68

Figure 4. Schematic diagram of calculation of the placement error distance between the "home" location and the placed location 69

Figure 5. Example of an original image (left) and its saliency map (right) copied from Parkhurst, Law and Niebur, 2002 (pp. 109). 104

Figure 6. The modified stimulus photograph presented to Group A. Group B saw the same photograph but all boxed objects were unboxed, and all unboxed objects were boxed. 116

Figure 7. The serial position curves found by Postman and Phillips (1965) 136

Figure 8. The probability of correct object selection as a function of the order in which objects were presented 160

Figure 9. Placement error (cm) as a function of the order in which objects were presented 162

Figure 10. Gaze duration (ms) as a function of the order in which objects were presented. 164

Figure 11. Number of fixations made during gaze duration as a function of the order in which objects were presented. 165

Figure 12. Object identity confidence as a function of the order in which objects were presented. 169

Figure 13. Object location confidence as a function of the order in which objects were presented. 171

Figure 14. Gaze duration (ms) as a function of the number of fixations made during gaze duration for box-following and free-viewing participants 174

Figure 15. Scatterplot showing the correlation between average fixation duration (ms) made on each of the 12 objects in Experiment 1 and the probability that each object was correctly selected in Experiment 3. 181

Figure 16. Panel a: The original stimulus photograph. Panel b: The spatial map corresponding to that photograph with slot locations for each of the objects. 194

Figure 17. Panel a. The original stimulus photograph. Panel b. The spatial map corresponding to the photograph with one of the slot locations occupied by an object file for "football". 197

Figure 18. Panel a: The original stimulus photograph. Panel b: The spatial map corresponding to that photograph with one of the slot locations occupied by an object file containing identity detail. 199

Figure 19. Panel a: The spatial map with an image of an actual object linked to a location. Panel b: The spatial map conceptualised as a semantic network with a node representing an object. 200

Figure 20. Illustration of the hierarchical semantic network model of memory representation 201

Figure 21. Illustration of a node for "football" with its defining characteristics. 202

Figure 22. Placement error and best fit error as a function of the number of fixations made on objects during the encoding phase (error bars represent standard error) ... 219

Figure 23. Block confidence and identity-location confidence as a function of the number of fixations made on objects during the encoding phase 222

Figure 24. Probability of correct object selection as a function of the number of fixations made on objects during the encoding phase for both Experiment 1 and Experiment 5 239

Figure 25. Identity and location confidence ratings as a function of the number of fixations made on an object 243

Figure 26. Probability of correct marker placement as a function of the number of fixations made on an object 245

Figure 27. Illustration of the exact positions of objects being encoded (left panel) and the corresponding vicinities of error (right panel) 249

Figure 28. Illustration of the exact positions of objects being encoded (left panel) and a range of differently sized vicinities of error corresponding to the number of fixations each object has received (right). Smaller regions result from more fixations on that object 250

Figure 29. Illustration of no competitor markers in the present study (top panel), and 13 competitor markers in a potential future experiment (bottom panel). In the bottom panel the unfilled circle denotes the actual position of the object. 251

List of tables

Table 1 Means and Standard Errors (in Parentheses) for Eye Movement Measures in the Encoding Phase 76

Table 2 Means and Standard Errors of Recall Measures in the Test Phase....... 77

Table 3 Means and Standard Errors (in Parentheses) for Probability of Correct Object Selection when Objects were and were not Fixated, and when they were Fixated just Once or Twice or more 81

Table 4 Means and Standard Errors (in Parentheses) for Placement Error Scores when Objects were and were not Fixated, and when they were Fixated just Once or Twice or more 84

Table 5 Means and Standard Errors (in Parentheses) of Eye Movement Measures for all Control Participants in Experiment 2 (No Objects Boxed) and the Participants in Experiment 1 who Experienced a 5 s Viewing Duration and Immediate Test 118

Table 6 Means and Standard Errors (in Parentheses) of Eye Movement Measures across Both Conditions (Half Objects Boxed or Objects Boxed) 120

Table 7 Means and Standard Errors of Recall Measures in the Test Phase for both Conditions (Half Objects Boxed or No Objects Boxed)........ 121

Table 8 Means and Standard Errors (in Parentheses) of Memory Performance across Both Conditions (Half Objects Boxed or No Objects Boxed)........ 122

Table 9 Means and Standard Errors (in Parentheses) for Memory Measures Within Each Condition 123

Table 10 Means and Standard Errors (In Parentheses) for Probability of Correct Object Selection 124

Table 11 Means and Standard Errors (In Parentheses) for Placement Error Scores........ 126

Table 12 Means and Standard Errors (in Parentheses) of Eye Movement Measures for the Free-viewing Condition and the Box-following Condition in the Encoding Phase 152

Table 13 Means and Standard Errors of Recall Measures in the Test Phase for the Free-viewing Condition and the Box-following Condition in Experiment 3 153

Table 14 Means and Standard Errors (in Parentheses) for Eye Movement Measures in the Encoding Phase for Experiments 1 and 3..........155

Table 15 Probability of Correct Object Selection as a Function of the Number of Fixations made during Gaze Duration..........167

Table 16 Placement Error as a Function of the Number of Fixations made during Gaze Duration..........168

Table 17 The Proportion of Participants that Selected each Object and Average Placement Error for each of the 12 Objects..........182

Table 18 Means and Standard Errors (in Parentheses) of Eye Movement Measures for the Block-placing Participants in Experiment 4 and the Control Condition from Experiment 3 in the Encoding Phase..........214

Table 19 Means and Standard Errors of Recall Measures in the Test Phase for the Block-placing Participants in Experiment 4 and the Control Condition taken from Experiment 3..........217

Table 20 Means and Standard Errors (in Parentheses) and *t*-test results for the Eye movement measures for the experimental and control groups..........237

DECLARATION OF AUTHORSHIP

I, David Corck-Adelman, declare that the thesis entitled "The Relationship between Encoding and Recall for the Identities and Locations of Objects" and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as:
    - Liversedge, S. P., Meadmore, K. L., Corck-Adelman, D., Shih, S., & Pollatsek, A. P. (2011). Eye movements and memory for objects and their locations. *Studies of Psychology and Behavior, 9(1),* 7-14.
    - Corck-Adelman, D., Shih, S., Pollatsek, A. P., & Liversedge, S. P. (under review). The role of visual encoding in object identity and location memory. *Journal of Experimental Psychology: Learning, Memory and Cognition.*

Signed: ....

Date: ....

Acknowledgements

First, I would like to thank my supervisor Simon Liversedge. He has provided me with the best assistance, encouragement and inspiration that I could have wished for. I have learnt so much from him and can never thank him enough for his support and guidance.

I would also like to thank Shui-I Shih for supporting me in applying for post-graduate study, for guiding me through my MSc and for teaching me so many valuable academic skills.

My colleagues in room 3109 made my experience as a PhD student so enjoyable and fulfilling. I could not have hoped for a better group of friends to share this experience with. I thank Katie Meadmore, Kathryn Willerton, Katie Gray, Julie Kirkby, Matt Parker, Hayward Godwin, Dan Schoth and Elle Boag for their support and kindness in the early days, and Natalie Mestry, Jenny Josephs, Sara Robertson, Iona Kerrigan and Louise Leyland for the same in the later days.

I have been lucky to work with some very supportive fellow PhD students and I thank them all for their encouragement; thanks to Itsek Alfasi, Phil Cordaro, Helen Richards, Sara Luedekke-Carr, Abi Lucas, Greg Neil, and Akif Guzel.

Thanks also to all my friends in Jazzmanix with whom I shared so many enjoyable musical and social experiences.

My family has always been incredibly supportive of everything I have done, and I am eternally grateful for the love and guidance they have always shown me. Thank you to Mum, Dad, Jon, Alex and Nanna.

Finally, I would particularly like to thank Shauna Ellis for her unflinching support, her calm and patient nature, and her kindness and love. She has been my shoulder to cry on, my friend to laugh with and a constant source of inspiration. I could not have achieved this without her.

1 Literature Review

1.1 Research on picture recognition

Studies conducted in the 1960s and 1970s provided evidence that observers have good memory for previously viewed pictures. In one study, Nickerson (1965) had participants examine 200 photographs for 5s each in preparation for a subsequent memory test. The test involved viewing a series of photographs and deciding which were old (photographs viewed as part of the original 200 stimulus set) and which were new (a photograph which had not previously been viewed) – a test also known as a two alternative forced choice test (2AFC). The results showed that in general, recognition accuracy was 95%. Nickerson manipulated the number of photographs that intervened between initial viewing and test, and found that accuracy decreased somewhat as the number of intervening photographs increased. However, even when 200 photographs were viewed in between the initial study of an image and the presentation of that image at test, recognition performance only dropped to 87%. Thus, Nickerson demonstrated that participants have good memory for previously viewed photographs, and that this memory exists even after a large number of other photographs have been viewed before the test occurs.

Similarly, Shepard (1967) showed participants 612 pictures, each for 6s. At an immediate memory test, participants correctly selected the viewed picture with 98% accuracy on 2AFC tests. Again, this provided evidence that observers retain at least some information about visual scenes, and that this information is explicitly available for report at a later time. Clearly some aspect of the images was retained, and this allowed participants to accurately recognise previously viewed images at test.

Standing, Conezio and Haber (1970) presented participants with one viewing of over two thousand photographs for 5s or 10s. When participants were later tested for their memory of their photographs, they scored 90% correct when completing a 2AFC test, even if the test happened a day later. This provided further evidence that participants can accurately recognise previously viewed photographs. It also demonstrated that this memory is robust, and exists even after a 24 hour delay. However, this study did not probe memory for specific scene details (so it was not known which aspects of the photographs were remembered), nor did it measure eye movements, thus neglecting to investigate which aspects of the photographs were looked at.

Standing (1973) compared memory for pictures with memory for words. In Experiment 1, participants were randomly assigned to viewing normal pictures (images that were not unusual in any way e.g., a dog), vivid pictures (defined as interesting or unusual subject matter e.g., a dog smoking a pipe) or words. Each image was presented for 6s, and participants were also randomly assigned to view 20, 40, 100, 200, 400, 1,000, 4,000 or 10,000 images. Participants were instructed to study each image in preparation for a 2AFC memory test that would occur after a two day delay. Standing found that participants correctly identified more pictures than words, and demonstrated better memory for vivid pictures than normal pictures. There was a small decrease in test accuracy as the number of viewed images increased. Despite a two day delay, Standing found that participants who viewed 1,000 images correctly identified 88% of the vivid pictures and 77% of the normal pictures (in comparison, participants correctly identified 62% of the words). This demonstrates that some information about pictures is stored in long-term memory and can be accessed even after two day delay. It also suggests that some types of visual information (vivid pictures) are remembered better than others (normal pictures). Potentially, the vivid pictures were remembered more accurately because their “unusualness” led to them being encoded more deeply in memory. Alternatively, their unusualness meant they were more easily distinguishable when compared to the “new” (or distractor) images. At this point it is important to note that similarity to distractor images (or, more generally, how easily stimuli are distinguishable from each other) is a critical issue in this literature review and as such is discussed below.

The studies presented above demonstrate convincingly that participants can accurately recognise previously viewed images, even if a large number of other images have been viewed in between original viewing and test. In turn, this suggests that some type of memory representation is retained from one brief view of an image (ranging between 5s and 10s). All of the studies used the same memory test: showing participants a mixture of previously viewed stimuli (old) and unseen stimuli (new) and having participants categorise these stimuli as old or new. Thus, this task could be simplified as a “spot-the-difference” task between image A (the previously viewed stimulus) and image B (the new stimulus).

However, in the studies discussed above, an issue that was not explored relates to the degree to which each image was encoded. It may be that just the gist of the image was retained, but no information about specific details was encoded. Alternatively, participants may have encoded specific details about certain objects and used this information to guide their responses. However, the methodologies employed did not explore this. A second critical issue relates to the similarity between A and B (or how easy it is to discriminate between the two images). In the studies by Nickerson (1965), Shepard (1967), Standing, Conezio and Haber (1970) and Standing (1973), the "new" stimuli were completely different images compared to the "old" stimuli (i.e. not modified versions of the "old" images). However, it was neither stated (nor quantified) how different they were. Potentially, the more similar A and B are, the easier it is to discriminate between them. For example, an "old" image might be a kitchen scene with red walls containing a kettle, sink and fridge. A "new" image might also be a kitchen scene but with blue walls and containing a stove, saucepan and plate. Clearly the two images are different, yet they are also both kitchens. It is logical to suggest that, as the kitchens contained different objects, participants would generally be able to correctly identify which "kitchen" they had previously studied. However, if the red kitchen was the "old" image, but the "new" image was a beach scene, accurate discrimination would be more likely as the images do not share any features. In the studies mentioned above, the degree to which the "old" and "new" stimuli were different was not stated. It was not known how difficult it was to distinguish between old and new images. With this in mind, it is difficult to draw conclusions based on the accuracy with which "old" images were correctly identified.

Consequently, one might expect that given a potential experiment in which participants discriminated between "old" full colour images and black and white "new" images, a very high degree of accuracy would be unsurprising, as the only information needed to accurately complete the task would be the presence or absence of colour in the image (e.g., if I see a colour photograph, it must be "old"). A slightly more difficult task would present old and new colour images, but an absence of semantically identical images (e.g., if a kitchen was "old", a different kitchen would never be "new" – a bathroom would be a "new" image). In this instance, successful performance would occur even if participants only remembered that "I saw a kitchen, but I did not see a bathroom". No specific details about the kitchen would be needed; instead, just knowing that a kitchen was presented would likely lead to accurate classification of old and new images.

In turn, a more difficult task would be to distinguish between two different kitchens. To successfully identify which was old and which was new, it is likely that specific detail(s) about the "old" kitchen would need to be stored in memory, even if these details were relatively superficial ("the kitchen I saw did not have windows"). Finally, if participants were asked to distinguish between two kitchens which are identical except for one small change (e.g., one object is rotated 90º), the task becomes much more difficult. In this example, it is not enough to know that a kitchen was viewed. Neither is it sufficient to remember some superficial details about the kitchen as a whole. Instead, highly detailed memory representations at the level of individual objects are required to lead to accurate discrimination between images (e.g., "I saw a small red kettle next to the tall white fridge").

The critical point here is that the degree to which two stimuli are similar is likely to affect how successfully participants can discriminate between them. The picture memory studies presented above all report exceptionally high levels of recognition accuracy leading to the conclusion that memory for visual scenes is excellent. However, it can be speculated that this conclusion is due partly to the large differences between the old and new images presented to participants. When considering these results, it is important to note that the degree to which the images were discriminable was neither controlled for, nor experimentally manipulated.

This issue of discriminability was assessed by Mandler and Ritchey (1977) who investigated whether various aspects of scenes were remembered differently. As in the studies reported previously, a 2AFC test was used to assess memory for previously viewed images. However, rather than having participants discriminate between old and new images, Mandler and Ritchey made one of eight possible changes to the old images. In this way, it can be summarised as searching for differences between A and A' (where A' is a modified version of A), rather than A and B (where B is a different image to A). During the encoding stage, participants viewed eight black and white slides containing six objects for 10s each, and were instructed to study them carefully. Additionally, half of the participants viewed organised pictures, in which the objects within the scene were arranged in a congruent manner (e.g., a chair, table and a mug arranged to look like a lounge) whilst the other participants viewed the same objects, but arranged in disorganised pictures (e.g., a chair, table and a mug arranged randomly). The memory test (a 2AFC task) occurred either immediately, after a delay of one day, a delay of one week, or a delay of 4 months. At test, participants viewed 64 slides. Half were "old" (i.e. identical to those viewed in the encoding stage) and half were "new". These new slides were identical to the "old" slides except for they contained one of eight possible changes: an object deletion, an object addition, a type change (where an object was replaced by a conceptually different but same sized object e.g., replacing a telephone with a notebook), a token change (where an object was replaced by a semantically identical but different shaped object e.g., switching a teddy bear for a different teddy bear), a rearrangement (where two objects switched position), an orientation change, the move of one object to a novel location, and a size change (where an object increased or decreased in size on the slide). Participants were required to state which of the 64 test slides were old, and which were new.

First, the results showed that overall accuracy on the old/new task was comparable between immediate and 1 day test. However, performance declined after a week delay, and declined further after 4 months. This demonstrates that generally memory for visual scenes is fairly robust after a one day delay, but decays over time. More interestingly, there was a three-way interaction between type of image (organised or disorganised), test delay period (immediate, 1 day, 1 week, 4 months) and type of change (the eight mentioned above). For some types of changes, the layout of the image at encoding (organised or disorganised) affected the rate of decay. For additions, deletions and type changes, and spatial configuration changes, there was no significant decrease in performance across the 4 month delay for participants who viewed organised images. However, for those participants who viewed disorganised images, memory for those types of changes decayed across the delay. This demonstrates that the way stimuli are arranged during the encoding period affects the rate at which information is retained. If visual information is arranged in a congruent configuration then additions, deletions or type changes are noticed accurately even after a 4 month test delay. However, if the same stimuli are arranged in an unexpected or meaningless way, those same changes are not noticed.

Additionally, changes made to the spatial configuration of scenes (i.e. switching in the locations of objects) were detected accurately even after 4 months for organised images, but poorly identified if the images were disorganised. Interestingly, some changes were often missed regardless of the configuration of the initial scene (for example size changes).

Broadly, these findings demonstrate that some memory for visual scenes is retained over a delay of 4 months. More critically, it suggests that different aspects of scenes may decay differentially. For organised pictures, changes made to the identities of objects (type changes) and changes which added a new object or removed an old object were accurately detected after a delay. However, for organised images, when objects increased or decreased in size, these changes were typically not accurately detected after a delay. There are at least two explanations for this finding. First, it could be that participants encode some aspects of scenes accurately (e.g., object identities) and other aspects less accurately (e.g., the sizes of objects). This might lead to successful detection of changes made to the accurately encoded aspects and unsuccessful detection of changes made to the poorly encoded aspects.

Alternatively, some changes might be inherently more difficult to detect than others, even if all aspects of an object (name, colour, location, size) are encoded equally accurately. As discussed above, changes made to individual objects in scenes might require more detailed memory representation than changes made to the overall scene.. Additionally, regarding changes made to objects themselves, certain changes made to objects may be more easily discriminable than others, and Mandler and Ritchey demonstrated that this may be true; the switching of object locations was detected well after a 4 month delay, but an object increasing in size was not accurately detected. What this suggests is that the different properties of an object might be encoded differentially. More generally, other similar studies have shown that recognition performance of briefly presented pictures remains at above-chance levels even after a delay of a year (Nickerson, 1968; Fajnstzejn-Pollack, 1973).

One study into memory for pictures attempted to understand how participants made successful old/new judgements. Parker (1978) had participants view seven drawings containing six objects for an unlimited time period in preparation for a 2AFC test. At test, 30 images were old (identical to the previously viewed stimuli) and 30 were changed in one of five ways: object deletion, object size increase, object size decrease, type substitution or token substitution. Importantly, eye movements were recorded both during initial viewing of the original scene, and during the test trials.

Parker reported that identification errors occurred on less than 5% of the trials (meaning that overall accuracy was 95% - a performance level comparable to Nickerson, 1965). The eye movement data demonstrated that during "same" trials (where no change was present), objects were usually fixated in the same order on every trial, and usually all six objects were fixated before a response was made. However, during "different" trials this order was often violated, and the changed object was typically fixated early in viewing. Parker suggested that information about the pre-change version of the object was stored in detail in memory, and also that information about the post-change object was encoded to some degree in the periphery, leading to that object being fixated out of order. Parker reported that on average, "subjects fixated a changed object more than one fixation earlier than the same object when it had not been changed" (Parker, 1978, pp. 289).

The finding that objects were fixated earlier in viewing when they had changed compared to trials where they had not changed provides some evidence that detailed information about objects seems to be stored in memory. Parker argued that the changed objects were fixated earlier than during no change trials because participants had used their peripheral vision to identify that the visual display differed in some way to the memory they had previously encoded about the display during the presentation trials. It is important to note that, despite recording eye movements made during initial inspection of the image, no analysis of this behaviour was reported.

The main finding from the studies reported above was that participants have good memory for scenes, and this contributed to the dominant perspective at the time that, as observers, we develop a complete visual representation of the visual environment and this representation allows us to interact successfully with the world, as well as allowing us to experience a richly detailed experience of the world - known broadly as the composite *scene representation hypothesis* (Castelhano & Henderson, 2005) and also as the *spatiotopic fusion hypothesis* (Irwin, 1992, as cited in O'Regan, 1992). However, the research by Mandler and Ritchey (1977) suggested that some aspects of scenes may be encoded and remembered better than other aspects, and that if memory for the poorly encoded or remembered aspects was probed, participants would not show a high level of correct identification.

Despite this, the dominant perspective was generally that memory for a previously viewed stimulus is detailed, and supports accurate performance on memory tests.

In direct contrast, a body of research that developed in the 1990s appeared to provide a challenge to the conclusion that a richly detailed representation of the visual world is held in memory. This research was named *change blindness.* Change blindness is the failure of observers to notice a change made to a visual scene when the change is hidden by some type of disruption or occlusion. The change is typically introduced either during a brief flicker that occurs between successively presented images on a computer screen, during a cut in camera angle during a video, or during a disruption or occlusion incident in a real-life interaction. Central to change blindness paradigms is that the change is hidden by a disruption or occlusion. If not, the change would be obvious. The disruption hides the motion transient that would otherwise attract attention to the change.

As a result, the changes do not attract attention automatically (attention is not pulled to the change as it occurs); instead the observer needs to search the scene for each object, comparing its current state with the state stored in memory (if one exists). Successful *change detection* therefore is a reflection of accurately encoded and stored memory for the precise pre-change details of an object, which supports a comparison with the currently available post-change details. Importantly, change blindness research always uses the A and A' comparison (rarely seen in the picture memory literature). Changes are made to individual objects in scenes, rather than the global scene itself. In this way, a standard change blindness task could be considered more difficult than a standard picture recognition task, as in change blindness the changes made are less easily discriminable. The overall finding of change blindness (that changes made in a variety of paradigms were often missed by observers) seemed to demonstrate that, contrary to the conclusions one may form on the basis of the picture recognition literature, memory representations of scenes are relatively impoverished and as a result, seemingly obvious changes made to scenes can pass unnoticed.

1.2 Change blindness

Studies that investigated change blindness have used a variety of paradigms to investigate observers' ability to detect changes. The overall conclusion from these change blindness studies was that, under certain conditions, observers will consistently fail to detect changes; therefore memory for visual scenes is poorly detailed. These studies provided a challenge to the composite scene representation hypothesis, and eventually led to a new theory to explain visual perception of scenes – Coherence Theory (Rensink, 2000) which is discussed later. An early change blindness study was conducted by Grimes (1996). Participants viewed scenes on a computer in preparation for a memory test (which was not administered; it was used to encourage participants to view the whole scene). They were also instructed that while viewing some of the images, one aspect of the scene would change (e.g., a building in a city scene increased in size by 25%; two cowboys sitting on a bench switched heads). Eye movements were recorded while participants viewed the scene, and all changes were made during a pre-defined saccade (intended to hide the motion signal associated with the change that would likely attract attention to the change). It is important to note that during a saccade the eyes are functionally blind and therefore changes cannot be detected if they are made during a saccade. A change was made in 80% of the trials, and overall participants failed to detect 67% of these changes. Grimes reported details of ten types of changes (although 40 changes occurred in the experiment). Change blindness ranged from 100% (e.g., when two men at a wedding exchanged hats) to only 18% (e.g., a large parrot changed from green to red). Clearly, some changes were more noticeable than others. Grimes reported that each trial lasted for 10 s, but did not report when the change occurred, only that it happened during a pre-defined saccade. Potentially, changes made earlier in the trial were less likely to be noticed as the pre-change detail had not been looked at.

Alternatively, the variability in change blindness suggests that the ease with which pre- and post-change versions are discriminable influences the likelihood that the change will be detected. Missing a change in the colour of an object from dark red to light red would be less surprising than missing a change in colour from red to yellow. Nevertheless, the results demonstrated that changes made to aspects of scenes that were relevant and interesting (rather than changes made to minor or peripheral details) could be missed by observers. Grimes suggested that this finding implied a failure to encode visual details in memory, or failure to consolidate those details, or a failure to retrieve the details to support successful change detection.

Levin and Simons (1997) demonstrated change blindness using a video of two actors having a conversation. Participants watched the video, but were not asked to look for changes. Nine changes were made during cuts in the camera angle (e.g., a scarf disappeared from an actor’s neck). Participants were asked if they had noticed any unexpected changes and if so, to describe the changes in detail. In Experiment 1, only one of the ten participants reported seeing a change whilst watching the video. Even when participants watched the video again and were explicitly told to look for changes, on average participants noticed only 2 of the 9 changes.

In Experiment 2A, participants watched one of eight possible videos. In each of the videos a single change was made to an actor across a cut in camera angle. For example, in one of the videos a man was sitting at a desk when the phone rang in another room. The man stood up and left the room to answer the phone. The man who was shown answering the phone was different to the man who was sitting at the desk. Across all eight videos, 33% of participants failed to notice that the actor had changed identity across the cut in camera angle. However, some observers who failed to report the change gave descriptions of the actors e.g., “a man in blue shirt stood up, walked to the door and answered the phone”. It was neither discussed, nor quantified, which of the two actors these descriptions referred to. Nevertheless, it suggests that participants possessed a memory representation of one, or both, of the actors. Levin and Simons concluded that even when attention is paid to an object (or person) it may not be represented in memory in detail, or sustained across time. However, some objects are represented well, and are preserved in memory, leading to successful change detection. Therefore, the authors claimed that successful change detection requires effortful processing of scene properties. If this process does not occur, change blindness is the likely outcome. This suggestion was corroborated by the findings of Experiment 2B (below).

However, it is possible that change blindness would occur *despite* detailed processing of both the pre- and post-change stimulus (e.g., the actor in Experiment 2A). If the two detailed memory representations were not compared, it is possible the change could pass undetected. Alternatively, the change may pass undetected because it is too difficult to discriminate between the pre- and post-change versions (as discussed in the previous section on picture recognition). Clearly, it would be unsurprising if observers failed to detect a change between two identical twins wearing the same clothes, on the basis that discrimination between the two would be very difficult. It is logical to speculate that as discriminability between two stimuli becomes easier, change blindness becomes less prevalent. However, in Experiment 2B, all participants watched all of the eight videos presented in Experiment 2A, as well as eight control videos in which no change occurred. Critically, in Experiment 2B participants were instructed that changes would occur in some of the sixteen videos and had to state which videos contained changes. Levin and Simons reported that change detection was almost at ceiling, suggesting that the failure to detect changes in Experiment 2A was not due to difficulties in discriminating between the actors. On this basis, Levin and Simons concluded that participants failed to notice the changes in Experiment 1 and 2A because richly detailed memory representations of viewed stimuli may not be formed for viewed stimuli.

There is an important distinction that must be drawn at this point between the literature on picture recognition and change blindness. The results from Experiment 1 and 2A (i.e. poor performance) are in direct contrast to the picture recognition literature (i.e. good performance). This contrast can be explained, in part, by the issue of discriminability as discussed above (picture recognition – easier to discriminate between competing stimuli; change blindness – more difficult to discriminate). However, task instructions may also play a part. In the picture recognition literature, participants are explicitly instructed to view the images in preparation for a memory test. However, in initial change blindness studies (although not in Experiment 2B above, Grimes, 1996, and not in flicker paradigms), participants are not aware that they need to remember details of the scene – they are not viewing the stimuli in preparation for a task that will probe their memory for what they have seen (typically participants are misled and told to “pay close attention” or receive no instructions in the real-world interaction studies).

Therefore, a critical issue is the extent to which visual information is encoded in memory intentionally (when instructed to by an experimenter; when trying to learn a new route to work etc), and the extent to which this process occurs automatically or implicitly without cognitive effort. It is logical to suggest that memory for scene details will be more richly detailed and readily available if the scene has been deliberately and carefully encoded into memory in preparation for a later memory test probing specific details. However, during normal viewing (and in everyday living) it is unlikely, as well as unnecessary that observers deeply encode all the details of the visual environment. As long as we understand the visual environment sufficiently at an abstract level ("there is a busy road on my right"), it is not normally necessary to know exactly which make and model of car is next to us from moment to moment. In contrast, it is important to remember where we put our keys if we know we will need them the next day. Knowing they are "somewhere on the desk" is unhelpful; remembering that we put them underneath our diary will facilitate finding them again. Therefore, the lack of intentional encoding instructions, coupled with stimuli that are more difficult to discriminate between, both contribute to the poor performance level observed in change blindness studies when compared to the picture recognition literature.

Change blindness research was taken out of the laboratory by Simons and Levin (1998) who introduced a change in conversation partner during a real-world interaction. In Experiment 1, participants were approached and asked to give directions. After 10-15 seconds of the interaction, two experimenters pushed between them carrying a door. This caused a short visual disruption. During this disruption, the conversation partner switched with a second experimenter, who then continued the conversation. Once the conversation had finished, participants were asked if they had noticed anything unusual. Only 7 of the 15 participants noticed the switch in conversation partner. As in Levin and Simons (1997), the authors concluded that attention was needed to detect changes. However, the presence of attention did not guarantee that a change would be noticed. This led the authors to the conclusion that as observers, we retain only abstract details from visual environments, not specific visual details.

Levin, Simons, Angelone and Chabris (2002) also used real-world scenarios to demonstrate change blindness. In one experiment, an actor switched with another actor while the participant prepared to take a photograph, and in a second experiment, two actors switched as participants waited at a desk to receive some forms. In both experiments, participants exhibited high levels of change blindness.

Importantly, in both studies, participants were not aware that a change would occur. Potentially, this lack of intentional encoding contributed to the change blindness level (as details of the stimuli would not be encoded in detail). Additionally, the degree to which the two experimenters could be discriminated may have contributed to the change blindness (it was reported that their clothing was different, and that there was a height difference of 5cm). However, research in other areas has shown that adults are typically poor at estimating their own height. Men tend to over-report by 1.4cm; women by 0.6cm (Rowland, 1990). This suggests that adults are poor at estimating the height of people in general, and may contribute to participants failing to notice the height difference. However, it is important to note that participants tend to *underestimate* the height of others (Flin & Shepherd, 1986) suggesting that people are generally poor at identifying basic characteristics of people such as height. Another possible explanation for the failure to notice changes may be that participants simply categorised the first experimenter as "a tall man with dark hair". As long as the second experimenter does not violate this abstract description, a change may not be detected.

Theoretically, this study suggests several possible reasons for change blindness. First, it may be that, despite fixating (and encoding) various aspects of visual scenes, detailed representations are not formed in memory whatsoever. Thus, change detection cannot occur as there is nothing to compare the currently fixated object to. If this is true, it is in direct contrast to the perspective influence by the picture recognition literature that a detailed memory representation of the environment does exist. Alternatively, a detailed memory representation may exist in memory for the previously viewed stimuli, but this information may not be used to support change detection. This may occur in situations where the change does not violate the meaning of the scene, and as a result, the detailed memory representation is not consulted (as the new image is similar enough not to warrant a comparison).

A third possibility is that the pre- and post-change stimuli may be too difficult to distinguish between, even if detailed memory for both exists *and* a comparison is made (although Experiment 2B conducted by Levin and Simons suggests that, in that study, the change could be noticed if prompted). A fourth possibility is simply that, at the moment the change occurs, the pre-change version has not been encoded into memory in sufficient detail to support accurate change detection (although note that, unlike the first possibility, the pre-change version has been encoded, although it lacks detail). If a changed aspect has not been fixated before changing, it is unsurprising to miss a change made to it. This suggests in turn that, if detailed memory representations for visual stimuli exist, they do not form immediately upon encountering the scene. Rather, the process of building a detailed representation may accumulate over time. Or, an object may have been fixated, but not encoded in detail e.g., "I saw a kettle, but I do not know the colour, the size, or the orientation of it". A critical issue here is the degree to which an object has been encoded and stored in memory. A detailed memory representation of an object is more likely to lead to successful change detection than a sparsely detailed (or non-existent) memory representation. However, as eye movements were not recorded in the video-based and real-world interaction studies (and the change latencies were not reported or manipulated) conclusions on this basis were not possible.

As well as introducing changes during saccades (Irwin, 1991; McConkie & Currie, 1996; Grimes, 1996; Henderson, 1997), eye-blinks (O'Regan, Deubel, Clark, & Rensink, 2000), motion picture cuts (Levin & Simons, 1997), real-world interactions (Simons & Levin, 1998; Levin, Simons, Angelone & Chabris, 2002), and in a paradigm where the whole image (including the target) moved gradually in a smooth fashion (Schofield, Bishop, & Allan, 2006), the presence of change blindness was also demonstrated in a display which did not involve the change occurring during a disruption. Simons, Franconeri and Reimer (2000) had participants view a scene for 12s, and during that period one object either gradually disappeared from the image, or gradually appeared. As in other change blindness studies, participants were poor at detecting the changes (typically noticing 64% of the changes). Simons et al. (2000) concluded that, consistent with other change blindness research, a small amount of visual detail is retained across successive views of a scene. They also suggested that the failure to detect gradual changes may occur because participants expect to notice the change easily, therefore choosing not to encode specific details of the scene.

This *change blindness blindness* (failure by participants to predict that they would miss large changes in change blindness tasks) has been demonstrated by on several occasions (Levin, Momen, Drivdahl & Simons, 2000; Levin, Drivdahl, Momen & Beck, 2002; Beck, Levin & Angelone, 2007) and is discussed later.

In summary, the dominant perspective on change blindness following the publication of the articles mentioned above was that participants do *not* encode and retain detailed memory representations for previously viewed stimuli. Instead, an impoverished and abstract memory representation exists, and this lacks sufficient detail to allow for successful change detection. Rensink (2000) proposed a Coherence Theory to explain these findings, and this is presented later.

1.3 Flicker paradigms

The studies presented so far used implicit change blindness tasks where participants were not aware that a change was about to happen. Around the same time, other researchers were using more explicit methodologies to study change blindness. Rensink, O'Regan and Clark (1997) developed the flicker paradigm which was based on two different areas of previous research. The first area was the study of the limitations of visual memory. Pashler (1988) investigated participants' ability to notice changes made to displays of letters that were presented for 100, 300 or 500 ms, briefly withdrawn (for 67 ms), then shown again. On half of the trials, a change was made to one of the consonants during this interval. Pashler found that higher change detection performance was found with longer exposure duration, although this difference was small (but statistically significant) – 59.6% for 100 ms compared to 69% for 500 ms exposure.

In Experiment 2, Pashler presented all stimuli for 300 ms and manipulated the inter-stimulus interval (ISI) – 34, 67 and 217 ms. During the ISI, the consonants were masked, either with a chess board-like black and white display, or a plain white display. Results showed that longer ISIs resulted in poorer detection performance – performance was best with a 34 ms ISI. There was also an effect of the mask, where performance was better without the mask than with it. There was also an interaction which showed that at 34 ms ISI and with no mask, performance was excellent. When a mask was presented at 34 ms ISI, performance dropped to similar levels of the 217 ms trials.

The results showed that visual information is retained in enough detail for at least a brief period, which allows participants to detect a change. If information was not retained, the changes would not be noticed. Pashler reported detection performance at a level above chance, providing evidence that at least some of the stimuli were represented in memory. However, eye movements were not recorded during this task, and therefore conclusions regarding the influence of visual processing on change detection were limited.

The second area of research that influenced the flicker paradigm did use eye movements. Earlier research had shown that changes made to the cases of letters in sentences were mainly detected if they coincided with the saccade target area. Otherwise, the changes passed undetected (McConkie & Zola, 1979). To notice the change, observers must simultaneously hold a representation of both A *and* A' (the pre- and post-change version). A comparison was required to detect the change (as noted earlier, a more difficult task than comparing A and B, as in the picture recognition literature). Rensink et al (1997) suggested that this processing demand exceeded normal human abilities, leading to change blindness. Poor change detection in saccade-contingent studies was attributed to blurring of the retinal image during saccades. Rensink et al. (1997) argued that this blurring masked the transient motion signals that indicated a change had occurred, resulting in a failure to notice changes.

In the flicker paradigm, scene A was presented for 240 ms, then a blank screen for 80 ms (a similar duration to Pashler), followed by a modified version of the original scene (A') for 240 ms. It utilised the free viewing conditions of the eye movement studies (e.g., McConkie & Zola, 1979). The sequence was presented in rapid alternation for either 60 s or until the participant detected the change and responded by pressing a button. The time taken to notice the change and the details of the change were recorded. When no blank screen was presented between scenes, or when the ISI was less than 80 ms, change detection was 100% as the change created a motion signal which was recognised by the sensory system (Shore, Burack, Miller, Joseph & Enns, 2006). Importantly, participants knew that a change would occur. This contrasts sharply with the real-world and video-based paradigms in which participants were *not* aware that a change would occur.

As mentioned earlier, this is a critical issue as potentially it influences the manner in which participants view and encode the scene. Change blindness tasks that neglect to inform participants that a change will occur potentially result in participants failing to encode specific details about the scene in memory (as it is unlikely that these details will be needed at a later stage). In this way, these tasks are a more accurate reflection of real life, as we do not usually need to remember the specific details of everything that we see. In contrast, the flicker paradigms explicitly instruct participants to search for changes. This is a more artificial way to view a scene (as during normal every day viewing we do not attempt to encode every detail about our visual environment), but also means that, potentially, participants are deliberately encoding specific details about the scene, and actively check if the currently available visual information is different to what is stored in their memory. Note though that the issue of target discriminability is still important; if the target is changed only in some very small way, it may still be missed if that particular feature has not been encoded. In a more general sense, giving specific instructions may change the way participants encode the visual information into memory. Potentially, specific task instructions may lead to deeper and more detailed memory representations than misleading (or the absence of) instructions.

Rensink et al. (1997) found that changes made to objects or areas independently rated as of marginal interest to the meaning of the scene took 10.9 s to detect, but changes made to areas of central interest were noticed after 4.7 s. However, when the blank screen was removed, both marginal and central changes were detected after only 0.9 s. When display duration for each image was increased from 240 ms to 560 ms, the same pattern was found – changes took approximately the same amount of time to detect, regardless of task duration.

The finding that central interest changes were noticed more quickly than marginal interest changes suggests a strong top-down component to change detection. Potentially, participants paid more attention during the task to those objects which they considered to add meaning to the scene. Furthermore, this suggests that participants searched for changes by looking first at the parts of the scene that were most meaningful, then searching less informative regions. This pattern of findings was demonstrated by Yarbus (1967) in which eye movement recordings showed convincingly that participants looked at the regions of a painting which were most relevant to the task instructions they had been given. For example, when instructed to estimate the ages of the people in the painting, most fixations were made on the faces of the people. By contrast, when asked to estimate the wealth of the people, fixations were mostly made on the objects in the room (e.g., paintings on the wall) and on clothing. This suggests that top-down influences direct attention to areas of a scene that are most informative.

Rensink et al. (1997) concluded that visual information may be retained only briefly in a short-term store and that potentially, the brief display periods (240 ms or 560 ms) meant that for all objects to be fixated once, the scene needed to flicker several times. On the assumption that when participants view scenes, average fixation durations are approximately 330 ms (Rayner, 1998), a scene containing 10 different objects presented for 560 ms on each "flicker" would need to flicker from A to A' for a period of 10 x 560 ms = 5.6 seconds before all 10 objects could be fixated. As objects typically need to be re-fixated in order to assess whether they have changed, it is perhaps unsurprising that changes sometimes took a long time to detect.

Importantly, eye movements were not monitored while participants searched for changes. Therefore, any conclusions made by Rensink et al. were based on response latencies and task accuracy, not on the oculomotor behaviour of observers. The researchers did not know which portions of the scene had been fixated, or the length of time each portion was fixated for. Critically, it was not known whether the changed object had been fixated at all before it changed. Potentially, the changes took a long time to detect because the brief display durations meant that objects took a long time to be fixated both in their pre- *and* post-change versions. Alternatively, participants may have fixated both versions but failed to encode the information in detail in memory. The findings that participants often took a long time to detect changes, and sometimes failed to detect changes at all, led directly to a new theory on scene perception called Coherence Theory (Rensink, 2000).

1.4 Why does change blindness occur? – Coherence Theory

The various methods used to explore change blindness led to the development of a new perspective which asserted that memory for scenes is relatively transient and impoverished, consisting mainly of abstract, not visually rich, details about the scene. A theory that encompassed these views was suggested by Rensink (2000) who proposed the Coherence Theory to explain the change blindness findings discussed previously. This theory can be distilled into three main claims. First, it states that immediately upon encountering a scene, a low level representation of the scene is formed. This low level representation contains a number of proto-objects. Proto-objects are volatile representations of objects in the scene, and lack detail. The formation of this low level representation occurs within a few hundred milliseconds. It can be considered as an abstract map of the whole scene, and occurs without the need for attention to be focused on different aspects of the scene. If this happens within a few hundred milliseconds, and we assume that average fixation durations whilst viewing scenes are typically 330 ms (Rayner, 1998), it would seem that the representation forms without the need for separate fixations on different portions of the scene. This claim is consistent with the idea that the gist of a scene can be extracted very effectively from brief presentations of scenes (Biederman, Mezzanotte & Rabinowitz, 1982; Schyns & Oliva, 1994; Intraub, 1997). This research indicates that some information about the meaning of a scene can be extracted very quickly, without the need for multiple fixations on different regions. According to Coherence Theory, as this map-like representation forms without attention being directed to different parts of the scene, it lacks coherence and is liable to decay.

Next, in order for this representation of the scene to become more detailed, focused attention must be directed to individual objects within the scene to allow them to become more stable, less susceptible to decay, and allow more information about them to be encoded. When attention (presumably via eye fixations, although the theory does not state this) is directed to an object, the proto-object representation is transformed into a more stable representation. Rensink states that focused attention can be thought of as a metaphorical hand which grasps proto-objects and holds them, allowing strengthening of the object representation. Coherence Theory states that four to six objects can be attended to (and therefore four to six proto-objects can be held) simultaneously. According to Coherence Theory, this limited amount of information that is stable at any one time explains why change blindness occurs. If the changed object is attended as it changes (meaning that a stable and detailed representation of it exists), then awareness of the change will occur. However, if the changed object is not the focus of attention as it changes, the change will likely pass unnoticed.

The third main claim is that once visual information is shifted to a different object then there is “little or no ‘after-effect’ of (an object) having been attended” (Rensink, 2000, pp. 20). Consequently, the change will likely pass unnoticed if the object is not the focus of attention at that time. As a result, Coherence Theory considers the change blindness findings, and claims that for a change to be noticed, focused attention must be on the changed aspect as the change happens. As only a small number of proto-objects can be held by attention at any one time, it follows that most aspects of a scene are not represented in detail at any one time. A change will only be detected if it happens to be the focus of attention at the moment of change. If it is not, the change will typically pass undetected.

Initially, it may seem as though Coherence Theory cannot possibly be correct. As observers, we generally exist successfully in our visual environment. How can it be that we do not represent visual information in a detailed way in memory? However, the change blindness literature has demonstrated that we do not automatically form a detailed and complete memory representation of the world from a single glance of a scene (or even after a longer period of viewing – changes in flicker paradigms often take more than 10 s to detect). Indeed, research by Nickerson and Adams (1979) demonstrated that participants are very poor at recalling and recognizing a penny coin, even though such an item is highly familiar. The perspective that visual memory is sparse and volatile was characterized by an earlier theory on visual perception, that of the world itself serving as an “outside memory” (O’Regan, 1992). This view on visual perception states that we use the visual environment as “a continuously available external memory” (O’Regan, 1992, pp. 484) and rather than representing what we see in memory, we just characterize what we see in abstract terms that allow us to understand what we see. If we need specific details, we need only move our eyes to the relevant part of our environment.

The results from the change blindness literature seemed to support Coherence Theory. These asserted that poor change detection was due to a limited number of objects being mentally represented at any one time, and the residual level of detail failed to support subsequent successful change detection. However, the research utilizing eye movement recording with flicker paradigms (Henderson & Hollingworth, 1999; Hollingworth, Schrock & Henderson, 2001; Hollingworth, Williams & Henderson, 2001; Hollingworth & Henderson, 2002) provided a challenge to the change blindness findings. The data did not support Coherence Theory, and Hollingworth and Henderson (2002) suggested an alternative view of scene perception to explain change blindness which they termed the Visual Memory Model (discussed in section 1.6).

1.5 A challenge to Coherence Theory

As stated above, Coherence Theory was drawn directly from the change blindness studies, where large and seemingly obvious changes were missed in a variety of different displays. Coherence Theory seemed to provide an explanation for this phenomenon. However, a further shift in perspective on change blindness occurred following a study by Simons, Chabris, Schnur and Levin (2002), who extended on the work by Simons and Levin (1998) by investigating whether preserved representations of either conversation partner existed in detail once the interaction had finished (recall that Simons and Levin switched conversation partner during a brief moment of occlusion). Simons et al. (2002) used the same paradigm as Simons and Levin (1998), but the change that occurred during a disruption was made to objects the experimenter was holding, not the experimenter himself. The experimenter was either (a) carrying an orange basketball which was removed during the disruption, or (b) not carrying a ball initially, but was holding one after the disruption, which lasted 2-3 seconds. However, the total length of the interaction was not reported, nor was the average time when the disruption occurred. Both of these things may contribute to the performance of participants. There may be a relationship between change detection rate and disruption onset, whereby a later disruption is more likely to be associated with a higher rate of change detection. This idea is supported by Agostinelli, Sherman, Fazio and Hearst (1986), who demonstrated that the accuracy with which participants detected additions and deletions in presentations of line drawings increased as the number of exposures to the stimuli before testing increased.

Six participants were randomly assigned to experience the addition of a ball, and five different participants experienced the removal of a ball. Once the interaction had finished, participants were asked if they had noticed anything strange, then asked if anything was different about the experimenter's appearance. If they answered "no" to both questions, they were considered to have not seen the change. Only three of the 11 participants reported noticing a change.

In Experiment 2, to ensure that participants had not simply pretended to have noticed the switch, an unusual red and white striped ball was used. Two control conditions (ball never there/ball always there) were included, and participants in these two groups were asked questions about a change that did not occur. Once again, the change occurred during the passing of a crowd of experimenters. Participants were randomly assigned to one of three groups: basketball removal, basketball addition, or no change. For half of the "no change" participants the ball was there all the time and in the other half it was never there.

Once the disruption had occurred, participants were systematically questioned to probe their memory for details of the interaction. Only 4 of the 13 participants in the removal condition noticed that the ball was no longer present after the disruption. In the addition condition, only 1 of the 13 participants noticed that following the disruption, the experimenter was holding a ball. In the no change condition, no participants reported a false alarm. However, following the systematic questioning, 69% of participants across the two change conditions *discovered* the memory of the pre-change image, despite failing to spontaneously report it. In this experiment, "discovered" means that despite failing to state that they had noticed the change, when asked about the presence/absence of the ball, participants could give accurate descriptions of the ball, demonstrating that they had some detailed memory for it.

This methodology allowed for a direct test of Coherence Theory. If, as predicted by Coherence Theory, only the currently attended object is stable and detailed, then a change to the ball (which was hidden, and not the focus of attention at the moment of change thanks to the visual disruption) should pass unnoticed. Despite the initial high rate of change blindness which suggested that participants were not aware of a change, the subsequent questioning of participants demonstrated that participants could accurately remember its pre-change form and compare that to its post-change form, leading to correct change detection. This directly contradicts Coherence Theory, which states, "there is little or no "after-effect" of (an object) having been attended." (Rensink, 2000, pp. 20).

Simons et al. (2002) concluded that, contrary to previous conclusions, participants do appear to retain detailed visual information in memory. They stated, "the finding that some observers appear change blind even when they are subsequently able to recall some details provides a strong argument against the use of change blindness to infer minimal representations of scenes" (pp. 92). Therefore, the evidence suggests that a lack of *comparison* causes change blindness, not the lack of detailed memory representations. Consistent with these findings, Mitroff, Simons and Levin (2004) used displays of line drawings to demonstrate that change blindness occurred despite participants exhibiting above-chance recognition of both pre- and post-change targets, again demonstrating that a lack of comparison, rather than a lack of memory, was influencing change blindness.

This shift in views concerning the contribution of change blindness to the scene perception and memory literature was also addressed in a review paper by Simons and Rensink (2005) who expressed doubt over the claims previously made by change blindness research. They stated that it was possible that change blindness may occur even with representations of visual information – for example if the two representations (pre-change and post-change) were never compared. Simons and Rensink acknowledged research by Hollingworth, stating that "mounting evidence suggests that some information is preserved even when observers fail to detect changes" (pp. 19).

Subsequent bodies of research have provided strong evidence that a large amount of detailed visual information *is* retained from previously viewed stimuli, and that the change blindness findings are most likely a product of unusual and unrealistic viewing conditions and tasks. Despite this, the change blindness literature is interesting and influential, as it challenged traditional views on attention and memory for visual stimuli. Coherence Theory (discussed above) is almost the complete opposite to the composite scene representation hypothesis. Whereas traditional thinking claimed that observers quickly form a highly detailed memory representation of scenes, Coherence Theory asserted that only impoverished and abstract representations are formed. In response to these earlier theories, the Visual Memory Model was proposed. Theoretically, it sits somewhere between the two more extreme views discussed in 1.4 above. Section 1.7 presents more evidence against Coherence Theory, and suggests an alternative perspective.

1.6 Basic characteristics of eye movements

A further challenge to Coherence Theory, and the perspective that memory representations lack detail and decay quickly, was posed by the use of eye movements in change detection tasks. Whereas the initial research using flicker paradigms did not record eye movements as participants searched for changes, later research that introduced changes as participants viewed scenes did. This enabled researchers to investigate the encoding processes that occurred as observers actively searched for changes. It also allowed researchers to investigate at which point during scene viewing participants detected the change. Subsequently, this allowed researchers to understand whether participants were missing changes simply because they did not fixated the changed object both before and after it changed, or whether the changes passed undetected *despite* being fixated before and after (suggesting that failure to encode and/or retain detailed information caused change blindness).

The utilisation of eye movements in studying scene perception was a very important step in understanding how observers visually process stimuli. It allowed researchers to know whether participants were looking at critical aspects in scenes, and therefore to draw different conclusions as to the nature of scene perception, memory, and change detection. Before presenting the research which used eye movements to further understanding in visual processing, it is important to first consider some basic characteristics of eye movements.

It is well established that during a variety of tasks (including scene perception and reading), the eyes regularly make rapid movements (called *saccades*). The periods in between the saccades when the eyes are still are called *fixations.* It is during these fixations that information about the visual environment is extracted and encoded. During scene perception, average fixation durations typically last for 330 ms (Rayner, 1998). Research has also demonstrated that extraction of visual information is greatly reduced during saccades. This is known as *saccadic suppression* (Rayner, 1998).

Consequently, during scene inspection visual information is not encoded in a continuous stream (although as observers it may feel that way).

Rather, each time a saccade is made to move the eyes to a new location and allow new information to be fixated, visual input is temporarily interrupted as saccadic suppression does not allow information to be encoded during the saccade. As a result, the visual input received by observers consists of brief (approximately 330 ms) periods where the eye is still and information can be extracted, and these brief periods during which information can be encoded are continually interrupted by saccades. Despite this, our perception of the visual world is that of a continuous visual input (e.g., during normal viewing we do not notice that saccadic suppression occurs).

A second key issue regards the function of saccades. Why do we move the eyes when this consistently suppresses visual input? The answer lies in the biological layout of the eye. The human visual field is typically divided into three distinct areas. First is the *fovea*, the central part of the visual field. The fovea processes the central two degrees of the visual field. Visual acuity is clearest at the fovea, meaning that information can be most clearly viewed when the eyes move and place the fovea directly towards whatever we want to look at. In the simplest terms, to see something clearly, it is best to look directly at it, rather than look near it. The region of the visual field outside the fovea is called the *parafovea* and visual acuity is poorer here. The parafovea typically processes the part of the visual field from two to 10 degrees.

The third region which has the poorest acuity is the *periphery.* The periphery processes any part of the visual field outside the central 10 degrees. Processing is typically most successful if a saccade is made to put that information into the fovea. This explains why saccades are made so often. Saccades allow information to be viewed in the clearest possible way, and as a result we constantly move the eyes to allow the visual environment to be seen accurately. Although research has demonstrated that participants can process information and detect changes presented in the parafovea and periphery when they are not allowed the make saccades and have to fixate a central cross (e.g., Hollingworth, Schrock & Henderson, 2001), this typically results in slower responses and more errors. Hence, the most optimal way to view visual stimuli is to make regular saccades which bring the information required into the centre of the visual field allowing the fovea (with the highest acuity) to scrutinise the stimuli.

This is a central issue in scene perception research. If it is correct that visual information is best viewed at the centre of the visual field, it then follows that for a scene to be encoded in detail into memory, each part of that scene must be fixated in turn. This means that a saccade is made to that region, and subsequently a fixation. If this does not happen, it is unlikely that the information at the unvisited regions will have been encoded. Although some information can be extracted without being fixated, moving the eyes to directly fixate an object is the behaviour most likely to allow information to be extracted in detail. This should be kept in mind when considering change blindness.

A final critical issue regards the relationship between attention and eye movements. Although it has been demonstrated in specific tasks that observers are able to decouple attention and eye movements (Posner, 1980), with complex images such as reading and scene perception the optimal way to process and encode information is to keep these two processes linked together by moving the eyes to the region that is being attended. There does not appear to be any benefit during scene perception to fixate one area of the scene but pay covert attention to a different region. To extract information in the most efficient way, we attend to the location where our eyes are fixated. This is also an important issue as during scene perception experiments where participants are allowed to move their eyes freely, eye movements are an excellent indicator of what is being attended to (Liversedge & Findlay, 2000). If this is true, then using eye movements to study where people look when they view visual scenes allows us to make predictions about which objects will be retained in memory, and potentially the degree of detail of that memory.

1.7 Further challenges to Coherence Theory: Eye movements and change detection tasks

In this next section, studies are reported which record eye movements while participants complete change detection tasks similar to the flicker paradigms presented above. However, it is important to note that most of the studies that will be presented did *not* use flicker paradigms. Traditional flicker paradigms introduce the change during an artificial visual disruption (e.g., a blank screen inserted between A and A') and also repeat the change until it is detected, or a pre-defined interval passes. Instead, most of the following studies present the change during a more natural visual disruption (during a saccade rather than a blank screen) and the change only occurs once. The image presentation order is image A (for a pre-determined time period or until the n^{th} saccade), a change made during a saccade, then A' until response. In this way, researchers can empirically control when the change occurs, and subsequently can deliberately introduce a change only once the to-be-changed object has been fixated, or alternatively introduce the change before it has been fixated. This allows for direct investigation into the reasons that participants miss changes in scenes.

In one such change detection task, Henderson and Hollingworth (1999) had participants view scenes on a computer in preparation for a memory test. The test did not occur; the mention of it encouraged participants to explore the scene. Participants were also told to report any changes (which were made to pre-determined objects in the scene). Each scene consisted of a number of different objects arranged to form realistic-looking environments (e.g., a living room). The target object that changed could be any of the objects in the display. Two types of changes were employed. First, a change where the target object (e.g., a clock) was rotated 90° during a saccade. Second, an object was deleted from the scene during a saccade. These changes were either made during the first saccade *toward* the target, or during the first saccade *away* from the target. A control condition was also included in which the change occurred during a saccade to a non-target object (i.e. when the eyes were in a different region of the screen to the changed target). Catch trials (with no change present) were also included.

The results showed that changes were poorly detected when neither the pre- nor post-change fixation was made on the target (the control condition). Rotations were missed more than 90% of the time, and deletions were missed more than 60% of the time. As in the change blindness research discussed earlier in section 1.2, changes made to objects in visual environments often passed unnoticed. However, what Henderson and Hollingworth (1999) demonstrated was that changes were much more likely to be detected if the pre- or post-change fixation was made to the changed object. This was especially true for object deletions, which were detected twice as often in trials where they were fixated directly before or after deletion, compared to trials where they were not fixated before or after the change. Thus, Henderson and Hollingworth demonstrated that fixation position is crucial in detecting changes. Fixating an object greatly increases the likelihood that a change made to it will be detected. Conversely, changes made to non-fixated objects are less likely to be detected.

Hollingworth, Schrock and Henderson (2001) also recorded eye movements in a change detection task. Note that on this occasion, a standard flicker paradigm was used with a 250 ms display duration and an 80 ms blank interval. Two types of changes were made; one similar to Henderson and Hollingworth (1999) where an object was deleted and another in which an object was rotated 90° between images A and A'. Participants were required to state when they saw a change, and to report the identity of the change.

The results showed that object deletions were detected significantly faster than rotations (2.76 s vs.4.37 s respectively). This is not a surprising result; firstly, if one considers the issue of target discriminability, it could be argued that it is easier to distinguish between an object being present and then disappearing, and more difficult to distinguish between two objects which are identical in every way except for the direction they are facing. In that way, deletions could be considered to be a more easily discriminable change than an orientation shift. Additionally, deletions result in larger visual disruption to the overall gist. There was a range in detection latency – the quickest change was spotted on average after only 1.1 s whereas the most difficult change typically took 12.9 s.

As reported, the rotations were detected after 4.37 s, a response time comparable to the time taken by participants in the Rensink et al. study (4.7 s). This suggests that the rotations used by Hollingworth et al. (2001) were similar in terms of attracting attention to the central interest objects used by Rensink et al. The results also showed that, at response, 74.5% of fixations were directly on the target object. This suggests that when changes were detected, the eyes were more likely to be fixating the target object than anything else – strong evidence for a link between fixation position and correct target detection.

In Experiment 2, half of the participants had to fixate centrally whilst trying to detect the changes. The other half of the participants could freely view the scene. The data showed that when participants were not able to move their eyes but had to covertly shift their attention around the scene, performance was poorer (91% detection accuracy) than participants who could freely view the scene (99.4% detection accuracy). These participants detected fewer changes, made more false alarms, and took longer to find changes than participants who could freely view the scene. These results suggest that fixation position plays a crucial role in detecting changes in scenes, although it is crucial to note that it is not essential to fixate an object in order to detect a change to it. Nonetheless, directly fixating objects leads to quicker and more accurate change detection.

Hollingworth, Williams and Henderson (2001) also recorded eye movements during a change detection task in which a token replacement was made during the first saccade away from the target. Participants were instructed to search for any changes that occurred during viewing. Overall, 26.6% of the changes were detected. Further analysis showed that of all the occasions on which a change was noticed during viewing, 41% of these correct responses were made more than 1500 ms after the change had occurred. Of these, 94% occurred only once the target had been re-fixated. The results suggest that detailed information about the target object was retained, and was used to support a comparison with the post-change version upon re-fixation of the target object, but rarely referred to before the target had been re-fixated. In turn, this suggests that to successfully notice a change, the changed object must be fixated both prior to, and after the change is made.

In line with the other studies using eye movements to study changes made in change detection tasks, the results supported the conclusion that detailed visual information *is* stored in memory, and used to support accurate change detection. Critically, successful change detection is very closely linked to fixation position. Changes are generally detected only when the changed aspect has been fixated in its pre- and post-change state. Once this has occurred, successful change detection usually occurs. However, it is important to note that fixating an object does not guarantee that a change will be noticed. In the studies discussed above, change detection rates were not at 100%, demonstrating that some changes still pass undetected despite participants moving their eyes towards or away from objects as they changed. However, what they do strongly indicate is that fixating an object (both before and after the change) means that a change is much more likely to be noticed than if the object is not fixated.

These studies also demonstrate that if an object is fixated both before and after the change, and the change is detected, some detailed memory for the pre-change version must exist in order to allow for successful change detection. Note, however, that other studies (e.g., Tatler, Gilchrist & Land,

2005) show that some object information can be remembered even without making a fixation directly on an object, suggesting that directly fixating an object is not essential. However, the likelihood with which an object will be remembered increased when the object was directly fixated, demonstrating objects are much more likely to be remembered if they are fixated than if they are not.

Hollingworth, Williams and Henderson (2001) reconsidered the change blindness literature in light of their findings. In previous change blindness studies, the conclusion had been widely drawn that visual representations of scenes were relatively impoverished, hence poor change detection. Hollingworth et al. (2001) argued that in change blindness research, change blindness may occur for one of several reasons. First, the changed feature may fail to be fixated prior to the change occurring. If this happens, change blindness is a likely outcome as there has been no opportunity to form a detailed memory representation for that object. Second, if the changed object is not re-fixated after the change happens, that change will likely be missed (as only detailed memory for the pre-change version exists in memory). Third, if no comparison between the pre- and post-change version is initiated, the change will be missed even if post-experimental questioning demonstrates accurate memory for both forms.

A further study conducted by Hollingworth and Henderson (2002) had participants view computer-rendered scenes for 20 s in preparation for a subsequent memory test whilst recording eye movements. Similar to Henderson and Hollingworth (1999), changes were made during a saccade, and participants were told to try to identify any changes that occurred during scene viewing. However, instead of using rotations and deletions, objects were either replaced by a different type of object (e.g., a notepad was replaced by a different notepad) or a token change, in which it was replaced by a similar type of object (e.g., a book). In the *change-after-fixation* condition, the change was only made once the participant had spent at least 90 ms looking at the target (to ensure that the participant had viewed the target before it changed). In the *change-before-fixation* condition, the change was made almost immediately after the trial onset (before the participant had fixated the target). There was also a control condition in which no changes occurred (and these control scenes were the subject of the memory test).

When the change occurred after fixation, 51.1% of type changes were detected compared to 28.4% of token changes. For both type and token changes, it was found that more changes-after-fixation were detected than changes-before-fixations. The results provided evidence that a change is significantly more likely to be detected if it has been fixated before it changes than if it has not been fixated before it changes. In turn, this also suggests that detailed visual information is retained in memory after fixating an object, and this memory enables change detection to occur.

Additionally, participants completed a 2AFC memory test after viewing all the scenes which probed their memory for the control scenes. Participants viewed each control scene, and a modified version in which either a token or type substitution had been made. This test happened at the end of the experiment (the delay between initially viewing the control scenes and being tested on them ranged from 5 – 30 minutes). Participants in the type-discrimination condition accurately identified 93.1% of the original scenes, and 80.6% in the token-discrimination condition (both significantly above chance). This demonstrates that, despite the test delay, participants could accurately discriminate between the original viewed version of a scene, and a modified version in which one aspect had been changed. Thus, the findings posed a problem to the notion that memory representations are limited to details regarding the currently attended object. The finding provided evidence that detailed information about previously fixated objects is retained, and is explicitly available for report, even after a delay of up to 30 minutes.

Hollingworth and Henderson concluded that, contrary to Coherence Theory, information about previously fixated objects *was* retained in memory rather than decaying once they were no longer the focus of attention. This information about objects was used to make a comparison if the object was re-fixated, and the eye movement data showed that changes were typically only detected once re-fixation had occurred. All of these studies using eye movements in flicker paradigms came to the same conclusion: in flicker paradigms, a change will most likely be detected only once three things have happened. First, an object is fixated (and encoded and retained) in its pre-change state. Second, it is fixated (and encoded and retained) in its post-change state. Last, the two memory representations are compared. If the changed aspect has been encoded and retained in memory, successful change detection is likely to occur.

1.8 Why does change blindness occur? – Visual Memory Model

Having conducted research that proved problematic for Coherence Theory, Hollingworth and Henderson (2002) proposed a new theory to explain scene perception and the existence of change blindness called the Visual Memory Model (VMM). It begins by claiming that once attention is paid to an object in a scene, representations are formed for that object that include a low-level sensory representation (based upon low level features), and a higher level detailed representation (based on higher level features). For example, these representations may consist of basic details (such as shape or colour), and more complex details (such as higher-level conceptual representations regarding meaning and identity). Once these representations are formed, they are bound to a position in an overall spatial map which codes their specific location within the overall scene. A short-term object file is therefore formed for that object. This engages both visual short-term memory and conceptual short-term memory.

If attention (via eye fixations) is focused on an object, more information will be added to the object file, and it will become more richly detailed. It is important to note that these short-term object files are quite weak, and are liable to be replaced if attention is diverted to another object. However, at some point during viewing (potentially during the first fixation, although the model does not specify) the information about the object is consolidated into long-term memory, forming what Henderson and Hollingworth call a "long-term object file". This includes the low and high level information extracted initially about the object, plus its location within the spatial map, in addition to any extra information that has been extracted whilst attention has been focused on it.

Once that object is no longer the focus of attention (because the eyes have moved to a new location) the initial short-term representation decays almost immediately, leaving behind only the long-term version. This long-term version is more stable and robust than the short-term memory representation. Importantly, the VMM is not clear on whether the decay of short-term representations is due to the withdrawal of visual attention, or due to the representation being overwritten or replaced by subsequent stimuli. However, the model suggests that the STM representations are retained at least briefly (approximately four items downstream). That is, the currently fixated object as well as the previously fixated three objects may co-exist in short-term memory.

Therefore, if a scene is viewed for an extended period of time, multiple fixations are likely made to different regions. Object and scene information builds in LTM from each fixation and is indexed to the overall spatial map. Only a small amount of information is kept in STM (potentially four objects). This information in STM is swiftly replaced by new information, but the long-term versions remain. These long-term memory representations of objects contain a high level of detail, and can be accessed at a later time if needed, hence successful change detection when they are accessed.

The model also proposes that any retrieval of the LTM representations for previously attended objects (and, in the case of change blindness paradigms, comparison of the stored information with the information currently available to perceive) is influenced heavily by visual attention, and this is driven by fixation position. Returning to a previously viewed spatial location is proposed to engage the existing LTM object file. Hollingworth and Henderson (2002) provided evidence that supports this final claim. They demonstrated that changes made to previously viewed objects were typically detected upon the first re-fixation of that object (which was still in the same spatial location). Also, research by Henderson and Hollingworth (1999) demonstrated that participants sometimes detected object deletions only when they fixated the original spatial position in which the object had previously been placed – obviously the original object had not acted as a cue to detect the change as it no longer was present. This suggests that attending to the marked spatial location acted in itself as a cue, and consequently to the retrieval of the LTM object file. Additionally, Richardson and Spivey (2000) demonstrated that when participants were presented with to-be-remembered video stimuli in a 2x2 array on a computer screen, at a later test they often made a saccade to the corresponding quarter of the screen, even though the screen was now blank. This suggests that moving the eyes to the previously occupied region of space facilitates memory recall.

The VMM also proposes that retrieval from LTM of the higher-level representation information that relates to the specific viewpoint of the scene explains how participants are able to notice rotation and token changes, and also to perform well on orientation discrimination tasks. Finally, when the scene is no longer available to view (or attention is no longer paid to objects in the scene), the LTM representation that exists consists of the overall spatial map with individual indexed objects attached. During any later perceptual experiences with the scene, the original scene map is retrieved. If the observer fixates a particular area of the scene that contains information about a specific object, that information becomes available to access.

The VMM clearly differs from Coherence Theory in terms of the way it describes the acquisition and retention of visual information. Coherence Theory states that once attention is withdrawn from an object, only an abstract and impoverished version of it remains. The VMM, however, states that a detailed version exists in LTM, and this can be accessed later if needed. Additionally, the VMM states that information from temporally separate fixations can be integrated and help accumulate information across time. If an object is re-fixated, information can be added to the already existing memory representation. At a fundamental level, the two theories offer different predictions as to the outcome of attended stimuli. Whereas Coherence Theory states that during normal scene viewing, memory representations are impoverished, lack detail and decay quickly, the Visual Memory Model asserts that a detailed memory representation is formed. This representation has a long-term component, and as such, accurate memory for specific visual details will be exhibited even after a delay.

1.9 Why are change blindness results interesting?

In all of the studies presented in this chapter, failure to detect changes in scenes has been consistently observed. Although the findings of change blindness led to the perspective that memory for visual scenes is relatively impoverished, the literature discussed in the next section, coupled with the research utilising eye movements in flicker paradigms, provides evidence which strongly challenges this notion. So what made change blindness findings interesting? Change blindness research challenged the assumption that as observers encounter a scene, they automatically form a highly detailed and robust memory representation for that scene. Rather, any memory that forms seems to accumulate somewhat gradually in detail across the extended viewing period. If a change is made to part of a scene that has not been fixated, it is perhaps unsurprising if the change passes unnoticed.

Accordingly, Noe, Pessoa and Thompson (2000) concluded that change blindness findings are only surprising if one adheres to the constructionist view of visual perception. That is, we automatically form a complete and detailed representation of everything that we see. In support of this, Levin, Momen, Drivdahl and Simons (2000) demonstrated that naive undergraduate observers consistently believe they will notice changes made in the environment (such as the real-world change blindness studies) although the literature convincingly demonstrates that this is not the case. This change blindness blindness has also been demonstrated by Levin, Drivdahl, Momen and Beck (2002). Noe et al. (2000) questioned why change blindness results have been considered to be so surprising, and concluded that this surprise is the same as the wonderment felt by observers watching a skilful magician. As observers, we tend to feel that we will notice anything that happens in front of us. However, this is not true, and our perceptual limitations are exposed equally by magicians, and change blindness paradigms.

Indeed, Kuhn, Tatler, Findlay and Cole (2007) demonstrated that experimentally naïve observers watching a video of a magician performing a simple magic trick in which objects “disappeared” detected the disappearance less often than participants who were told that the objects were about to disappear. When participants were told what was about to happen, they typically noticed it happening. In change blindness studies, participants are not told specifically which objects will change (and in some paradigms, they are not aware that any change will occur). This ambiguity is likely to contribute to the high rates of change blindness. It is logical to suggest that in flicker paradigms, if participants were told which object would undergo a change, they would quickly and easily detect that change.

However, when participants have to search for the changed object, successful change detection takes longer. Therefore it appears incorrect to conclude that our memory representations are sparsely detailed. Instead, change blindness findings should be taken as a demonstration that we do not immediately represent an entire scene in memory. Rather, this information is argued to accumulate gradually across time (and presumably, fixations), and if a change occurs before that aspect has been encoded, it is highly likely to pass undetected. Change blindness allows researchers to gain a deeper understanding of how memory and attention interact to contribute to our perception of the world, as well as challenging assumptions about the visual system.

1.10 Research into memory for object identities

The studies discussed in the next section all provide strong evidence that detailed visual information regarding object properties (e.g., presence, shape, colour) is retained robustly in memory. One such study was conducted by Hollingworth (2004) which investigated whether VSTM and VLTM both contribute to memory representations of visual scenes. This study also sought to investigate whether Coherence Theory or the Visual Memory Model was a more accurate explanation for scene perception. Hollingworth used a follow-the-dot paradigm which allowed for precise control over the order in which stimuli were fixated. Participants followed a neon green dot as it moved around a computer-rendered image, shifting location from object to object and visiting each object in turn. In each scene, a target object was pre-selected, and the serial position of the target in each sequence was manipulated. For example, in a 4-back condition, 4 other objects were visited between the target and the test. In a 0-back condition, the currently fixated object was tested (0 items were visited in between target and test). At the end of each sequence, the target object was masked and memory for the object was tested using a two-alternative forced choice test (2AFC). Two versions of the scene were presented; the original viewed scene, and a modified version of the scene in which the target was either rotated 90° or replaced by a similar, plausible object. Participants were required to select which of the two options was the original scene. This same 2AFC test was carried out for all of the experiments reported in the Hollingworth (2004) paper.

Several experiments were conducted. In Experiment 1, three serial positions were tested: 1-back, 4-back and 10-back. The 1-back condition was chosen as it falls within existing estimates of STM capacity – the most recently fixated object (Pashler, 1988). The 4-back condition was selected to test the limits of VSTM, as suggested by Coherence Theory (Rensink, 2000). The 10-back condition was chosen as being a serial position well outside even the most liberal of VSTM limits. Hollingworth stated that a recency effect (high performance in the 1-back condition compared to the other two conditions) would provide evidence of a STM component to visual scene memory. Similarly, a pre-recency (or primacy) effect, shown by accurate performance in the 10-back condition, would suggest a LTM element to memory for objects in scenes.

The results showed that a higher percentage of changes were detected in the 1-back condition than the 4-back condition. The 4-back and 10-back conditions were not significantly different, and there was a trend towards higher performance in the 1-back condition compared to the 10-back condition. The accurate performance in the 10-back condition suggested a contribution of VLTM to online scene representation as 10 objects had been viewed in between viewing of the target and test, yet participants could accurately select which of two versions of an object they had previously viewed (despite not knowing which of the 10 objects they would be tested on). These results support the Visual Memory Model suggested by Hollingworth and Henderson (2002) but cause problems for Coherence Theory which would predict poorer performance for the 10-back condition.

Experiment 3 compared 0-back, 2-back and 4-back using identical methodology to Experiment 1. Results showed 0-back performance was significantly better than both 2-back and 4-back, and that 2-back and 4-back were not reliably different. This suggests that the VSTM component may be limited to 2 serial positions; the currently fixated object, and the previous object. Experiment 4 corroborated these results by demonstrating a high level of performance for 0-back and 1-back compared to 4-back.

Experiment 5 showed that even if the test was left until the end of the session (an average of 12.1 minutes and 402 trials) performance was still above chance (although inferior to immediate testing). The overall picture of results that emerged from this research posed a problem for Coherence Theory, the results from change blindness studies, and the perspective that poorly detailed visual information is retained once a visual scene is no longer visually available. Hollingworth provided strong evidence that not only is visual information regarding objects in scenes retained and used in subsequent memory tasks, but that it can last at least 12 minutes, and 402 trials. Therefore, he concluded that visual representations are supported by both visual short-term memory and long-term memory.

Hollingworth (2005a) also investigated the contribution of long-term memory to memory for object identities. In Experiment 1, participants freely viewed each of a series of 42 computer-rendered images in preparation for a change detection task. Each image was inspected for a 20 s period. For half of the scenes, the change detection task for that image occurred immediately after viewing the image for 20 s (an immediate test). The change detection task for the other half of the scenes was delayed until all 42 images had been viewed (delayed test). On average, the delayed test occurred after a 9.2 minute delay. One of two possible changes was made: either an object was rotated 90º, or was replaced by another version of the same object (a token change). All changes occurred during a 200 ms patterned screen visual disruption. Participants were required to state which object had changed. The results showed only a non-significant reduction in accuracy when test occurred with a delay. Therefore, the results provide evidence that memory representations of object identities are retained in long-term memory (supporting the picture recognition literature). Additionally, this finding demonstrates that long-term memory for scenes consists of specific object details (e.g., the ability to correctly remember which orientation an object was presented at) rather than lacking detail as suggested by change blindness findings.

In Experiment 3, identical methodology was used except that the test delay was extended to 24 hours. The results showed that performance on both the rotation and token change detection tasks was significantly poorer after a 24 hour delay than when tested immediately, although performance after 24 hours was still significantly better than chance. This demonstrates that even after 24 hours delay, participants retain specific information about object identities.

Another study that demonstrated accurate memory for object identities was conducted by Melcher (2001). Participants viewed computer-generated scenes containing 12 scene-congruent objects in preparation for a memory test on the details of the scenes. Some scenes were viewed for 1, 2, 3 or 4 continuous seconds (continuous-presentation trials). Other scenes were viewed on several occasions for brief views of 0.25, 1 or 2 seconds (re-test trials). For example, this allowed for a comparison of memory performance between scenes viewed for 4 consecutive seconds, or 4 separate one second views. At test, participants were required to recall as many objects as possible. Interestingly, the results showed that performance for trials where the same scene was viewed on more than one occasion (re-test trials) was comparable to continuous-presentation trials. Memory performance was comparably good regardless of whether scenes were viewed for a continuous viewing period (e.g., 4 s) or several interrupted shorter viewing periods (4 x 1 s). Melcher suggested that not only did this provide evidence that detailed visual information regarding object identity is accurately encoded and retained in memory, but also that these memory representations accumulate across separate views (and, presumably, separate fixations).

In related work, Melcher and Kowler (2001) showed participants computer-rendered scenes which contained 12 objects for 1, 2, or 4 seconds at a time whilst recording eye movements. After each trial, one of several possible tests was administered. First, for free recall tests, participants were simply required to list objects they had seen in the image. Results showed that increased scene viewing duration led directly to a higher amount of accurately reported objects. After 1 second of scene viewing approximately 3-4 objects were recalled, and after 4 seconds of viewing, typically 5 objects were recalled. Similar to Melcher (2001), this demonstrates that information regarding object identities accumulates with extended viewing time.

On some trials, the scene was presented on several different occasions before the test was presented. The results showed that performance improved when a previously viewed scene was tested. Memory performance was facilitated by being tested on a scene that had been viewed on more than one previous occasion. This provides evidence that memory for object identities within a scene is robust enough to persist beyond one single view, meaning it is presumably stored in long-term memory.

Melcher and Kowler (2001) also provided evidence to suggest that the representation is visually-based, and not verbally-based. In a separate experiment, the objects in the scenes were replaced by verbal labels, for example “umbrella”. They suggested that if the representations were verbally-based then performance on the tests would be comparable, or superior, to trials using the objects themselves. Results showed that despite comparable performance for tests that immediately followed the first presentation of the image, tests that followed repeated viewing of the scenes did not show accumulation of the verbal labels. This suggests that memory representations are visually-based, not verbally-based. This finding corroborates the results of Hollingworth (2005a, Experiment 2B) which found no reduction in change detection performance when a concurrent verbal working memory task was carried out.

Similarly, Liu and Jiang (2005) presented photographs containing 10 objects for 250 ms, followed by a coloured mask for 250 ms. An interval of 1000 ms followed, then the test. The memory test presented a display of 20 objects containing the 10 observed objects and 10 plausible distractor objects. Participants were required to indicate the 10 objects they had viewed in the observation stage, and also to report their confidence for each object. Half of the participants were also required to repeat “teddy bear” aloud throughout the experiment to investigate the effect of a concurrent verbal task.

The results indicated a significant effect of recognition order (the order in which participants chose the 10 objects from the display of 20) on both accuracy and confidence. This demonstrated that objects selected early in the test phase were associated with high confidence and high accuracy. The articulatory suppression task had no effect on performance whatsoever, suggesting that verbal encoding did not play a role in the forming and maintenance of visual representations. Again, these findings are in line with Melcher and Kowler (2001) and Hollingworth (2005a). The results also showed that the first object selected was accurate 75% of the time - well above chance. However, objects 2-10 were selected no better than chance.

In Experiment 2, a 2x2 design was used where encoding duration and background type were manipulated. Half the participants viewed the scene for 250 ms, whereas the other half were allowed to view the scene for as long as they wanted (self-paced condition). Additionally, half the participants viewed natural scenes containing 10 objects (as in Experiment 1) whereas the other half viewed the 10 objects on a plain white background, removing the context of the scene. The test phase consisted of a display of 4 objects containing 1 object from the original display and 3 plausible distractors. Participants were required to select the previously viewed object.

The results showed better performance in the context-absent than context-present condition and also better performance in the self-paced than 250 ms condition. In the self-paced conditions, participants viewed the scene on average for 13.7 s when context was present and 13.2 s with a blank background. A positive correlation was observed between viewing time and accuracy. Viewing the scene for a longer period of time resulted in better performance at test. As Experiment 2 only tested memory for 1 object, the total amount of objects retained was estimated. It was estimated that in the context-present condition, 0.67 objects were remembered when the scene was presented for 250 ms (similar to Experiment 1) and 5.33 objects were remembered for the self-pacing condition. In the context-absent condition, 250 ms viewing time resulted in 2.1 objects and 7.41 for self-pacing. Liu and Jiang estimated the number of objects remembered by using the following formula: Accuracy = [N x 100% + (10 – N) x 25%]/10. This formula used the overall accuracy of each participant to estimate how many objects they had remembered correctly.The results demonstrated that a single 250 ms view does not allow all 10 objects to be fixated. Therefore, it is unsurprising that changes made using the flicker paradigm (e.g., Rensink, O'Regan & Clark, 1997) take a long time to detect, as they typically present the images for 250 ms at a time. The results suggest that one object can be encoded in that period. Therefore, for successful change detection, participants must serially search the image until they have encoded the pre- and post-change versions of the target. Assuming that one object is encoded every 250 ms, it is unsurprising that successful change detection in flicker paradigms can require a long viewing period. However, unlimited viewing time allows for encoding of more objects, suggesting that the details of a visual scene can be maintained in memory once the scene is no longer present.

The results of Liu and Jiang are interesting as they suggest that when participants are allowed an unlimited amount of time to view a display of 10 objects, on average they only correctly remember 5.33 of them. However, Liu and Jiang did not consider the way in which the scene was visually processed, as eye movements were not recorded. Potentially participants fixated only 6 of the 10 objects in the scene. If that were true, then remembering 5.33 demonstrates excellent memory for viewed stimuli. By contrast, if all 10 objects were fixated, then remembering 5.33 reflects that more than 4 objects have been forgotten. This issue is addressed by the present research.

Hollingworth (2006) showed participants 3D-rendered natural scene for 20 s, followed immediately by a forced-choice recognition task. Additionally, the target object was either presented within its surrounding context (background-present condition) or superimposed onto a plain background (background-absent condition). Participants had to either identify a token replacement (i.e., the target object was now a different object) or 90º rotation of the target. Hollingworth suggested that in order to perform either task well, memory for the original visual form of the target was required. If this did not exist, then performance would be equal to chance. Hollingworth predicted that if object representations are bound to the scene context then performance should be better for background-present trials. However, if object representations are stored independently then performance should be facilitated for background-absent trials.

The results demonstrated that, contrary to the findings of Liu and Jiang (2005), performance was significantly better when the background was present as opposed to absent (88.3% versus 79.8%) and both were at a higher rate than chance (50%). Token replacements were detected significantly more often than rotations (87.3% versus 79.8%). These results showed that detailed visual information regarding object identity is stored in memory, and able to be accessed when needed. Also, memory for an individual object seems to be stored as part of a larger overall scene representation. This is an interesting finding as it suggests that memory for object identity is not isolated from its surroundings. This suggests that the environment in which the object is located contributes to memory for that object. It may also be that the location in which an object is situated is encoded along with the identity of the object. This is an issue that the research presented in this thesis will explore in more detail.

In Experiment 2, identical methodology was employed, except that the test was delayed so that one trial came between the viewing of the scene and the test. The results showed that despite the delay in performing the test, accuracy rates were comparable to Experiment 1. The same patterns of results (better detection of token changes; better performance in background present trials) were observed. This led Hollingworth to conclude that visual object memory is robust, and that the scene context plays a central role in object perception. Objects are not remembered in isolation, but rather as part of the overall scene. Similarly, Hollingworth (2005b) demonstrated that a 20 s scene preview facilitated later placement of an object on a blank screen relative to a no-preview condition, and that a target-present scene preview led to more accurate object placement than a target-absent scene preview. These findings are important as they consider the location of an object within a scene, as well as simply the object itself. Memory for the location of an object within a scene has been investigated substantially, and this area of literature is discussed in section 1.11, but there has been very little research which has employed the eye movement recording of the object identity literature, and applied it to the object location research.

Another study which investigated memory for object identities was carried out by Melcher (2006) who investigated the persistence of memory for visual scenes. Participants viewed photographs of real-world scenes, and were aware that a memory test would follow. Stimuli were presented either for 20 s continuously, or for a combined period of 10 s (5 s followed by a further viewing of 5 s) or 15 s (5 s followed by a later period of 10 s, or vice versa). Combined views were separated by between 4-6 trials. After viewing the scenes, participants completed a memory test which consisted of three multiple choice questions testing memory for various aspects of the scene such as identity, colour, shape and location. These trials investigated whether information from scenes is accumulated from successive viewing periods.

The results showed that increased viewing time led to more accurate performance on the questions regarding scene detail. The results indicated that performance in the 5 s + 5 s condition was comparable with performance in the 10 s condition; viewing an image for two separate periods of 5 s resulted in almost identical performance to viewing the scene for a continuous period of 10 s. Overall, memory improved over time. This suggests that visual information is retained in memory, and that the degree of detail accumulates as viewing time increases. Additionally, this memory representation can be accessed and used to support accurate performance in a later memory test.

Similar to Rensink et al. (1997), Melcher had independent observers characterise aspects of the scene as being of central or marginal interest. The results showed that performance was better for questions about central aspects than marginal aspects. However, as eye movements were not monitored during the experiment it is not known how much processing each aspect of the scene received. It may have been that superior performance for central aspects was due to longer visual processing durations, but without monitoring eye movements this conclusion cannot be reached. However, it is possible that participants spent more time looking at areas judged as central interest than marginal interest, leading to better performance at test.

In Experiment 2, Melcher found that if participants viewed a scene for 1 s, memory performance was unsurprisingly poor (scoring approximately 60% on questions about the scene). However, if a scene was previewed for 10 s and then viewed again later for 1 s, performance improved by over 20%. This provides evidence that information about the scene was stored in memory from the initial 10 s view, and could be accessed later during the 1 s view.

This suggests that the memory representation did not “start again” when encountering the scene for the second time, but rather continued from where it left off. In other words, the memory representation built up from the first view of the image was robust enough to exist across 4-6 other trials, and then be accessed accurately to answer questions about specific details. In one condition, participants had to wait 60 s after the 10 s + 1 s viewing time, and also to read aloud a section of text. This was theorised to engage their working memory. The 60 s delay and reading task had no effect on performance, suggesting that working memory does not contribute to the maintenance of the visual information – instead this information must be stored in long-term memory, adding further support to the Visual Memory Model.

Melcher concluded that detailed information about objects in scenes is retained in long-term memory. More interestingly, these details seem to accumulate across separate views, and it does not matter if viewing of scene A is interrupted by viewing scene B. When scene A is presented again, accumulation of object properties continues where it left off previously. This is consistent with the Visual Memory Model. However, without investigating the oculomotor behaviour of observers, it is difficult to understand how memory for the objects built. Melcher concluded that, “the exact nature of scene memory and its underlying mechanisms remain a topic of debate” (pp. 13).

Hollingworth (2009) investigated whether a preview of a scene prior to searching for an object in the same scene facilitated search for the target. In Experiment 1a, participants were randomly assigned to either experience a 10s scene preview before searching for the target (preview condition) or to simply search for the target without the preview (no preview). Once the scene preview had been shown (or a blank screen with “no preview” had been presented) the target object was presented on the screen in isolation for 1500ms. This was done to inform participants which object they had to search for in the scene. Finally the scene was displayed until the participant located the target and pressed a button. Participants had to decide whether the target was the same as the object presented in isolation, or different. Two possible changes were used: a 90º rotation or a The results showed that participants found the target quicker when they had a 10 s preview than when they had no preview (1232 ms vs 1487 ms). The eye movement data showed that time taken to make the first fixation on the target object was faster with preview than without (390 ms vs 591 ms).

In Experiment 1b, preview times of 2000 ms, 500 ms and 0 ms (no preview) were used. The methodology was otherwise identical. Again the data showed that the target object was fixated earlier if the scene had been previewed than if it had not. The target object was reported earlier with a preview than without. Across Experiments 1a and 1b, the effect of the preview increased with longer preview duration. This suggests that during the preview period, visual information accumulated for participants. When the preview was seen for 10 s, the target was fixated after only 390 ms – this suggests that participants knew where to look for the target object as they had visual information stored from the preview. Even when the preview duration was only 500 ms, search was superior to the no preview condition, suggesting that some visual information was extracted from the preview, and that it was explicitly available to facilitate search.

In Experiment 2, the target preview duration was always 10 s, and a third type of preview was introduced. This preview showed the scene, but the target object was absent (target-absent-preview). The target-present preview and no preview conditions were also tested. A reliable effect of preview was observed for the time taken to fixate the target object in the test phase. The target object was fixated earliest in the target-present-preview condition, slower in the target-absent-preview condition and slowest in the no preview condition. Reaction time to report the target followed the same pattern.

The results suggest that a scene preview allows a visual representation to form and accumulate (as in Hollingworth, 2005b), and that this representation is explicitly available to facilitate search for a target object. The results also demonstrate that the facilitatory effect of the preview increased with longer preview durations, providing further evidence that visual information accumulates in memory over the course of viewing a scene (Melcher, 2001; Melcher & Kowler, 2001; Hollingworth & Henderson, 2002; Hollingworth, 2004; Melcher, 2006). The target-present-preview allowed participants to encode the specific location of objects within the scene, which resulted in superior performance in the test phase. As the target was present in the preview, it meant that it was found quickly. The target-absent-preview did not contain the target object, so participants took longer to find the target upon test. However, it did allow participants to encode the scene layout and context, thus facilitating search for the target in areas it would most likely be found. For example, a kitchen scene preview with a kettle target would likely result in participants searching the worktop area first. Unsurprisingly, this would result in shorter RTs than for no preview conditions, but longer RTs than the target-present-preview. Presumably the knowledge of the layout of the kitchen in the preview, coupled with prior knowledge about kitchens and where kettles are typically found would result in the correct area being searched first.

Overall, the results of a number of studies provide compelling evidence that detailed representations are encoded and retained in long-term memory. This finding is in stark contrast to the change blindness findings discussed earlier. This conclusion also suggests that the level of detail associated with these memory representations accumulates with extended viewing duration. This detail associated with the memory representations is not restricted to information about object shape or presence, but also includes the specific location of an object within a scene. Thus, this research demonstrates that memory for object location within a scene is remembered accurately and can be used to aid search for that target in a subsequent test.

Overall, the research into memory for object identities is supported by the Visual Memory Model which predicts the both short-term and long-term memory representations form for previously viewed objects. As time spent processing an object or scene increases, memory performance improves. It is likely that this is because increased fixations on an individual object allow information to accumulate and become more detailed, leading to improved memory. However, this has not yet been conclusively demonstrated. Nevertheless, this research has demonstrated that participants do retain detailed information about object identity for previously viewed objects and scenes, and has also investigated memory for object locations. Therefore, this research leads neatly into a separate, yet related, area of research which has investigated memory for the locations of objects. This research has used a variety of paradigms (the locations of objects in computer-based arrays, line drawings and real-world environments) to demonstrate that not only do participants retain accurate memory for object identity, but also for object location.

1.11 Research into memory for object locations

Uttl and Graf (1993) investigated memory for objects and their locations in scenes. Both younger and older adults were tested, but for the purposes of this review only the results from the younger adults will be considered. Participants were required to enter an office environment containing 40 objects. They were given instructions which forced them to interact with each object in some way (e.g., "move the *newspaper* from the *desk* to the *table*"). Some participants were explicitly told that a memory test probing their memory for the specific locations of objects would occur later in the experiment (intentional learning condition). Other participants were simply told that their memory would be tested (incidental learning condition). After a brief delay, participants re-entered the room. The objects had been collected together and were now all randomly arranged on one of the surfaces in the room. Participants were required to place the objects back on the item of furniture on which they were originally positioned. It did not matter whether the object was placed in the correct position on that surface, just that the surface was correct.

Participants who had received intentional learning instructions placed more objects correctly than those participants who received incidental instructions. Uttl and Graf concluded that participants have good memory for the locations of objects in scenes, especially in real-world situations. They also demonstrated that even if participants are not explicitly told to remember the locations of objects, they still demonstrate memory for the correct locations. However, despite participants interacting with each object, it was not known how much visual processing each object required. It is logical to suggest that each object was directly fixated, but it is difficult to understand the mechanisms involved in forming memory representations for object locations without measuring how much processing each object received. Nonetheless, the results demonstrated that participants retain information regarding the location of an object within an environment.

Postma and de Haan (1996) used a computer-based display to investigate memory for spatial locations. In Experiment 1, participants viewed a square frame for 30 seconds which contained a series of characters arranged in a random fashion. After viewing the image, participants were subsequently presented with the now-empty frame outside of which all the characters were positioned. Participants were required to use the mouse, and relocate the objects in their original positions. Participants viewed one of the three possible types of stimuli: all the same stimuli (e.g., all Ds), all different letters (e.g., D, F, L) or all different nonsense shapes (e.g., characters that had no meaning).

Postma and de Haan argued that relocating the same stimuli required simply *positional encoding* (i.e. which positions were previously occupied), whereas relocating different types of letters, and nonsense shapes required both positional encoding and *object-to-position assignment* (i.e. which specific objects were in which previous positions).

Postma and de Haan suggested that verbal factors might play a role in the differences between these two types of encoding. For object-to-position assignment, verbal factors may help (*the dog* was next to *the wall*). Alternatively, objects could be sorted serially (from left to right, or top to bottom). However, for positional encoding, it is less likely that verbal factors would contribute. Postma and de Haan argued that it is difficult to describe a location that needs to be filled if all the objects are identical. Having participants perform a concurrent verbal overshadowing task would help to tease these two types of encoding apart. Therefore, half of the participants completed the task with verbal overshadowing, and half completed it in silence. Postma and de Haan used two dependent measures: *absolute relocation error*, which was the distance between original position and its replaced position, and the *best fit solution,* which measured the distance between each replaced object and the object it was placed nearest to.

The results showed a number of interest findings. First, nonsense stimuli were replaced least accurately. Different letters (e.g., D, F, L) were replaced more accurately and most accurate placement was observed for the same letters (all Ds). Therefore, being required solely to replace the overall layout appears to be easiest, and having to replace specific nonsense letters in their specific location was most difficult. This may be because the nonsense letters were difficult to verbalise; this was supported by the finding that verbal overshadowing inhibited performance, more so for the letter and nonsense stimuli (both stimuli which required specific object-to-position knowledge). Also, participants often reported that they tried to name the nonsense letters or made up labels for them; verbal overshadowing appears to have had a negative effect on this process.

In Experiment 2, the locations of the stimuli were pre-marked for the replacement task. This meant that participants did not have to remember the location of stimuli – they only had to remember which stimulus was in each pre-defined position. Each participant therefore completed three different sets of trials: a) replacing specific stimuli to their correct pre-marked location (assignment only), b) reconstruct the overall spatial layout using X placeholders (position only) and c) replace specific objects to their specific locations (combined). Half of the participants completed the task whilst performing an articulatory suppression task (repeating "blah").

The results showed that for assignment replacements (placing the 12 stimuli in their pre-marked positions) performance was poorer when articulatory suppression was carried out. Results showed that for positions and combined replacements, position replacements were significantly more accurate than combined replacements. It was also found that positions only replacement was extremely robust; articulatory suppression had no effect. This suggests that there is a small or non-existent role of verbalisation for overall spatial layout – perhaps unsurprising as verbalising the overall spatial layout is difficult.

The evidence suggesting that articulatory suppression worsened performance in the combined replacements demonstrates that there is a verbal role in the encoding of specific objects to their locations; perhaps participants "rehearse" the locations (e.g., by repeating "the clock was on the table") sub-vocally. Conversely, overall spatial configuration seemed resistant to both articulatory suppression and set size. This suggests that the overall spatial layout of the scene did not require explicit encoding, and occurred more automatically.

Taken together, Postma and de Haan suggested that these results indicate that two processes may function when scenes are perceived. First, overall spatial accuracy is encoded and this is not affected by articulatory suppression. This builds an overall spatial layout or map, as suggested by Rensink (2000), and Henderson and Hollingworth (2002). The second process involves extraction of specific object properties (e.g., identity) and their locations within the overall map. Qualitative reports from participants appeared to support this view; many participants mentioned that they first tried to get the positions correct, and subsequently placed the objects in their correct places. The findings reported previously regarding accumulation of scene information (Melcher, 2001; 2006) seem to support this idea.

Overall, the findings of Postma and de Haan (1996) demonstrated that participants have good memory for the positions of characters in visual displays, that verbalisation plays an important role in the formation of these memories, and that memory for scenes can be considered in terms of memory for objects, memory for locations, and memory for the specific location of specific objects. However, as in much of the research on memory for locations, the encoding process was not investigated.

James and Kimura (1997) investigated location memory for line drawings of objects. Participants viewed an array consisting of 27 objects for 1 minute. A memory test was carried out immediately afterwards which probed spatial memory, as one of two types of changes was made to the displays. Either the positions of some of the original objects were switched (location-exchange condition), or some of the objects were moved to previously empty locations (location-switch condition). In the location-exchange condition, participants were instructed to indicate which objects were in their original position, and which objects were in a new (switched) position. In the location-switch condition, two different displays were created in which some of the objects were moved to previously unoccupied locations. Once again, participants were instructed to indicate which objects were in their original position, and which objects were in a previously unoccupied position.

The results provide evidence that observers have good memory for the location of objects in scenes. In the location-switch condition, participants correctly indicated the positions of 82% of the objects. In the location-exchange condition, 75% of the objects were correctly marked. The primary variable of interest was gender and the results demonstrated that women outperformed men. However, in both conditions, participants consistently correctly identified over three-quarters of the objects that had changed position, demonstrating that visual information regarding the locations of objects was retained from the initial display, and this information was explicitly available for access in a subsequent memory test. However, this study did not investigate the encoding that each object received. Also, it did not probe memory for the exact location of objects, but for the position of objects relative to other objects in the scene.

Postma, Izendoorn and De Haan (1998) tested memory for displays of 10 everyday objects presented on a computer for 30 s. After 30 s, the objects disappeared and were presented in a random order above the blank frame. Participants were instructed to place the 10 objects into their correct position. Participants were randomly assigned to one of three conditions. First, the object-to-position condition which at test pre-marked the 10 filled locations, and participants had to use the mouse to move items into their correct pre-marked location. The second was the positions-only condition. In this condition, the 10 objects presented above the frame were identical and participants were required to mark the 10 correct locations. The final condition was the combined condition in which participants had to place the correct objects in the correct locations. Half of the participants were required to perform a concurrent verbal overshadowing task whilst viewing the image.

Data analysis differed depending on the condition. For the object-to-position condition, the percentage of objects that were misplaced was calculated. For the positions-only condition, the average best-fit error for all 10 objects was calculated. Finally, for the combined condition the absolute error was calculated (the distance between the original position an object was located and its subsequent replaced position).

The results for the object-to-position condition showed that performance was significantly more accurate when the image was viewed in silence than when performing the articulatory suppression task. The silent group placed on average 7.5 objects in their correct pre-marked condition whereas the articulatory group placed approximately 6.3 objects accurately. This suggests that articulatory suppression interferes with spatial processing, resulting in less accurate performance. Potentially, participants verbally labelled the objects they viewed. Results for the positions-only condition showed that the silent group replaced objects with an average of 12 mm error, but the articulatory suppression group replaced the objects on average 14 mm away from original locations. Once again, the influence of articulatory suppression resulted in significantly poorer performance. Last, the combined condition showed that the silent group replaced objects on average 15 mm from their original location, but the articulatory suppression group replaced objects 16 mm away. Performance in the positions-only condition was reliably better than performance in the combined condition. Other studies have used paper-based methods to demonstrate that participants have good memory for not only the identities of objects, but also their locations (Kohler, Moscovitch & Melo, 2001; Levy, Astur & Frick, 2005).

Iachini, Sergi, Ruggiero and Gnisci (2005) used a real-world paradigm to investigate memory for the location of objects. Their primary interest was whether males or females exhibited superior memory for the location of objects. Participants entered a circular room which contained seven objects, and were instructed to memorise as much as possible. There was no time limit.

Once the participants had finished inspecting the room, they were led outside. Upon re-entry after a delay of approximately 1 minute, the seven original objects had been moved to one side of the room. Seven distractor objects had now been introduced to the environment, and were placed amongst the original objects. Each distractor object was similar in appearance to one of the original objects (e.g., a slipper and a shoe; a book and a notepad). Participants were required to select the objects they had viewed in the original configuration, and place them in the correct locations.

The results showed that overall, both males and females were very accurate in their memory for the original objects (both at 98% accuracy) and that the average object placement error was very small (1.96 cm for females and 1.85 cm for males). As the diameter of the cylindrical room was 3.5 m, these low levels of error represent very accurate memory for the locations of objects. This study demonstrates that given unlimited opportunity to encode a real-world environment, participants are able to accurately select and place seven objects, but that males are slightly better than females at this task. It is important to note that much research has been carried out on the differences between males and females in terms of spatial memory, and although some research has provided evidence that males are superior for some aspects of spatial memory (see above, and also Miller & Santoni, 1986; Orsini, Chiacchio, Cinque, Cocchiaro, Schiappa, & Grossi, 1986), this finding is not consistently observed. Accordingly, there are also studies that suggest that females have superior spatial memory (Voyer & Bryden, 1990; Uecker & Obrzut, 1993; Eals & Silverman, 1994; De Goede & Postma, 2008). However, this debate is beyond the scope of the present thesis.

The mechanisms underlying encoding processes for the location of objects are not well understood. What is known, however, is that participants can show good memory for the location of objects in real world environments (Uttl & Graf, 1993), as well as the location characters in computer-based visual displays (Postma, 1996; Postma & de Haan, 1996; Postma, Izendoorn, & De Haan, 1998) and line drawings (James & Kimura, 1997; Kohler, Moscovitch, & Melo, 2001; Levy, Astur, & Frick, 2005). There are a large number of other studies which have investigated memory for the locations of objects. Many of these have focused on patients with different types of memory impairments such as Alzheimer's (Kessels, Feijen, & Postma, 2005), amnesia (Shoquierat & Mayes, 1991) and also on the development of memory for object locations

(Schumann-Hengsteler, 1992). However, in theoretical terms, it is still not well understood specifically *how* memory for the locations of objects is encoded and retained in memory. The present research will investigate this issue.

1.12 Summary of previous literature

There has been extensive work to investigate the encoding processes in relation to memory for object identity, but this work has neglected memory for object locations. By contrast, work that has investigated the component processes of object location memory has failed to consider the encoding processes. Thus, there is a clear gap in the literature; there is a substantial lack of research that has investigated the role of encoding processes in memory for the both identities and locations of objects in a real-world environment.

To date, only Tatler, Gilchrist and Land (2005) have attempted to address these issues simultaneously. However, they employed paper-and-pencil based memory tasks which provided limited information about memory for the specific locations of objects. Also, they only used a 5 s viewing duration which may not have allowed enough time for objects to receive more than one non-consecutive fixation. To address this gap in the literature, the work presented in this thesis will record eye movements during encoding of visual scenes, and will investigate the influence of these encoding processes on memory for the identities and locations of objects within those scenes. The memory test will take place in the actual environment depicted in the photograph, allowing for precise measurement of memory for object location.

In their study, Tatler, Gilchrist and Land (2005) had participants view real world environments for 5 s whilst having their eye movements recorded with a head mounted eye tracker. Participants stood in the doorway of real rooms, and viewed the rooms in preparation for a memory test which would potentially ask questions about any aspects of the room. Immediately afterwards, participants completed a 4AFC questionnaire which probed memory for object presence, shape, colour, and position. The results provided evidence that object presence, colour and position (location) were encoded and retained in memory. Moreover, the eye movement data provided evidence that objects need to be directly fixated in order to extract information about location, but that information about identity and colour can be extracted without the object being directly fixated. As a result, memory for object position accumulated with fixations. That is, as the number of fixations made on an objects increased during the 5 s encoding period, so did the accuracy with which questions regarding its position were subsequently answered.

However, there was no relationship between the number of fixations and the accuracy on questions probing object identity or colour. These properties seemed to be encoded into memory without needing to be directly fixated. The same pattern of results was found in Experiment 2 when participants viewed photographs of the rooms for 5 s, rather than viewing the actual rooms.

1.13 Proposed research and introduction to the novel paradigm

As stated above, it is clear that there is an area of research that is yet to be investigated substantially. Namely, understanding how the eye movements made by observers when viewing scenes containing objects influence the memory that observers subsequently demonstrate for both the identity and location of those objects. In light of this, a novel paradigm has been developed to directly investigate this issue. It addresses the fundamental theoretical question of how visual encoding contributes to memory for the identity and location of objects.

The basic paradigm is as follows: participants sign-up for an experiment which they are informed will test memory for a visual scene. After giving informed consent, participants complete the *encoding phase* of the experiment. In this, they view one photograph on a computer screen on one occasion while their eye movements are recorded. They are instructed to look at the photograph and try to remember it, as a memory test will follow after a pre-defined interval (both immediate tests and test delayed by 24 hours are used).

Participants then complete the second stage of the experiment; the *test phase.* This probes memory for the contents of the photograph. Participants are taken to the actual room shown in the photograph which is now empty. All 12 objects presented in the photograph have been removed from the surfaces and placed randomly on the floor intermixed with 12 distractor objects.

Participants are told to "make the room look like the photograph". To do this, in the basic version of the paradigm, participants are required to perform two tasks. First, they have to select the objects that they remember from the photograph (up to a maximum of 12). If they select less than 12 objects, they are instructed to select more objects until 12 objects have been chosen, guessing if necessary. Second, participants place each object in the position that they remember it being located in the photograph. Hence, upon completion of the test phase, the room should be participants' best approximation of the photograph.

This novel paradigm allows for three separate, yet complementary, areas of investigation. First, recording the eye movements made by participants during the encoding phase allows detailed analysis of the oculomotor behaviour exhibited by participants. The data show which objects are fixated and which are not. For fixated objects, the number of fixations made to each object can be calculated, as well as the number of separate visits made to that object. The eye movement data also allow calculation of the duration of each single fixation, the duration of each visit, known as gaze duration (by summing the fixations) and the total fixation duration. The order in which objects are fixated can also be examined.

These measures allow insight into what memory for the scene is likely to consist of, based upon several assumptions from the literature. Objects that are looked at are more likely to be remembered than objects that are not looked at. Of the objects that are looked at, objects looked at more often and/or for longer are more likely to be remembered than objects that are looked at less often and/or for less time (in these experiments, *remembered* can refer both to possessing memory for the identity of an object, and/or memory for the location of an object, and in the results section these will be considered separately). As mentioned in the literature review, during scene viewing attention is typically very closely linked to fixation position. Hence, it is highly likely that the object which is being fixated is also the object which is being encoded into memory.

Second, the test phase of the paradigm allows for precise assessment of participants' memory for the photograph, both for the identities of the objects, and the locations. If participants possess memory for the identity of objects, it is likely that they will select those objects. Conversely, if participants do not have memory for the identity of any objects, it is unlikely that those objects will be selected. The same assumptions hold for the placing of objects. If participants possess memory for the location of objects, it is expected that objects will be placed close to their original location. However, if no memory for the location of objects is stored then objects are likely to be placed in a random fashion, leading in all probability to a high degree of error.

Last, and most interestingly, the paradigm also allows investigation into how the eye movements that participants make in the encoding phase influence the memory they exhibit in the test phase. In other words, it permits an exploration of the relationship between variability in the eye movements participants make during encoding and variability in memory for the identity and location of objects in the visual environment.

To conclude, the literature review has shown that while a large amount of research has been dedicated to memory for scenes, there are still some questions that remain unanswered. Critically, the relationship between encoding and retrieval of memory for the locations and identities of objects in scenes is not well understood, and a novel paradigm has been proposed to allow investigation into this area.

2 The role of visual encoding in memory for object identity and object location

2.1 Introduction

Chapter 1 identified an area of literature that has not received substantial investigation. The research presented in the following chapters was designed to directly address this paucity of research. Not only is it important to understand how memory for the identities and locations of objects forms from a theoretical perspective, but also from an applied perspective. Accordingly, object location memory is vital for everyday activities necessary for functioning optimally within society. Identifying objects in a scene, and then retaining the identities of the objects and their locations within that scene in memory is a fundamentally important aspect of visuo-cognitive function. Without this ability, it would be difficult to learn new routes, remember where objects are located around the house, and eye-witness reports of crimes would be severely compromised.

As discussed in Chapter 1, there is a fairly large area of research which has investigated memory for the location of objects within a variety of different environments (Uttl & Graf, 1993; Postma & de Haan, 1996; Iachini, Sergi, Ruggiero, & Gnisci, 2005). However, there has been only limited research which has sought to explore the influence of eye movements on this memory for object locations. Consequently, the encoding mechanisms for object location memory are not well understood. The experimental work carried out for this thesis directly addresses this issue. In this chapter, the findings of Experiment 1 are reported. In Experiment 1, participants' eye movements were recorded as they looked at a photograph containing 12 objects for either 5 s or 10 s. Participants were then tested for their memory for the identity and location of the objects. Critically, the relationship between the encoding carried out while looking at the photograph, and the memory exhibited at test was explored.

As reported in Chapter 1, Uttl and Graf (1993) conducted an experiment that tested younger and older participants' memory for objects and their locations in a real-world environment. Uttl and Graf concluded that memory for the location of objects in real-world situations was good. This ability of participants to accurately relocate objects has also been demonstrated by Iachini, Sergi, Ruggiero, and Gnisci (2005) using objects in a real-world environment, Postma and de Haan (1996) using letter arrays on a computer screen, and Kohler, Moscovitch, and Melo (2001) using arrays of line drawings of objects. Studies have also demonstrated sex differences in object location memory (Postma, Izendoorn, & De Haan, 1998; Postma, Jager, Kessels, Koppeschaar, & van Honk, 2004; James & Kimura, 1997; Levy, Astur, & Frick, 2005), and, like Uttl and Graf, that object location memory declines with age (Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000; Kessels, Hobbel, & Postma, 2007; Chalfonte & Johnson, 1996). Clearly, a significant amount of work has been carried out to investigate memory effects for objects and their locations and this research overwhelmingly shows that memory for the locations of objects is fairly accurate.

As well as this research on memory for the locations of objects, there has also been a substantial amount of research which has investigated memory for the identities of objects. Importantly, much of this research has considered the role of visual encoding. Henderson and Hollingworth (1999) made changes to the identity of an object during an 80 ms flicker either before that object had been fixated, or after it had been fixated. The change was noticed much more often when it was made after the object had been fixated than when it was made before the fixation. This provides evidence that directly fixating an object seems to play a critical role in the identity of an object being stored in memory (see also Hollingworth, Shrock & Henderson, 2001; Hollingworth, Williams & Henderson, 2001).

Hollingworth and Henderson (2002) used a change detection task to show that in a 3D-rendered image participants could detect changes at an above-chance level, indicating that precise representations of fixated objects are stored in memory, and can be explicitly accessed if required. Better than chance performance was observed for token replacements in Experiment 1, and for rotations in Experiment 2.

Hollingworth (2004) forced participants to view objects in a scene in a pre-defined presentation order and found that participants could decide whether a subsequently presented object was identical, or was instead a token replacement of a fixated object, at above-chance levels, even if 9 intervening objects had been viewed between initial inspection of the object and the presentation of the object at a forced choice decision phase. Thus, even the identities of objects fixated early during scene inspection are represented in memory and can be accessed later if required, suggesting a long-term component to object identity memory. Hollingworth (2005) found that participants were even able to correctly select an object in its appropriate orientation from identical competitors that were oriented differently from the original.

Two other aspects of memory for objects and their locations that are relevant to the present experiment have been examined in the literature; test delay and viewing duration. Pezdek, Whetstone, Reynolds, Askari and Dougherty (1989) demonstrated that objects viewed in real-world environments can be recalled and recognized at above-chance levels, even after a test delay of 24 hours. In their experiment participants entered a room containing 16 objects. Recall performance for the identities of the objects was superior when testing occurred immediately than after a 24 hour delay, though note that this study did not consider the influence of test delay on memory for object locations, nor did it consider the encoding mechanisms that contributed to the memory that was exhibited. However, it clearly suggests that some information about the locations of objects decays with a test delay.

The role of viewing duration on memory for object identities has also been studied. As presented in Chapter 1, Melcher and Kowler (2001) presented participants with computer-rendered images of rooms containing 12 objects and found that the number of objects participants recalled increased with longer encoding durations. Although the eye movement data showed that when a scene was viewed for 4s, a higher proportion of the recalled objects were fixated than when it was viewed for shorter durations (e.g., 1 s), Melcher and Kowler did not directly investigate the relationship between eye movements during encoding and memory for object identities and their locations.

A later study by Melcher (2006) did explore memory for different object characteristics (e.g., colour, location, shape). However, eye movements were not recorded. Melcher investigated whether memory for visual information accumulated across successive inspections of a scene. Participants viewed each scene either continuously (i.e. for 15 s), or separated by other trials (i.e. 10 s of scene A, followed by forced text reading or 10 s of scene B, followed by 5 s of scene A again). Melcher found that memory for a range of object properties including identity and location, was improved for longer inspection periods, even when those periods were interrupted. This suggests that detailed memory for objects in scenes is retained even when the scene is no longer available to view.

Tatler, Gilchrist and Land (2005) investigated the role of eye movements in memory for object properties. In their experiment, participants viewed real-world environments for 5 s in preparation for a memory test whilst having their eye movements recorded with a head-mounted eye tracker. Immediately afterwards, participants completed a questionnaire which probed memory for object presence, shape, color, and location. Results showed that information about object presence, color and position was encoded and retained in memory. Information about object location accumulated with fixations, and direct fixation of an object was necessary for encoding of location information, but not for encoding of information about object identity and color. This finding is interesting as it suggests that information about object identity and location may be encoded differently. Specifically, it suggests that object identities may be encoded somewhat more automatically than object locations, as object identities may be encoded without the need for an object to be directly fixated. The present experiment directly investigates whether information about object identities and locations is encoded differently for each as found by Tatler et al. (2005).

What should be clear from the literature discussed above, and in chapter 1, is that there has been considerable research to investigate memory for objects in scenes. However, this research falls quite neatly into one of two separable areas: those studies focusing on object location memory (but neglecting encoding processes), and those focusing on memory for object identities, but not in relation to their precise locations within scenes, or in relation to visual sampling behavior associated with information encoding. To date, only the study by Tatler et al. (2005) has attempted to address these issues simultaneously. However, they employed an off-line paper-based questionnaire to assess memory performance. To thoroughly explore these issues, a novel paradigm was developed in Experiment 1.

Before the specific details relating to the methodology used in Experiment 1 are explained, the general methodology used throughout this thesis will be presented. The reason for this presentation of general methodology is that in all experiments in this thesis, the methodology consists of the same two distinct phases; an *encoding phase* and a *test phase.* As all experiments use these same two phases, a general description will be provided at this point and any details that differ from experiment to experiment will be provided in each chapter. During the encoding phase, participants viewed a photograph of a room (see Figure 1) in which 12 objects were arranged on a variety of surfaces. As it is well established that there is a strong relationship between eye movements and cognitive processing during scene perception (Rayner, 1998; Liversedge & Findlay, 2000) participants' eye movements were recorded as they viewed the photograph. This allowed insight to be gained into the encoding process likely to contribute to the memory that was formed.

Once the encoding phase had finished, the test phase commenced either immediately, or after a 24 hour delay. During the recall phase of the experiment, participants were taken to the room that appeared in the photograph. The 12 objects that had appeared in the photograph were now randomly arranged on the floor interspersed with 12 other distractor objects. In Experiment 1, participants were required to select the 12 objects that had appeared in the photograph that they had inspected (guessing if they could not remember). Participants were then required to replace the objects in the room as accurately as possible in the positions that they thought they had occupied in the photograph. Measuring the number of objects correctly selected, and the deviation of the replaced objects from their original position allowed the accuracy of memory for object identity and location to be assessed. Furthermore, the relationship between oculomotor behavior during visual encoding and degree of recall accuracy could be explored.

Two independent variables were manipulated, namely viewing duration and test delay. This allowed examination of their general influence on memory for object identity and object location. Most crucially, it allowed investigation into the relationship between visual encoding of the scene and subsequent memory for objects and their locations in that scene. Participants were forewarned that they would view the scene for either a short or a long duration (5 s or 10 s). On the assumption that most fixations during scene inspection are approximately 250-500 ms (e.g., see Rayner, Smith, Malcolm, & Henderson, 2009), then 5 s should be approximately long enough for participants to fixate each of the objects in the scene if they so wished, while a viewing duration of 10s would provide enough time for at least one, if not two fixations on each of the objects in the scene.

Figure 1. The stimulus photograph viewed by participants in the encoding phase

It was expected that the independent variables would lead to differences in the participants' oculomotor behaviour. For example, one area of interest concerned how participants chose to allocate fixations across the scene when they were under differing time limitations (5 s or 10 s). A second area of interest was whether participants moved their eyes differently when anticipating a different length of time before the test occurred (no delay or a 24 hour delay).

In turn, it was anticipated that the independent variables would lead to differences in memory performance at test. Regarding the influence of viewing duration on test performance, following Tatler, Gilchrist and Land (2005) and Melcher (2006) it was expected that recall performance, both in terms of object identity and location, would improve with increased viewing duration. In addition, consistent with the findings of Pezdek et al., (1989), it was anticipated that reduced probability of correct object selection and less accurate spatial replacement of objects in the room would be observed after a 24 hour delay than when replacements were made immediately. Finally, at a more general level, it was predicted that recall performance would be related to eye movement patterns during visual encoding. Specifically, an exploration was carried out into the possible relationship between the likelihood of fixating an object and the time spent fixating an object, and the probability of correct object selection and accuracy of spatial replacement of objects.

As presented in Chapter 1, the Visual Memory Model (Hollingworth & Henderson, 2002) provides a descriptive framework which is supported by the findings of Melcher (2006) and Tatler, Gilchrist, and Rusted (2005), as well as Pezdek et al. (1989). This model addresses the issue of how objects in scenes are encoded and remembered and as such it is very relevant to the present paradigm and will be used to make predictions.

The Visual Memory Model has a number of clearly defined stages. First, it is implied (although not explicitly stated) that upon encountering a visual scene, a spatial map forms quickly in memory. This spatial map marks out the locations that are filled by objects. Second, when an object is directly fixated, the identity of that object is encoded and a short-term memory object file is created. Initially, this object file contains information that pertains both to the lower-level characteristics of the object (e.g., colour) and higher-level characteristics of the object (e.g., meaning). However, if further separate fixations are made on that same object, more and more specific information pertaining to object identity is encoded, leading to a more richly detailed object file.

Third, the model states that at some point during a fixation on an object, the identity of that object is linked to one of the positions within the spatial map. It is important to note that the Visual Memory Model does not specify the rate at which information about object identities is encoded into memory, nor does it state when the identity is linked to a location. However, it does state that information about objects in scenes (e.g., colour, shape) accumulates across separate fixations. On each separate fixation, more information is encoded, increasing the degree of detail associated with the memory for that object. Therefore, as time spent processing an object increases, the strength and degree of detail associated with the representation of an object also increases.

Fourth, the short-term memory object file is consolidated into a more robust long-term memory object file which is linked to one of the positions within the spatial map. This means that when fixation is moved to another object within the scene, the initial short-term memory representation decays and what remains in memory is the spatially indexed long-term memory object file. This long-term memory representation is robust enough to support accurate memory even after a delay. As stated, if that same object is re-fixated later, there is no sense of "starting again" with objects; the object file is simply accessed and supplemented with any further information.

Therefore, this model allows clear predictions to be generated for Experiment 1. First, more accurate recall of both object identity and object location information was expected for participants with 10 s viewing duration compared to those with only 5 s. This is on the basis that the object file associated with each object is augmented with information from separate fixations, and the assumption that longer viewing duration will allow for more fixations to be made on objects. This has already been demonstrated by Melcher (2006) and Tatler et al. (2005).

Second, the Visual Memory Model predicts that memory for object identity and location would be more accurate for participants tested immediately than for participants who were tested after a 24 hour delay, as the degree of detail would become less readily available and more sparsely detailed. However, memory for objects after 24 hours would still be accurate, due to the information about the object being transferred into long-term memory, but not as accurate as participants tested immediately. Despite the memory for that object becoming less detailed and readily available over the delay, there would still be enough detail remaining to perform the memory task at a good standard. This pattern of results was demonstrated by Pezdek et al. (1989).

The third, and most important prediction, relates to the potential relationship between visual encoding and memory for object identities and locations. If information about object identity and location accumulates across separate fixations, as stipulated by the Visual Memory Model, and as per Rayner (1998) and Liversedge and Findlay (2000), then there would be a close relationship between the number of fixations made on an object, and the accuracy for the identity and location of that object. Specifically, it would be predicted that more fixations would lead directly to more objects being selected and being placed more accurately.

However, in sharp contrast to the Visual Memory Model, an alternative theory on scene perception was proposed by Rensink (2000) called Coherence Theory. This theory was drawn partly from the change blindness literature (studies demonstrating that large-scale changes made in a variety of visual displays often pass undetected – see Chapter 1 for a more detailed description and discussion of change blindness). Coherence Theory states that, rather than information about objects being stored in a long-term form and accumulating across separate fixations, information about objects is encoded when focused attention is paid to that object, but disintegrates rapidly when attention is withdrawn, leaving only a sparse and abstract representation. Importantly, there is no detailed long-term memory for previously attended objects.

Consequently, the predictions based on Coherence Theory are quite different. Unlike the Visual Memory Model, Coherence Theory states that information about object identity and location does not accumulate across separate fixations. Rather, detailed information is retained about the currently fixated object, but this information decays rapidly once attention is directed to a new object, leaving only an abstract representation in memory. This abstract memory is not robust enough to support accurate performance on a memory task probing specific details about the object. Therefore, Coherence Theory predicts that participants who viewed the photograph for 5 s would correctly select the same number of objects as participants with 10 s viewing duration, and place them equally as accurately. Hence, objects that were fixated early during viewing would not be retained in any great detail. There would be no advantage in viewing the scene for 10 s compared to 5 s; the quality of memory for objects would be the same under both viewing durations. Coherence Theory would also predict that performance after 24 hours would be very poor as no detailed long-term memory record would exist for previously fixated objects.

Finally, Coherence Theory would state that there would be no overall relationship between the number of fixations made on objects, and the quality of memory for the identity and location of those objects. On the assumption that only information about the currently fixated object is stored in any great detail, then it would be expected that memory for object identity and object location would be poor regardless of whether an object had been fixated once, twice, three times or not at all. No enduring detail would remain in memory for any object except the currently fixated object.

To reiterate, if it is found that memory for object identity and location is better after 10 s than after 5 s, that memory for identity and location is more accurate when tested immediately than tested after a 24 hour delay, and that a strong relationship exists between the number of fixations made on objects and the subsequent accuracy with which they are recalled and placed, then strong support will be given to the Visual Memory Model.

However, if memory performance is comparably poor between the 5 s and 10 s viewing duration conditions, the immediate test and 24 hour test delay groups, and no relationship is observed between the number of fixations made on objects and the identity and location accuracy, Coherence Theory will be supported.

2.1.1 Explanation of key procedures used in the thesis: General methods

This next section provides details about some of the key analysis procedures that are used in dealing with the data in this thesis and that are relevant to more than one chapter. It is important that these procedures are explained clearly as they are used throughout the experiments in this thesis. In addition, the placement error and best fit error data from different experiments are compared to provide evidence of replication of effects from one experiment to another.

2.1.2 Calculation of placement error

For all of the experiments in this thesis, object placement error was calculated in the same way. Once the stimulus photograph had been created, the "home" position of all the objects was recorded to the nearest centimetre using X and Y co-ordinates (see Figure 2 below). These co-ordinates gave the exact location of each object relative to the *surface* it was placed on (chair, small table, desk or shelf).

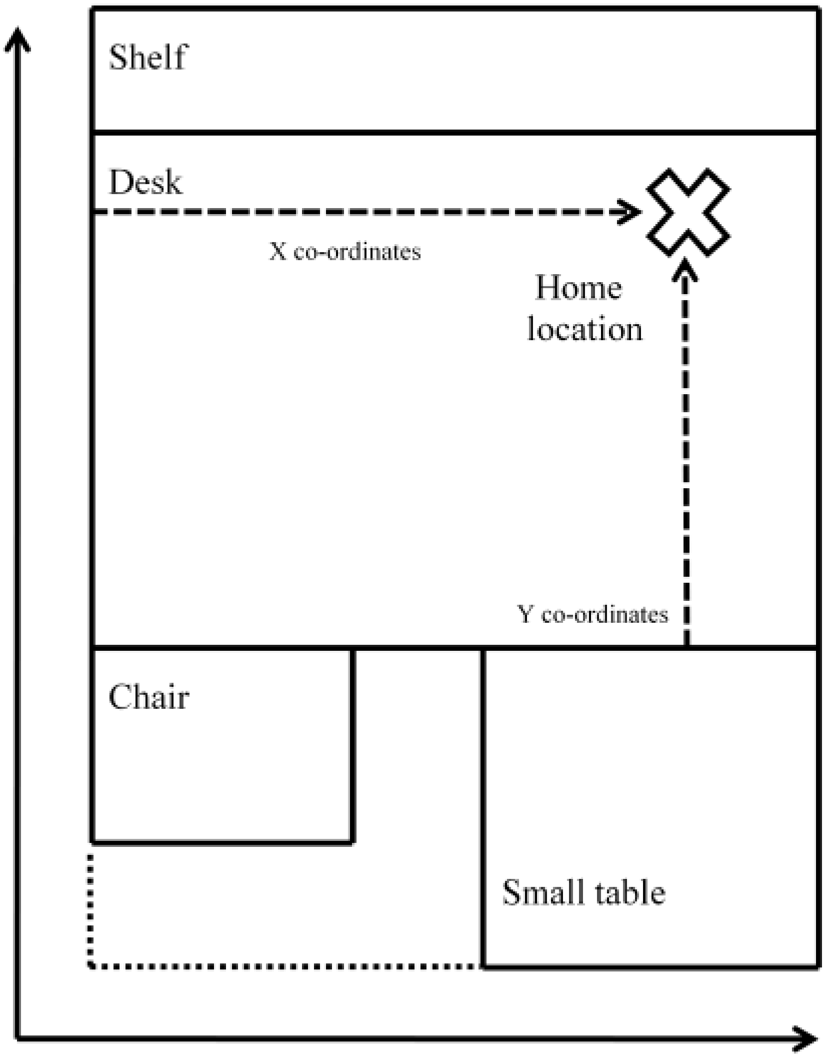

Figure 2. Schematic diagram of calculation of the surface-specific co-ordinates for "home" locations

The X and Y co-ordinates for each of the surfaces was also calculated. This meant that each pair of X and Y co-ordinates for objects could subsequently be transformed from a *surface-specific* co-ordinate pair to a *generalised* co-ordinate pair. This gave the location of each object relative to all four of the surfaces, rather than the surface it was placed on (see Figure 3 below).

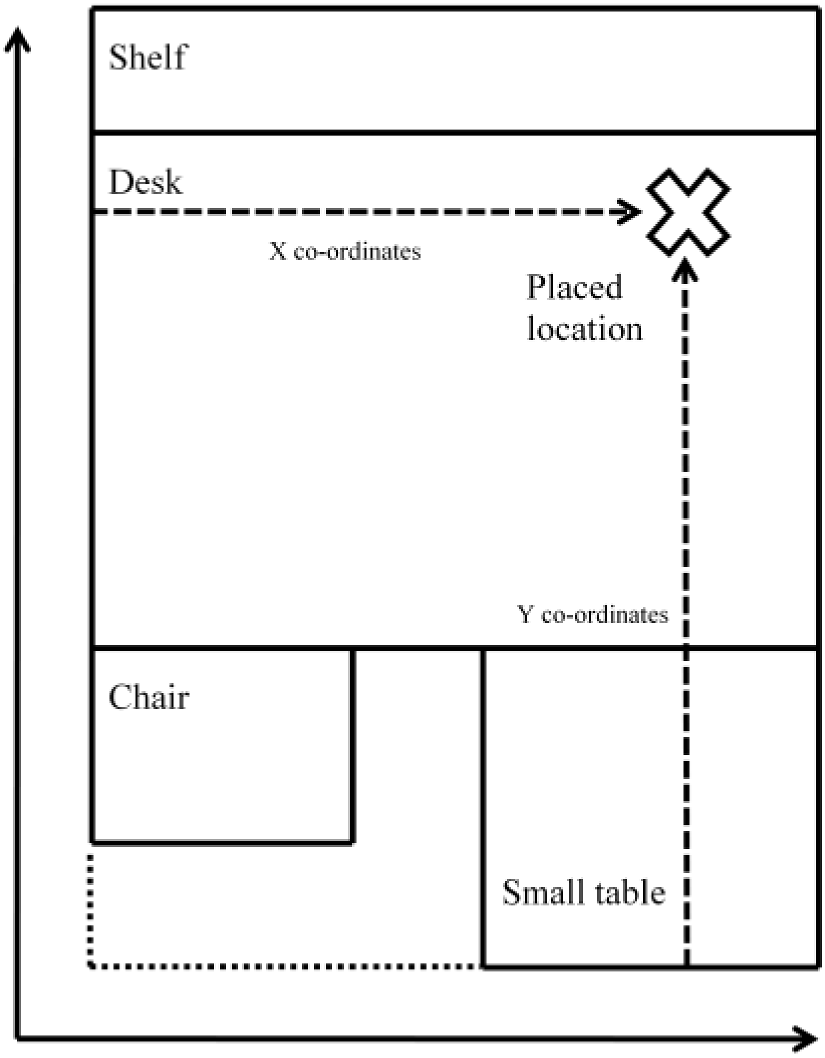

Figure 3. Schematic diagram of calculation of the generalised co-ordinates for "home" locations

This transformation of all co-ordinate pairs from surface to generalised allowed calculation of the *placement error* for each object. This was defined as the distance in centimetres between the home location of the object (where it was located in the photograph) and the placed location (where the participant chose to place that object during the memory test). It also meant that this placement error distance could be calculated even if an object was placed on an incorrect surface. It is important to note that calculation of these placement error distances assume that all four surfaces are on a level horizontal plane (see Figure 4).

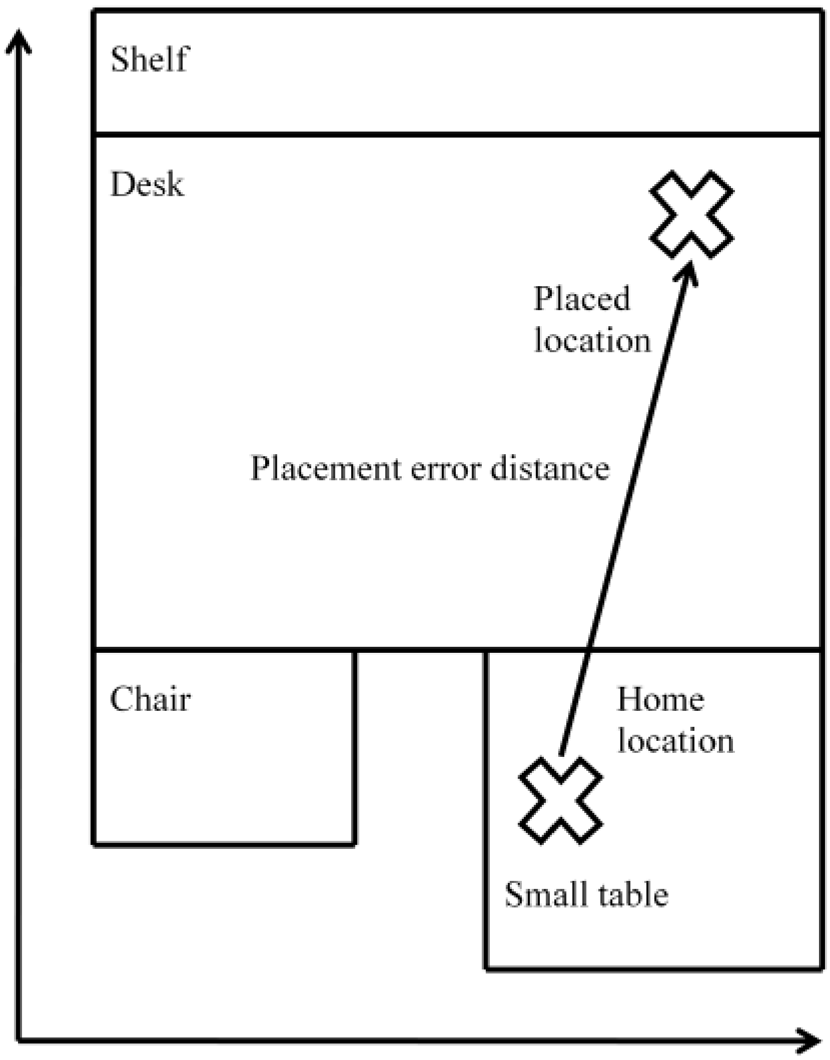

Figure 4. Schematic diagram of calculation of the placement error distance between the "home" location and the placed location

This method of calculating placement error is one simple way of characterising the accuracy with which objects were placed, thus allowing assessment of the accuracy with which the locations of objects were stored in memory. In this thesis, the placement error for each correctly selected object was calculated, and then the average was found for all correctly selected objects. It is important to note that placement error could not be calculated for incorrectly selected objects (distractor objects) as they did not have a home location. Therefore they were all excluded from any placement error analysis (but included in best fit analyses, see below).

2.1.3 Calculation of best fit error

While the object placement error described above allowed examination of the accuracy with which participants recalled the *specific* location of objects in the photograph, it did not allow examination of memory for the *overall spatial configuration* of the objects. That is, it rewarded participants for placing objects in or near the correct home location, but penalised them for switching the locations of two objects. For example, the correct location of the lava lamp was the top right corner of the photograph, and the correct location of the plate was the bottom right corner of the photograph. If a participant placed the lava lamp in the location of the plate and vice versa, this would contribute to a high average placement error (reflecting poor memory for the specific locations of objects). No credit would be given to the participant for correctly recalling that *something* was in the top right corner, and *something* was in the bottom right corner.

To address this issue, the best fit error was calculated for each participant. This was calculated for all 12 placed objects, regardless of whether the objects were present in the photograph or not. This measure of memory for object location ignores the identity of objects, and instead looks to reward placement of objects near to the closest correct location, whatever it might be. So, in the example above, even though the lava lamp and plate were placed a long way from their correct locations, they were placed close to *one* of the correct locations. Hence, this measure of location memory allows examination of how well participants remembered the overall layout of the photograph, or their overall spatial configuration memory.

To calculate the best fit error, the Hungarian Algorithm (Munkres, 1957), also known as the *Linear Assignment Problem*, was employed using Microsoft Excel 2007. The Hungarian Algorithm has a number of clearly defined stages. First, the distance between each of the 12 home locations and each of the 12 placed locations was calculated (a total of 144 different distances). These 144 distances were arranged into a 12 x 12 matrix.

A macro was written that completed the following steps: The smallest value in each row was found and then subtracted from each value in that row (leaving at least one zero in each row). This was done for all 12 rows. Then the smallest value in each column was found and subtracted from each value in that column (leaving at least one zero in each column). Once these subtractions had occurred, at least 12 zeroes were present in the matrix (representing the smallest distances in each row and column). A line was drawn through all appropriate rows and columns to cover all the zeroes such that the minimum number of lines was used. If exactly 12 lines were required, an optimal solution had been found. If less than 12 lines were required, it meant that some placed locations shared home locations. In that case, the smallest uncovered value was found and subtracted from each remaining uncovered value. This same value was also added to each value that was already covered by both a vertical *and* horizontal line. Then the lines were drawn through the zeroes again.

This process was repeated until the optimal solution was found. The average of all 12 values was calculated to give the best fit error; the average distance each object was placed from any of the home locations, regardless of whether they were placed near to their actual home location.

2.1.4 Monte Carlo simulation to calculate random placement error

To examine whether participants were randomly placing objects, or whether they were demonstrating accurate memory for the locations of objects, a Monte Carlo simulation was carried out to generate a placement error value which represented random performance. A macro was written in Microsoft Excel 2007 which simulated the actions made by human participants as closely as possible. The macro followed six clearly defined steps in order to provide a random simulation of a) object selection and b) object placement.

The first three steps aimed to simulate *object selection behaviour.* First, the numbers from 1-24 were listed with each number representing an object. Objects number 1-12 represented correct object selections (i.e. those objects that were present in the photograph) whilst objects numbered 13-24 represented distractor selections (i.e. objects not present in the photograph).

Second, the list of 24 numbers was shuffled. From this shuffled list, the first 12 numbers were selected, and the last 12 numbers were discarded. The first 12 numbers contained a random mix of numbers from 1-12 (correct object selections) and numbers from 13-24 (incorrect distractor selections).

Third, the number of correct object selections (i.e. numbers from 1-12) was counted. This represented random performance for the identity of objects. 10,000 repetitions of this process produced an average value of 6.00

objects. Clearly, when human participants were presented with 24 objects, of which half were correct, it would be expected that, by chance, they would correctly select 6 objects, and 6 distractors. In that respect, the Monte Carlo seemed to be accurately simulating actual behaviour.

The following three steps attempted to replicate *object placement behaviour.* First, all selections numbered 1-12 (representing correct object selections) were randomly assigned an X and Y co-ordinate. The range of these values was restricted to the locations human participants could place objects. In turn, this was restricted by the dimensions of the room.

The potential X values ranged from 0 – 135cm. A low X value meant that the object was placed near the left-hand edge of the room. A high X value meant that the object was placed near the right-hand edge of the room. The potential Y values ranged from 0 – 182cm. A low Y value reflected placement nearer to the small table (the bottom of Figure 4 above), whereas a high Y value represented placement near the shelf (the top of Figure 4 above). It is important to note at this point that the maximum X value of 135 and maximum Y value of 182 reflect the width of the cubicle (135 cm) and the 2D distance between the back of the cubicle (where the sill ends) and the front of the cubicle (where the small table ends) which is 182 cm. It encompasses all four of the surfaces and also the floor area underneath the chair and between the chair and the small table. Participants were allowed place objects anywhere in the cubicle, including the floor, if they so desired. However, only one participant placed a single object on the floor. Using these dimensions (135 X 182) gives a total area of 24420 cm^2. However, when the cumulative area of the four surfaces is considered, the total area is 21684 cm^2.[1]

For the second step, just as for the actual participant data, the distance between the randomly generated location and the known home locations (the location where the object was situated in the stimulus photograph) was calculated. Third, the average placement error for the randomly selected objects was calculated. 10, 000 repetitions of this process generated an average placement error of 84.51cm (*SE* = 0.04cm). This value is the average placement error expected if participants placed objects randomly within the room.

---

[1] Desk: 135 x 85 = 11475 cm^2. Sill: 133 x 40 = 5320 cm^2. Small table: 57 x 57 = 3249 cm^2. Chair: 40 x 41 = 1640 cm^2. Total surface area = 21684 cm^2.

It is important to note that for human participants, there was no restriction on the configuration of objects. That is, two (or more) objects could completely, or partially, share the same space. Accordingly, two participants in the pilot study placed objects into configurations which led to objects sharing space (e.g., the scissors were put into the desktidy). Therefore, for the purpose of the Monte Carlo simulation, two objects could overlap in space; that is, an object could be randomly assigned to a location that was already partly or completely filled by another object.

2.2 Method

Participants and design

All of the participants were experimentally naive University of Southampton undergraduates and postgraduates who participated voluntarily or in exchange for course credit. All participants had normal or corrected-to-normal vision. In total, 65 participants completed the experiment. There were 36 males and 29 females. The average age was 23.23 years (*SD* = 3.20) and ages ranged from 19 – 33 years. Of these 65 participants, 21 were excluded from the analyses. The criterion of exclusion was any loss of eye movement data during the encoding phase. For a participant to be included in the data analysis, 100% of the eye movement data was required. Participants were also removed from the data analysis if they could not be calibrated on the eye tracker. This strict criterion was applied to all experiments in this thesis. Removing these participants left data from 44 participants.

The design was factorial with two independent variables: Viewing Duration (either 5 or 10 seconds) and Test Delay (either immediate or 24 hours). The design was between-participants, with 11 participants in each of the four conditions.

Apparatus

A Fourward Technologies Dual Purkinje Image eye tracker was used to record eye movements in the encoding phase. The eye tracker was interfaced with a Pentium 4 computer. The stimulus photograph was displayed on a 21 inch colour monitor. The viewing distance was 100 cm. A research cubicle was used for the test phase and this was the same research cubicle seen in the photograph. The cubicle contained the 12 objects seen in the photograph, and 12 plausible distractor objects.

Stimuli

Participants viewed a photograph of a small research cubicle with 4 surfaces containing 12 objects (see Figure 1).

Procedure

There were two distinct phases in the experiment. In the *encoding phase*, participants viewed the photograph for either 5 s or 10 s in preparation for a subsequent memory test whilst having their eye movements recorded. Participants were informed at the start of the experiment how long they would have to view the photograph.

The *test phase* occurred either after a 24 hour delay, or immediately after the encoding phase. Participants were informed prior to the start of the experiment when the memory test would occur. The test phase required participants to enter the actual cubicle shown in the photograph. All 12 objects that had been present in the photograph were now arranged on the floor where they were randomly intermixed with the 12 distractor objects. Participants were required to perform two tasks. First, they had to select the objects they remembered from the photograph and were required to select 12 of the 24 objects. Second, participants had to place these objects in the correct locations within the room.

2.3 Results

In the results section, three sets of analyses are reported. First, the results that relate to the encoding phase are reported. Then the results from the test phase. Finally, analyses are presented which investigated the relationship between the encoding phase and the test phase. For all analyses presented in this thesis, any fixations shorter than 80 ms or longer than 1200 ms have been excluded. This exclusion criterion is widely used in eye movement experiments (e.g., Liversedge, Rayner, White, Findlay & McSorley, 2006).

Analyses of Encoding Behaviour

For all subsequent analyses, any fixations shorter than 80ms or longer than 1200 ms were judged to be outliers and were not included (a standard practice in many scene perception experiments; see Hollingworth & Henderson, 2002). Table 1 presents the means and standard errors for a series of eye movement measures. A series of 2 (Viewing duration: 5 s vs. 10 s) x 2 (Test delay: 0 delay vs. 24 hour delay) ANOVAs were carried out. The data was averaged for each participant and then entered separately into the ANOVA, leaving a dataset with 44 rows, one for each participant.

The ANOVAs showed that, unsurprisingly, a longer viewing duration led to more thorough exploration of the photograph. Participants with 10s looked at more of the objects and made longer fixations than participants with 5s. Specifically, the extra 5 s viewing duration in the 10 s viewing duration condition allowed participants to spend more total time fixating objects, $F(1, 40) = 179.80$, $p < .001$, fixate more objects, $F(1, 40) = 24.77$, $p < .001$, make marginally longer first fixations, $F(1, 40) = 3.48$, $p = .07$, make longer average fixations, $F(1, 40) = 8.76$, $p < .01$, make longer gaze durations, $F(1, 40) = 7.77$, $p < .01$, have longer average total times on objects conditional on the objects being fixated, $F(1, 40) = 18.88$, $p < .001$, make more fixations on average on each object conditional on the object being fixated, $F(1, 40) = 6.12$, $p < .05$, and finally refixate more objects, $F(1, 40) = 29.96$, $p < .001$.

An interesting finding is that participants who knew they had 10s to view the photograph had longer gaze durations and longer fixation durations than participants with 5s. This suggests that participants were able to change the way that they looked at the photograph as a direct result of the instructions they had received. Specifically, being forewarned of a longer viewing duration led participants to adopt a less urgent scanning strategy. In contrast, participants that knew they only had 5s to view the photograph displayed more urgent eye movements and this was characterised by shorter fixation durations, shorter gaze durations and re-fixation of fewer objects. For all measures there were no main effects of test delay, and no significant interactions between test delay and viewing duration (all Fs < 2.71). This suggests that knowing when the test was going to occur had no effect on the way participants looked at the photograph.

Table 1

Means and Standard Errors (in Parentheses) for Eye Movement Measures in the Encoding Phase

Viewing Duration:	5s				10s			
Test Delay:	0 delay		24 h delay		0 delay		24 h delay	
Measure	M	SE	M	SE	M	SE	M	SE
Total time fixating objects (ms)	3104	(283)	3235	(119)	6859	(299)	6501	(324)
Proportion of viewing duration fixating objects	0.79	(0.07)	0.84	(0.03)	0.84	(0.03)	0.84	(0.03)
Number of objects fixated	7.09	(0.67)	7.36	(0.61)	9.64	(0.58)	10.36	(0.31)
Average first fixation duration (ms) *	292	(13)	286	(23)	349	(27)	306	(17)
Average fixation duration (ms) *	271	(13)	267	(16)	338	(22)	299	(14)
Average gaze duration (ms) *	400	(31)	378	(30)	578	(68)	451	(41)
Number of fixations made during gaze duration *	1.48	(0.14)	1.37	(0.08)	1.69	(0.15)	1.51	(0.08)
Average total time (ms) *	449	(35)	477	(54)	751	(72)	631	(40)
Number of fixations made during total time *	1.70	(0.17)	1.81	(0.20)	2.22	(0.17)	2.13	(0.14)
Number of objects refixated	1.09	(0.31)	1.64	(0.31)	3.09	(0.34)	3.27	(0.36)

* denotes that this measure is contingent on an object being fixated

Analyses of Recall Behaviour – the Number of Objects Selected Correctly

As stated in the Method section, the first task participants had to complete in the test phase was to select the objects they remembered seeing in the photograph. If participants selected fewer than 12 objects, they were instructed to keep selecting objects until they had selected 12. The mean values are shown in Table 2.

Table 2

Means and Standard Errors (in Parentheses) of Recall Measures in the Test Phase

Viewing Duration:	5s				10s			
Test Delay:	0 delay		24 h delay		0 delay		24 h delay	
Measure	M	SE	M	SE	M	SE	M	SE
Number of objects correctly selected	7.45	(0.49)	8.09	(0.37)	9.09	(0.41)	8.09	(0.34)
Placement error (cm) for correctly selected objects	40.50	(5.25)	60.08	(4.16)	38.40	(5.30)	40.11	(4.41)
Best fit error (cm) for all 12 selected objects	25.72	(2.42)	24.31	(1.05)	23.13	(2.10)	25.11	(2.24)

As predicted by the Visual Memory Model, more objects were selected when participants had 10 s than when they had 5 s, $F(1, 40) = 4.01$, $p = .05$. However, there was no effect of test delay, $F(1, 40) = .20$, $p > .05$. There was a significant interaction between viewing duration and test delay, $F(1, 40) = 4.01$, $p = .05$. This interaction showed that in the 10s viewing condition, 1 more object was selected after an immediate memory test than after a memory test delayed by 24 hours, although this different was not statistically significant, $t(20) = 1.86$, $p > .05$ (note, however, that the p value was .08 and therefore was marginally significant). In the 5s viewing condition, 0.64 fewer objects were selected, $t(20) = -1.03$, $p > .05$, but this difference was not significant. What this interaction suggests is that being tested immediately was only beneficial for participants with a 10 s viewing duration. There was no effect of test delay for participants with a 5 s viewing duration. Clearly, the best memory for object identities was observed for participants with 10s to view the photograph *and* those who had an immediate memory test. Further support for this is shown by a independent *t*-test between the 10 s/0 delay condition and the other three conditions combined; significantly more objects were correctly selected in the 10s/0 delay condition (M = 9.09) than the average of the other three conditions (M = 7.88), $t(42) = 2.58$, $p < .05$.

To investigate whether memory for object identities was better than chance (= 6 objects), four one-sample *t*-tests were carried out; one for each condition. In all four conditions, participants correctly selected more objects than would be expected by chance (see Table 2). Performance was better than chance for the 5 s/0 delay condition, $t(10) = 2.95$, $p < .05$; for the 5 s/24 hour delay condition, $t(10) = 5.68$, $p < .001$; for the 10 s/0 delay condition, $t(10) = 7.46$, $p < .001$; and for the 10 s/24 hour delay condition, $t(10) = 6.10$, $p < .001$.

Analysis of Recall Behavior - Placement Error for Correctly Selected Objects

Once participants had selected objects from the pool of 24 objects, they were then required to perform the second part of the test phase. Participants had to place the 12 objects in the locations they remembered the objects filling in the photograph. The objects could be placed on any of the surfaces in the cubicle. For each object, the placement error was calculated using the method described earlier in this chapter. Note that the placement error could not be calculated for any incorrectly selected objects. The reason for this was that these incorrectly selected objects did not have a “home” location. As before, a 2x2 ANOVA was carried out on these placement error scores.

A main effect of viewing duration was observed on placement error for the correctly selected objects, $F(1, 40) = 5.27$, $p < .05$. As expected, memory for locations of objects was more accurate for participants with 10s viewing duration than 5s viewing duration. There was also a main effect of test delay on placement error, with objects placed more accurately after an immediate test than after a test delayed by 24 hours, $F(1, 40) = 4.91$, $p < .05$. In addition, there was a marginally significant interaction between viewing duration and test delay, $F(1, 40) = 3.45$, $p = .07$. As this interaction was not statistically significant, any interpretations of the following tests must be treated with a degree of caution. Closer inspection of this interaction using an unpaired post-hoc *t*-test showed that in the 5 s viewing condition, objects were placed 19.58cm more accurately at an immediate test than at a delayed test, $t(20) = -2.92$, $p < .01$. However, in the 10 s viewing condition, objects were placed only 1.72cm more accurately at immediate tests than at delayed tests, $t(20) = -.25$, $p > .05$ and this difference was clearly not significant. Table 2 demonstrates that memory for the location of objects was poorest for the 5 s/24 hour delay condition than any of the other conditions. This was further demonstrated with an independent post-hoc *t*-test between the 5 s/24 hour delay condition and the other three conditions combined. This *t*-test showed that object location memory was significantly less accurate in the 5 s/24 hour condition than the other three conditions, $t(42) = 3.76$, $p < .01$.

When the findings for object identity and object location are considered together, it becomes apparent that the pattern for each is markedly different. Regarding object identity, the ANOVA showed that delay only had an effect for participants with 10 s viewing duration. However, for object location, delay only had an effect for participants with 5 s viewing duration.

Analysis of Recall Behavior – Best Fit Error for all Objects

The next analysis investigated whether viewing duration and test delay affected memory for the overall locations of objects. This was assessed using the best fit error (more detailed discussion of how the best fit error was calculated are presented earlier in this chapter). As such, the best fit error gave a measure of how well participants remembered the locations that were filled by objects without having to specify which object was positioned at each location. A 2x2 ANOVA showed that there was no effect of viewing duration on best fit error, $F(1, 40) = .20$, $p > .05$. In addition, there was no effect of test delay, $F(1, 40) = .02$, $p > .05$ and no interaction between viewing duration and test delay, $F(1, 40) = .71$, $p > .05$. Taken together, these findings suggest that memory for the overall locations of objects did not improve with an extra 5 s of viewing duration, nor did it decay appreciably after a 24 hour delay. In turn, this suggests that memory for the overall locations of objects forms in memory relatively early during scene viewing, and is robust.

The Relationship between Encoding and Recall Behavior – Memory for Object Identity

The previous analyses have demonstrated that memory for the identity of objects is more accurate after 10 s than 5 s viewing duration. To further investigate how information about object identities accumulates in memory, the eye movement data were considered. First, objects were categorised as being either fixated or not-fixated. A mixed 2 (Fixation status: fixated vs. not fixated) x 2 (Viewing duration: 5 s vs. 10 s) x 2 (Test delay: 0 delay vs. 24 hour delay) ANOVA was carried out on the probability of correct object selection. Table 3 shows the means and standard errors.

Of the 44 participants that completed the experiment, three of them fixated all 12 objects, meaning that only 41 participants were included in the following analysis. The 2x2x2 ANOVA showed that there was a main effect of fixation status, with fixated objects ($M = 0.77$) more likely to be correctly selected than non-fixated objects ($M = 0.54$), $F(1, 37) = 17.43$, $p < .001$. However, there was no significant effect either of viewing duration, $F(1, 37) = .02$, $p > .05$ or test delay, $F(1, 37) = .01$, $p > .05$.

Additionally, there was no interaction between fixation status and viewing duration, $F(1, 37) = .68$, $p > .05$, no interaction between fixation status and test delay, $F(1, 37) = .95$, $p > .05$, and no interaction between viewing duration and test delay, $F(1, 37) = 2.35$, $p > .05$. Additionally, memory for non-fixated objects was obviously not different from chance (0.50), $t(40) = .74$, $p > .05$, whereas memory for fixated objects was significantly better than chance, $t(43) = 12.22$, $p < .001$. These findings provide evidence that whether or not an object is fixated during encoding predicts the likelihood that an object will or will not be remembered at test. However, it also shows that an increased viewing duration (10 s vs. 5 s) simply makes it more likely that an object will be fixated, and does not change the fundamental relationship between encoding and recall.

Table 3

Means and Standard Errors (in Parentheses) for Probability of Correct Object Selection when Objects were and were not Fixated, and when they were Fixated just Once or Twice or more

Viewing Duration:	5s				10s			
Test Delay:	0 delay		24 h delay		0 delay		24 h delay	
	M	SE	M	SE	M	SE	M	SE
Probability of correct selection for non-fixated objects	0.47	(0.08)	0.65	(0.08)	0.54	(0.12)	0.47	(0.12)
Probability of correct selection for fixated objects	0.76	(0.05)	0.74	(0.04)	0.84	(0.04)	0.73	(0.04)
Probability of correct selection for objects fixated once	0.70	(0.08)	0.62	(0.09)	0.78	(0.07)	0.55	(0.06)
Probability of correct selection for objects fixated twice or more	0.79	(0.06)	0.84	(0.05)	0.88	(0.05)	0.82	(0.04)

The next analysis is similar to the previous one, but instead only considers fixated objects. These fixated objects were then coded as either being fixated only once, or fixated twice or more. Of the 44 participants, one of them did not fixate any of the objects twice or more, and three of them did not fixate any objects exactly once, meaning that only 40 of the 44 participants were included.

A mixed 2 (Fixation status: fixated once vs. fixated twice or more) x 2 (Viewing duration: 5 s vs. 10 s) x 2 (Test delay: 0 delay vs. 24 hour delay) ANOVA was carried out on the probability of correct object selection. There was a main effect of fixation status, $F(1, 36) = 12.14$, $p < .01$, with objects fixated twice or more ($M = 0.83$) correctly selected significantly more often than objects fixated once (*M* = 0.66). However, as before there was no significant effect of viewing duration, $F(1, 36) = .14$, $p < .01$, and no effect of test delay, $F(1, 36) = 2.15$, $p > .05$.

The interaction between fixation status and viewing duration was not significant, $F(1, 36) = .04$, $p > .05$, nor was the three way interaction between fixation status, viewing duration and test delay, $F(1, 36) = .01$, $p > .05$. However, there was a marginally significant interaction between fixation status and test delay, $F(1, 36) = 3.78$, $p = .06$. As before, as this interaction was not significant at the .05 level, any potential subsequent effects must be treated with a degree of caution. Closer inspection of this interaction showed that for objects fixated only once, memory for object identities was much more accurate when the test was immediate (*M* = 0.74) than when it was delayed (*M* = 0.58), $t(39) = 2.12$, $p < .05$. However, for objects fixated twice or more, memory for object identities was accurate regardless of whether the memory test happened immediately (*M* = 0.84) or after a 24 hour delay (*M* = 0.83), $t(41) = .12$, $p > .05$.

This interaction suggests that for information about object identity to be consolidated into long-term memory, objects need to be fixated at least twice. If an object is fixated only once, enough information is encoded to allow successful recall at immediate testing intervals, but memory for object identities after a 24 hour delay is no more accurate than chance, $t(20) = 1.50$, $p < .05$. Possible explanations for this interaction are presented in the Discussion section.

The Relationship between Encoding and Recall Behavior – Memory for Object Location

The analyses used to investigate the relationship between encoding and object identity memory were also used to investigate the relationship between encoding and memory for object location. As before, the first analysis coded objects as either being fixated or not fixated. A 2 (Fixation status: fixated vs. not fixated) x 2 (Viewing duration: 5 s vs. 10 s) x 2 (Test delay: 0 delay vs. 24 hour delay) ANOVA was carried out on the placement error data.

It is important to note than of the 44 participants who completed the experiment, six were not included in this analysis. As stated before, three participants fixated all 12 objects, meaning they had blank cells for non-fixated objects. In addition, three more participants did not correctly select any of the non-fixated objects. There was a main effect of fixation status, $F(1, 34) = 8.89$, $p < .01$ on placement error, with fixated objects ($M = 42.0$cm) placed more accurately than non-fixated objects ($M = 62.1$cm). Table 4 shows the means and standard errors for all conditions.

The interaction between fixation status and viewing duration was not significant, $F(1, 34) = .01$, $p > .05$, and nor was the interaction between fixation status and test delay, $F(1, 34) = .20$, $p > .05$. In addition, the three way interaction between fixation status, viewing duration and test delay was not significant, $F(1, 34) = .01$, $p > .05$. The interaction between viewing duration and test delay was not significant, $F(1, 34) = 1.01$, $p > .05$. There was no main effect of viewing duration on placement error, $F(1, 34) = .47$, $p > .05$ but there was a significant main effect of test delay on placement error, $F(1, 34) = 4.28$, $p < .05$. This further supports the finding presented earlier that memory for the locations of objects was more accurate after an immediate test than after a delayed test.

The next analysis only included fixated objects and coded them as receiving one fixation or two or more fixations. Of the 44 participants who completed the experiment, only 39 were included in this analysis. As mentioned above, four participants were not included; one participant did not fixate any objects twice or more, and three did not fixate objects exactly once. In addition, one participant fixated one object once, but did not correctly select it. This meant that five participants were excluded from the analysis. A 2 (Fixation status: fixated once vs. fixated twice or more) x 2 (Viewing duration: 5 s vs. 10 s) x 2 (Test delay: 0 delay vs. 24 hour delay) showed that, unlike for object identity, there was no main effect of fixation status on placement error, with objects fixated twice or more ($M = 36.0$cm) placed only slightly more accurately than objects fixated once ($M = 44.1$cm), $F(1, 35) = 3.00$, $p = .09$. There was no main effect of viewing duration, $F(1, 35) = .65$, $p > .05$ and no main effect of test delay, $F(1, 35) = 1.00$, $p > .05$. There was also no interaction between viewing duration and test delay, $F(1, 35) = 2.71$, $p > .05$, no interaction between fixation status and test delay, $F(1, 35) = .18$, $p > .05$, no three way interaction between fixation status, viewing duration and test delay, $F(1, 35) = .46$, $p > .05$. However, there was a significant interaction between fixation status and viewing duration, $F(1, 35) = 4.68$, $p < .05$.

Table 4

Means and Standard Errors (in Parentheses) for Placement Error Scores when Objects were and were not Fixated, and when they were Fixated just Once or Twice or more

Viewing Duration:	5s				10s			
Test Delay:	0 delay		24 h delay		0 delay		24 h delay	
	M	SE	M	SE	M	SE	M	SE
Placement error for non-fixated objects	52.4	(12.1)	75.1	(9.5)	54.7	(10.2)	65.0	(13.5)
Placement error for fixated objects	35.4	(8.0)	52.8	(5.1)	35.8	(6.2)	38.8	(4.1)
Placement error for objects fixated once	44.1	(13.5)	62.3	(11.0)	41.6	(8.7)	31.6	(4.0)
Placement error for objects fixated twice or more	24.7	(6.6)	40.3	(8.7)	38.0	(7.3)	39.7	(5.1)
Monte Carlo value (cm)	84.4	(0.1)	84.5	(0.1)	84.6	(0.1)	84.5	(0.1)

Closer inspection of this interaction showed that for participants with 5 s to view the photograph, objects fixated twice or more (M = 32.53cm) were placed more accurately than objects fixated only once (M = 53.20cm), $t(17) = 2.11$, $p = .05$. However, for participants with 10 s viewing duration, objects fixated only once (M = 36.35cm) were placed no less accurately than objects fixated twice or more (M = 38.92cm), $t(20) = -.54$, $p > .05$. The most likely explanation for this is that fixation durations were shorter in the 5 s condition than the 10 s condition. Therefore, one fixation in the 10 s condition is not equivalent to one fixation in the 5 s condition. In turn, this suggests that one (long) fixation in the 10 s condition allows information about object location to be encoded, and this location information does not improve with a second fixation. However, in the 5 s condition, one (short) fixation allows some information about object location to be encoded, but a second (short) fixation is required if location information is to be adequately encoded.

Assessment of Whether Object Location Memory is Better than "Chance"

The final analyses investigated whether participants' memory for the locations of objects was better than chance. To obtain a meaningful chance value for object location, a Monte Carlo simulation was carried out. For full details of the procedure used to generate the value, see the discussion earlier in this chapter. The mean Monte Carlo value for each condition can be seen in Table 4. In all four conditions, fixated objects were placed significantly more accurately than would be expected by chance; in the 5 s/0 delay condition, $t(10) = -6.10$, $p < .001$; in the 5 s/24 hour delay condition, $t(10) = -6.23$, $p < .001$; in the 10 s/0 delay condition, $t(10) = -7.88$, $p < .001$; and in the 10 s/24 hour delay condition, $t(10) = -11.04$, $p < .001$.

For non-fixated objects, an interesting pattern of results was found. In both delayed testing conditions, non-fixated objects were placed no better than the respective chance values; in the 5 s/24 hour delay condition, $t(10) = -.98$, $p > .05$; in the 10 s/24 hour delay condition, $t(7) = -1.45$, $p > .05$. However, when tested immediately, non-fixated objects were placed significantly more accurately than chance; in the 5 s/0 delay condition, $t(10) = -2.65$, $p < .05$; in the 10 s/0 delay condition, $t(7) = -2.91$, $p < .05$. This results pattern suggests that when tested immediately, participants are able to place non-fixated objects somewhat accurately. However, when the test is delayed, this is not demonstrated. Possible explanations for this finding are explored in the Discussion.

2.4 Discussion

The first finding of note was that, compared to participants with 5 s viewing duration, participants with 10 s viewing duration looked at more of the objects, had longer gaze durations and fixation durations, and also re-fixated more objects. Put simply, more time to look at the photograph led directly to more thorough exploration of the photograph. Taken together, these results suggest that participants with 10 s viewing duration changed the way they looked at the photograph in direct response to the task instructions. Recall that participants were told before the experiment started whether they would have 5 s or 10 s to look at the photograph. It seems, therefore, that in preparation for the upcoming memory test, participants demonstrated some meta-cognitive awareness as to the most appropriate way to look at the photograph. Participants with only 5 s exhibited a more urgent scanning strategy, characterised by shorter fixations on fewer objects and fewer re-fixations. However, participants with 10 s demonstrated a less urgent scanning strategy, characterised by longer fixations on more objects and more frequent re-fixation of objects.

On the assumption that the eyes and attention are typically closely linked (Rayner, 1998; Liversedge & Findlay, 2000) it is likely that these differences in encoding behaviour would result in differences in performance at the subsequent memory test. This was correct. At test, participants with 10 s viewing duration correctly selected more of the objects than participants with a 5 s viewing duration (as found by Melcher & Kowler, 2001; Melcher, 2006), and placed those objects more accurately in the cubicle than participants with only 5 s. However, the improvement for both object identity and object location was not observed in the same way at both test intervals. A 10 s viewing duration only led to improved object identity memory when the test was immediate. In contrast, a 10 s viewing duration led only to improved object location memory when the test was delayed by 24 hours.

Importantly, the finding that memory for the identity of objects was better after 10 s than after 5 s does not support the claims of Coherence Theory (Rensink, 2000). According to this theory, information about previously viewed objects (other than the last object that was processed or attended to) is not stored in detail in memory. Given this, Coherence Theory would predict that memory performance would be no better after 10 s than 5 s. However, this was not what was found. Instead, memory for object identities and locations was more accurate after 10 s than 5 s.

Two further findings that do not support the predictions of Coherence Theory are that memory for identities after a 24 hour delay, while somewhat poorer than after an immediate test, was still better than expected by chance. In addition, memory for the location of fixated objects was better than chance even after a 24 hour delay. Coherence Theory would predict that detailed information about the identity of objects would not be remembered at a delayed test because such information about specific object properties is not stored in long-term memory.

The final finding that did not support the claims of Coherence Theory was that there was a close relationship between encoding and recall for the identities and locations of objects, with the detail associated with object identity memory increasing as the number of fixations made on an object increased. As Coherence Theory does not stipulate that information about object identities and locations develops over fixations in memory, then it provides no means to explain those effects that occurred in the current data. Taken together, it seems that the results of the present experiment do not fit within the account offered by the Coherence Theory.

Interestingly, however, the overall pattern of results does seem to support the framework of the Visual Memory Model and for this reason, the results will be discussed within this framework. The finding that a longer viewing duration only led to better memory for the identities of objects at immediate testing conditions whereas a longer viewing duration was only beneficial for location memory after a delayed test is potentially of great importance as it provides evidence that information about object identity and object location may be encoded differently. These findings seem to suggest that information about objects is not encoded simultaneously. Instead, perhaps different object properties are encoded at different rates. This possibility will be explored in more depth.

A further interesting finding relates to the best fit error. This measure was considered to represent the accuracy with which participants remembered the overall layout of the objects in the scene. The pattern of results showed that best fit error was comparable after 5 s and 10 s viewing duration, and also after an immediate and a delayed test. Put simply, information about the overall spatial layout seems to be encoded early during scene viewing, and is robust over a delay. From a theoretical perspective, this fits neatly within the Visual Memory Model and Hollingworth and Henderson's idea of a spatial map. In the Visual Memory Model, the spatial map forms rapidly and is used to guide the encoding of individual objects. The notion that the overall layout or gist of a scene forms rapidly during scene viewing is also consistent with Hasher and Zacks (1979). They suggested that information about the locations of objects is encoded somewhat automatically. This automatic encoding of overall spatial layouts has also been demonstrated by Zechmeister, McKillip, Pasko, and Bespalec, (1975), von Wright, Gebhard, and Karttunen (1975), Mandler, Seegmiller, and Day (1977), and Aginsky and Tarr (2000).

Thus far, the pattern of results seems to indicate that object identities and object locations are encoded somewhat differently. The analyses that probed the relationship between encoding and recall directly addressed this issue. For identity memory, the relationship was very simple and apparent; if an object was not fixated then the probability of it being remembered was equal to chance. However, if an object was fixated, it was more likely than chance to be correctly selected. This finding fits within the framework of the Visual Memory Model, and has been demonstrated in earlier studies (Henderson & Hollingworth, 1999; Hollingworth, Shrock, & Henderson, 2001; Hollingworth, Williams, & Henderson, 2001; Hollingworth & Henderson, 2002). In addition, objects selected twice or more were more likely to be correctly selected than objects fixated just once. This demonstrates that it is not only important whether an object is fixated or not, but also the number of times an object is fixated. This finding suggests that information about object identity continues to accumulate across separate fixations. Again, this finding is consistent with the Visual Memory Model.

There was also a marginally significant interaction between fixation status (one fixation vs. two or more fixations) and test delay. The interaction showed that objects fixated once were remembered accurately when the test was immediate, but were remembered no better than chance if the test was delayed. However, objects fixated twice or more were remembered accurately regardless of when the test occurred. This finding is important as it relates to another part of the visual Memory Model; namely the consolidation of information into long-term memory.

The Visual Memory Model states that at some point during fixation on an object, the encoded information is transferred from short-term memory to long-term memory. The model does not state how or when this process occurs. However, the interaction presented above seems to suggest that for object identities, this consolidation happens during the second fixation on an object. Of course, this does not mean that two separate fixations are required in order for an object to be consolidated into long-term memory. Instead, consolidation could happen after a specified period of time. However, in the present experiment, "encoding" was analysed using the number of fixations as delimiters, rather than time per se. Despite this, the results seem to suggest that consolidation into long-term memory does not happen immediately upon fixation of an object. Instead, it occurs sometime later during fixation on an object and crucially, without this consolidation process, only a short-term memory representation seems to exist for object identities. Importantly, this pattern was found for both the 5 s and 10 s viewing conditions. This suggest that memory for object identities is linked closely to the amount of visual processing each object receives, rather than a more global factor such as the overall amount of time that the scene was viewed for.

Perhaps the most critical finding of Experiment 1 was that information about object locations was encoded differently in relation to object identity. The initial finding was the same as for identity; fixated objects were placed more accurately than non-fixated objects. This finding also supports the Visual Memory Model and provides evidence that to accurately remember the location of an object, that object needs to be fixated. However, there was no clear evidence that information about object location continued to be encoded across separate fixations. There was no difference in placement error between objects fixated only once, and objects fixated twice or more. This suggests that, as per the Visual Memory Model, the location of a specific object is encoded during the first fixation on an object. Unlike object identity, subsequent fixations did not seem to lead to more detailed location information being encoded. This distinction between memory identities and locations has been demonstrated by some previous researchers (Postma & de Haan, 1996; Simons, 1996; Jiang, Olson, & Chun, 2000; Alvarez & Oliva, 2007) although these studies have not examined the encoding mechanisms involved.

There are at least two possible explanations as to the difference in encoding patterns for object identity and object location. The first relates to the participants' expectations of the memory task. Before the experiment began, participants were told that they would have to complete a memory task after viewing the photograph. They were also instructed as to the length of the test delay (either no delay or 24 hour delay). However, they were not told what the memory test would consist of. Therefore, a possible explanation for the difference in encoding for identities and locations is that participants chose to prioritise encoding identities on the basis that they expected to be tested on object identities. Importantly, this explanation rests on the assumption that participants can, and do, choose which property of an object they pay attention to during a fixation.

Informal post-experimental questioning revealed that many of the participants expected at test to be required to list the objects they had seen in the photograph. These participants subsequently expressed surprise when tested on their memory for the locations of objects. Therefore it seems, to some degree at least, that participants chose to encode the identities of the objects ahead of the locations of objects. Note that if this explanation is correct, it demonstrates that participants can encode information about object locations even if they are prioritising the identities of objects. It is important to note, however, this explanation is based solely on anecdotal evidence gained from participants and therefore must be treated with a degree of caution.

The second explanation for the difference in encoding patterns between object identities and object locations rests on the fundamental differences between the degree of detail associated with each. The Visual Memory Model states that when an object is fixated, low-level and high-level information about the identity of that object is encoded. This means that the object is identified and basic information about its form is also encoded (e.g., a pen). As the object continues to be fixated (or is fixated again at a later time) then the degree of detail associated with that object also increases. Over time, more and more information about the identity of that object can be encoded. For example, after two fixations, one might be able to state that not only is the object a pen, but also have encoded information about its size, shape, colour, and make or model. Therefore, with extended viewing duration, the memory representation of that object has progressed from simply “a pen” to “a large thick black Parker pen”. If this is correct, it is easy to see how information about object identities might accumulate across separate fixations. It is important to note that the degree to which this happens depends on the complexity of the object; more complex objects (e.g., a calculator) have more potential properties to encode compared to a more simplistic object (e.g., a ruler). Clearly, at some point, encoding of object properties must cease (either because the object has been encoded completely, or because a fixation is made elsewhere), but before encoding stops, it seems that object identity information is continually encoded.

In contrast, information about object locations might be encoded quite differently. According to the Visual Memory Model, the object identity short-term memory object file is bound to its position within the spatial map quickly. The data from the present experiment suggest that this happens during the first fixation on an object. If this is true, then the specific location of an object may be encoded fairly rapidly. Once this location is encoded, it is somewhat difficult to see how that location could be improved or fine-tuned with subsequent fixations. Merely looking at the same location continually would not seem to add any detail to that location. If one was to try to embellish the location, the optimal strategy would be to make fixations on the object and on other landmarks around that object (e.g., walls, edges of surfaces) to try to encode a more finer-grain position. This sort of scanning pattern was not observed in the present experiment. Therefore, it seems as though repeated fixations on an object do not lead to improved memory for the location of that object, but do lead to improved memory for the identity of that object.

There is one final aspect of the location data that merits further discussion. Recall that non-fixated objects were placed significantly more accurately than chance when the memory test was immediate, but no better than chance when the test was delayed. This pattern of results is interesting as it suggests that some information about the locations of objects was encoded, even for objects that were not fixated. There are at least two possible explanations for this pattern of results. First, it could be that objects do not need to be directly fixated in order for information about their location to be encoded. This is perhaps somewhat unsurprising; in order to select the location of the next saccade, observers must have some awareness of potentially interesting areas they wish to fixate. As such, when fixating an object, observers are aware of other filled locations. However, as these other locations have not been fixated, there is no opportunity for them to be consolidated into long-term memory, thus explaining why memory for non-fixated objects after a 24 hour delay is no better than chance. Only a weak, short-term memory representation is formed for the non-fixated objects (if at all).

The second explanation is based upon the link between candidate positions in the spatial map and object files. According to the Visual Memory Model, when an object is fixated, the short-term memory object file is created for that object, and that object file is linked to one of the positions in the (already created) spatial map. If this is true, then once the viewing duration is finished, each participant is left with the following things in memory: a completed spatial map with all 12 positions marked out; some of those positions in the spatial map are strongly linked to specific objects; other positions are weakly linked to specific objects; some of the positions are unfilled and not linked to any of the objects. This means that at test, participants may know that certain locations need to be filled, but do not know which objects belong in which candidate locations. It is sensible to suggest that participants chose to first place objects for which they were most sure about the locations. This is supported by a significant positive correlation between placement order and placement error, $r(361) = .40$, $p < .001$.

Consequently, towards the end of the memory test, participants are likely to be left with some objects that they are unlikely to have fixated, and an identical number of available locations within the spatial map. If a large number of locations are still available, the average placement error for those non-fixated objects is likely to be high. However, if only a few locations are still available, the average placement error is likely to be lower. Hence, in actual fact, the accuracy with which non-fixated objects are placed is likely due directly to the accuracy with which fixated objects have already been placed. When the links between locations and objects are strong, the available degrees of freedom for the remaining objects may decrease such that placement error is small. Conversely, when the links between locations and objects is weak, there are more degrees of freedom available for the remaining objects, and as such, placement error is higher.

Regarding Experiment 1, in summary the pattern of results suggests that memory for object identities is better for fixated than non-fixated objects, and is also better for objects fixated twice or more than objects fixated only once. More specifically, objects fixated only once were remembered well only at immediate tests, whereas objects fixated twice or more were always remembered accurately, regardless of when the test occurred. However, memory for object locations, while better for fixated than non-fixated objects, did not improve from one to two or more fixations. Instead, object location information seemed to plateau after a single fixation. This critical dissociation seems to suggest that the fundamental nature of object identities and object locations is different, and that the two are encoded in different ways during encoding of a scene.

3 Using attentional capture to modulate the relationship between encoding and recall for object identity and object location

3.1 Introduction

The aims of Experiment 2 were first to try to replicate the findings of Experiment 1, and second (and more importantly) to investigate whether the relationship between the number of fixations and memory for identity and location could be fundamentally influenced by using a lower-level attentional capture technique to draw attention preferentially to half of the objects in the stimulus photograph. A modified version of the stimulus photograph was presented to participants and in this photograph half of the objects were surrounded by a bright green box ("boxed objects"). It was expected that the green boxes would render those boxed objects highly visually salient. As previous research has demonstrated that visual attention is often drawn to the most visually salient portions of scenes, it was expected that this manipulation would lead participants to preferentially fixate the boxed objects ahead of the non-boxed objects. If so, this would demonstrate that attention can be driven around the scene to certain objects at the expense of others, leading those objects to be preferentially encoded. In turn, it is likely that this preferential encoding would lead to improved memory for those objects at test.

Previous research has demonstrated that both higher-level factors (e.g., task instructions, previous experience) and lower-level factors (e.g., colour, luminance etc) can be used to drive oculomotor behaviour around scenes. Examples of both types of factors are discussed in detail below. Note that these two influences are not mutually exclusive, and in fact the degree to which an object is salient within a scene is determined by the task instructions in relation to lower level characteristics of objects. In Experiment 2 both higher-level and lower-level influences are combined to modify the way in which observers look at the stimulus photograph. As in Experiment 1, the task instructions to "try to remember as much about the scene as possible" are expected to lead to the 12 objects being fixated for the majority of the viewing duration. However, the primary focus is whether a lower-level attentional capture technique can modulate the eye movements observed in Experiment 1, and the analyses will examine whether this changes the relationship between encoding and recall. The technique used to capture attention involved making half of the objects visually salient by placing a green box around them. It is likely that this will have a modulatory effect on object saliency, making those boxed objects highly salient for at least two reasons. First, they are likely to be highly visually salient because the task instructions to "remember as much about the photograph as possible" were shown in Experiment 1 to lead participants to prioritise fixating the objects rather than elsewhere in the scene. Second, the boxed objects are likely to be highly visually salient due to the green boxes. As a result, the boxed objects are highly likely to be fixated more often and for longer during encoding, whereas the non-boxed objects are less likely to be fixated.

3.1.1 Higher-level influences on eye movements: Task instructions

An early example of the influence of higher-level factors on visual attention was provided by Buswell (1935) who investigated how task instructions affected eye movements. Generally, participants were asked to simply look at pictures and the primary subject of investigation was whether there are cultural differences in eye movements. However, on one occasion, eye movements were recorded for two separate views of the same picture by the same observer. The picture contained, amongst other objects, a tower. When simply asked to look at the picture, Buswell found that fixations were spread fairly evenly around the picture. However, when the observer was asked to look for a person at one of the windows of the tower, fixations were almost exclusively clustered on the tower, ignoring the other portions of the scene. Buswell also reported that when the observer was performing the search task, fixations were generally longer than during the free viewing condition. These results demonstrated that when asked to search for something rather than freely view a picture, fixations were longer and made to the most task-relevant portion of the scene, suggesting that participants were able to exercise cognitive control over their eye movements, and allocate their visual attention to the most informative scene regions.

A second seminal study that provided evidence that higher-level factors influence eye movements was conducted by Yarbus (1967) who demonstrated that where an observer chose to look in a scene could be influenced by the task instructions they were given. The observer viewed a painting of a room containing several people who were looking towards the door where another figure had just entered. Yarbus asked the observer to look at the painting under seven different viewing instructions. The eye movements showed that when the observer was asked to estimate the material circumstances of the people in the painting, fixations were mainly made to the clothes and the objects in the room. However, when the same observer was asked to estimate the ages of the people, fixations were typically made to faces and bodies of the people in the painting. Clearly, the instructions given to the observer influenced the regions of the scene that the observer chose to allocate their visual attention to (via fixations). In turn, this suggests that observers do not randomly look at scenes. Instead, they look at the regions that are most useful and informative to the relevant task.

These two studies have been very influential as they are widely considered to be the first empirical demonstrations of task-specific eye movements. However, despite the study by Yarbus (1967) being widely cited, it is not often reported that the results were gained from the eye movements of only one observer (DeAngelus & Pelz, 2009). Additionally, the observer looked at the painting on seven occasions, each time for a 3 minute period. DeAngelus and Pelz replicated the findings of Yarbus using a sample size of 17. They allowed participants unlimited time to view a selection of images which included the same image used by Yarbus. The same pattern of eye movements was found. For example, when asked to "give the ages of the people", the eye movements showed a "uniform distribution across the faces, with very little attention given to background elements." (DeAngelus & Pelz, 2009, pp. 802). This corroborates the findings of Yarbus that the observer focussed attention on the faces in the painting when told to estimate ages of people in the painting.

More recently, Castelhano, Mack, and Henderson (2009) compared participants' eye movements when asked to either memorise a scene, or search for a specific object in that scene. Participants viewed each scene for 10 s. Fixations were approximately evenly distributed across the scene during the memorisation condition. However, in the visual search condition, fixations were mostly made on regions of the scene which were most likely to contain the target. For example, when searching for a bucket in a scene consisting of a hardware shop located on the corner of a snowy street, participants tended to exclusively fixate the shop windows, and not make fixations to other parts of the scene. Castelhano, Mack and Henderson argued that participants used their previous knowledge about the context of the scene to guide eye movements to the areas in which the target was most likely to be. Put simply, participants knew from previous experience that a bucket would be highly likely to be found in the shop window, and therefore prioritised fixating that portion of the scene. In a more general sense, this provides further evidence that participants' eye movements differed depending on whether they were searching for an object, or memorising the scene.

In the research presented above, participants chose to look at certain objects or regions of scenes at the expense of others in response to specific task instructions. However, a critical issue in scene perception is how participants choose which objects or regions are most relevant to the task instructions without looking at the objects first. In other words, how do participants know whether that object is worth looking at before making a fixation on it? The most likely explanation is that participants use prior knowledge about the context of a scene to guide their eye movements to the locations which are most likely to contain the relevant information. Previous research has demonstrated how prior knowledge of a scene can influence where people choose to look. Castelhano and Henderson (2007) instructed participants to search for a target object in a test scene. However, prior to searching for the object, participants received a 250 ms scene preview. This preview was identical to the test scene, different to the test scene, or was a meaningless patterned mask preview.

The results showed that at test, participants fixated the target object significantly faster if they had seen an identical scene preview than if they had seen a different scene preview or a meaningless mask preview. This suggested that the 250ms identical scene preview had allowed participants to form some memory for the layout and context of the scene, and this memory helped guide eye movements to the most likely target location at test. As studies on gist have demonstrated that the gist of a scene can be encoded within 150 ms (Biederman, Mezzanotte & Rabinowitz, 1982), 125 ms (Potter, 1976) or even 26 ms (Rousselet, Joubert & Fabre-Thorpe, 2005), a 250 ms scene preview presumably allowed participants to encode the gist of the scene and this knowledge about the gist was likely to have facilitated search for the target object. Research on gist is discussed in more detail in Chapter 5 in this thesis.

3.1.2 Higher-level influences on eye movements: Prior knowledge

Higher-level influences which affect where attention is allocated have also been demonstrated using head-mounted eye-trackers in everyday tasks. It is likely that prior knowledge about the task influences where participants look in these tasks. For example, Land, Mennie, and Rusted (1999) recorded eye movements as participants made a cup of tea in a kitchen. The most relevant finding was that participants mostly fixated task-relevant objects (e.g., kettle, mug) and ignored task-irrelevant objects (e.g., window, cooker) even though there were a large number of irrelevant objects in the scene. This suggests that participants were using their previous knowledge about the correct way to make a cup of tea to guide their fixation position on to the task-relevant objects in the environment.

Similar results were demonstrated in a sandwich-making task by Hayhoe (2000). Land and Hayhoe (2001) also used a sandwich making task where participants sat at a table and prepared a peanut butter sandwich. On the table was a mix of task-relevant objects (e.g., knife) and task-irrelevant objects (e.g., pliers, scotch tape). Before the sandwich making task began, the eye movements made by participants as they scanned the table in preparation were analysed. It was found that on average the proportion of irrelevant objects that were fixated was 52%. However, once the sandwich making task began, the task-irrelevant objects was fixated less often; the proportion decreased to only 18%. Land and Hayhoe suggested that this demonstrated that higher-level factors (specifically how relevant each object was to the task) were driving eye movements leading participants to fixate objects related to the task, and ignore objects that were not relevant. Taken together, these findings show that participants can and do modify their eye movements contingent on the task they are performing and their previous knowledge.

Other related research has found that participants use previous knowledge about motion to fixate empty, yet informative, regions of the visual environment. Land and McLeod (2000) analysed the eye movements people made when playing cricket. In their experiment, cricket players wore a head-mounted eye-tracker whilst batting. It was found that as the ball was bowled towards them, participants initially fixated the ball, and then made a saccade to fixate the location where they anticipated the ball would bounce before it did so. Land and McLeod stated that "the fovea thus 'lay in wait' for the bounce" (pp. 1341). This finding further demonstrates that higher-level factors (e.g., previous experience about the typical motion pattern of a cricket ball relative to the motor requirements of successful batting) can influence where in a scene participants choose to focus depending on the task. Clearly, there were no lower-level factors attracting attention towards the bounce location as it was a blank region of floor. Instead, previous knowledge was used to predict where the ball would bounce. The finding that participants fixate a blank region when anticipating the bounce of a ball (approximately 200 ms before it bounces) has also been found with table-tennis players (Land & Furneaux, 1997).

3.1.3 Higher-level Influences on eye movements: Linguistically-driven eye movements

Despite the host of previous research that participants tend to fixate informative regions of scenes, a line of research has developed which demonstrates that under certain conditions, participants choose to fixate blank regions of the scene. As stated in Chapter 1, Richardson and Spivey (2000) presented participants with videos of people reading out facts. The videos were presented in one of four possible positions in a 2x2 array. Once the video had finished, participants listened to facts, some of which they had already heard in the videos. During this listening period, participants often looked at blank regions of a display if the fact they heard had been spoken by someone in a video presented in that region.

Similarly, Altmann (2004) investigated whether participants move their eyes to portions of space that were currently empty, but were previously filled with relevant information. In this study, participants completed 20 experimental trials. On each trial, participants saw a scene for 5 s. Each scene consisted of drawings with four aspects: two people (protagonists) and two items. Participants were simply instructed that they would see some scenes. They were not asked to perform any particular task. After the scene was removed, it was replaced by a blank screen and participants heard a sentence. On experimental trials, the sentence referred to one of the protagonists and one of the items in the scene (e.g., the *man* will eat the *cake*). The results showed that as the name of a particular object was spoken, participants often looked towards the region of the blank screen where that object was previously located, even though the screen was now blank.

Additionally, Altmann found that whilst the verb was spoken (e.g., "*eat*"), participants often made a saccade towards the region of the blank screen which had previously contained the object that the verb preceded. For example, when the sentence "the man will eat the cake" was spoken, a saccade was made towards the cake region whilst "eat" was spoken, suggesting that participants could accurately predict the end of the sentence.

A critical question is why did participants look at the blank regions? Altmann suggested that when the objects are looked at, an episodic trace is formed for each one. When this trace is activated (via the spoken sentence), an eye movement is automatically made towards the known location of that object, even though it is no longer present. Importantly, both the research by Richardson and Spivey (2000) and Altmann (2004) demonstrate that in certain tasks, the dominant factor in leading participants to fixate a blank portion of the visual array is language. Under such circumstances, rather than visual attention being influenced by task instructions or previous experience, it is driven by the linguistic content of the utterance that is heard and fixations are very rapidly made on the portion of the scene to which the word or words are referentially associated. This is in contrast to the research by Land and McLeod (2000) and Land and Furneaux (1997) where the dominant factor in leading participants to fixate blank portions of space was previous experience about the typical motion pattern of a ball and the batter's attempt to hit it accurately. However, regardless of this, both areas of research demonstrate that higher-level factors can strongly influence where observers look.

Further evidence of participants fixating blank, yet informative, regions of a scene comes from Theeuwes, Kramer and Irwin (2011). In this experiment participants were presented with a display containing four different coloured circles on a white background. One circle was located in each of the four corners. Participants saw this display for 100 ms while carrying out a verbal overshadowing task (repeating a two-digit number out loud). After 100 ms, the circles disappeared and all that remained on the screen was the plain white background. After a 900 ms delay, participants were asked whether a specific coloured circle had been present in the delay (e.g., "was red present?") and they had to respond as quickly and accurately as possible. On some trials, after the question was asked, a dot was presented in one of the four previously filled corners for 100 ms. The most relevant results were obtained on trials where the coloured circle probed in the question had been present in the display. On a quarter of these trials, the location of the dot coincided with the location of the coloured circle referred to in the question. Response time to the question was significantly faster on these trials than on trials when the dot was in a different location to the location of the coloured circle referred to in the question. What this suggests is that when participants try to remember information about a visual display they move their eyes to the location that was filled by the information they are trying to remember.

Another example of participants fixating blank regions is an experiment by Spivey and Geng (2001). In their experiment participants wore a head-mounted eye tracker and completed an experiment in which they had to move objects around on a table. Participants were then told that they were going to have a break from the experiment and that they should face the wall to face a blank projector screen and would listen to some descriptions of visual imagery for 5 minutes, then continue with the object moving experiment. They were also told that they eye tracker would be switched off during this break. However, the object moving experiment was a sham experiment, and the data of interest was the eye movements made by participants during the “break”. During this part of the experiment, the eye tracker was not switched off and participants listened to five different descriptions. Four of the descriptions included specific directional content (leftwards, rightwards, upwards and downwards). One was a control description in which no directional content was included.

The eye movement data taken from the “break” showed that, despite the projector screen being blank, participants tended to make eye movements that matched the description that they were hearing. For the four directional descriptions participants were more likely to make saccades in the corresponding direction than in the control condition. Furthermore, participants were more likely to look in the prescribed direction than any of the other directions during each of the directional descriptions. This provides evidence that under certain conditions participants fixate blank and uninformative regions of scenes.

In a similar experiment, Johansson, Holsanova and Holmqvist (2006) had participants listen to directional descriptions and recorded their eye movements as they did so. Similar to the experiment by Spivey and Geng, the participants were facing a plain white board. After hearing each description, the participants had to re-tell the descriptions in their own words. The eye movement data showed that participants made relevant directional saccades as they re-told the story even though they were looking at a plain white board. Therefore there is a wealth of evidence that demonstrates that, under certain experimental conditions, participants look at blank and seemingly uninformative regions of scenes or the environment.

3.1.4 Lower-level Influences on eye movements: Saliency maps

In contrast to the research that has shown that higher-level factors influence where participants look in scenes, research has also shown that lower-level factors can be used to predict where participants look. However, these studies typically use paradigms which do *not* involve participants taking part in any active looking tasks (examples of active looking tasks include visual search or memorisation). Instead, participants are typically told to just look at the images. For example, Reinagel and Zador (1999) showed participants a series of black and white images and found that participants looked more often at the regions of the images that contained higher spatial contrast. Mannan, Ruddock, and Wooding (1996) demonstrated that when participants viewed images for 3 s, the regions of the scene where they fixated tended to contain higher spatial frequency content (with high levels of visual contrast) than non-fixated regions. This suggests that when given a brief view of a picture, participants tend to fixate regions that attract attention due to their lower-level properties.

Parkhurst, Law, and Niebur (2002) had participants view a series of coloured images each for a 5 s period. They instructed participants simply to look around the images. For each image, a saliency map was created, and the eye movements made by participants were compared to the saliency map. The saliency map for each image represented the initial image in terms of the extent to which different areas of the image contained the greatest degree of orientation, colour and intensity. As a result, for each image, a black and white image was created which showed the most salient regions in white and the least salient regions in black. Figure 5 below shows an example of saliency map and its original image (Parkhurst, Law & Niebur, 2002).

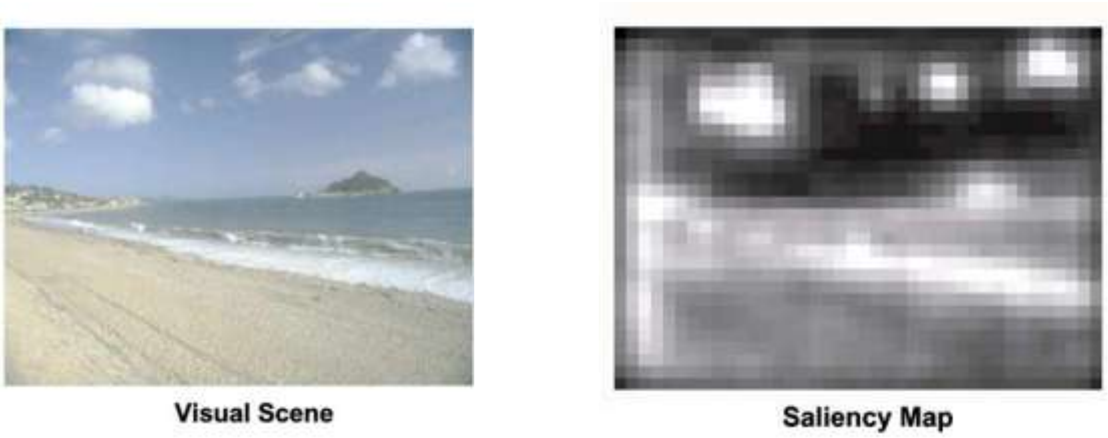

Figure 5. Example of an original image (left) and its saliency map (right) copied from Parkhurst, Law and Niebur, 2002 (pp. 109).

The results showed a significant correlation between the locations participants fixated, and the degree of salience at that location, with participants tending to fixate the most visually salient regions whilst making few fixations to the least salient regions. Parkhurst, Law, and Niebur reported that the location of the first fixation was typically more salient than expected by chance, providing evidence that the location that participants first fixate seems to be driven by lower-level influences. In addition, a number of other studies have provided evidence that participants often fixate regions of scenes that contain high levels of contrast (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Torralba, Oliva, Castelhano, & Henderson, 2006).

The idea of a map detailing where the areas with highest saliency were located was an influential way of considering how observers view scenes. Perhaps the most dominant model was the Saliency Map (Itti & Koch, 2000; 2001). As in Parkhurst, Law and Niebur (2002), the Saliency Map model produced a modified version of each image which represented the extent to which colour, intensity and orientation were highly salient. The aim of creating a saliency map is to form an objective distinction between different areas of the scene. An assumption that underpins this approach is the portions of scenes that are relatively uniform are less interesting, and therefore are less likely to attract attention via fixations, than those portions which contain more low-level variation (e.g., colour, contrast). Once the saliency map is calculated, its properties can be correlated with the eye movements observers made on the original image to see if more fixations are made to the regions of greatest interest. Another model of visual attention that stated that lower-level factors influenced where observers look in a visual scene was proposed by Findlay and Walker (1999). They presented a model of saccade generation which focused on where and when participants choose to move their eyes when viewing a scene. According to their model, there is competition between different areas of a scene and a winner-takes-all process selects the location to which each saccade is directed. As one location is selected and fixated, the other areas of the scene are inhibited so that attention can be focused on the fixated region.

It is important to note that the current dominant perspective is that low-level salience models are not the most effective way to explain how observers view scenes. Accordingly, Tatler (2009) stated that, "the case against a low-level salience model now seems overwhelming" (pp. 785). Indeed, research that analysed the relationship between fixations and visual properties has asserted that these correlations are small. Tatler, Baddeley, and Gilchrist (2005) asked participants to view a series of 48 images in preparation for an immediate memory task. Each image was presented for a random period, ranging from 1 s to 10 s. For analysis, saliency maps were created for each image. The results showed that early on in scene viewing, participants tended to fixate the most salient regions of the scene (which were typically located in the centre of the image), but as scene viewing continued, the location of fixations became less predictable. Tatler, Baddeley, and Gilchrist stated that the most plausible explanation for these findings was a strategic divergence model whereby attention (and fixation) is initially drawn to the most salient portion of a scene, but at some point during scene viewing participants choose where to look, leading to variability in the fixation positions. As participants knew they would be probed on their memory for the scene, but did not know specifically which aspects would be probed, presumably each participant adopted a different strategy. This led to an unpredictable spread of fixations across the scenes.

Tatler et al. considered saliency maps, and stated that two factors may influence potential correlations between fixation positions and saliency. First, many photographs of natural scenes have higher salience in the centre of the image and lower salience around the periphery. This is often due to the tendency to place the most interesting aspect of a scene in the centre. Second, there is often a bias in scene viewing to fixate the central portion of an image. It may be that this tendency is driven by the higher level of salience often found at the centre of images, or it may reflect prior knowledge that the most interesting aspect of images is often found in the centre. Either way, these factors must be considered when analysing the relationship between fixation and salience.

Further evidence against the saliency map perspective of visual attention was provided by Henderson, Malcolm and Schandl (2009) who recorded eye movements as participants completed a visual search task for a pre-defined object. On critical trials, the target object was chosen deliberately to be visually non-salient. However, in each scene, there were a number of visually salient distractors that were not relevant to the search task. The results showed that participants fixated the non-salient target on 90% of trials, but only fixated the salient distractors on fewer than 10% of the trials. Participants tended to fixate task relevant, rather than visually salient, objects. These results provide evidence that participants tend to fixate regions of the scene that are relevant to the task, rather than fixating the most visually salient region of the scene. A critical point to make here is that whether an object is or is not visually salient, is determined to a significant degree by what the task instructions are.

3.1.5 Lower-level influences on eye movements: Colour and visual search

In the present experiment, bright green rectangles were placed around selected objects to draw attention to those objects. These green boxes were chosen and used as a lower-level attentional capture technique. This technique was employed because there is substantial evidence to demonstrate that bright colours can strongly attract visual attention. Most notably, this has been demonstrated in the Visual Search paradigm. In a typical visual search paradigm, participants are presented with a display which contains a range of distractor objects. Participants are instructed to search for a pre-defined target object, and to respond when they find it. Some trials do not contain a target (catch trials) and both speed and accuracy of response are emphasised. Note that eye movements are typically not recorded in these experiments.

Theeuwes (1991) presented participants with displays consisting of several distractors which were angled lines and a single target which was a line oriented either horizontally or vertically. All stimuli were arranged in a circle around a central point. On each trial, participants were instructed to search for the target line but were not allowed to move their eyes around the display. Instead, participants had to keep fixating the centre of the display. The stimulus lines were all located within coloured circles. Participants had to search for the target and respond by indicating whether the target was a horizontal or vertical line. On some of the trials, the target was located within a green circle whilst all of the distractors were located within a red circle. On other trials, the target was in a red circle while the distractors were in green circles. In addition, on some of the trials the target and distractors were located in the same colour circles. The set size was also manipulated; trials either contained 5, 7, or 9 stimuli.

The results showed that for control trials where participants had to shift their attention around all the stimuli to find the target, response time and error rate increased as the set size increased. However, on trials where the target was located within a uniquely coloured circle (red amongst green or green amongst red) there was no effect of set size. Instead, response time and error rate did not increase as set size increased. This suggests that locating the target within a uniquely coloured circle led to that target "popping out" of the display and made it easy to find. In other words, the uniquely coloured circle strongly attracted visual attention.

This was further demonstrated by Theeuwes (1992) in a modified version of the visual search task described above. In trials where the target was located in a circle that was the same colour as the distractors, but a single distractor was located within a uniquely coloured circle, response time and error rate increased as the set size increased, suggesting that the uniquely coloured distractor interfered with the search for the target which did not "pop" out. Instead, one of the distractors "popped out" by virtue of its unique colour. This further demonstrates that colour can attract attention strongly (see also Northdurft, 1993; Theeuwes, 1994; Mounts, 2000).

In a more recent experiment, Turatto and Galfano (2000) asked participants to search for a vertical line target amongst 5 slanted line distractors arranged in a circular pattern around a central point. In Experiment 1, all lines were presented within a coloured circle. On every trial, 5 of the circles were red and one was green. Accordingly, on one sixth of the trials, the vertical line target was presented within a green circle, rendering the target extremely salient. The display was presented for 180 ms, and results showed that when the target coincided with the salient green circle, accuracy was highest. Even though colour was not relevant to the line detection task, the results suggest that attention was preferentially drawn to the green circle, leading to extremely accurate performance when that coincided with the target.

It is important to note, however, that there is evidence that colour does *not* capture attention. Gibson and Jiang (1998) conducted a visual search experiment in which participants had to search displays of eight letters for an H or a U. On each trial, the eight letters were presented for 86 ms. For the first 192 trials, all eight of the letters were white and, after being presented for 86 ms, were masked by a white rectangle. These trials were standard conjunction trials as finding the target was a demanding task which required participants to serially search each of the letters to find the target. After 192 of these trials, there followed immediately a surprise singleton trial where the target was the only red letter amongst the white distractors. After the single surprise trial, 192 standard feature trials followed. In each of these trials, the target was red.

As expected, participants correctly identified the target more often in the 192 feature trials (M = 0.93) than in the 192 conjunction trials (M = 0.69). However, critically, accuracy for the single surprise trial was only 0.78 which was no different to the average accuracy for the conjunction trials. Gibson and Jiang argued that this provided evidence that colour does not immediately capture attention. The surprise appearance of a uniquely coloured target did not lead to more efficient target recognition accuracy in the first instance.

Another experiment that provided evidence that colour does not capture attention was conducted by Irwin, Colcombe, Kramer and Hahn (2000). They had participants search for a target (the letter C) amongst other distractor letters. All letters were enclosed within grey circles which surrounded a central fixation point. At the beginning of each trial, no letters were presented in the circles. Instead, each grey circle contained a figure-8 placeholder. After 1500 ms, several things occurred simultaneously. The figure 8s changed into letters (one of which was the target letter C). All except one of the grey circles changed colour and became red. At the same time, a new circle appeared in the display in a previously unoccupied location.

On some trials, the target letter C was presented in one of the initial circles which did not change colour while every other circle became red. This meant that the new circle, which was also red, was an irrelevant onset distractor. In other words, a new object appeared but it was not uniquely coloured. Therefore, the target was located in a uniquely coloured existing circle. On other trials, the target appeared in the new circle, but this new circle was *not* uniquely coloured. Thus, there were two possibilities for the target. It was either located in an existing uniquely coloured circle (which was competing with the onset of a new object), or the target was located in a new non-uniquely coloured circle (which was competing with an old uniquely coloured circle). Importantly, for both types of trial, there was also a counterpart control condition where nothing competed with the target letter.

The results demonstrated that accuracy was very high (over 99%). The critical finding was that when the target was presented in a new circle, response time was no different when there was an existing uniquely coloured distractor present in the display or not. However, when the target was presented in an existing uniquely coloured circle, response time to the target was slower if there was a new grey circle present than if there was not. In other words, even though the target was presented in a uniquely coloured circle, the presence of a new object appearing in the display attracted attention and led to a slower response time to the target. Irwin et al. argued that the onset of a new object captures attention in an involuntary way and can override the potential attention-capturing characteristics of the uniquely coloured object. Importantly, when the target was presented in a new circle, the presence of an existing uniquely coloured circle did not lead to slower response times to the target, indicating that a uniquely coloured distractor did not capture attention in the same way as a new object. What this seems to suggest is that colour does not necessarily attract attention, although in the experiment by Irwin et al. colour was competing with the onset of a new object and this new object onset seemed to capture attention very strongly.

Irwin et al. also recorded eye movements throughout the experiment and reported how participants' eye movements changed depending on the target location i.e. in a new grey circle or an existing red circle. Saccade latency (the time it took for the eyes to shift from the central fixation point to one of the stimuli) to the colour target was slower when a grey onset object appeared than when no onset object appeared. However, when the target was in a grey onset object, saccade latency was comparable when an existing circle turned red or when no existing circle turned red. The eye movement behaviour closely matched the response time behaviour and provides further evidence that colour may not capture attention involuntarily.

Other experiments have provided evidence that suggests that colour might not capture attention (Jonides & Yantis, 1988; Todd & Kramer, 1994; Franconeri & Simons, 2003). What is important to note is that studies that demonstrate that colour does not capture attention typically compare colour against a competing attention-capturing technique e.g., new object onset. In the present experiment presented in this thesis, no such competition occurred. Instead, some of the objects were surrounded by a coloured box and the rest of the objects were not. On this basis, it was expected that the coloured boxes would lead participants to fixate the objects in the coloured boxes preferentially.

3.1.6 The present experiment

Clearly, a host of previous research has demonstrated that visual attention can be allocated to certain aspects of visual displays for at least two particular reasons. First, visual attention might be allocated to certain objects due to the visual saliency of those elements of the display. For example, a uniquely coloured object in an array is likely to be attended to preferentially at the expense of the other objects. Second, attention might be allocated to certain objects because they are highly relevant to the task e.g., a visual search task. . It should be clear from the preceding discussion that the extent to which an aspect of the display is visually salient can be determined by either just one, or possibly both of these mechanisms. What is important to acknowledge is that if in the present experiment participants preferentially fixate the boxed objects ahead of the non-boxed objects as expected, it will not be possible to distinguish between two possible explanations for this preferential attentional allocation. It is possible that participants will allocate more attention to the boxed objects than the non-boxed objects because the boxed objects attract attention more readily due to the presence of the bright green boxes. Alternatively (or perhaps additionally) the boxed objects may be preferentially attended to because the green boxes lead participants to prioritise attending to those objects instead of the non-boxed objects. Even though participants are not informed that some of the objects will be boxed, participants may interpret the boxing to mean that some objects are more important than others and change their oculomotor behaviour accordingly.

Regarding the present experiment,, participants were instructed to look at the photograph and remember as much as possible in preparation for an immediate memory test. As in Experiment 1, it was expected that these task instructions would result in participants choosing to primarily fixate the 12 objects in the scene. In addition, lower-level characteristics of half of the objects were manipulated to preferentially draw attention to those objects in the photograph at the expense of others. This was achieved through the use of the green boxes. Importantly, the boxed objects were expected to be fixated preferentially for two possible reasons; because they were surrounded by brightly coloured boxes and therefore attracted attention, and because participants chose to pay more attention to them due to the presence of the green boxes leading participants to prioritise the boxed objects ahead of the non-boxed objects.

The purpose of placing boxes around half of the objects was to see whether the fundamental relationship between encoding and recall for object identity and location observed in Experiment 1 could be influenced. As the results of Experiment 1 suggested that there is a relationship between the amount of visual encoding an object receives, and the accuracy with which that object is selected and placed, it follows that if a manipulation was employed that could cause some objects to be preferentially processed over others during encoding, then there would be a corresponding difference in recall behaviour. In turn, it was expected that those preferentially fixated objects would be preferentially selected and placed, providing further evidence of a strong relationship between encoding and recall. To modulate the relationship, bright green rectangles were superimposed around half of the objects in the photograph (boxed objects), leaving the other half of the objects without rectangles around them (non-boxed objects).

It is important to note that the methodology of the test phase was modified slightly from Experiment 1. In the present experiment, participants gave confidence ratings that represented how confident they were that each of the 12 selected objects had been present in the photograph. The reason for this addition to the procedure was to investigate how confidence for object identity related to memory for object identity. Previous research (Shaffer & Shiffrin, 1972) has suggested that confidence for memory of visual scenes is very closely related to memory for visual scenes, and therefore the procedure was modified to investigate the degree to which participants possess veridical insight into the quality of their memory for the identity and locations of objects. In the study by Shaffer and Shiffrin, participants viewed 120 images in preparation for a later recognition test. Each image was viewed for 200 ms, 500 ms, 1000 ms, 2000 ms or 4000 ms. At test, participants viewed a series of images, some of which they had seen before ("old") and some of which they had never seen before ("new"). Participants had to state how confident they were that they had seen each test picture before. Confidence was measured on a 6-point scale where 1 meant "not at all confident" and 6 meant "completely confident".

The results showed that confidence increased as the presentation time of an image increased. In addition, the results demonstrated that the likelihood of an image being correctly categorised as old or new increased as the presentation time increased. What this suggests is that the degree to which people are confident about their memory for visual stimuli is closely linked to their actual memory for visual stimuli. However, it is important to note that the issue of memory confidence is not the primary focus of this thesis. Whilst the confidence ratings are reported in the results section for completeness, the results that will be considered in most detail will be the data pertaining to the relationship between encoding behaviour and object identity and location memory.

Predictions for Experiment 2 were based on the Visual Memory Model (Hollingworth & Henderson, 2002), the literature presented above and the findings of Experiment 1 in this thesis. The first expected finding was that the basic characteristics of saccades and fixations made during encoding for the participants in Experiment 2 who viewed a photograph without any green boxes (the control condition) would be very similar to those in the corresponding condition from Experiment 1 (participants who had a 5s viewing duration and an immediate test). Because participants experienced identical encoding and test phases, and eye movements were recorded on the same eye tracker, it was expected that these data would be similar, thereby providing a direct replication of the basic eye movement findings observed in Experiment 1.

The second results pattern that was expected was that the basic characteristics of eye movements would be similar between participants who saw a photograph where half the objects were boxed and participants who saw a photograph where none of the objects were boxed. The reason for this was that the green boxes would not fundamentally change the nature of the eye movements participants made when they looked at the photograph. Instead, it was expected that the green boxes would change which of the objects were looked at.

The third prediction was that participants would preferentially fixate the boxed objects rather than the non-boxed objects on the basis of the boxed objects being more visually salient than the non-boxed objects. This would be reflected by participants fixating more boxed objects than non-boxed objects, fixating these boxed objects for longer than non-boxed objects, and fixating boxed objects before fixating non-boxed objects. This prediction was based on the findings presented earlier that, under certain task instructions, highly visually salient aspects of scenes are often fixated preferentially due to them capturing attention.

Fourth, it was also expected that this prioritisation of boxed objects during the encoding phase would be reflected in the recall phase, with more accurate memory for identities and locations shown for boxed than non-boxed objects. In addition, it was expected that, as in Experiment 1, an increased number of fixations on objects would lead to more accurate memory for the identity, but not the location, of objects. This pattern of results was found in Experiment 1, and therefore it is sensible to predict that it will be replicated in Experiment 2. This prediction was based on the claims of the Visual Memory Model and the findings of Experiment 1 in this thesis. According to the Visual Memory Model, when an object is fixated, a memory representation is formed for the identity of that object. This memory representation can be accessed at a later time if needed (e.g., at test) and is also known as a short-term memory object file. It initially contains abstract details about the object, but as the time spent fixating the object increases, more and more information about the identity of the object is encoded, such as its shape, colour and size etc. Hence, as the number of fixations made on an object increases, the degree to which the object file is embellished with specific details about the object also increases. However, the Visual Memory Model states that memory representations are not formed for the identity of non-fixated objects. Therefore, it was predicted that as non-boxed objects were less likely to be fixated, it was less likely that a memory representation for those objects would be formed, and therefore it was less likely they would be recalled at test.

Regarding object location memory, it was predicted that memory for the locations of objects would be more accurate for fixated than non-fixated objects, but would not improve as the number of fixations made on an object increased. This pattern of results was found in Experiment 1. However, it was not known whether the relationship between encoding and recall would be the same for both boxed and non-boxed objects. It could be argued that boxing objects *per se* may have influenced the encoding-recall relationship, rather than any modulation occurring due to attentional factors (such as preferential encoding of the boxed objects).

Finally, regarding confidence in memory for object identity and locations, it was expected the boxed objects would be selected more confidently than non-boxed objects. This prediction was based on the expectation that boxed objects would receive more visual processing during the encoding phase than non-boxed objects. In turn, as found by Shaffer and Shiffrin (1972), it was predicted that confidence would be closely linked to memory performance. Therefore, as boxed objects were more likely to be fixated, and also selected, than non-boxed objects, they were also likely to be selected more confidently.

3.2 Method

Participants and design

The 38 participants (*M* = 19.47 years, *SD* = 2.31 years) who volunteered to participate voluntarily or in return for course credit were all undergraduates or postgraduate students from the University of Southampton. Three were male and thirty-five were female. All participants had normal or correct-to-normal vision, and were naïve to the aims of the experiment. None of the participants had taken part in Experiment 1. Data from 8 participants was excluded from the analyses due to tracker loss during the encoding phase. This left data from 30 participants.

A mixed 2x2 design was employed with one between-groups variable (Photograph A vs. Photograph B) and one within-groups variable (boxed objects vs. unboxed objects). Objects that were boxed in Photograph A were unboxed in Photograph B, and vice versa. There was also a control condition. Participants were randomly assigned to view one of the three photographs. In total, 10 participants saw Photograph A, 10 participants saw Photograph B and 10 participants saw the control photograph.

Apparatus

The apparatus was identical to that used in Experiment 1.

Stimuli

The stimulus photograph presented to the control condition was identical to the stimulus photograph presented to all participants in Experiment 1. However, participants in the two experimental conditions (Photograph A and Photograph B) saw a modified version of the stimulus photograph (see Figure 6).

The modified version had six bright green rectangles superimposed on the photograph, and these rectangles formed boxes around six of the 12 objects in the photograph. In Photograph A, six of the objects had a green rectangle around them for the duration of the display period whilst the other six objects did not have a green rectangle around them (see Figure 6 below). In Photograph B, the green rectangles were around the six objects that had previously not had a rectangle around them (i.e. the six previously boxed objects were non-boxed, and the previously non-boxed objects were now boxed).

Procedure

The general procedure was the same as in Experiment 1, except that in all three conditions, the photograph was only displayed for 5 s, and the memory test always occurred immediately after viewing the photograph. As demonstrated in Experiment 1, a 10 s encoding phase resulted in participants fixating on average 10 objects (a 5 s encoding phase resulted in 7.23 objects being fixated on average).

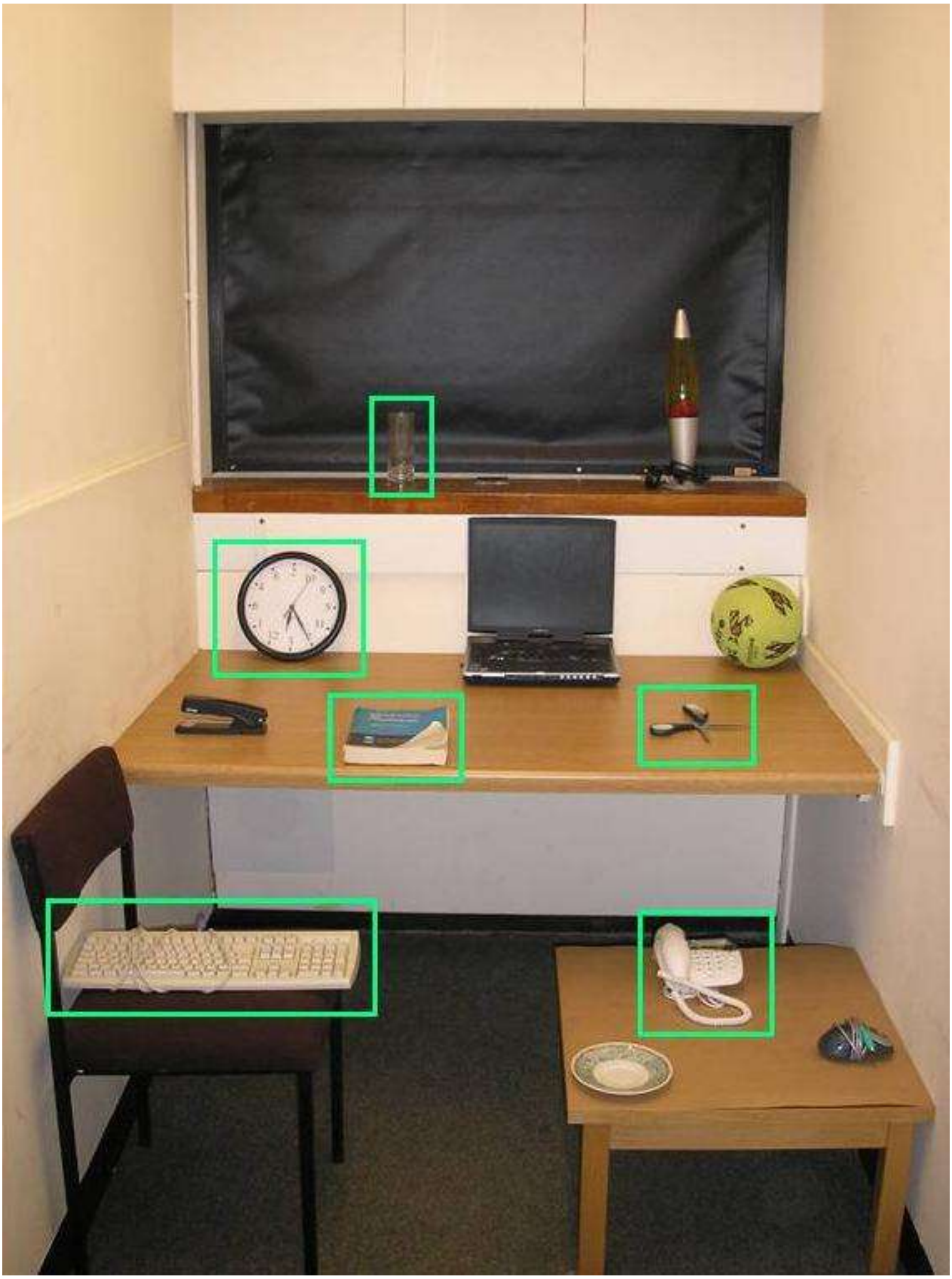

Figure 6. The modified stimulus photograph presented to Group A. Group B saw the same photograph but all boxed objects were unboxed, and all unboxed objects were boxed.

Therefore, it was expected that the brief 5 s encoding period would maximise any potential encoding advantage for the boxed objects versus the non-boxed objects. Additionally, after completing the memory test, participants were asked to indicate how confident they were that each object selection was correct on a rating scale from 0 – 10 (0 = complete guess, 10 = complete confidence).

3.3 Results

The same three sets of analyses were carried out as in Experiment 1; (1) Encoding behaviour, (2) Recall behaviour and (3) the Relationship between Encoding and Recall Behaviour. For all analyses, any fixations shorter than 80 ms or longer than 1200 ms were excluded from the analysis.

Analyses of Encoding Behaviour – Comparison of Experiment 2 with Experiment 1

The first analysis compared the encoding behaviour made by participants in Experiment 2 who viewed the photograph without any green boxes (control condition) with the participants from Experiment 1 who had a 5 s viewing duration and an immediate test. Methodologically, these participants experienced identical encoding and test phases and therefore it was expected that there would be no differences between them for any of the encoding measures. In addition, if no differences were observed, it would mean that Experiment 2 directly replicated the encoding behaviour found in Experiment 1.

A series of between-groups *t*-tests found that there were no significant differences between the two groups for any of the encoding measures (all *p*s > .10). This demonstrates that the eye movements observed in the control condition of Experiment 2 directly replicated the corresponding condition from Experiment 1. As there were no differences between Experiment 1 and Experiment 2, the data from the 11 participants who viewed the stimulus for 5s and were tested for recall immediately in Experiment 1 were combined with the data from the participants in the control condition from Experiment 2 (no objects boxed). This created an enlarged control condition with 21 participants (11 from Experiment 1 and 10 from Experiment 2), and a boxed condition with 20 participants (all from Experiment 2).

Table 5

Means and Standard Errors (in Parentheses) of Eye Movement Measures for all Control Participants in Experiment 2 (No Objects Boxed) and the Participants in Experiment 1 who Experienced a 5 s Viewing Duration and Immediate Test

Condition:	Experiment 2 (no boxes)		Experiment 1 (5s/0 delay)	
Measure	M	SE	M	SE
Total time fixating objects (ms)	3339	(123)	3104	(283)
Proportion of viewing duration spent fixating objects	0.84	(0.02)	0.79	(0.07)
Number of objects fixated	7.40	(0.43)	7.09	(0.67)
Average first fixation duration (ms) *	276	(20)	292	(13)
Average fixation duration (ms) *	269	(15)	271	(13)
Average gaze duration (ms) *	377	(33)	400	(31)
Average number of fixations made during gaze duration *	1.40	(0.06)	1.48	(0.14)
Average total time (ms) *	466	(33)	449	(35)
Number of fixations made during total time *	1.74	(0.09)	1.70	(0.17)
Number of objects refixated	2.00	(0.42)	1.09	(0.31)

* denotes that this measure is contingent on an object being fixated

Additionally, eye movement data from the two experimental conditions (Photograph A – in which half the objects were boxed, or Photograph B – in which the other objects were boxed) were combined. A series of between-groups *t*-tests showed no significant differences for any of the encoding measures (all *p*s $> .10$), and in addition, the Photograph variable was not theoretically of interest. Therefore, for the subsequent analyses, the participants who saw a photograph with half of the objects boxed were combined to create an enlarged experimental group.

Analyses of Encoding Behaviour – Comparison of Half Objects Boxed With No Objects Boxed

A series of between-groups t-tests demonstrated that of the 10 encoding measures, no difference was found between the Experimental condition (half objects boxed) and the Control conditions (no objects boxed) for eight of the eye movement measures. However, a difference was found for two of the measures. For one of the *t*-tests reported below, the Levene's test for equality of variances was significant. This means that the variances of the two groups were not equal and as a result, the degrees of freedom associated with the test were not N – 2 but instead were decimal numbers.

There was no difference between the Experimental and Control conditions for the following variables: Total time spent fixating objects, $t(39) = 1.61$, $p > .05$; Proportion of viewing duration spent fixating objects, $t(39) = 1.10$, $p > .05$; Number of objects fixated, $t(39) = -.18$, $p > .05$; Average gaze duration, $t(39) = 1.03$, $p > .05$; Average number of fixations made during gaze duration, $t(39) = -1.03$, $p > .05$; Average total time, $t(39) = 1.51$, $p > .05$; Average number of fixations made during total time, $t(39) = -.53$, $p > .05$; and the number of objects re-fixated, $t(33.12) = -.39$, $p > .05$.

However, there was a significant difference between the Experimental and Control conditions for average first fixation duration, $t(39) = 2.02$, $p < .05$, and average fixation duration, $t(39) = 2.31$, $p < .05$. It is likely that these two differences are due to participants making longer fixations on the boxed objects in the Experimental condition, and this possibility is examined more closely below. However, overall, the pattern of results demonstrates that the presence of green boxes in the Experimental condition did not substantially change the way in which participants looked at the photograph. The fundamental characteristics of saccades and fixations during encoding were broadly similar for participants with half of the objects boxed and participants with none of the objects boxed. However, it is possible that the boxed objects influenced specifically *which* objects were fixated and later analyses directly probe that issue.

Table 6

Means and Standard Errors (in Parentheses) of Eye Movement Measures across Both Conditions (Half Objects Boxed or Objects Boxed)

Condition:	Half objects boxed		No objects boxed	
Measure	M	SE	M	SE
Total time fixating objects (ms)	3512	(90)	3216	(158)
Proportion of viewing duration spent fixating objects	0.86	(0.02)	0.81	(0.04)
Number of objects fixated	7.15	(0.29)	7.24	(0.40)
Average first fixation duration (ms) *	325	(17)	284	(11)
Average fixation duration (ms) *	314	(17)	270	(10)
Average gaze duration (ms) *	419	(18)	389	(22)
Average number of fixations made during gaze duration *	1.34	(0.06)	1.44	(0.08)
Average total time (ms) *	506	(22)	457	(24)
Number of fixations made during total time *	1.65	(0.08)	1.72	(0.09)
Number of objects refixated	1.40	(0.17)	1.52	(0.27)

* denotes that this measure is contingent on an object being fixated

Analyses of Recall Behaviour – Performance in the Memory Test

A series of unpaired *t*-tests showed that there were no differences observed between the Experimental and the Control conditions for any of the memory measures: for the number of objects correctly selected, $t(39) = .67$, $p > .05$; for the average placement error, $t(39) = -.57$, $p > .05$ or for best fit error, $t(39) = .36$, $p > .05$. This demonstrates that the presence of green boxes in the Experimental photographs did not lead to improved overall memory for object identity, object location, or memory for the overall spatial configuration of objects (see Table 7).

Table 7

Means and Standard Errors (in Parentheses) of Recall Measures in the Test Phase for both Conditions (Half Objects Boxed or No Objects Boxed)

Condition:	Half objects boxed		No objects boxed	
Measure	M	SE	M	SE
Number of objects correctly selected	7.95	(0.32)	7.67	(0.28)
Placement error (cm) for correctly selected objects	34.39	(2.13)	36.64	(3.26)
Best fit error (cm) for all 12 selected objects	26.87	(1.23)	26.11	(1.72)

Analysis of the Boxing Manipulation – Encoding Measures

Next, the control participants were removed from the analysis and only data from the 20 experimental participants was considered. A mixed 2 (Photograph: A vs. B) x 2 (Object status: boxed vs. non-boxed) ANOVA was carried out to see if boxed objects received differential visual encoding compared to non-boxed objects. Photograph was included as a variable; however this was theoretically uninteresting as the objects were randomly assigned to be boxed or non-boxed and any potential effects of the

Photograph were not relevant theoretically. As predicted, more of the boxed objects were fixated than non-boxed objects, $F(1, 18) = 48.27$, $p < .001$ (see Table 8).

Table 8

Means and Standard Errors (in Parentheses) of Memory Performance across Both Conditions (Half Objects Boxed or No Objects Boxed)

Condition:	Boxed objects		Non-boxed objects	
Measure	M	SE	M	SE
Number of objects visited (/6)	4.85	(0.24)	2.30	(0.22)
Gaze duration (ms) *	445	(21)	358	(23)
Number of fixations during gaze duration *	1.36	(0.07)	1.26	(0.07)
Average fixation duration (ms) *	352	(24)	280	(12)
Total Time *	2502	(146)	1010	(105)
Number of fixations made in total time *	7.95	(0.65)	3.65	(0.37)

* denotes that this measure is contingent on an object being fixated

It is important to note that some of the non-boxed objects were also fixated, but participants clearly fixated the boxed objects preferentially. In addition, gaze durations were longer for boxed than non-boxed objects, $F(1, 18) = 10.31$, $p < .01$, total times were longer for boxed than non-boxed objects, $F(1, 18) = 44.45$, $p < .001$ and average fixation durations were longer for the boxed objects compared to the non-boxed objects, $F(1, 18) = 14.42$, $p < .001$. Also, boxed objects received more fixations than non-boxed objects, $F(1, 18) = 24.44$, $p < .001$. However, there was no difference between boxed and non-boxed objects in terms of the number of fixations received during gaze duration, $F(1, 18) = 1.28$, $p > .05$. Taken together, these findings provide strong evidence that placing green boxes around some of the objects in a scene led to those objects receiving a greater degree of encoding than objects that were non-boxed.

Analysis of the Boxing Manipulation – Recall Measures

Unsurprisingly, more boxed objects were correctly recalled than non-boxed objects, $F(1, 18) = 11.56$, $p < .001$ (see Table 9). Although this difference was highly significant, it only reflected a difference of about 1 object. Numerically, boxed objects were placed more accurately than non-boxed objects, but this difference was not statistically significant, $F(1, 18) = 1.51$, $p > .05$.

However, there were other advantages to objects being boxed; on average the boxed objects were placed significantly earlier than non-boxed objects, $F(1, 18) = 10.23$, $p < .01$ and they were also placed more confidently, $F(1, 18) = 9.84$, $p < .01$. It is likely that this higher confidence and earlier selection is due to the increased degree of cognitive processing the boxed objects received during the encoding phase.

Table 9

Means and Standard Errors (in Parentheses) for Memory Measures Within Each Condition

Condition:	Boxed objects		Non-boxed objects	
Measure	M	SE	M	SE
Number of objects correctly selected (/6)	4.40	(0.22)	3.55	(0.20)
Placement error (cm) for correctly selected objects	45.8	(4.0)	51.3	(3.7)
Average placement order (/12)	4.44	(0.37)	6.63	(0.35)
Average placement confidence (/10)	5.76	(0.51)	3.41	(0.33)

Relationship between Encoding and Memory for Object Identity

As in Experiment 1, it was investigated whether memory for object identity was better for fixated objects than for non-fixated objects (see Table 10). A mixed 2 (Fixation status: not fixated vs. fixated) x 2 (Photograph: A vs. B) x 2 (Object status: boxed vs. non-boxed) ANOVA was carried out on the probability of correct object selection. As in Experiment 1, objects were more likely to be correctly selected if they had been fixated than if they had not been fixated, $F(1, 12) = 26.36$, $p < .001$.

Table 10

Means and Standard Errors (In Parentheses) for Probability of Correct Object Selection

Condition:	Boxed objects		Non-boxed objects	
Measure	M	SE	M	SE
Probability of correct selection for non-fixated objects	0.45	(0.11)	0.48	(0.07)
Probability of correct selection for fixated objects	0.82	(0.05)	0.91	(0.04)
Probability of correct selection for objects fixated once	0.87	(0.07)	0.83	(0.11)
Probability of correct selection for objects fixated twice or more	0.81	(0.07)	0.80	(0.13)

No effect of Photograph, or Object status was found. Regarding performance relative to chance, for both boxed, $t(19) = 7.12$, $p < .001$ and non-boxed objects, $t(19) = 6.49$, $p < .001$ fixated objects were significantly more likely to be selected than chance (0.5) but non-fixated objects were selected no better than chance (both *p*s > .05).

Next, non-fixated objects were discarded and fixated objects were coded as having received only one fixation, or two or more fixations. A second mixed 2x2x2 ANOVA showed that, unlike in Experiment 1, there was no effect of fixations on memory for object identity, $F(1, 8) = .11$, $p > .05$. Memory for object identity did not improve from one to two or more fixations. This is probably because in Experiment 2 all participants had only 5s to view the photograph and consequently there was a lack of objects that received two or more fixations. Closer examination of the data showed that of the 20 participants, only 14 fixated both boxed and non-boxed objects twice or more and only 15 participants fixated both boxed and non-boxed objects only once. Overall, only 10 of the 20 participants fixated both boxed and non-boxed objects both once, and twice or more. Unsurprisingly, participants were less likely to have fixated non-boxed objects than boxed objects. This explains the low degrees of freedom associated with this ANOVA.

Relationship between Encoding and Memory for Object Location

As before, a mixed 2 (Fixation status: not fixated vs. fixated) x 2 (Photograph: A vs. B) x 2 (Object status: boxed vs. non-boxed) ANOVA was carried out on the placement error. Unlike in Experiment 1, object location memory was not statistically better for fixated than non-fixated objects, $F(1, 7) = .70$, $p > .05$ although the numeric trend was as expected with the mean placement error smaller for fixated than non-fixated objects. The lack of a significant effect is again likely due to a lack of data as a result of the short encoding period. Specifically, boxed objects were rarely ignored during encoding. This led to a severe lack of power in this analysis.

However, when the boxing manipulation was ignored and a 2 (Photograph: A or B) x 2 (Fixation status: Not fixated vs. fixated) ANOVA was carried out, there was a main effect of Fixation status, with fixated objects placed more accurately than non-fixated objects, $F(1, 18) = 5.72$, $p < .05$. There was no missing data because the data were collapsed across both boxed and non-boxed objects. The final analysis was a 2x2x2 ANOVA with the number of fixations (one vs. two or more) included. Due to the short 5s encoding period, very few objects were fixated twice or more and this lack of data likely contributed to the lack of a significant effect of the number of fixations on placement error, $F(1, 6) = .22$, $p > .05$.

Table 11

Means and Standard Errors (In Parentheses) for Placement Error Scores

Condition:	Boxed objects		Non-boxed objects	
Measure	M	SE	M	SE
Placement error for non-fixated objects	56.3	(6.7)	57.1	(8.1)
Placement error for fixated objects	35.4	(8.3)	55.2	(10.1)
Placement error for objects fixated once	56.1	(12.8)	43.9	(13.7)
Placement error for objects fixated twice or more	39.6	(11.6)	42.8	(14.3)

3.4 Discussion

The results of Experiment 2 are important for two reasons. First, the findings replicate the basic encoding behaviour and relationship between encoding and recall observed in Experiment 1. Second, the results extend the findings of Experiment 1 by demonstrating that the basic relationship between encoding and recall can be modulated by influencing which objects participants look at. Each of the findings presented in the Results section will be discussed in turn.

The first finding of note is that the eye movements made by participants who saw a photograph with no green boxes directly replicated the eye movements made by participants in the 5 s/0 delay condition from Experiment 1. This replication is very important because it validates the findings from Experiment 1.

The second finding is that the overall manner in which participants in Experiment 2 looked at the photograph when half of the objects were boxed was not substantially different to the control condition where none of the objects were boxed. Only two of the encoding measures were different; first fixation duration and fixation durations were longer for the photographs where half of the objects were boxed. These two differences probably occurred because participants tended to spend more time looking at the boxed objects than the non-boxed objects. This explanation was confirmed by subsequent analyses. In a more general sense, the lack of substantial differences between the conditions suggests is that the green boxes themselves did not fundamentally change the way participants looked at the photograph.

Regarding performance at test, the results showed that memory was comparable in both conditions, both in terms of the number of objects correctly selected, and the accuracy with which objects were placed. On the basis that there is a close relationship between fixation and attention (Rayner, 1998; Liversedge & Findlay, 2000) and a close relationship between encoding and recall (see Experiment 1 in this thesis) it is unremarkable that memory performance was approximately the same in both conditions. This is because the overall encoding behaviour was not substantially different. To be clear, as there were no fundamental differences in the basic characteristics of fixations and saccades between the half-boxed and no-boxed participants, it is unsurprising that that there were no differences in participants 'memory performance.

The next analyses excluded the control participants (who saw a photograph without any boxes) and only included the participants who saw a photograph in which half of the objects were boxed. As previous research has shown that lower-level factors (Mannan, Ruddock, & Wooding, 1996; Reinagel & Zador, 1999; Parkshurt, Law, & Nieber, 2002) and in particular colour (Theeuwes, 1991; 1992; 1994; Northdurft, 1993; Turatto & Galfano, 2000; Mounts, 2000) can strongly influence where visual attention is allocated during scene viewing, it is unsurprising that the objects surrounded by a green box received preferential encoding compared to the non-boxed object. The presence of green boxes around half of the objects possibly led to those objects being visually more salient than the non-boxed objects thereby receiving more attention during the encoding phase. It is also likely that participants chose to fixate the boxed objects more often and for longer than the non-boxed objects because the presence of the green boxes around half of the objects meant that participants chose to fixate them preferentially because they had prioritised the boxed objects ahead of the non-boxed objects. Regardless of the explanation, the boxed objects were looked at more often, and for longer periods than non-boxed objects. The finding that fixation durations were longer for boxed than non-boxed objects demonstrates is that the green boxes were very effective at capturing attention, and also holding fixation on the boxed objects for longer periods than on the non-boxed objects.

Boxed objects were subsequently more likely to be correctly selected at test than non-boxed objects. This finding is as expected, as boxed objects were more likely than non-boxed objects to have been fixated. An advantage in memory for the identities of fixated objects over non-fixated objects has been demonstrated in Experiment 1 in this thesis and also by several other researchers (Henderson & Hollingworth, 1999; Hollingworth, Shrock, & Henderson, 2001; Hollingworth, Williams, & Henderson, 2001; Hollingworth & Henderson, 2002). In addition, this finding supports the claims of the Visual Memory Model (Hollingworth & Henderson, 2002).

Furthermore, boxed objects were selected earlier and selected more confidently than non-boxed objects which were less likely to have been fixated during the encoding phase. It is likely that that this higher confidence for boxed objects resulted from the greater amount of visual encoding that boxed objects received relative to non-boxed objects. This finding corroborates the pattern of results found by Shaffer and Shiffrin (1972) in which confidence was highest for scenes that had been looked at for longest. Therefore, it seems as though confidence for object identity memory is closely linked to the actual memory for object identity.

In addition, as boxed objects were looked at more often and for longer, it is likely that the memory representations for boxed objects contained more detail (having been embellished on each separate fixation) and as a result were more likely to be remembered than non-boxed objects. As a result, at test, the boxed objects tended to be selected earlier than non-boxed objects. There was also a numeric trend towards boxed objects being placed more accurately than non-boxed objects. However, importantly, there was no significant difference between boxed and non-boxed objects for object placement error.

The most likely explanation for this superior memory for the identities of boxed vs. non-boxed objects is based directly on the claims of the Visual Memory Model (Hollingworth & Henderson, 2002). According to this model, the way in which object location information is encoded relies upon the formation of a spatial map early in scene viewing. In this experiment, as in Experiment 1, it is likely that the spatial map is formed rapidly upon the scene being presented. Once the spatial map has formed, the 12 available candidate locations are now available in memory to be filled by specific object identities. As each object is directly fixated, information about its identity begins to be encoded and at some point during fixation of that object the identity is bound to one of the locations within the spatial map. Once this binding has occurred, the location does not improve with any extra encoding. This means that any extended fixation duration on an object is used to embellish and add detail to the memory representation for the identity of the object, not the location. If this is correct, it explains why longer fixation durations on boxed objects led to higher probability of correct object selection, but no more accurate memory for the object location. Thus, as in Experiment 1, a clear relationship between encoding and recall was observed for the identity of both boxed and non-boxed objects. A higher amount of processing seemed to lead to more accurate memory for those objects.

Importantly, there was no effect of Photograph on either the encoding and recall measures. This is important because it demonstrates that it was not the inherent qualities of the objects themselves that captured attention. Instead, the green boxes captured attention, regardless of which objects they surrounded. If the dominant influence had been the objects themselves, then one would have expected that some objects would have been regularly fixated regardless of whether they were boxed or non-boxed. However, as there was no main effect of Photograph, and no significant interactions, it demonstrates that what influenced whether an object was looked at was whether it was surrounded by a green box, not whether it was inherently visually salient.

To further examine the relationship between encoding and recall, ANOVAs were carried out which were very similar to the ANOVAs conducted in Experiment 1. The first analysis investigated whether memory for the identities of objects was better for fixated objects than non-fixated objects. As in Experiment 1, fixated objects were more likely to be selected than non-fixated objects. Importantly, this pattern was basically identical for both boxed and non-boxed objects. This is a critical finding as it suggests that both boxed and non-boxed objects were encoded in a similar way. The presence of green boxes did not seem to facilitate encoding for object identities. Unlike Experiment 1, there was no evidence of improvement in object identity from one to two or more fixations. However, the most likely explanation for this lack of a significant effect is that there was a lack of objects that were fixated twice or more, especially for non-boxed objects. Recall that all participants had only 5 s to view the photograph. This brief encoding period meant that participants rarely made two or more fixations on objects. This led to a lack of power in the analysis.

Similar ANOVAs for object location did not find that fixated objects were placed significantly more accurately than non-fixate objects, although again this is likely due to a lack of data (boxed objects were rarely ignored, and non-boxed objects were less likely to receive a lot of fixations). When the data were collapsed across both boxed and non-boxed objects, it was found that fixated objects were placed more accurately than non-boxed objects (and none of the data was missing). However, as in Experiment 1, there was no improvement in memory for object locations from one to two or more fixations.

In summary, the findings of Experiment 2 demonstrate that the presence of green boxes did not make objects any more memorable than non-boxed objects, but it did make them more visually salient (as demonstrated by Theeuwes, 1991; 1992; 1994; Northdurft, 1993; Turatto & Galfano, 2000; Mounts, 2000). This increase in saliency meant that boxed objects were more likely to be fixated and in turn were more likely to be correctly selected at test (see also Henderson & Hollingworth, 1999; Hollingworth, Shrock, & Henderson, 2001; Hollingworth, Williams, & Henderson, 2001; Hollingworth & Henderson, 2002). However, due to the differential encoding patterns of identities and locations, boxed objects were placed no more accurately than non-boxed objects.

Thus, Experiment 2 has provided two important findings. First it has replicated the results patterns from Experiment 1, both in terms of the basic characteristics of eye movements observed for participants who saw a photograph without any green boxes, and also for the differential relationship between encoding and recall found for identity and location memory. As discussed in Experiment 1, this dissociation is interesting because it suggests that identity and location information is encoded differently for objects and scenes. It also suggests that the Visual Memory Model may not be entirely accurate in its present state. Later experiments in this thesis further explore this potential dissociation in more detail.

The second key finding of Experiment 2 is that the relationship between encoding and recall for objects in scenes can be fundamentally influenced by using a lower-level attentional capture technique to lead participants around the objects in a scene, resulting in participants fixating some objects at the expense of others. In this way, Experiment 2 has demonstrated that by using a lower-level attentional capture technique, participants' attention can be driven around the photograph and as a direct result, this leads to quantifiable differences in their memory performance.

4 The influence of fixation order on memory for object identity and object location

4.1 Introduction

In Experiments 1 and 2, an issue that arose, but was not analysed formally, was whether the order in which participants chose to fixate the objects had any effect on the accuracy with which the objects were (a) selected and (b) placed. This is an interesting issue to investigate as previous research has provided evidence that, under certain conditions, the order in which stimuli are presented directly affects the accuracy with which those stimuli are later remembered. For example, in a classic study, Postman and Phillips (1968) found that words presented at the beginning and ends of lists were remembered better than words presented in the middle of lists; a primacy and recency effect respectively. If this pattern of results is considered in the context of the findings of Experiments 1 and 2 (that object identity memory improves as the number of fixations increases but that object location memory does not improve with subsequent fixations), it may be that this basic relationship had been modulated by the order in which objects were fixated by the participants. In other words, the number of fixations made on an object may not be the only factor that influences how well an object is selected and placed. Instead, the order in which objects are looked at may also have an effect on accuracy of memory.

Importantly, the free-viewing encoding phase in Experiments 1 and 2 meant that this issue could not be formally explored. During the encoding phase, objects were often re-fixated, especially for participants with the 10 s viewing duration. Any instances of objects being re-fixated would have caused difficulties if objects were to be indexed into a serial order that determined which object had been fixated first, which object had been fixated second etc. If objects were to be ordered using the first visit to an object as an index, any potential primacy effect would possibly be due to the subsequent re-fixation of that object, rather than that object's position in the fixation order (if indeed that object *was* re-fixated). For example, consider the ball. For one participant, the ball might be the third object fixated (after the clock and the laptop) meaning it would be assigned as object 3. However, if the ball was later re-fixated on two more separate occasions (including it being the last object to be fixated), one could either categorise the ball as the 3rd object viewed, or the last object viewed. However, although both of these categorisations would be technically correct, any conclusions drawn about primacy and recency memory effects on the basis of the position of the ball in the fixation order would be flawed.

At test, if the ball was selected and placed accurately, one might draw the conclusion that this accuracy of memory was due to the ball being fixated early in the trial (at position 3), and would be evidence for a primacy effect. However, the true explanation for the accurate memory might instead be due to the fact that the ball was fixated several times. Alternatively (or possibly additionally) it may have been because the ball was also fixated *last* in the sequence, providing evidence instead for a recency effect (an advantage in memory for stimuli presented late in a sequence).

The same problem would also occur if the fixation order of objects was approached using the most recent fixation on an object (i.e. which object was fixated last? Which object was fixated next-to-last?); any potential recency advantage would be compromised if objects had also been fixated earlier in the encoding phase. Therefore, for the potential influence of fixation order to be investigated thoroughly, a modified version of the initial paradigm was needed.

In this modified version, the fixation order of objects was influenced, and to some degree, controlled, by the experimenter. A bright green box was superimposed around one object at a time in the photograph to lead participants around the objects in a pre-defined quasi-random order. This modification allowed exploration of whether the order in which objects were fixated influenced the accuracy with which their identities and locations were recalled.

4.1.1 Serial order and memory for words in lists

As presented in the introductory paragraph, Postman and Phillips (1965) found that the accuracy with which a previously viewed word was recognised was influenced by the position of the word in a list. In their study, participants were presented with word lists containing 10, 20 or 30 words. Each word was presented one at a time for a 1 s duration and was only presented once.

Participants were required to look at the words and try to remember them in preparation for a free recall memory test in which participants had to list as many words as they could remember. Each participant completed 18 memory tests on 18 different word lists: six tests occurred immediately after the last word on that list was presented, six tests happened after a 15 s delay and six after a 30s delay. During the 15 s and 30 s delay periods, participants were required to count backwards to inhibit any sub-vocal rehearsal of the words. This was theorised to engage working memory and prevent any stimuli that were stored only in short-term memory being consolidated into long-term memory across the test delay.

The most relevant finding was that for immediate tests, words presented at the end of the lists were remembered most often; a recency effect. For example, the last word presented was recalled 76% of the time, and the second-to-last word was recalled 69% of the time. In addition, words presented at the beginning of the lists were remembered quite often; a primacy effect. For example, the first word presented was recalled approximately 74% of the time, and the second word was recalled approximately 56% of the time. However, words presented in the middle of the list were remembered less often. The fifth word presented was remembered only 40% of the time and the sixth word only 47% of the time. However, when the memory test occurred after a 15 s or 30 s delay, no advantage was observed for words at the end of the list, although the primacy effect remained. This same results pattern was also found by Glanzer and Cunitz (1966) and Murdock (1962).

The dominant explanation for this pattern of results was that words presented early were rehearsed sub-vocally (in short-term memory). This repetition of words led to them being encoded deeply and consolidated into long-term memory. Presumably, the more times that a word was repeated, the more likely it was to be consolidated into long-term memory. In turn, these words were highly likely to be retrieved at test. Conversely, words presented late in the list did not receive this rehearsal and therefore were not consolidated into long-term memory. However, due to the immediate test, these words remained in short-term memory and were therefore also easily accessible. Evidence to support this perspective came from the occasions when the memory test was delayed. In those instances, words presented at the end of the sequences (stored only in short-term memory) decayed and the

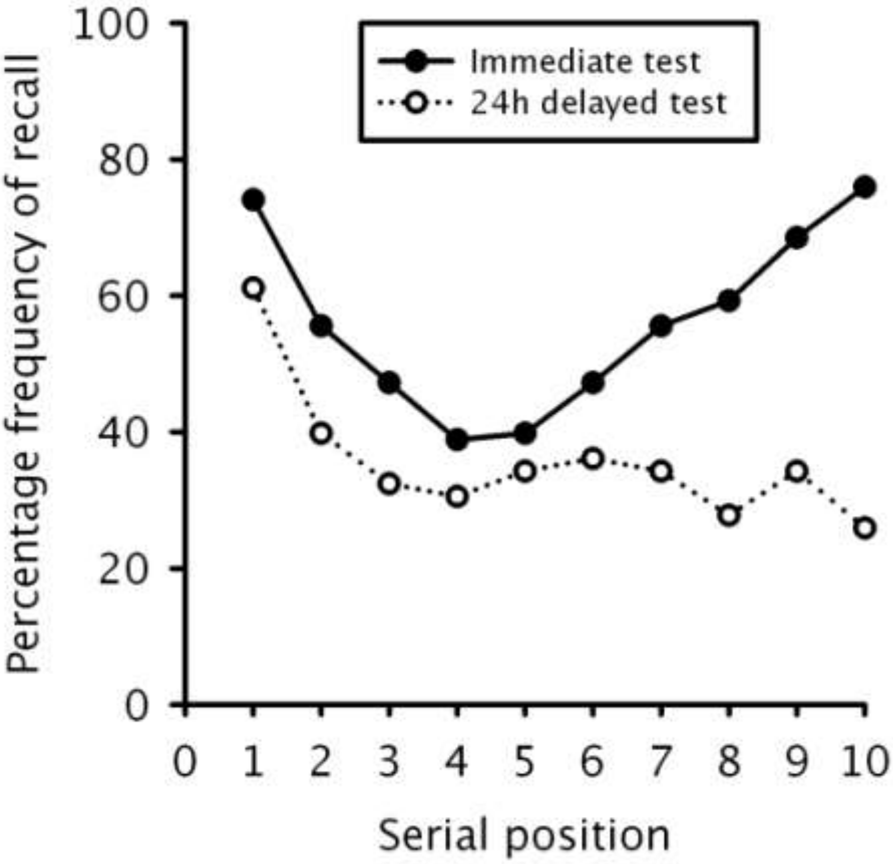

Figure 7. The serial position curves found by Postman and Phillips (1965)

recency effect was diminished. Last, for both immediate and delayed tests, words presented in the middle of the sequence did not receive enough (or any) rehearsal to encode them into long-term memory, nor did they stay in short-term memory, instead being displaced by subsequent words. This led to words presented in the middle of the sequence being poorly remembered no matter when the test occurred. Figure 7 shows the typical serial position curves found by Postman and Phillips.

4.1.2 Serial order and memory for the identity of objects in scenes

Other studies have investigated the influence of serial order on memory, but for objects in scenes, rather than words in lists. As presented in Chapter 1, Hollingworth (2004) used a follow-the-dot paradigm to investigate memory for the identities of objects in visual scenes. Participants viewed 48 scenes containing a variable number of objects. On each trial, the scene was previewed for 1000 ms, followed by the start of the moving dot sequence. A bright green dot moved from one object to the next in a pre-defined order. The dot appeared and remained on each object for 300 ms and then disappeared for 800 ms to encourage participants to closely inspect the object. Then the dot appeared on a different object for 300 ms, and then disappeared for 800 ms. This cycle of 300 ms dots and 800 ms object viewing was repeated until the end of the trial and the dot had appeared on each of the objects in the scene in turn. Once the cycle had finished, the target object was masked, then the first of two test objects was presented; the original version of the object, or a modified version. The modified version was either rotated by 90° or was a token substitution (a different object from the same category).

On each trial, one of the objects was pre-defined as the target object, and the position of this target object in the dot order was manipulated. In Experiment 1, the target object was either the penultimate object (the 1-back condition, meaning that 1 object was fixated between the target and the end of the trial), the 5th from last (4-back) or the 11th from last (10-back).

The most relevant finding from the results was that recognition accuracy was highest for the 1-back trials (on average 86.5% correct). Performance was no different between the 4-back (78.9%) and 10-back (81%) conditions. This provides evidence of a recency effect for objects in visual scenes, with highly accurate memory shown for the visual form of objects fixated towards the end of the fixation sequence. It also showed that objects fixated at the beginning of the sequence were remembered as accurately as objects fixated 4th from the end of the sequence. Hollingworth described this as a “robust prerecency effect” (Hollingworth, 2004, pp. 525). However, unlike in Postman and Phillips (1965) there was no clear primacy effect for the first few objects that were fixated. Instead, the first 8 objects were all remembered equally often (between 76% and 82% of the time).

In further experiments within the same article, Hollingworth investigated memory for objects in many different positions within the fixation order. The overall results pattern remained the same; memory for the two most recently fixated objects was highly accurate (over 85%) and memory for the other objects was approximately 80%. In Experiment 6, all 10 fixation positions were tested. Participants did not know which object would be tested, or which fixation position would be tested. Therefore it is likely that participants attended to all objects. The findings of Experiment 6 corroborated the earlier studies by showing a strong advantage to objects presented at the end of the sequence, but stable and accurate memory for objects presented early and during the middle of the sequence. Again, in contrast to the word list findings, there was no clear primacy effect for the first objects fixated.

The most likely explanation for the lack of a clear primacy effect is that visual information seems to be more difficult to rehearse than words, a finding demonstrated by Shaffer and Shiffrin (1972). Potentially, words are easier to rehearse because there is just a single stimulus to repeat and remember – the word itself. In contrast, objects within scenes are comprised of a multitude of characteristics, or properties, that uniquely define what the object is. Thus, the range of information associated with an object makes it more complex to encode and rehearse. In the study by Shaffer and Shiffrin, participants were presented with 120 visual stimuli one at a time. Each stimulus was presented on a slide for 200 ms, 500 ms, 1 s, 2 s or 4 s. Participants were explicitly instructed to "concentrate on each and every slide as it appears. When a slide leaves the screen, think about it and try to remember it during the period before the next slide appears" (Shaffer & Shiffrin, 1972, pp. 293). The purpose of these instructions was to encourage participants to visually rehearse each object while waiting for the next object to be presented. The duration of the inter-trial interval was also varied; participants had to wait 1 s, 2 s or 4 s in between each stimulus presentation. Participants were informed that they would have to provide confidence ratings for a series of visual stimuli at a later test.

At test, participants saw another set of pictures and had to give a confidence rating (1 – 6) for each picture that reflected how certain they were that they had seen each slide before. At test, half of the test slides were "old" (i.e. had been seen before by participants) and half were "new" (i.e. had never been seen by participants). The results showed that participants gave higher confidence ratings for the old stimuli (4.95) than the new stimuli (1.73). Unsurprisingly, as presentation time of the stimuli increased, so did confidence. However, critically, there was no difference in confidence between the three inter-trial intervals. Confidence was comparable between 1 s (4.92), 2 s (4.99) and 4 s (4.95). The probability that a stimulus was correctly recognised was calculated using the confidence ratings and there was no difference in recognition accuracy between the inter-trial intervals (0.79, 0.80 and 0.79 respectively). These results suggest that participants did not seem to rehearse visual information. If participants did rehearse visual information, one might expect that confidence would be higher when the inter-trial interval was longer, as this would have allowed more rehearsal to have taken place, but clearly this was not the case. However, instead it could indicate that participants only need 1 s to rehearse an object before it is encoded into long-term memory and any additional rehearsal time is not required.

In terms of the methodology used by Hollingworth (2004) to assess the influence of fixation order on memory for object identity, a critical issue is that eye movements were not recorded. Participants were told to look at the object that was highlighted by the green dot and to keep looking at that object until the dot appeared on a new object. Hollingworth stated that, "participants did not have any difficulty complying with this instruction" (Hollingworth, 2004, pp. 523). However, the degree to which participants actually followed the dot was not reported. It may have been the case that participants followed the task instructions perfectly. Alternatively, participants may have moved their eyes to the highlighted object as the dot appeared (as they were instructed to do), but moved their eyes on to another object at some point during the 800ms period that the dot was not present. If that was the case, then the data presented by Hollingworth may have suffered from the same problem discussed above with Experiment 1 in this thesis; namely that any conclusions drawn on the basis of the presentation order may in fact be due to the periods of re-fixation made on the objects during the trial. In the present experiment, eye movements were precisely recorded and reported to allow comprehensive analysis of the extent to which participants fixated each object in turn and to evaluate how accurately participants followed the task instructions.

Additionally, Hollingworth (2004) only tested memory for the identities of objects, probing memory for orientation and object type in Experiment 1, and just object type in Experiments 2 – 6. In the present study, memory for the precise locations of objects was tested to allow for a more thorough investigation into the influence of fixation order on memory and to see whether this influenced object identity memory and object location memory in the same manner.

Other studies have also investigated the influence of fixation order on memory for objects in scenes. Irwin and Zelinsky (2002) presented participants with 7 objects arranged in a semi-circle and instructed them to look at the objects in preparation for a subsequent recognition test. On each trial the recognition test occurred after a pre-defined number of fixations had been made (1, 3 or 5 fixations, or 3, 9 or 15 fixations). As the first saccade after the critical fixation was initiated, the screen went black for an average of 153 ms. Then one of the seven positions in the semi-circular display was cued, and participants were presented with the actual seven stimulus objects and had to state which of them had been present at the cued location. The most relevant finding was that recognition accuracy was highest for the three most recently fixated objects. If the last fixation in the trial happened to be made on the test object, recognition accuracy was approximately 90%. When the test object was the second-to-last or third-to-last object to be fixated, it was correctly identified on approximately 80% of the trials. This is clear evidence of a recency effect. However, there was no evidence for a primacy effect; objects fixated early in the sequence were typically recognised at levels of around 66% accuracy.

It is important to note that participants completed 147 trials and saw the same 7 objects on every trial. This is likely to have led to high levels of familiarity with the objects. Potentially, this familiarity meant participants did not need to directly fixate the objects in order to identify them as they were so familiar with their identities. For this reason, in the present study, participants completed only one trial to eliminate the possibility that the objects did not need to be directly fixated in order to accurately identify them.

In another study, Zelinsky and Loschsky (2005) showed participants an array containing 9 objects laid out in a random pattern. Participants were instructed to remember the identity and location of the objects, and their eye movements were recorded. On each trial, unbeknown to the participants, one object was randomly chosen as the target. Once the target object was fixated, the trial continued until the participant fixated a pre-defined number of other objects (rather than after a pre-defined number of fixations as in Irwin & Zelinsky, 2002). The number of objects fixated between the last fixation on the target and the end of trial was manipulated, and intervals ranging from 1 – 7 objects were used. At the end of each trial, the objects disappeared and one of the locations was highlighted for 1 s to indicate the location of the target object. Four objects were then presented on the screen, one of which was the target. Participants had to indicate which object was the target object.

Results showed that identification accuracy when only one object was fixated after the target was 87%, dropping to 65% when three objects intervened. There was no further decline in identification accuracy across the intervals from 4 – 7 objects. Notably, identification accuracy was still well above chance (25%) even when 7 objects were fixated between the target and test. Taken together, these results suggest a strong recency effect for intervals of 1 – 3 objects. However, unlike Postman and Phillips (1965) there was no clear primacy effect for the objects fixated first (intervals 6 and 7). Instead, there was a stable primacy effect which spanned the intervals 4 – 7. This results pattern, coupled with the findings of Hollingworth (2004) and Irwin and Zelinsky (2002), is interesting as it provides a different pattern of results to that found for word lists. The object memory studies consistently find a recency effect for the 3 most recently fixated objects. This is analogous to the recency effect found for word lists. However, object identity research has not found a clear primacy effect for the earliest fixated objects.

4.1.3 Serial order and memory for the locations of objects in scenes

In comparison to the literature presented above, there has been less work that has investigated how serial order influences memory for locations. However, there is some research which has investigated the effect of serial order on memory for the locations of dots and simple shapes. Jones, Farrand, Stuart and Morris (1995) presented participants with sequences of dots on a computer screen. Each sequence contained 4, 7, or 10 dots and within each sequence each dot appeared in a different position on the screen. Each dot appeared for 1 s. At the end of each trial, participants were presented with all of the dots on the screen at the same time and had to indicate the order in which the dots had been presented.

The number of errors made was calculated, and the results showed that for sequences of 4 dots memory was very accurate across all 4 serial positions. However, for sequences of 7 and 10 dots, an inverted U-shape distribution was produced. This pattern of results showed that the smallest amount of errors occurred for dots presented at the beginning and the end of the sequence, with errors higher for dots presented in the middle of the sequence. However, the primacy effect was more pronounced than the recency effect. This pattern of results is similar to that found by Glanzer and Cunitz (1966), Murdock (1962) and Postman and Phillips (1965). As a result, Jones et al. (1995)

suggested that spatial information was encoded and stored in a very similar way to the word list stimuli presented earlier.

Smyth and Scholey (1996) also investigated the effect of serial order on memory for locations. Participants saw 9 squares presented on a computer screen simultaneously. On each trial, one of the squares was highlighted one at a time for 1.5 s. The length of each trial was also varied: trials lasted for 4, 5, 6 or 7 squares and during each trial no square was highlighted twice. At the end of each trial, participants were tested on their memory for the order in which the squares had been presented.

The results showed a similar pattern to Jones et al. The smallest number of errors was observed for squares presented at the beginning of the sequence, and slightly more errors were made for squares presented at the end of the sequence. Errors were made more frequently for squares presented in the middle of the sequence, demonstrating a clear primacy and recency effect for memory for the locations of squares. Therefore unlike memory for identities of objects which found only a clear recency effect, these studies investigating serial order and memory for locations found both a clear primacy and a recency effect.

4.1.4 Comparison of memory for word lists and memory for objects in scenes: Manner of encoding

At this point it may be useful to try to draw a distinction between the words used as stimuli by Postman and Phillips, and the visual images used as stimuli in the present experiment and in other relevant research discussed above (e.g., Irwin & Zelinsky, 2002; Hollingworth, 2004; Zelinsky & Loschsky, 2005). There are at least two clear distinctions in relation to processing of scene stimuli and words to be made. First, it is possible that the way in which words and images are encoded and consolidated into memory is different. Postman and Phillips suggested that words are continually rehearsed throughout the trial, meaning that during the presentation of the second word (e.g., *cat*), participants are likely to be repeating the name of the first word (flower) as well (e.g., *flower, cat*). If true, it follows that during presentation of the third word (e.g., *table*) participants are sub-vocally rehearsing the previous two words as well (e.g., *flower, cat, table*). Importantly, this suggests that while attention is focused on a particular stimulus, some attention is also allocated to the rehearsal of some (or all) of the previously attended words. If this is true, then it seems that, to some degree at least, to-be-remembered words that are presented in a list are maintained, to some degree at least, in parallel.

In contrast, it is possible that objects in a scene are encoded in a more serial fashion. Evidence for this was provided by Shaffer and Shiffrin (1972). Their results suggested that instead of maintaining a mental image of the previously fixated object, maintenance of a visual image seems to cease once the image is no longer available to visually inspect. This may explain why primacy effects tend not to be found for visual stimuli, but are found for verbal stimuli. Accordingly, the respective results patterns of Postman and Phillips, and Hollingworth share a common finding (a recency effect at immediate test) but also differ (Postman and Phillips found a primacy effect, Hollingworth did not). A possible explanation for this may be due to the difference in the way objects and words are identified by the visual system.

4.1.5 Comparison of memory for word lists and memory for objects in scenes: Identification period

A second important distinction between words in lists and objects in scenes relates to the respective amounts of time that words and images need to be accurately identified by the visual system. Prior research suggests that words are identified substantially faster than objects. In the reading literature, there is evidence that participants can read sentences without disruption to their comprehension or speed of reading, even when words are masked (Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981) or disappear after only 50-60 ms (Rayner, Liversedge, White, & Vergilino-Perez, 2003; Liversedge, Rayner, White, Vergilino-Perez, Findlay, & Kentridge, 2004). This suggests that the word identification system can very quickly compute the identity of a word. It is important to note, however, that this does not mean that a word is completely processed in the 50-60 ms period. Instead, the visual information necessary for subsequent identification can be obtained during the first 50-60 ms of a fixation.

In contrast, scene perception research has demonstrated that objects require longer inspection periods than words before they can be identified by the visual system. Rayner, Smith, Malcolm and Henderson (2009) presented participants with scenes containing a number of objects. On each trial, participants were instructed to search the scene for a pre-specified target object and to report the location of the target object within the scene. However, the entire scene was masked after each fixation. The mask appeared after one of four possible fixation periods (25, 50, 75 or 150 ms). The results showed that total scene search time and fixation duration on objects decreased as the mask onset increased, and consequently search accuracy increased. However, performance after 150 ms was not quite as efficient or accurate as the free-viewing condition where no mask was displayed. In a second experiment using different mask onsets (75, 150, 200 or 250 ms) it was found that performance was comparable to the free-viewing condition in the condition with a 150 ms exposure prior to the mask. Search accuracy did not improve with longer mask onsets indicating that typically 150 ms is sufficient time for the visual information necessary for object identification to be encoded. It is important to note that, as for words, this brief period does not mean that an object is fully identified after 150 ms. Rather, the information needed for that object to be processed is acquired during that brief period.

Taken together, this research suggests that the visual system identifies words quicker than objects. Of course, the time taken to identify an object or a word does not signify the end of encoding for that stimulus. Further processing is needed for the word or object to be adequately encoded and consolidated into memory.

4.1.6 The present experiment

In Experiment 3 the main aim was to investigate whether a recency and/or a primacy effect were present in memory for objects in scenes. In particular, the main question of interest was: does the order in which objects are fixated influence the accuracy with which their identities and locations are recalled? To do this, a modified version of the paradigm used in Experiments 1 was employed. Participants were told that they would have to fixate each of the 12 objects in the scene in a pre-defined quasi-random order. To do this, a bright green rectangle was superimposed around the 12 objects one at a time and participants were told to follow the green box as it moved from object to object. The box was superimposed around each of the objects once only and spent 810 ms on each object. No objects were "revisited" by the green box. In this way, the order in which participants looked at objects could be controlled to a significant degree (although not completely). The test phase was identical to the test phase in Experiment 2. Participants had to select 12 objects from the pool of 24, then place those 12 objects in the location they remembered them being located in the photograph. As in Experiment 2, participants gave confidence ratings from 0-10 that represented their confidence that each object had been present in the photograph. In addition, they gave confidence ratings from 0-10 that represented their confidence that each location had been filled by an object in the photograph.

Several predictions were made. First, it was predicted that there would be substantial differences in eye movement behaviour between the box-following and free-viewing conditions. This prediction was based on two factors. First, it was predicted on the basis that participants can, and do, change the way in which they move their eyes around a scene contingent on the task instructions (Buswell, 1935; Yarbus, 1967; Castelhano, Mack, & Henderson, 2009 – see Chapter 3 for a more detailed discussion of this literature). Specifically, it was predicted that, due to the task instructions, box-following participants would spend more of the viewing period looking at the objects, fixate more of the objects, and fixate the objects for longer than the free-viewing participants. As well as the task instructions influencing participants to change the way they looked at the photograph, it was also expected that the moving green box would have a substantial effect on the eye movement behaviour. Recall that in Experiment 2 in this thesis participants looked at objects that were surrounded by a green box more often and for longer than objects that were not surrounded by a green box. These findings seem to suggest that the green boxes used in Experiment 2 were somewhat effective at capturing visual attention. As a result, a similar box was used in Experiment 3. It was anticipated that participants' attention would be allocated to whichever object the green box was highlighting due both to the saliency of the green box, and the task instructions to follow the green box as it moved from object to object.

As Experiment 1 demonstrated that memory for the identities of objects improved as the number of fixations made on objects increased, it was predicted that box-following participants would correctly select more objects than free-viewing participants because the box-following participants were likely to have looked at more of the objects and made longer fixations on the objects than the free-viewing participants. However, no difference was expected for memory for object locations. This is based on the findings of Experiment 1 that increased fixations during encoding did not lead to more accurate memory for the locations of objects.

Regarding the influence of presentation order on memory for identity and location, there were two possible predictions. If the order in which objects are fixated *does* influence the accuracy with which they are remembered (as found by Postman & Phillips, 1965), then a primacy and a recency effect would be expected in the data probing memory for object identity and/or object location. Specifically, it would be expected that memory for the identity and location of objects would be more accurate for the first and last objects fixated than the objects fixated in the middle of the sequence.

However, if the fixation order does *not* influence the accuracy with which the identities and locations of objects are remembered then a primacy and/or recency effect would not be expected. In fact, rather than predicting primacy and recency effects, the Visual Memory Model (Henderson & Hollingworth, 2002) would predict that objects that received the most visual processing would be remembered most often and objects that received the least visual processing would be remembered least often. The Visual Memory Model states that information about objects accumulates in memory as visual processing time increases, meaning that the memory representation associated with each object becomes further embellished as the number of fixations increases. If participants closely follow the task instructions to follow the green box around the scene, then each object should receive approximately the same amount of visual processing. In that case, the Visual Memory Model would predict that all objects would be remembered approximately as accurately because the memory representation for each object would be equally detailed. Therefore, each object would be highly likely to be remembered, leading to a pattern of data without primacy or recency effects. Instead, memory performance would be comparable across all presentation positions.

Finally, regarding identity confidence and location confidence, it was expected that this would closely match the actual memory performance exhibited by participants. This is based on the findings of Shaffer and Shiffrin (1972), and the findings of Experiment 2 in this thesis. That is, if a primacy and recency effect were observed for the identity and location memory, then confidence ratings would be consistent with these primacy and recency effects. However, if memory for identities and locations was comparable across all presentation positions, then confidence for identities and locations would match this pattern. In short, as found by Shaffer and Shiffrin and in Experiment 2, participants' confidence for their memory should closely mirror their memory performance.

4.2 Method

Participants and design

Forty participants were recruited from the University of Southampton community and either participated voluntarily or received course credit. All participants had normal or corrected-to-normal vision. The average age was 20.31 years (*SD* = 2.00 years) and age ranged from 18 to 27 years. There were 13 males and 27 females. Participants were experimentally naive as to the aims of the experiment. None of the participants had taken part in any of the earlier experiments in this thesis. Data from seven participants was removed leaving data from 33 participants. The criterion for exclusion was any tracker loss during the encoding phase of the experiment. The design was between-participants with 10 of the participants in the free-viewing condition and 23 participants in the box-following condition. Of those 23 participants, for 12 of them the green box moved around the objects in Order A (ball, stapler, clock, plate, lava lamp, mouse, phone, laptop, glass, book, keyboard, scissors) and for the other 11 participants the box moved around the objects in the reverse order, Order B. The purpose of this design was to ensure that any potential advantages in memory for certain objects were counter-balanced across the two presentation orders.

Apparatus

Monocular eye movement recordings from the right eye were taken using an EyeLink 1000 eye tracker. The position of the participant's right eye was recorded every millisecond. The colour photograph of the cubicle was presented on a 21 inch colour monitor (see Chapter 2 for the photograph). The viewing distance was 57 cm and participants leaned on chin and forehead rests during the experiment to minimise head movements. The test phase took place in a research cubicle which contained the 12 objects seen in the photograph, and 12 plausible distractor objects.

Stimuli

The stimulus used was identical to that used in Experiment 1, except for the addition of the green boxes for participants in the box-following condition (see Procedure for more details).

Procedure

The procedure for the participants in the free-viewing condition was identical to Experiment 2 except that in Experiment 3 all participants viewed the stimulus photograph for 10 s and the memory test always took place immediately. However, box-following participants experienced a modified encoding phase (but an identical test phase). In the encoding phase, box-following participants first completed a practice trial in which 12 letter Xs were displayed on the screen and a bright green box moved from one X to the next. The box highlighted each X for 810 ms. This practice trial was carried out to allow participants to get used to following the green box from point to point and to familiarise themselves with the speed that the box moved. Examination of these data showed that all participants accurately followed the green box and had no problems doing so.

In the experimental encoding trial, participants either saw the photograph from Experiment 1 for 10 s (free-viewing condition) or experienced the moving green box (box-following condition). In the box-following condition, participants were forewarned that a green box would appear to "jump" from object to object during the trial in the same manner it had moved from X to X in the practice trial. They were instructed to look at the object that was highlighted by the green box and to keep looking at that object until the box moved to a new object. They were then to move their eyes to the object that the box was now highlighting. Participants were also told not to look at objects unless the box was highlighting that object.

Once the trial began, a fixation cross was presented for 1000 ms at the centre of the screen. The photograph from Experiment 1 was presented but with the addition of a bright green box around one of the objects (in presentation order A it was around the ball and in presentation order B it surrounded the scissors). After the green box was present for 810 ms it disappeared from that first object and immediately re-appeared around the next object in the pre-defined sequence. The pattern continued for all 12 objects in the scene and the trial lasted for 9720 ms. This time period allowed comparison with the free-viewing participants who saw the original photograph from Experiment 1 for 10 s. This 10 s encoding duration was used to ensure that each of the 12 objects could be fixated. A 5 s encoding duration would not have allowed enough time for each of the 12 objects to be fixated for a substantial duration. The pre-defined order was selected randomly and then reversed in the counterbalanced condition. Participants saw either Presentation order A or the reversed counter-balanced Presentation order B. The test phase for all participants was identical to the test phase in Experiment 2 except for the addition of having to give confidence ratings that represented participants' confidence that each of the locations had been filled in the stimulus photograph.

4.3 Results

As in Experiments 1 and 2, any fixations shorter than 80 ms or longer than 1200 ms were excluded from all analyses.

Between-Groups Analysis – Encoding Phase

The first analysis compared the eye movement behaviour of participants in the box-following condition with participants in the free-viewing condition. Between-groups *t*-tests showed a large number of differences between the two groups in terms of oculomotor behaviour (see Table 12). For some of the *t*-tests reported in this chapter, the Levene's test for equality of variances was significant. This means that the variances were not equal and as a result, the degrees of freedom associated with the test were not N – 2 but instead were decimal numbers. This difference in variance is probably because there were more than twice as many participants in the box-following condition (n = 23) than in the free-viewing condition (n = 10[2]).

[2] Unlike Experiment 2, where control participants were combined with the corresponding condition from Experiment 1, the control participants in the present experiment were not combined with the corresponding condition from Experiment 1 (10s/0 delay) because a different eye tracker was used in Experiment 3.

First, participants in the box-following condition spent longer in total fixating objects than did participants in the free-viewing condition, $t(9.72) = -3.91$, $p < .001$. This is unsurprising as box-following participants were instructed to look only at the objects, whereas free-viewing participants were given no instructions regarding their oculomotor behaviour. This finding suggests that box-following participants closely followed the task instructions to follow the green box around the photograph.

Further evidence that box-following participants accurately followed the instructions was shown by a significant difference in the amount of time participants spent looking at blank regions (i.e. regions not containing any of the 12 objects). Box-following participants spent significantly less time looking at blank regions than free-viewing participants, $t(9.18) = 3.79$, $p < .001$. It is important to note here that of the box-following participants, 14 of the 23 did not spend any time looking at blank regions (and so did 1 participant in the free-viewing condition), hence the comparatively low average total time. Of the participants who did fixate blank regions, the average total time was 163ms which was significantly lower than the free-viewing participants, $t(8.32) = 3.75$, $p < .01$.

Also, box-following participants made significantly fewer fixations on blank regions than free-viewing participants, $t(9.49) = 3.77$, $p < .001$. Of the box-following participants who did fixate a blank region, on average only 1.22 fixations were made, which was significantly fewer than the free-viewing condition who made 4.11 fixations on blank regions, $t(8.51) = 3.45$, $p < .01$.

The average fixation duration on blank regions was significantly shorter for box-following participants than free-viewing participants, $t(31) = 5.17$, $p < .001$ (though note again that less than half of the box-following participants fixated blank portions). The average blank region fixation duration when participants did fixate blank regions was 131.33 ms (which was still significantly lower than the free-viewing condition, $t(16) = 3.77$, $p < .01$). Box-following participants spent a higher proportion of their viewing duration fixating the objects than the free-viewing participants, $t(9.17) = -3.83$, $p < .001$.

Taken together, these results suggest that box-following participants exhibited substantially different eye movements compared to the free-viewing participants. Box-following participants spent most of the available viewing duration fixating the objects (on average 99% of the available time), and only fixated blank regions very occasionally. When box-following participants did fixate blank regions, they did so only briefly. The box-following participants also made fewer fixations in total, and these fixations were longer than the fixations made by free-viewing participants. In comparison, the free-viewing participants made more fixations, made these fixations over a wider range of areas in the photograph (looking both at objects and blank regions), and generally made shorter fixations. Consistent with the task instructions, all participants except one in the box-following condition fixated all 12 objects (one participant fixated 11 objects). Consequently, box-following participants fixated significantly more objects than the free-viewing participants did, $t(9.12) = -2.50$, $p < .05$. Despite this, there was no difference in the number of fixations that were made on objects, $t(31) = 1.54$, $p > .05$. Therefore, despite looking at more of the objects and spending more time looking at objects, box-following participants did not make more fixations on the objects. The most plausible explanation is that average fixation durations were longer for box-following participants than for free-viewing participants. However, compared to free-viewing participants, box-following participants made fewer fixations overall, $t(31) = 4.36$, $p < .001$. Clearly, the reason for this difference was that free-viewing participants also made a number of fixations on blank regions, which box-following participants did not do as often.

As suggested above, box-following participants made longer average fixation durations than free-viewing participants, $t(31) = -4.40$, $p < .001$, longer gaze durations, $t(11.15) = -3.32$, $p < .001$ and made more fixations during total time, $t(31) = 4.33$, $p < .001$ than free-viewing participants. However, there was no difference between the box-following and free-viewing participants in the number of fixations made during gaze duration, $t(31) = -1.41$, $p > .05$ or total time, $t(9.47) = -0.59$, $p > .05$.

These results show that box-following and free-viewing participants made different eye movement patterns during the encoding phase. Compared

Table 12

Means and Standard Errors (in Parentheses) of Eye Movement Measures for the Free-viewing Condition and the Box-following Condition in the Encoding Phase

Condition:	Free-viewing condition (no green box)		Box-following condition (moving green box)	
Measure	M	SE	M	SE
Total time fixating objects (ms)	7332	(334)	8662	(66)
Total time fixating blank (ms)	818	(198)	64	(20)
Fixations made on blank regions	3.70	(0.84)	0.48	(0.14)
Average blank region fixation duration	203	(29)	51	(15)
Proportion of viewing duration spent fixating objects	0.90	(0.02)	0.99	(0.00)
Fixations made on objects	25.30	(1.35)	23.30	(0.63)
Fixations made in total	29.00	(0.92)	23.78	(0.68)
Number of objects fixated	10.60	(0.54)	11.96	(0.04)
Average fixation duration (ms)	292	(8)	374	(12)
Average gaze duration (ms)	469	(47)	635	(16)
Average number of fixations made during gaze duration	1.54	(0.11)	1.68	(0.05)
Average total time (ms)	701	(34)	722	(6)
Number of fixations made during total time	2.41	(0.12)	1.93	(0.05)

to free-viewing participants, box-following participants made fewer fixations in total, spent a higher proportion of the time looking at objects than blank regions, re-fixated fewer objects and made longer fixations both during gaze duration and in general. An interesting question, therefore, is whether these differences led to variation in memory for the identity and locations of objects.

Between-Groups Analysis – Test Phase

Interestingly, despite the differences in encoding behaviour detailed above, memory performance was virtually identical in the box-following and free-viewing condition (see Table 13 below). The finding that there was no difference in memory performance between the two conditions was not what was predicted, and a possible explanation is presented in the Discussion. Between-groups t-tests showed that there was no difference between the box-following group and the free-viewing group in the number of objects correctly selected, $t(31) = -0.79$, $p > .05$, no difference in the average placement error for correctly selected objects, $t(31) = -0.56$, $p > .05$, and no difference in the best fit error, $t(31) = -0.03$, $p > .05$. Clearly the moving green box did not facilitate memory relative to the free-viewing condition.

Table 13

Means and Standard Errors (in Parentheses) of Recall Measures in the Test Phase for the Free-viewing Condition and the Box-following Condition in Experiment 3

Condition:	Free-viewing condition (no green box)		Box-following condition (moving green box)	
Measure	M	SE	M	SE
Number of objects correctly selected	8.90	(0.38)	9.26	(0.25)
Average placement error (cm)	38.28	(4.02)	41.79	(3.71)
Best fit error (cm)	24.06	(2.25)	24.13	(1.40)

An important point to note here is that the box-following participants looked at significantly more objects than the free-viewing participants did (11.96 vs 10.60) yet there was no difference in the number of objects correctly selected. Despite looking at 1.36 more objects on average, there was only an increase of 0.36 objects. Possible explanations are explored in the Discussion. However, recall that the main motivation for the experiment was to see whether the fixation order influenced memory for the identities and locations of objects.

Comparison of Free-Viewing Condition from Experiment 3 with the Corresponding Condition from Experiment 1 – Overall Eye Movement Measures and Memory Test Data

The encoding phase experienced by the free-viewing participants in the present experiment was identical to the encoding phase experienced by participants in Experiment 1 who had 10 s viewing duration and an immediate test. On the basis that these participants experienced identical experimental conditions, it was a useful opportunity to establish that the effects found in Experiment 1 replicated in Experiment 3. Obviously it was expected that the effects would be replicated and that the fundamental characteristics of saccades and fixations would be very similar for these two groups of participants. To investigate this, a series of unpaired *t*-tests were carried out on both the eye movement measures and the recall measures to see if performance was similar. If performance was similar, the free- viewing condition of the current experiment would be a robust demonstration of reliability and replication of the effects reported in Experiment 1.

A difference was only found for two of the 16 measures, one of which was marginal. Participants in Experiment 3 made significantly more fixations on the objects than participants in Experiment 1, $t(16) = -2.10$, $p = .05$. Also, average fixation duration was marginally longer for participants in Experiment 1 than Experiment 3, $t(16) = 2.00$, $p = .06$. All other *p*s > .05. It is possible that these two differences occurred because different eye tracking machines and eye tracking software were used in Experiment 1 and Experiment 3. The reason for using a different eye tracker in Experiment 3 was that the software used in Experiment 1 did not support the methodology of the moving green box used in Experiment 3. On the basis of these results it is clear that the overall encoding behaviour and memory test performance results pattern

Table 14

Means and Standard Errors (in Parentheses) for Eye Movement Measures in the Encoding Phase for Experiments 1 and 3

Condition:	Experiment 1 (10s/0 delay)		Experiment 3 (Box-following)	
Measure	M	SE	M	SE
Total time fixating objects (ms)	6947	(327)	7332	(334)
Total time fixating blank (ms)	1296	(283)	818	(198)
Fixations made on blank regions	5.50	(1.30)	3.70	(0.84)
Average blank region fixation duration	222	(45)	203	(29)
Proportion of viewing duration spent fixating objects	0.84	(0.03)	0.90	(0.02)
Fixations made on objects	21.50	(1.10)	25.30	(1.35)
Fixations made in total	27.00	(1.13)	29.00	(0.92)
Number of objects fixated	9.88	(0.61)	10.60	(0.54)
Average fixation duration (ms)	326	(16)	292	(8)
Average gaze duration (ms)	558	(71)	469	(47)
Average number of fixations made in gaze duration	1.71	(0.21)	1.54	(0.11)
Average total time (ms)	735	(82)	701	(34)
Number of fixations made in total time	2.26	(0.23)	2.41	(0.12)
Number of objects correctly selected	9.25	(0.53)	8.90	(0.38)
Placement error (cm) for correctly selected objects	42.94	(6.31)	38.28	(4.02)
Best fit error (cm) for all 12 objects	23.59	(2.60)	24.06	(2.25)

found in Experiment 1 has been replicated by Experiment 3. The means and standards errors are presented in Table 14.

Comparison of Free-Viewing Condition with Experiment 1 – Relationship between Encoding and Recall for Object Identity

To enable more thorough comparison with Experiment 1, the data from the free-viewing participants were analysed to examine how memory for locations and identities accumulated across fixations made in total (as in the Results section of Experiment 1). In the present experiment, only five of the 10 participants looked at all 12 objects in the photograph. This meant that five participants had missing data in the "zero fixations" bin. This represented 12.5% of the data. To enable analysis to take place, a linear regression was carried out on the probability of correct selection data using the number of fixations as the predictor variable. The regression equation was significant at the .001 level and yielded an equation of $y = 0.19X + 0.39$. This meant that 0.39 replaced the blank cells (a value similar to the previous mean of 0.42).

In a direct replication of Experiment 1, a paired samples *t*-test showed that fixated objects ($M = 0.78$, $SE = 0.03$) were more likely to be selected than non-fixated objects ($M = 0.40$, $SE = 0.08$), $t(9) = -4.50$, $p < .01$. As all participants had 10s to view the photograph, there was substantial data in the "two" and "three or more" bins. This extra data meant that further investigation into how object identity memory accumulated across fixations could be carried out.

A repeated-measures ANOVA using number of fixations (0, 1, 2, 3+) as the independent variable showed that there was a significant main effect of the number of fixations on probability of correct object selection, $F(3, 27) = 17.56$, $p < .001$. There was also a significant linear contrast, $F(1, 9) = 39.97$, $p < .001$. Planned post-hoc Bonferroni-corrected paired *t*-tests showed that objects fixated twice ($M = 0.85$, $SE = 0.06$) and three times or more ($M = 0.96$, $SE = 0.03$) were correctly selected more often than non-fixated objects (both *p*s < .01). In addition, objects fixated twice and three times or more were correctly selected more often than objects fixated only once ($M = 0.49$, $SE = 0.07$) - both *p*s < .05. This provides further evidence that information about object identity accumulates in memory across separate fixations. Importantly, these results further replicate the findings of Experiments 1 and 2 in this thesis, demonstrating that the results observed in Experiments 1 and 2 are valid and robust.

Comparison of Free-Viewing Condition with Experiment 1 – Relationship between Encoding and Memory for Object Location

Regarding placement error scores, only four participants placed objects that had not been fixated. In addition, nine of the ten participants placed objects that had been fixated once. This meant that there were six empty cells in the "zero fixations" bin and one empty cell in the "one fixation" bin. This meant that 17.5% of the data were missing, primarily in the zero fixation bin. To try to replace the missing data to enable analysis to take place, a linear regression was carried out on the placement error scores using number of fixations as the predictor variable. However, the regression was not significant ($p > .1$) and yielded a regression equation of $y = (-5.66\ X) + 50.5$. As a result, the value of 50.5 could potentially be used to replace all blank cells in the "zero fixations" bin. Note that this value generated from the non-significant regression is a value not dissimilar from the previous mean value of 48.84. Importantly however, despite the value generated by the regression being very similar to the existing mean value, the regression was not significant and therefore it was not appropriate to use this value in the analysis.

It was judged that this analysis was important to carry out, however, so the mean value for object location for non-fixated objects was taken from Experiment 1 as this was a stable and established mean value. The average placement error from Experiment 1 for non-fixated objects for participants with a 10 s viewing duration and an immediate test was 54.7cm. Note that this value of 54.7 is very similar to the previous average of 48.84 and the value generated from the regression of 50.5.

The value of 54.7 was inserted into the blank cells in the "zero fixations" bin. First, a paired *t*-test showed that fixated objects ($M = 36.80$, $SE = 4.32$) were placed significantly more accurately than non-fixated objects ($M = 52.35$, $SE = 4.59$), $t(9) = 2.46$, $p < .05$ which directly replicates the findings of Experiment 1. However, as in Experiment 1, memory for locations did not improve with repeated fixations. A repeated-measures ANOVA showed that there was no significant effect of fixations on memory for object locations, $F(3, 24) = 2.34$, $p > .05$ although there was a significant linear contrast, $F(1, 8) = 9.42$, $p < .05$ suggesting that placement error generally decreased as fixations increased. Thus, the data from the control participants in Experiment 3 replicates the findings of Experiment 1. Object identity information accumulates across separate fixations, but location memory does not. Instead, location memory is better for fixated than non-fixated objects, but does not seem to accumulate across fixations.

The Influence of Fixation Order – Eye Movement Analysis

As mentioned in the Method section, participants in the box-following condition were instructed to follow the green box around the photograph, and to keep looking at the highlighted object until the box moved to the next object. Generally, participants followed the instructions accurately. However, performance was not perfect. The eye movement data were analysed using the Eyelink software on a fixation-by-fixation basis to generate a report stating the duration and location of each fixation. This list was then checked manually to evaluate whether each fixation was made on the correct object at the correct time, or whether fixations were made on objects out of turn.

This sequence of fixations was then compared to the pre-defined fixation order. Fixations were subsequently categorised as either being (a) on an object and in order, (b) on an object but out of order, (c) or not on an object. Furthermore, the fixations in (b) were classified as being (i) pre-fixations (looking at an object before it had been highlighted) or (ii) re-fixations (looking at an object after it had been highlighted).

As presented in Table 14 (above), on average participants in the box-following condition made 23.78 fixations during the encoding phase (*SE* = 0.67 fixations). Of these fixations, on average 0.48 fixations were made on blank regions of the photograph (*SE* = 0.14) which was 2.01% of the fixations made. Average fixation durations on blank regions were 203 ms.

Only 1.26 fixations were pre-fixations (*SE* = 0.24) which was 5.30% of the fixations. On average, these pre-fixations lasted for 239 ms. A one-sample *t*-test with the average fixation duration (374 ms) showed that on average pre-fixations were shorter than average fixation durations, $t(15) = -5.28$, $p < .001$. This suggests that when participants did make fixations on objects out of turn, they terminated these fixations quickly. Presumably this is because they realised that they had not followed the instructions properly.

Regarding re-fixations, on average 1 fixation was a re-fixation of an object (*SE* = 0.21) which made up 4.20% of the fixations. These re-fixations lasted for 215 ms on average. A one-sample *t*-test against the average fixation duration showed that re-fixations were shorter than average fixation durations, $t(14) = -8.88$, $p < .001$. This meant that on average, 21.04 fixations were made on objects and in the correct order ($SE = 0.62$) which represented 88.48% of the fixations[3]. What these data show is that in general, participants followed the task instructions accurately and fixated objects in the correct order.

The Influence of Fixation Order on Memory for Object Identities

To reiterate, the primary aim of Experiment 3 was to investigate whether the order in which objects were fixated influenced the accuracy with which those objects were (a) selected and (b) placed. Firstly, object identity memory was investigated. The probability of correct object selection was examined as a function of the presentation order to determine whether there was a primacy and/or recency effect as in the word list research presented earlier (e.g., Murdock, 1962; Postman & Phillips, 1965; Glanzer & Cunitz, 1966). Data were collapsed across both presentation orders (A and B) as this variable was just a counterbalancing device. A one-way repeated measures ANOVA on the probability of correct object selection with the 12 presentation positions (1st object presented, 2nd object presented etc) as the independent variable showed that there was no main effect of presentation order on probability of correct object selection, $F(11, 242) = 1.28$, $p > .05$. However, there was a significant quadratic contrast, $F(1, 22) = 5.55$, $p < .05$. This quadratic contrast provided evidence that is somewhat of a quadratic shape to the data.

A quadratic regression was carried out to investigate whether, as found in the word list research, there was a quadratic trend to the data. This regression was marginally significant, $F(2, 273) = 2.51$, $p = .08$ and yielded the following equation: $y = 0.005X^2 - 0.068X + 0.965$. Figure 8 shows the data with the quadratic curve fitted.

---

[3] Note that three of the 23 participants made 100% of their fixations on the correct object in the correct order, thereby following the instructions perfectly.

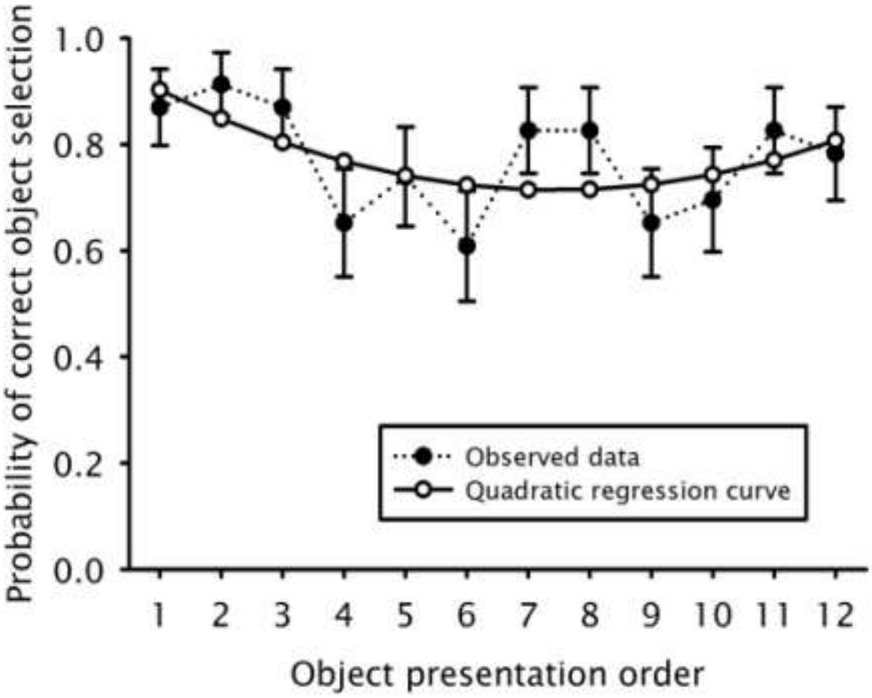

Figure 8. The probability of correct object selection as a function of the order in which objects were presented

As the quadratic regression equation was not significant at the .05 level, there is no clear evidence for a strong quadratic trend to the data as found in the word list studies (Murdock, 1962; Postman & Phillips, 1965; Glanzer & Cunitz, 1966). However, closer inspection of the pattern of data in Figure 8 could lead some to argue that there seems to be a primacy effect for the first three objects presented, and a recency effect for the last two objects presented. However, crucially, unlike the word list studies, there was no clear decline in memory performance for objects viewed in the middle of the presentation order. Instead, objects in positions 7 and 8 were remembered almost as well as the objects at the beginning and end of the sequence. This central peak was not found in the word list studies. Therefore, on balance, it seems reasonable to conclude that there was no clear effect of presentation order on memory for the identity of objects. What is clear, however, is that some objects are remembered better than other objects, despite all objects received approximately the same amount of visual encoding. 12 one-sample *t*-tests were carried out against chance (0.5) to investigate whether any of the objects were remembered more often than chance, and if so, which presentation positions the objects were presented in. The *t*-tests showed that objects in positions 1, 2, 3, 5, 7, 8, 11 and 12 were remembered more often than chance (all *p*s $< .05$) but objects in positions 4, 6, 9 and 10 were not remembered more often than chance (all *p*s $> .05$). The finding that some of the objects presented early in the sequence were remembered very often suggests that these objects may have been consolidated into long-term memory, as per Postman and Phillips (1965). However, it is more difficult to explain why some of the objects presented in the middle of the sequence were also remembered more often than chance. More detailed discussion of this pattern of results is found in the Discussion.

The Influence of Fixation Order on Memory for Object Locations

Next, memory for object locations was investigated. At this stage it is important to note that only two of the 23 box-following participants correctly selected all 12 objects. Therefore, 21 of the participants had missing data in some cells. It was impractical to remove all of the participants with missing data as this would have left only two participants in the data set. In total 22.83% of the data were missing. The missing data were spread across the data set. The missing values could either be replaced by using regression, or by mean value for each object and each condition (i.e. any missing values for the ball would be replaced by the group A ball mean or the group B ball mean, depending on where the missing values lay). However, a regression analysis found no significant linear regression for presentation order on object placement error. This is unsurprising as the pattern of data shown in Figure 8 (above) suggests that there is not a linear relationship between presentation order and memory for object identities. Because of this, it was deemed inappropriate in this instance to use regression to generate appropriate values to replace the missing values. Instead, the mean value for each presentation position was calculated for both presentation orders (A and B) and the relevant mean value replaced each missing value.

Again, data were collapsed across both presentation orders (A and B). A repeated-measures ANOVA showed that there was no main effect of presentation order on memory for object locations, $F(11, 242) = 1.56$, $p > .05$. There were no significant contrasts. Twelve one-sample *t*-tests with the Monte Carlo value of chance placement error from Experiment 1 (84.60) showed that objects at all presentation positions were placed more accurately than chance (all *p*s $< .001$).

A quadratic regression was not significant, $F(2, 273) = 0.47$, $p > .05$, and there was no significant quadratic contrast, $F(1, 22) = .001$, $p > .05$ suggesting that there is no evidence of a primacy or a recency effect in memory for the locations of objects. Figure 9 shows the pattern of results with the quadratic curve fitted. As with the object identity data presented earlier, there was no clear primacy or recency effect. Instead, memory for locations seems to be most accurate for objects in positions 1, 4, 5, 8 and 9 (note that accuracy for object location is reflected by low error scores).

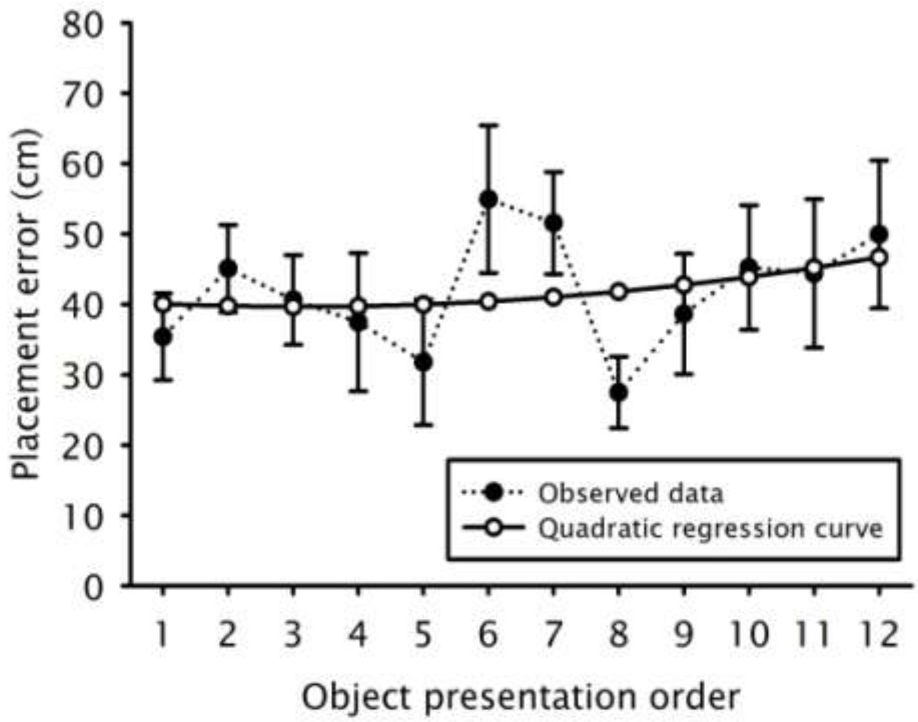

Figure 9. Placement error (cm) as a function of the order in which objects were presented

The influence of fixation order on gaze duration

A further analysis also probed in more detail how accurately participants followed the task instructions. This analysis investigated how long participants fixated each object during their first visit to the object (gaze duration). Earlier results suggested that participants followed the task instructions closely and rarely made fixations on objects out of turn. Recall that participants were told to keep fixating the highlighted object until the box moved to the next object and that the green box highlighted each object for 810 ms. If participants followed the instructions accurately, one might expect that the gaze duration on each object would be comparable across all objects. Therefore, the purpose of this analysis was to assess whether each object received approximately the same amount of visual processing during the encoding phase. In Experiments 1 and 2, it was found that the accuracy with which an object was selected and placed was closely linked to the amount of visual processing that object had received. On that basis, one might expect the same relationship to be observed in the present experiment. As the previous analyses have not demonstrated any clear serial position effects on memory for identities and locations, it is likely that the amount of visual processing that each object received during encoding was comparable.

A repeated-measures ANOVA showed that, as expected, there was no main effect of presentation order on gaze duration, $F(11, 242) = 1.20$, $p > .05$. There was no significant linear contrast. This suggests that each object received approximately the same amount of visual processing during the encoding phase. Both a linear regression and a quadratic regression were carried out to assess which regression line fitted the data better. Neither regression was significant (both *p*s > .05). The lack of a significant regression suggests that there was no difference between objects in the amount of visual processing that each object received. It therefore suggests that participants closely followed the task instructions and looked at each object for approximately the same amount of time. In addition, the average gaze durations were consistently approximate to the box presentation duration of 810ms, providing further evidence that participants followed the task instructions and looked at each of the boxed objects in turn, maintaining fixation on those objects until the box moved to another object. Figure 10 shows the pattern of data with the quadratic regression curve fitted.

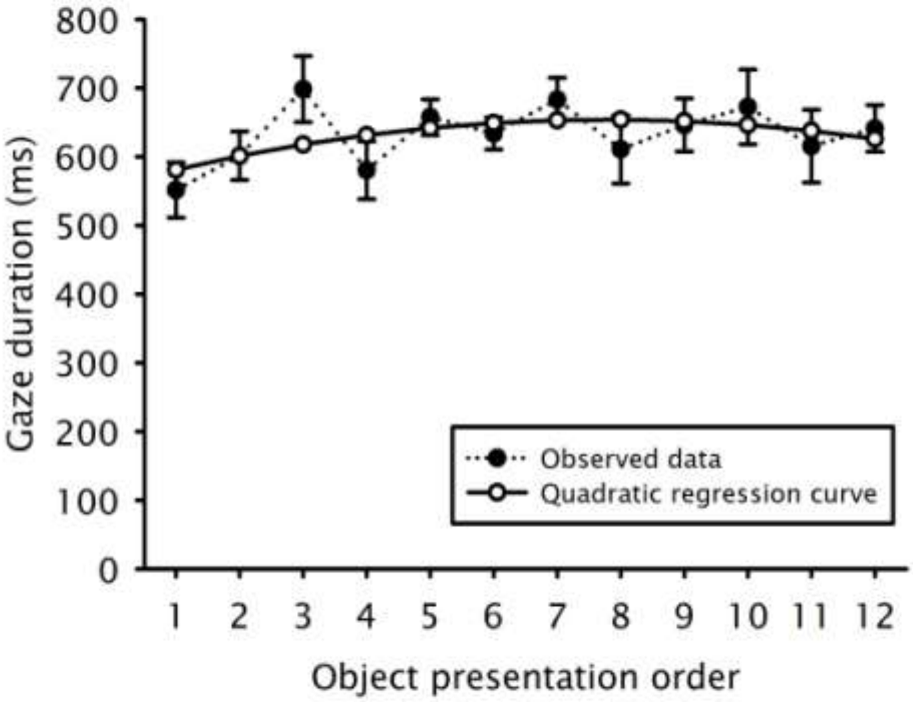

Figure 10. Gaze duration (ms) as a function of the order in which objects were presented.

The influence of fixation order on the number of fixations made during gaze duration

As in the previous analysis, the effect of presentation order on visual processing was investigated. In this analysis, the number of fixations made during gaze duration was used as the dependent variable. As before, there was no main effect of presentation order on the number of fixations made during gaze duration, $F(11, 242) = .86$, $p > .05$. Figure 11 shows the pattern of the data.

Again, both a linear regression and a quadratic regression were carried out. Figure 11 shows the pattern of data with the curve fitted. As before neither regression was significant (both ps > .05).

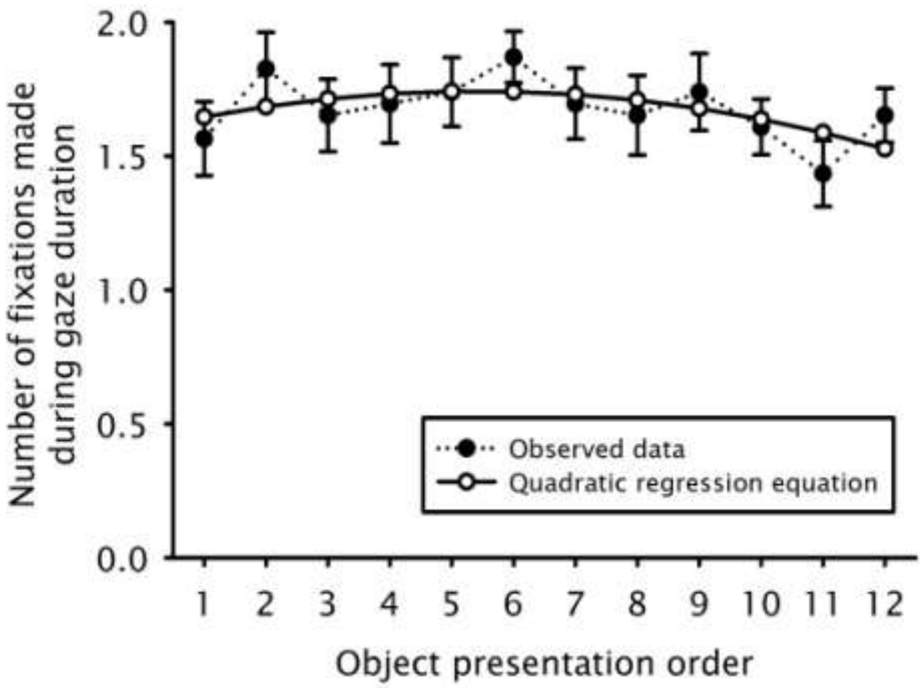

Figure 11. Number of fixations made during gaze duration as a function of the order in which objects were presented.

Relationship between Fixations made during Gaze Duration in the Encoding Phase and Memory for Object Identities

As in Experiments 1 and 2, analyses were carried out to investigate the relationship between encoding and memory. Recall that in Experiments 1 and 2 it was found that both the identity and location was recalled more accurately for fixated objects than non-fixated objects. In addition, in Experiment 1 it was demonstrated that memory for object identity improved from one to two or more fixations, but memory for the location of objects did not improve from one to two or more fixations. In the present experiment, only one object was not-fixated (by one participant) so consequently fixations were grouped into one, two or three or more bins and three-way repeated measures ANOVAs were carried out on the probabilities of correct object selection, and placement error scores.

As the methodology encouraged participants not to re-fixate objects (and this happened rarely, as quantified above) the number of fixations made during gaze duration was used as an independent variable unlike in Experiments 1 and 2 where the number of fixations made in total was used. However, of the 23 participants, 11 of them did not make 3 fixations during gaze duration on any objects. This meant that 15.94% of the data were missing and all of this data was from the three or more bin. To try to replace these blank cells, a linear regression was carried out using number of fixations during gaze duration as the predictor variable. A linear regression was judged to be an appropriate method of generating values to replace missing cells because in Experiments 1 and 2, there was a linear relationship between the number of fixations and probability of correct object selection. In addition, unlike the missing data from the placement error analysis reported earlier, data was only missing from one of the "bins". However, the regression model was not significant, and yielded an equation of $y = 0.005X + 0.764$. Rather than use the intercept of a non-significant regression, the data from the "two" and "three or more" bins were pooled leaving bins of one fixation and two or more fixations.

For these next analyses, the presentation order (A or B) was included in the analyses as a dummy variable but the data were collapsed across both orders in the tables. First, regarding object identity, a mixed 2x2 ANOVA showed that there was no improvement in memory for the identities of objects from one to two or more fixations, $F(1, 21) = 0.10$, $p > .05$. Instead, probability of correct selection was comparable regardless of the number of fixations made on an object during gaze duration. There was no effect of presentation order (A or B), $F(1, 21) = 0.10$, $p > .05$ and no interaction between the two, $F(1, 21) = 3.10$, $p > .05$. It was inappropriate to compare fixated and non-fixated objects because all except one participant fixated all 12 objects.

Table 15

Probability of Correct Object Selection as a Function of the Number of Fixations made during Gaze Duration

Fixation status	M	SE
Probability of correct selection for objects fixated once	0.76	(0.04)
Probability of correct selection for objects fixated twice or more	0.78	(0.03)

Relationship between fixations made during gaze duration in the encoding phase and memory for object locations

Next, the same analysis was carried out on placement error. 15.94% of the data were missing and all of this data was from the three or more bin. A linear regression analysis was carried out on the placement error scores using the number of fixations during gaze duration as the predictor variable. This was not significant and it yielded a regression equation of $y = -2.091X + 45.248$. As the regression was not significant, the missing data was not replaced by the intercept of a non-significant regression. Instead, data from the "two" and "three or more" bins were pooled.

Again, a mixed 2x2 ANOVA showed that there was no effect of the number of fixations, $F(1, 21) = 0.01$, $p > .05$. However, there was a main effect of condition, $F(1, 21) = 8.92$, $p < .001$ where objects were placed more accurately in condition A than B. The interaction was not significant, , $F(1, 21) = 0.90$, p .05. However, as stated before, the different presentation conditions are theoretically uninteresting.

Taken together, these analyses show that, unlike in previous experiments, memory for object identity did not seem to improve with an increasing number of fixations. Instead, memory for both object identity and object location was comparable across fixations. However, this may not be a surprising finding when one considers how the task instructions given to participants and the subsequent eye movement behaviour exhibited by participants in Experiment 3 differs to the earlier experiments. In Experiments

1 and 2, participants had no restrictions on the order or duration of their fixations. Participants were free to make as many or as few fixations on objects as they chose to (or indeed to not fixate an object at all). Furthermore there was no restriction on re-visiting objects. Participants were allowed to return to objects on as many occasions as they chose to.

In stark contrast, in the present experiment the order and duration of fixations were directed through the task instructions to follow the green box that surrounded each object in turn for an equivalent duration. Unlike Experiments 1 and 2 participants were instructed to look at the objects for only as long as they were surrounded by the green box, to not return to objects and, due to the nature of the experimental design, participants were encouraged to look at all 12 of the objects.

As a result of these quantitative differences in task instructions and quantitative differences in eye movement behaviour it is perhaps unsurprising that different data patterns relating to the relationship between encoding and memory occurred. In addition, there was no significant correlation between gaze duration and placement error, $r(213) = 0.03$, $p > .05$ and no significant correlation between total time and placement error, $r(213) = -.08$, $p > .05$.

Table 16

Placement Error as a Function of the Number of Fixations made during Gaze Duration

Fixation status	M	SE
Placement error for objects fixated once	41.92	(6.04)
Placement error for objects fixated twice	41.76	(3.66)

The Effect of Presentation Order on Confidence for Object Identity

The final analyses investigated the effect of presentation order on confidence for both identity and location. Recall that during the test phase, participants were asked for their confidence that each of their 12 selected objects had been present in the stimulus photograph and had to respond with a number between 0 and 10. 0 represented an absolute guess, and 10 represented absolute confidence. Figure 12 shows object confidence as a function of presentation order. As only two of the 23 participants correctly selected all 12 objects, 22.83% of the cells were empty. To replace this missing data, a linear regression was carried out. The linear regression was significant, $F(1, 211) = 8.06$, $p < .01$ and yielded the following equation: $y = -.166X + 9.132$ which was used to replace the missing data.

A 12-way repeated measures ANOVA on confidence for object identities with presentation order as the independent variable showed that there was a significant main effect of presentation order on confidence for object identity, $F(11, 242) = 2.96$, $p < .001$ (see Figure 12). This is in contrast to the memory analyses presented earlier which found no effect of presentation order on memory for identity or location. This finding suggests that participants were more confident about objects that they had looked at earlier in the presentation order than objects presented late in the presentation order. There was also a significant linear contrast, $F(1, 22) = 15.22$, $p < .01$.

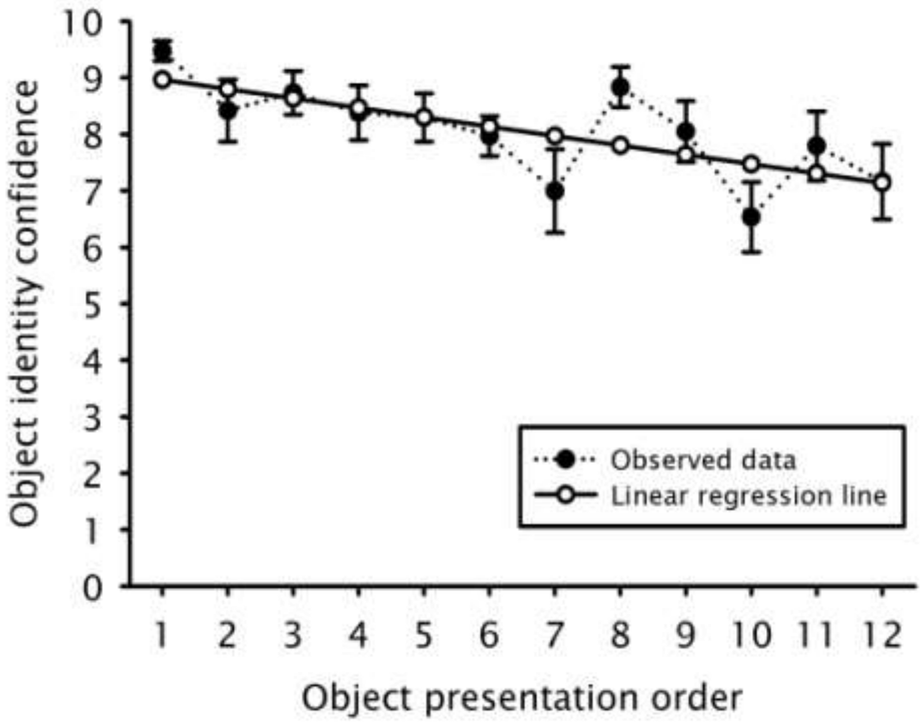

Figure 12. Object identity confidence as a function of the order in which objects were presented.

In addition to carrying out the regression above to replace the missing values, both a linear regression and a quadratic regression were carried out to find out which regression best fit the pattern of data. Both regressions were significant, but the linear regression yielded a lower p value, $F(1, 274) = 14.60$, $p < .001$, than the quadratic regression, $F(2, 273) = 7.48$, $p < .01$. Therefore, the regression model that was generated was $y = -.166X + 9.132$. Figure 12 shows the pattern of data with the linear regression line fitted. This confirms that confidence for the identity of objects was higher for objects presented early in the sequence than at the end of the sequence.

The Effect of Presentation Order on Confidence for Object Location

The final analysis investigated whether object location confidence was influenced by presentation order. As before, a linear regression was carried out to replace missing values. This linear regression was significant, $F(1, 211) = 5.29$, $p < .05$ and yielded the following equation: $y = -.126X + 6.888$ and this equation was used to replace missing values. A 12-way repeated measures ANOVA was carried out on the location confidence scores and there was a significant effect of presentation order on object location confidence, $F(11, 242) = 2.80$, $p < .01$. There was also a significant linear contrast, $F(1, 22) = 8.06$, $p < .05$ with object location confidence higher for objects presented early in the sequence, and lower for objects presented later.

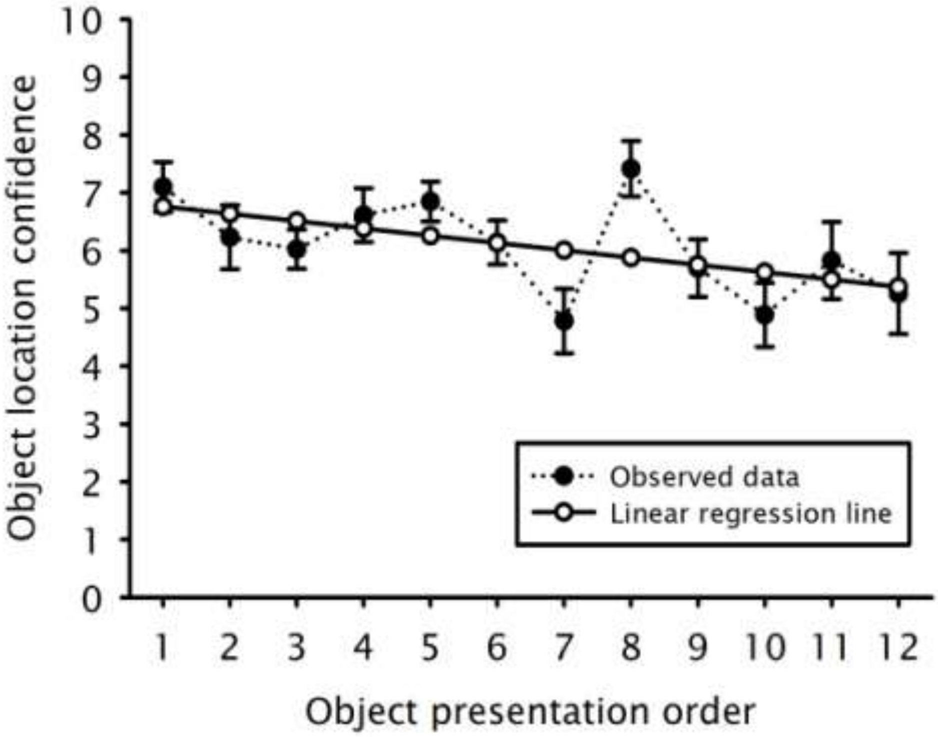

Figure 13. Object location confidence as a function of the order in which objects were presented.

A quadratic regression and a linear regression were carried out, and again both were significant, although again the p value for the linear regression was lower, $F(1, 274) = 8.44$, $p < .01$ than the p value for the quadratic regression, $F(2, 273) = 4.23$, $p < .05$. This generated the following linear regression model: $y = -.126X + 6.888$. As before, this suggests that confidence for the location of objects was higher for objects presented early in the sequence than for objects presented late in the sequence. Figure 13 shows the data pattern with the linear regression line fitted.

4.4 Discussion

Several interesting findings have been derived from Experiment 3 and each one will be discussed in turn. First, it was found that the eye movements made by participants differed depending on whether participants were in the free-viewing condition or the box-following condition. When told to follow a green box around the photograph as it "jumped" from object to object, participants made fewer and longer fixations over a smaller range of areas of the photograph compared to free-viewing participants who had been given the freedom to examine the photograph with no restrictions. This suggests that the presence of task instructions for the box-following participants directly led those participants to change the way in which they moved their eyes compared to the free-viewing participants. The box-following participants tended to look mainly at the objects instead of the (uninformative) blank regions, to make longer fixations on objects and to make fewer fixations in general. This provides evidence that the task instructions given to participants can strongly influence what they choose to do with their eyes, and that participants tend to make eye movements that are most appropriate to the task instructions, as found by Buswell (1935), Yarbus (1967) and Castelhano, Mack, and Henderson (2009).

Interestingly, despite the box-following participants making more fixations, making longer fixations and fixating more of the objects compared to free-viewing participants (11.96 vs. 10.60), no advantage was observed in memory for either the identity or the location of objects for box-following participants. However, box-following participants selected the same number of objects (9.26) as free-viewing participants (8.90) and placed objects no more accurately (41.79 cm vs. 38.28 cm).

There are at least two explanations for this. First, it may be the case that following the green box around the photograph required a certain level of attentional resources that would otherwise have been used for encoding the objects. Given the freedom to encode objects at one's own leisure in the free-viewing condition, it seems highly unlikely that a participant would shift their gaze to another object (object N + 1) until the object they were currently fixating (object N) had been encoded. However, in the box-following condition the green box highlighted each object for 810 ms and this may have been too brief a period for some of the objects to be encoded fully. Thus, even though box-following participants fixated more objects than free-viewing participants, they may not have finished encoding each object before transferring their gaze to the next object. If this is true, it might explain why the box-following participants did not correctly select more objects than the free-viewing participants, despite fixating more of the objects. Put simply, in the box-following condition, fixating an object may not have guaranteed that its identity and location were encoded.

Alternatively, the lack of a memory advantage for the box-following participants may reflect the natural limit of participants' memory for objects in scenes. As suggested by Miller (1956) short-term memory typically holds between 5 - 9 chunks of information. However, that finding was based on digit-span research, not memory for complex objects in scenes. It was found that some of the box-following participants did correctly select all 12 objects, demonstrating that this is not beyond the limits of normal human memory (although this was unusual – only two participants did so).

The next finding is that the eye movements and memory performance exhibited by the free-viewing participants in this study (Experiment 3) were very similar to the corresponding condition in Experiment 1 (10 s/0 delay). Of the 14 measures compared, one was significant at the .05 level, and one was marginally significant at the .06 level. All other measures were not significantly different. This finding is important because it replicates the eye movements and memory performance demonstrated by the participants in Experiment 1 who had 10s viewing duration and an immediate test i.e. those participants who experienced identical experimental conditions. This finding is also unsurprising as the free-viewing participants in the present study experience identical testing conditions to their counterparts in Experiment 1. Note, however, that different eye trackers were used in Experiment 1 and Experiment 3.

A second replication of results was also observed for the free-viewing participants in the present study. As in Experiment 1, it was demonstrated that memory for the identities of objects was better for fixated than non-fixated objects. In addition, memory for object identities was better for objects fixated twice than objects fixated once, and objects fixated three times or more than objects fixated once. This provides further evidence that object identity information accumulates in memory across multiple fixations. As in Experiment 1, memory for object locations was more accurate for fixated than non-fixated objects, but did not improve with subsequent fixations. As in Experiment 1, the most likely explanation is based upon the Visual Memory Model (Hollingworth & Henderson, 2002) as explained in detail in Chapter 2.

This same pattern of effects was not found for the box-following participants. The most likely explanation is that, due to the task instructions requiring participants to either follow the green box around the scene or simply look at the photograph freely, the duration of each fixation during gaze duration was different for box-following and free-viewing participants. A mixed 2 (Condition: box-following vs. free-viewing) x 3 (Number of fixations during gaze duration: 1, 2, or 3 or more) ANOVA on gaze duration showed a significant effect of number of fixations during gaze duration on gaze duration, $F(2, 26) = 63.59$, $p < .001$ with gaze durations unsurprisingly increasing as the number of fixations increased. However, critically, there was an interaction between the condition and the number of fixations made during gaze duration, $F(2, 26) = 17.54$, $p < .001$ (see Figure 14).

This interaction shows that for the box-following participants, gaze duration was not substantially different regardless of whether one, two, or three or more fixations were made on an object. For example, gaze duration for objects fixated once was 528 ms and for objects fixated three times or more it was 771 ms, an increase of only 243 ms. However, for free-viewing participants, gaze duration for objects fixated once was 336 ms but for objects fixated three times or more it was 1,093 ms, an increase of 757 ms.

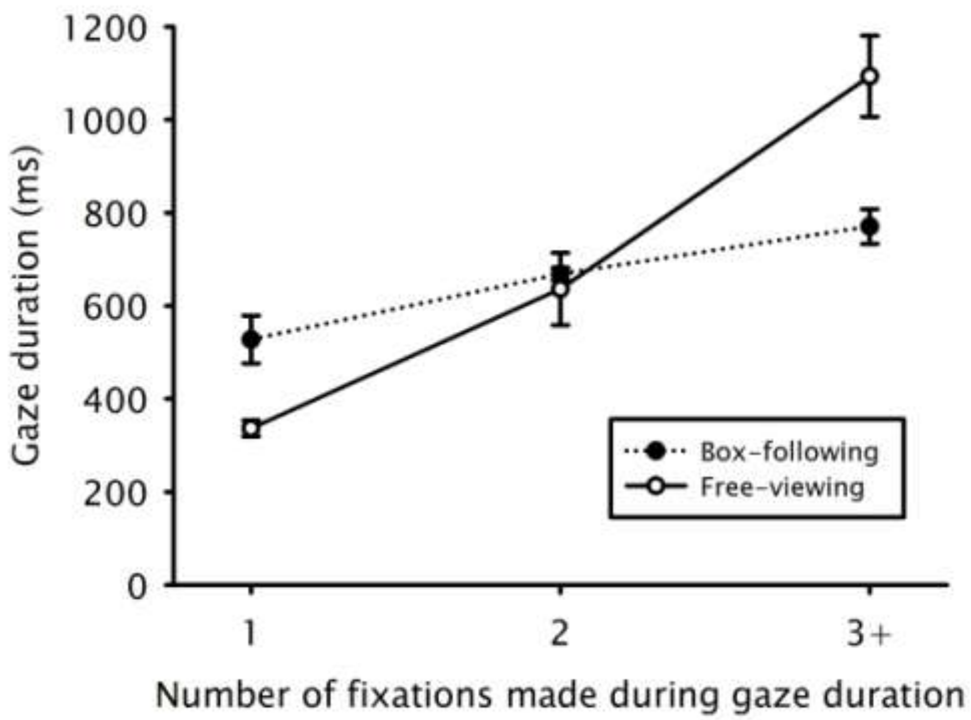

Figure 14. Gaze duration (ms) as a function of the number of fixations made during gaze duration for box-following and free-viewing participants

In addition, post-hoc *t*-tests showed that for objects fixated only once, gaze duration was longer for the box-following than the free-viewing participants, $t(10.83) = 3.53$, $p < .05$. For objects fixated twice, there was no difference in gaze duration between box-following and free-viewing participants, $t(4.24) = .41$, $p > .05$. However, for objects fixated three times or more, in contrast to objects fixated only once, gaze duration was longer for free-viewing than box-following participants, $t(13) = -4.05$, $p < .01$.

These results suggest that the duration of one fixation during gaze duration for the free-viewing participants was not equivalent to the duration of one fixation in gaze duration for free-viewing participants. The data show that for box-following participants, memory for object identities after one fixation was high (0.77) and did not improve with subsequent fixations. In contrast, memory for objects fixated once by free-viewing participants was lower initially (0.49) but improved as the number of fixations increased from two (0.85) to three or more fixations (0.96).

Therefore, two distinct patterns of encoding have been observed. Free-viewing participants replicated the encoding behaviour of participants in Experiments 1 and 2 in this thesis. This was characterised by shorter fixations (292 ms) and shorter gaze durations (469 ms). However, the box-following participants made longer fixations (372 ms) and longer gaze durations (635 ms). It is highly likely that this happened in response to the task instructions to "look at the boxed object and keep looking at it until the next object is boxed".

To clarify, the most likely reason for the lack of accumulation for object identity information for box-following participants was that the difference in visual encoding between one and three or more fixations did *not* allow for substantially more information to be encoded. However, the difference between one and three or more fixations for free-viewing participants *did* allow accumulation of object identity information to occur.

The next two findings relate only to the box-following participants, and the potential effects of the object presentation order. It was predicted that if presentation order did have an effect on memory for object identity, then probability of correct object selection would be higher for objects presented early and late, and substantially lower for objects presented in the middle of the sequence; in other words, a quadratic trend would have been expected (as found by Glanzer & Cunitz, 1966, Murdock, 1962, and Postman & Phillips, 1965). However, if the order in which objects were presented did not have an effect on memory for object identities, then it would be expected that a quadratic trend would not be observed.

The results showed first that there was no robust evidence for a clear quadratic data pattern for memory for object identities, suggesting that there is no clear primacy or recency effect on the accuracy of memory for object identities. Instead, some objects simply seemed to be remembered better than others, regardless of the order that those objects were fixated in. The one-sample *t*-tests showed that, as found by Hollingworth (2004), Irwin and Zelinsky (2002), and Zelinsky and Loschsky (2005), memory for the identity of the last two objects in the sequence was very accurate and significantly better than chance. However, objects that were presented in other positions within the sequence were also remembered more often than chance. Objects in positions 1, 2, 3 were remembered more often than chance, suggesting a primacy effect as found in the word lists studies by Glanzer and Cunitz (1966), Murdock (1962), and Postman and Phillips (1965). However, the critical finding that was not predicted was that objects in positions 5, 7 and 8 were also remembered accurately. This accuracy in memory for objects in the middle of the sequence was not found in the previous studies investigating presentation order and memory for objects, and memory for word lists. On this basis, the data are not consistent with data reported by Irwin and Zelinsky, (2002), Hollingworth, (2004), Zelinsky and Loschsky (2005). A similarly unconvincing pattern of results was observed for object location memory. There was no clear significant quadratic trend to the data, suggesting that object location memory was not related to the order in which objects were presented in the sequence. However, memory for the locations of objects was clearly poorest for objects presented in the middle of the presentation order (positions 6 and 7). Despite this, there were not clear primacy and recency effects as found by Jones et al. (1995) and Smyth and Scholey (1996). Instead, objects in presentation positions 1 – 5, and 8 – 12 were placed fairly accurately.

Taken together, these two findings suggest that overall the order in which objects are looked at does not clearly affect the memory for the identity or location of those objects. To be clear, there is no robust evidence for primacy or recency effects in memory for the identities and locations in scenes. Instead, perhaps there is an alternative explanation as to why the identities and locations of some objects were remembered better than others, regardless of their position in the sequence (e.g., the identities of objects presented in position 7 and 8 were remembered more often that objects in positions 4 or 10). A possible explanation for the finding that some objects are remembered better than others is that this memory for object identity and location is influenced by the amount of visual processing that each object received during the encoding phase (as found in Experiments 1 and 2 in this thesis, and also demonstrated by Tatler, Gilchrist, and Rusted (2005).

To investigate whether some objects received more visual encoding than others despite the experimental manipulation of the moving green boxes an analysis was carried out to see whether there was substantial variability between the objects in the amount of visual processing each object received. Based on the findings of Experiments 1 and 2, the Visual Memory Model (Hollingworth & Henderson, 2002) and the findings of Tatler, Gilchrist, and Rusted (2005) it was predicted that if there were differences in the amount of visual processing, then objects that received a lot of visual processing would be (a) selected more often and (b) placed more accurately than objects that had received less visual processing.

Indeed, when an analysis of the effect of presentation order was carried out on the amount of visual processing each object received, there was no reliable effect of presentation order. Instead, for all participants, each object received approximately the same amount of visual processing time, both in terms of gaze duration itself, and the number of fixations made during gaze duration. How then can the differences in accuracy of memory for identities and locations between objects be explained, given that there were no clear effects of presentation order (as predicted by Murdock, 1962; Postman & Phillips, 1965; Glanzer & Cunitz, 1966) and no clear effects of visual processing duration?

Possible Explanations for Variability in Memory for Object Identities and Locations

Object/Scene Congruency

There are at least three possible explanations for some objects being remembered better than others despite visual processing being broadly similar for all objects. The first is that some objects may have been considered somewhat unexpected in the scene, and as such they may have violated the gist of the scene, thereby making them more memorable. The objects were located in a University cubicle which contained a desk and a chair, and it is possible that participants may have interpreted the photograph to be an office scene. If that is true, then one might not expect certain objects (a ball, a clock or lava lamp) to be present in the scene and as such, these objects may have been remembered preferentially. This issue of scene congruency has been studied in depth by many researchers. Some have provided evidence that scene incongruent objects attract attention faster and for longer than scene congruent objects (Loftus & Mackworth, 1978; Underwood & Foulsham, 2006) while others have shown that scene incongruent objects are identified less readily than scene congruent objects (Davenport & Potter, 2004). Silva, Groeger, and Bradshaw (2006) recorded eye movements as participants looked at photographs of a woman carrying out everyday tasks. Each photograph was viewed for 15 s and contained a mix of task-relevant and task-irrelevant objects. When participants were told to name as many task-relevant objects whilst viewing the photograph, they spent more time looking at task-relevant objects and vice versa. However, participants who were told name objects beginning with the letter “c” (the control condition) spent an equal amount of time looking at both task-relevant and task-irrelevant objects. Participants were then given a surprise recognition memory task on the contents of the photographs and recognition data showed the same pattern of results: participants who named task-relevant objects recognised more task-relevant objects, and vice versa. However, there was no such difference for control participants who recognised task-relevant and task-irrelevant objects equally often. These results suggest that in the absence of specific instructions regarding the relevance of objects to the scene or task (as in the present experiment), there is no clear difference in the way task-relevant and task-irrelevant objects are looked at or remembered.

Object Size

The second possible explanation for some objects being remembered better than others given comparable amounts of visual encoding relates to the size of the objects. Uttl, Graf, and Siegenthaler (2007) demonstrated that when the same objects were presented in three different sizes (small, medium and large) participants correctly identified the large objects fastest and the small objects slowest, thereby suggesting that as the size of an object increases, the speed with which it is identified decreases. In the present experiment, there was variation in size between the objects with the smallest object (the glass) occupying 2379 pixels (which was 0.54% of the photograph) and the largest object (the keyboard) occupying 14129 pixels (3.21% of the photograph). Potentially, the larger the object, the easier it is to (a) identify and therefore (b) remember. This rests on the assumption that larger objects are easier to identify than smaller objects as found by Uttl, Graf and Siegenthaler and therefore are encoded and consolidated into memory more quickly. However, in the present dataset there were only very weak, non-significant correlations between the size of the objects and the probability of correct object selection, $r(276) = .05$, $p > .05$ and between object size and placement error, $r(213) = .01$, $p > .05$. Therefore, it seems as though object size did not influence the accuracy with which the identities and locations of objects were remembered.

Differential Nature of Objects in Scenes vs. Word Lists

The third and most likely explanation for the finding that some objects were remembered better than others, despite no clear effects of either presentation order or the amount of visual processing, is that processing by which participants remember the identities and locations of objects in scenes is fundamentally different to that by which they remember lists of words. In normal scene viewing, there is natural variation in the amount of time objects are looked at. The results of the experiments in this thesis and the Visual Memory Model suggest that this encoding variation typically influences memory for the identity and location of objects. According to the Visual Memory Model, information about the identity of objects is encoded across multiple fixations, and is embellished even if those fixations are not successive. As a result, memory for the identities of objects improves as the number of fixations on those objects increases. During normal scene viewing, participants can choose to ignore some objects, spend variable amounts of time on other objects whilst returning on multiple occasions to other objects. This leads to the variation in the amount of encoding time each object receives, and this in turn, leads directly to differences in memory performance.

This pattern of results has been consistently demonstrated throughout this thesis. However, the paradigm used in this experiment that controlled (at least to some degree) the amount of time each object was looked at, meant that there was very little variation in the time spent encoding the different objects. Even so, the results still showed that some objects were remembered better than others. Imposing very similar encoding durations for each object in the modified encoding phase used in the present experiment may mean that some objects were processed to a greater degree than they would have been during a free viewing situation, and conversely, some are processed to a lesser degree than would have occurred during free viewing. If this reasoning is correct, then a negative correlation would be expected between the accuracy with which objects were remembered in this experiment, and the average fixation duration made on each object under free-viewing conditions. In other words, objects that typically elicit long fixations during normal scene viewing in Experiment 1 will be remembered poorly in Experiment 3 because the boxing manipulation of Experiment 3 does not permit the viewer to process them sufficiently such that memory for them is effective. To investigate this further, the average fixation duration made on each of the 12 objects in Experiment 1 was correlated with the probability that each object was correctly selected in Experiment 3. This analysis showed that whilst there was a negative correlation, as predicted, it was not significant, $r(11) = -.47$, $p > .10$. Note, however that there were only 12 data points in this analysis (one corresponding to each object), and therefore, the analysis might be considered to be lacking in power. Figure 15 shows the scatterplot for the data.

It is clear that the current data set can only provide a limited opportunity to investigate this explanation of why different objects are remembered more or less effectively. In the spirit of caution, therefore, it is perhaps safest to conclude that there is little evidence for a relationship between the average fixation duration made on each object during free-viewing and the likelihood it will be remembered under sequential (boxed) viewing conditions. As a result, this issue requires direct investigation in future studies.

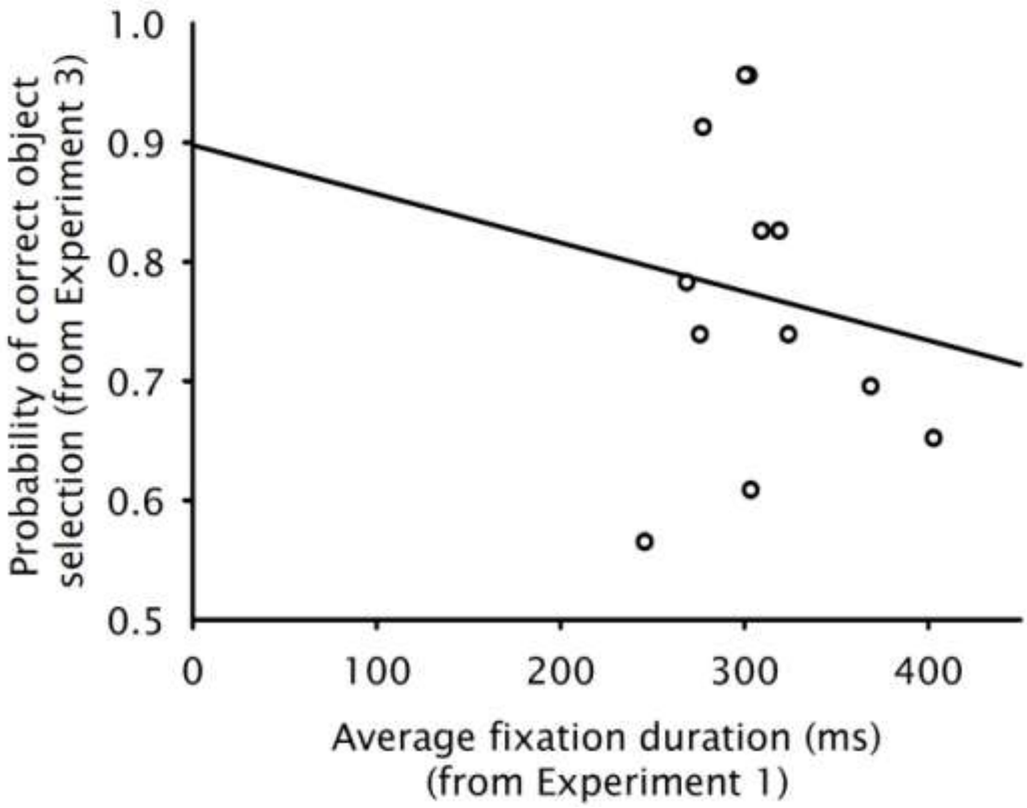

Figure 15. Scatterplot showing the correlation between average fixation duration (ms) made on each of the 12 objects in Experiment 1 and the probability that each object was correctly selected in Experiment 3.

It is important to note that different objects were selected with varying degrees of frequency in the test phase for box-following participants. Table 17 shows the proportion of box-following participants that selected each of the objects.

Table 17

The Proportion of Participants that Selected each Object and Average Placement Error for each of the 12 Objects

Object	Proportion of participants that selected the object	SE	Average placement error (cm)	SE
Ball	0.90	0.07	46.16	10.56
Book	0.90	0.07	42.63	7.93
Clock	0.75	0.10	41.13	9.10
Glass	0.60	0.11	28.13	6.09
Keyboard	0.85	0.08	33.04	6.57
Laptop	0.65	0.11	34.46	6.53
Lava lamp	0.95	0.05	30.83	7.52
Mouse	0.65	0.11	48.60	8.19
Phone	0.90	0.07	60.23	8.68
Plate	0.75	0.10	41.02	11.90
Scissors	0.80	0.09	30.70	6.93
Stapler	0.90	0.07	56.66	9.20

It seems as though there are some inherent differences between how often each object is remembered when objects have received approximately the same amount of visual processing and this may also be a contributing factor to the lack of a significant effect of presentation order on identity and location memory.

Confidence Data

The confidence data initially appear somewhat surprising as, unlike in Experiment 2, they do not seem to closely match the actual behaviour exhibited by participants in the test phase. Instead, for both identity and location confidence, there seems to be a linear relationship between confidence and presentation order with confidence decreasing over objects as they were viewed in turn. This was not what was predicted. Based on the findings of Experiment 2 and the data from Shaffer and Shiffrin (1972), it was expected that the confidence ratings would reflect the actual memory for objects and locations. However, it is important to note that confidence for both identities and locations did not vary substantially, with average confidence for identities ranging from 9.47 to 6.53; a difference of 2.94. Similarly, average confidence for locations ranged from 7.41 to 4.78; a difference of 2.63. This demonstrates that there was only a moderate effect of the presentation order on confidence and that confidence did not vary substantially between objects. Further evidence for this was shown by the moderate negative correlations between presentation order and identity confidence, $r(213) = -.20$, $p < .01$ and between presentation order and location confidence, $r(213) = -.16$, $p < .05$ (even though these effects were reliable).

In contrast, confidence seemed to be much more strongly related to the order in which participants chose to *select* objects at test. The correlation between selection order and identity confidence was highly reliable, $r(213) = -.53$, $p < .001$ as was that between selection order and location confidence, $r(213) = -.64$, $p < .001$. These correlations are much stronger than the correlations reported above for presentation order, suggesting confidence is more closely related to the order in which participants chose to select the objects, than to the order in which they looked at the objects.

It is plausible that participants chose to place objects they were most confident about first, and left objects they were less confident about until last. Presumably, confidence variability reflects the degree to which each object was encoded in detail, with participants choosing to place the objects that they had encoded in more detail first, and leaving the objects encoded in less detail until later in the test phase. If this was the case, then a positive correlation should exist between selection order and placement error, with participants first placing objects for which they were most sure about the location (represented by a low error score) and leaving the objects for which they were least sure about the location until the end of the test phase (a high error score). Consistent with this expectation, there was a positive correlation between selection order and placement error, $r(213) = .30$, $p < .001$, confirming that participants tended to place objects for which they were most sure about the location first, and least sure about the location last. These analyses, therefore, provide evidence that confidence *is* closely related to memory (as found by Shaffer & Shiffrin, and in Experiment 2 in this thesis). In short, participants encode some objects in more detail than others. The objects encoded in most detail are placed most accurately and most confidently, and objects encoded in least detail are placed least accurately and least confidently.

Summary

Unsurprisingly, the pattern of results for free-viewing participants closely matched the encoding behaviour, memory performance and relationship between encoding and memory observed by participants in Experiments 1 and 2 in this thesis, thereby directly replicating those findings. Specifically, it was found that memory for object identity was more accurate for fixated objects than non-fixated objects. In addition, objects fixated twice, and three times or more were selected more often than objects fixated once, demonstrating that information about object identity continues to accumulate in memory across separate fixations. This pattern of results further supports the explanation based on the Visual Memory Model as outlined in chapter 2.

An additional replication was found for object location memory. Fixated objects were placed marginally more accurately than non-fixated objects, but there was no clear accumulation in memory for object location. Again, this directly replicates the findings of Experiments 1 and 2, and the Visual Memory Model. However, the box-following participants exhibited substantially different encoding behaviour compared to the free-viewing participants. It is likely that this was in direct response to the different task instructions that they received, and also the different encoding phase that they experienced (the moving green box). As the experimental manipulation of highlighting each object with a green box for 810ms was so effective, it led to there being very little variation between objects in terms of visual encoding. As a result, the Visual Memory Model would predict that as all objects received the same amount of encoding, memory for all objects would be comparably accurate. However, this was not what was found. Instead, the identities and locations were simply remembered better for some objects than for others.

This suggests that, given identical amounts of visual encoding for each object, other factors (e.g., scene congruency, object size, or perhaps other unknown variables) may influence the accuracy with which an object is (a) selected and (b) placed. A possible explanation is that different objects require different amounts of visual processing. The methodology used in Experiment 3 that, to some degree, controlled the amount of time each object was looked at meant that this variation in encoding was significantly removed. As a result, some objects may have received more time under direct fixation than they would have received under free-viewing conditions, and conversely, some objects may have received less attention. This disparity may well have led to the variation observed for identity and location memory. This issue requires further research to directly examine factors that influence the encoding of objects.

What can be concluded quite firmly, however, is that there is no clear primacy and recency effect for the identity and location of objects when visual encoding is controlled. Instead, the identity and location of some objects are remembered better than others, and these accurately remembered objects are not restricted to objects at the beginning and end of the presentation order.

5 The relationship between the number of fixations on an object and memory for the configuration of objects and specific object locations

5.1 Introduction

The results of Experiments 1 and 2 in this thesis provided evidence that, as per the Visual Memory Model (Hollingworth & Henderson, 2002), detailed information about the identity and location of objects in scenes is encoded during scene viewing and is consolidated into long-term memory. In addition, those experiments added detail to the existing framework of the Visual Memory Model by demonstrating that information about the overall configuration of objects, object identities and specific object locations seem to be encoded differently. The data support the view that (a) information about the overall configuration of objects in the scene (as measured by the best fit error) is encoded robustly and very early during scene viewing; (b) information about the identity of each object is encoded initially during the first fixation on that object, and continues to be embellished and enriched during subsequent fixations on that object; and (c) information about the specific location of an object is encoded during the first fixation on an object but does not become more accurate with extended encoding via more fixations on that object. Thus it appears that three distinct processes are observed when the identities and locations of objects in a scene are encoded. At this point, it seems appropriate to re-consider Hollingworth and Henderson's Visual Memory Model as it could be argued that the model requires some development in order to be able to explain the findings reported so far in this thesis.

5.1.1 Hollingworth and Henderson's Visual Memory Model

The Visual Memory Model as specified by Hollingworth and Henderson provides a framework which aims to describe how a visual scene and the objects in that scene are encoded. Although the model has clearly defined stages, it does not seem able to account for some of the findings presented in this thesis, such as the dissociation between encoding of object identities and specific object locations. According to the model, the first thing to happen upon encountering a visual scene is that a short-term object file is created when an object is directly fixated. This object file initially contains abstract details about the object, such as the identity of the object e.g., a ball. Once this object file is created, it is indexed to a position in a spatial map. However, the model in its present form does not provide any detail about the nature of the spatial map. One might assume that the spatial map is formed *before* any object files are created, or certainly before an object file can be indexed to the spatial map, as the indexing of an object file to the spatial map necessitates the presence of the spatial map. If the spatial map is not created first, or in parallel with the object file, then there is nothing available for the object file to be indexed to. Therefore, one part of the Visual Memory Model that might require some re-consideration is the spatial map, and in particular the nature of the spatial map that is formed, as well as the time course of encoding for the spatial map. This is explored in more detail later.

Returning to the Visual Memory Model in its current state, it asserts that once an object file has been created and indexed to a position in the spatial map, the information associated with the object file is consolidated into long-term memory, forming a long-term memory object file. However, importantly, the model does not state when this consolidation occurs. The data from Experiment 1 suggest that consolidation of identity information happens during the second and subsequent fixations on an object. This is on the basis that the identities of objects fixated only once were remembered accurately only when the test was immediate, but not when the memory test took place after a 24 hour delay. However, the identities of objects fixated twice or more were remembered accurately no matter when the memory test occurred. Despite the model not specifying when this consolidation happens, it states that when overt visual attention is moved to another object within the scene, the initial short-term memory representation for the object decays and what remains in memory is the spatially indexed long-term memory object file. Evidence from the delayed memory test in Experiment 1 suggests that this long-term memory representation is robust enough to support accurate memory for that object even after a delay of 24 hours (and of course that any such long term representation for an object necessitated at least two fixations on it).

The final important stage of the current Visual Memory Model concerns the encoding of local object information across multiple fixations. The model does not define what "local object information" is, although one interpretation might be that "local" refers to information about the identity *and* the location of the object. What is made clear in the model is that this local object information can be encoded during more than one fixation on the same object. The model does not state whether these fixations need to be successive in order for object information to be successfully encoded; however the present data provide evidence that it is not important for fixations to be consecutive. In addition, the model does not state whether information about object identities and specific object locations is encoded similarly. The data from Experiment 1 seem to suggest that information about the identities of objects is encoded during at least two different fixations, whereas information about the specific location of an object is only encoded during the first fixation made on an object. It therefore seems apparent that while the Hollingworth and Henderson's Visual Memory Model adequately describes the basic processes associated with the encoding of objects in scenes, there are at least five important issues that it does not currently address:

1) The time course of encoding for the spatial map.
2) The nature of the spatial map.
3) Whether object identities and specific object locations are encoded in the same way.
4) The time course of encoding for object identities and object locations.
5) The time course of consolidation of information from short-term to long-term memory.

The data reported earlier in this thesis have provided some evidence to address issues (3), (4) and (5). Further discussion of the way in which the present data relate to those issues is presented later. What is clear, however, is that the present data do not speak directly to issues (1) and (2). Therefore what follows next is some consideration of relevant areas of literature that might relate to those issues.

5.1.2 *The time course of encoding for the spatial map: The link between the spatial map and scene gist*

The first issue that might require some more consideration is the *time course of encoding* for the spatial map. Based on Hollingworth and Henderson's model, it seems reasonable to assume that the spatial map is fully encoded *before* the first individual object file is created. The model clearly states that when the first fixation is made on an object in a scene, an object file is created and indexed to the spatial map. Therefore it seems to be a necessity that the spatial map is already fully formed by the end of the first fixation on the first object, otherwise the object file could not be indexed to anything. On the basis that the average fixation duration during normal scene viewing is approximately 330 ms (Henderson, 2003) one might suggest that the spatial map is formed somewhere in the region of 300 – 400 ms after the scene onset. Given this time course, that is, that the time to instantiate the spatial map is quite short, it is worth considering how the spatial map might be related to scene gist, as a substantial amount of research has demonstrated that the gist of a visual scene can also be encoded very early in scene viewing.

5.1.3 *Scene gist*

One such study on the acquisition of scene gist was carried out by Biederman, Mezzanotte and Rabinowitz (1982). In their Experiment 2, participants viewed 277 slides each for 150 ms. Before each slide was presented, participants were presented with the name of a target object and cued as to the location of that target object within the scene. Once the 150 ms presentation was over, participants had to state whether the target object was in a normal location, or had been violated in one of five possible ways. The five possible violations were: (a) the target object was not located on any surface and thus appeared to be floating in the air; (b) the target object overlapped with the background; (c) the target object was located in an unusual or unlikely context; (d) the target object was located in an unusual position within an expected context; (e) the target object was either too big or too small relative to other objects in the scene. The results showed that overall accuracy was 88% and the false alarm rate was 10.3%. It was not reported whether some types of violations were detected better than others. However, the reaction time data showed that targets located in unusual or unlikely contexts were detected fastest (~ 835 ms), and targets that overlapped with the background were detected slowest (~ 880 ms). As participants were able to accurately assess whether an object was in a normal location or not, it seems possible that some information about both the target object and some of the other objects in the scene had been encoded after only 150 ms. In turn, this information seems to be detailed enough to support accurate assessment of whether an object is in an expected location or if an object is the correct size.

Similarly, Potter (1976) showed participants sequences of visual scenes. In each sequence, each scene was presented once for 125, 167, 250 or 333 ms. Within each sequence participants were instructed to look out for a particular pre-defined scene and respond if they saw that scene in the sequence. Participants were cued to search for a particular scene either by seeing the scene before the sequence began, or by reading a description of the scene. The results showed that regardless of the cue type, accuracy was higher than 70% for scenes presented for 125 ms, and over 90% for scenes presented for 167, 250 and 333 ms. The false alarm rate was calculated to be less than 1%. These results demonstrate that participants can extract meaningful information about whether a scene fits a particular category during a single brief view of a scene.

More recently, Rousselet, Joubert and Fabre-Thorpe (2005) presented participants with coloured scenes each for 26 ms. Each scene was viewed only once. Scenes belonged to one of four categories: sea, mountain, indoor, or urban. The experiment was organised so that participants completed a number of blocks. Each block contained 96 trials. For each of the 96 trials, participants had to decide whether each image belonged to a specified category. For example, for the first 96 trials, participants had to decide whether each of the scenes belonged to the “sea” category or not. In the next block of 96 trials, participants had to decide whether each of the scenes belonged to the “mountain” category or not. The results showed that participants responded correctly over 95% of the time for all four categories, providing further evidence that the gist of a scene can be accurately extracted even if a scene is viewed for only 26 ms.

A number of other studies have demonstrated that the gist of a scene can be extracted quickly, typically within 100 ms (Metzger & Antes, 1983; Schyns & Oliva, 1994; Oliva & Schyns, 1997). Clearly there has been substantial research

on the speed at which the gist of a scene can be identified. It is important to note that in order to identify the gist of a scene it does not seem to be necessary for individual objects to be identified. Indeed, the speed at which images are presented in gist research typically do not allow more than a single fixation to be made indicating that the information required to extract scene gist is acquired within a single brief fixation (probably within 150 ms).
A dominant explanation for the speed at which gist is extracted is that participants use some previous knowledge about scenes (e.g., colour) to facilitate their categorisation of a scene based on gist (Oliva & Schyns, 2000; Goffaux, Jacques, Mouraux, Oliva, Schyns & Rossion, 2005; Castelhano & Henderson, 2008). For example, it has been suggested that participants can accurately identify the gist of natural scenes (e.g., mountains, fields, sea) because of the colours often associated with each (white, green and blue respectively). Such research typically asserts that the gist of a scene is encoded without the necessity for the identity of individual objects to be encoded thus explaining why the gist can be acquired even if the presentation duration is extremely short. An alternative explanation suggests that observers can identify the gist of a scene quickly by making decisions about the spatial layout of objects, and whether objects are congruent with the scene or not (Biederman, Mezzanotte & Rabinowitz, 1982). This explanation suggests that the identity of at least one individual object is encoded in order for a decision to be made as to its congruency with the scene.

As there seem to be differences in explaining how gist is acquired, it seems appropriate to briefly discuss the definition of scene gist as this also seems to vary within the literature. Rousselet, Joubert and Fabre-Thorpe (2005) used the word "category" as a synonym for gist, suggesting that gist refers simply to the type of scene that is portrayed (e.g., countryside). This is quite a broad definition of gist, referring to a general impression about the scene as a whole and seems compatible with an explanation that assumes that covert attention is paid to the whole scene simultaneously. In contrast, Castelhano and Henderson (2008) defined gist as decisions made about a scene on the basis of objects within the scene, and the layout of those objects within the scene. This definition of gist relies on processing of more detailed and fine-grained information, referring to specific objects within a scene. This perspective fits with the suggestion that overt attention may be paid to an individual object or objects to allow decisions about their scene congruency to be made.

Clearly, defining gist is difficult. However, despite vagueness associated with its definition, the suggestion that spatial layout plays a role in discerning the gist of a scene (Biederman, Mezzanotte & Rabinowitz, 1982; and see above) suggests that the spatial map and scene gist might be similar. Turning to the current experimental situation, there seem to be at least two ways in which scene gist and the spatial map might be similar. First, both gist of a scene and the spatial map seem to be formed early in scene viewing. Second, they both seem to be based on visually salient characteristics of the scene. However, despite these similarities, it is perhaps dangerous to assume that they are the same thing. Conceptually, it seems as though the spatial map is a psychological construct which plays an important role in the Visual Memory Model (used as a base to which object files are anchored). The gist of a scene, however, is a conceptual understanding about the meaning or context of a scene. Functionally, these may well be very different.

5.1.4 The nature of the spatial map: The link between the spatial map and saliency maps

Another issue that may be worth considering at this stage concerns the link between the formation of the spatial map, as stipulated in Hollingworth and Henderson's Visual Memory Model, and the formation of a saliency map used to guide saccadic orienting. This link relates to the *nature of the spatial map.* It is possible that the spatial map has two key functions:

(a) To act as a base to which individual object files can be linked
(b) To play a role in guidance of saccades around the scene (i.e. facilitating "where" decisions in relation to eye movements)

Regarding point (a) above, within the context of the present research, the spatial map might be conceptualised as an empty frame that contains 12 placeholders, or empty slots, each of which represents a location at which one of the objects in the photograph is positioned. It is proposed that initially, this spatial map contains the most basic of information about the scene (such as slot locations), and is used as a template to which more detailed information (such as the identities of objects) can be added as more fixations are made around the scene (see Figure 16).

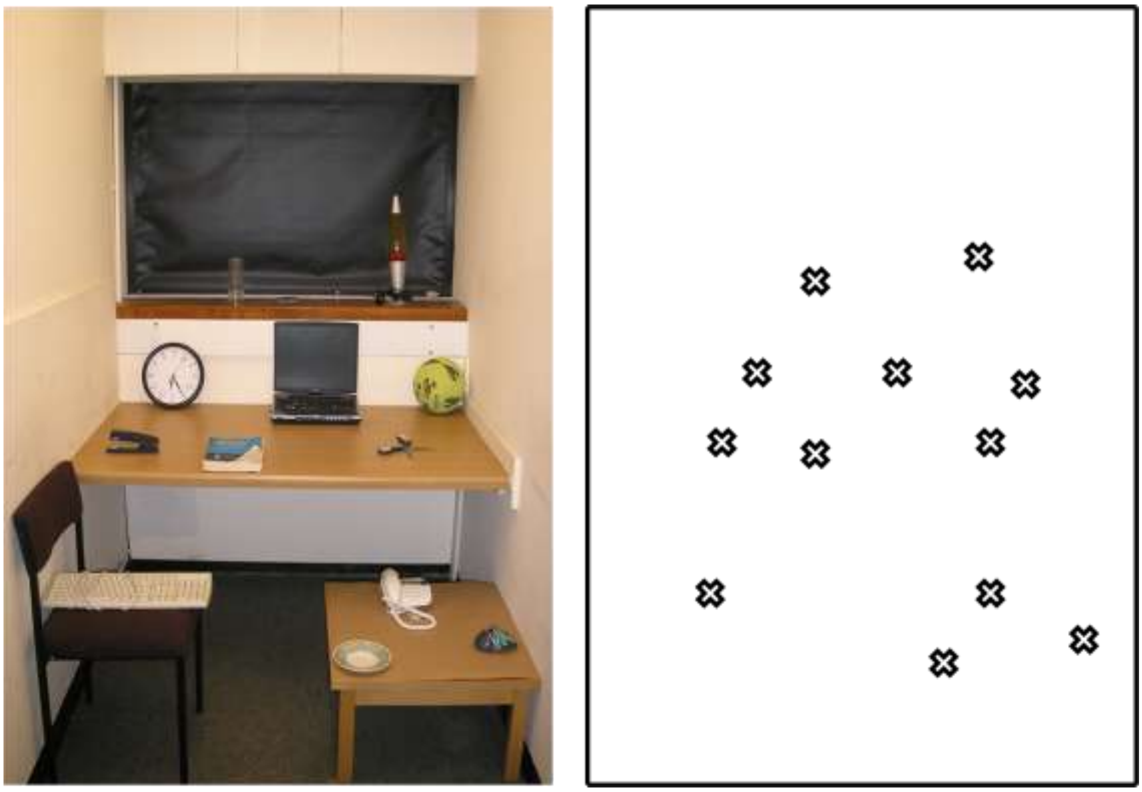

Figure 16. Panel a: The original stimulus photograph. Panel b: The spatial map corresponding to that photograph with slot locations for each of the objects.

It is likely, however, that the spatial map contains more detail about the scene than just a series of empty slot locations. For example, a key question is whether information about depth is captured in the spatial map. In principle, the spatial map may simply represent the layout of the scene two-dimensionally. If this were the case, then slot locations would essentially be represented as a series of x,y coordinates. Alternatively, depth information could be represented in the spatial map meaning the scene layout is represented three-dimensionally.

Although the present data do not speak directly to the issue of how depth information is represented in the spatial map, they do provide some evidence that suggests that depth information is represented within the spatial map. In Experiment 1, participants were able to accurately place objects, not only on the correct surface within the cubicle, but also in broadly the correct position on each surface. That is, participants were able to distinguish between objects which belonged towards the front of surfaces, and those which belonged towards the back of surfaces. Assuming that placements were made on the basis of the spatial map, then this strongly suggests that they had not formed a two-dimensional representation of the spatial layout of the scene (as in a series of x,y coordinates), but instead had encoded the positions of objects relative to other objects, and relative to the edges of the surfaces in depth. Turning now to point (b) noted above, it seems plausible that the spatial map is closely related to the notion of a saliency map. According to saliency map models (e.g., Mannan, Ruddock, & Wooding, 1996; Findlay & Walker; 1999; Reinagel & Zador, 1999; Parkhurst, Law & Niebur, 2002; Itti & Koch, 2000; 2001) some regions of scenes, or objects within scenes, possess high levels of intrinsic visual saliency on the basis of their inherent properties such as colour, luminance or contrast. Typically, this is because these regions or objects are markedly different from the surrounding environment or surrounding objects in one or more object property. Their individual (often) unique characteristics render them visually salient. Under free-viewing conditions (i.e. in the absence of any specific task instructions) regions or objects that are high in intrinsic visual salience are highly likely to be fixated, whereas regions or objects that are low in intrinsic visual salience are less likely to be fixated.

One might suggest that the spatial map reflects variability in saliency in a similar manner to saliency maps. As stated above, the spatial map is assumed to have a number of placeholders which represent the location of each of the objects in the scene, or areas of high saliency, which require direct fixation in order for detailed encoding. In the context of the present experiment, it seems likely that a spatial map is formed which contains 12 placeholders, one for each of the locations of the 12 objects. Participants are very likely to make saccades to locations in the spatial map which have placeholders, as these locations are associated with a visually salient object. In contrast, participants are unlikely to make fixations to regions of the spatial map which do not contain placeholders as these regions are unlikely to be associated with an object, and are therefore not very visually salient.

The important point to make here is that potentially the spatial map (within which object identity files are anchored) is closely linked to the saliency map (which guides the eyes around scenes). In other words, the saliency map drives the eyes to salient locations (which tend to contain objects) and any information about the object that is encoded at the salient location is linked to that respective position in the spatial map. Clearly, the spatial map and saliency map are not the same. Instead what is being suggested is that they seem to be similar and share some functional properties.

There is some evidence in the literature that the spatial layout of a scene can be used to guide oculomotor behaviour around salient aspects of that scene. For example, Chun and Jiang (1998) provided evidence that repeatedly presented spatial layouts facilitated search for a target compared to novel spatial layouts. In their study, participants were instructed to search for a target (the letter T rotated 90º to the left or the right) amongst a number of distractors (the letter L rotated 90º). Participants had to press a button when they found the target letter and state the orientation of the target. The critical dependent measure was response time. For each trial, participants either saw one of 12 possible "old" displays (in which the targets and distractors were in pre-defined and repeated locations) or a "new" display (in which the targets and distractors were in random locations). The results showed that response times were fasted for repeated displays than for novel displays. Chun and Jiang termed this the *contextual cueing effect.* This suggests that over repeated trials, participants learnt the spatial layouts of the "old" displays and used this information to guide their oculomotor behaviour. In turn this seems to support the notion discussed above that the spatial map and the saliency map might share some functional properties and may function in a similar fashion.

5.1.5 Are object identities and specific object locations encoded in the same way? What is the time course of encoding for object identities and object locations?

Hollingworth and Henderson's Visual Memory Model states that local object information is encoded during multiple fixations on an object. However, this account remains vague in defining what local object information consists of. One might speculate that local object information includes information about the identity of an object, as well as the specific location of that object. However, the data presented in this thesis strongly support the perspective that information about object identities and specific object locations are not encoded in the same way. Instead, it seems that information about the identity of objects is encoded across multiple fixations, with information about the specific locations of objects being encoded (and arguably reaching ceiling) during the first fixation on an object. Thus there appears to be a clear dissociation between the manner in which these two object properties are encoded and stored.

The data support the view that during normal scene viewing, overt visual attention (via fixations) is typically directed to one object at a time. Once an object has been fixated, an object file is created for that object. According to the Visual Memory Model, initially, this object file contains information about the basic identity of an object, such as the name of the object. At this point in scene viewing the spatial map could be conceptualised as a frame containing a number of empty placeholders and one placeholder containing an object file (see Figure 17).

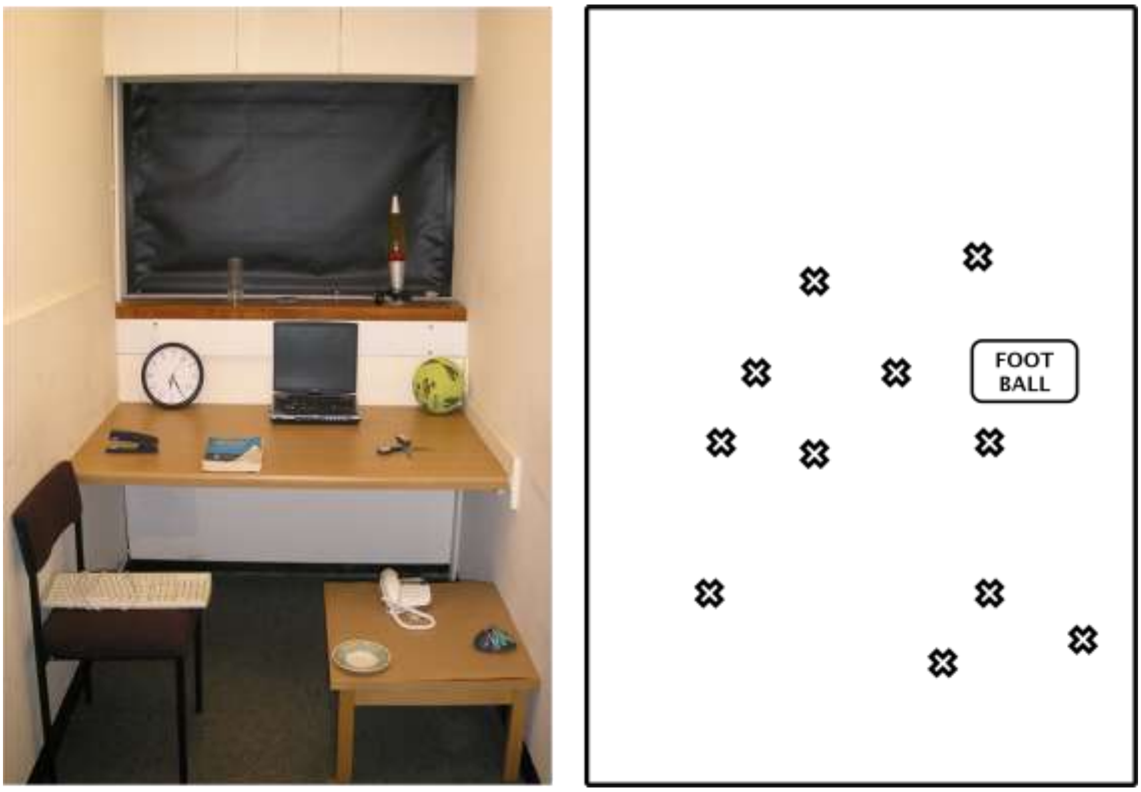

Figure 17. Panel a. The original stimulus photograph. Panel b. The spatial map corresponding to the photograph with one of the slot locations occupied by an object file for “football”.

As the number of fixations made on that object increases, more and more information about its identity is encoded. Depending on the complexity of the object and the degree of detail associated with its identity, several fixations may be made on the object. What is clear from the findings of Experiments 1, 2 and 3 is that information about the identity of an object is encoded over fixations (at least two fixations based on the present findings). In addition, the results thus far indicate that it does not matter whether these fixations are successive. Even if other fixations are made on other objects, once an object is re-fixated the object file for that object is accessed immediately and further information can be added to the information already stored in the object file. There is no sense of “starting again” (Melcher, 2001; 2006).

In contrast, the information about the specific location of that object seems to be encoded during the first fixation on that object and the accuracy with which the specific location of that object is remembered does not improve if more fixations are made on that object. It is proposed that this happens on the first fixation on an object because Experiments 1 and 2 have demonstrated that the objects fixated only once are placed as accurately as objects fixated twice, or three times or more. If additional fixations are made on an object, the accuracy with which an object is placed does not appear to improve. This is implied in the current Visual Memory Model, but it is not made explicit. To reiterate, the present data provide clear evidence that the identities and locations of objects are encoded differently. Specifically, the identities of objects are encoded over multiple fixations, embellishing the degree of detail in the object file relating to an object’s identity. However, the specific location that an object occupies is encoded during the first fixation on an object. Furthermore, if an object is fixated on several occasions, the object file accumulates more detail regarding object identity. Instead of simply containing abstract information about the meaning of the object, the object file can now store highly detailed and object-specific information about the identity of that object such as, say, the size, shape and type of object etc (see Figure 18).

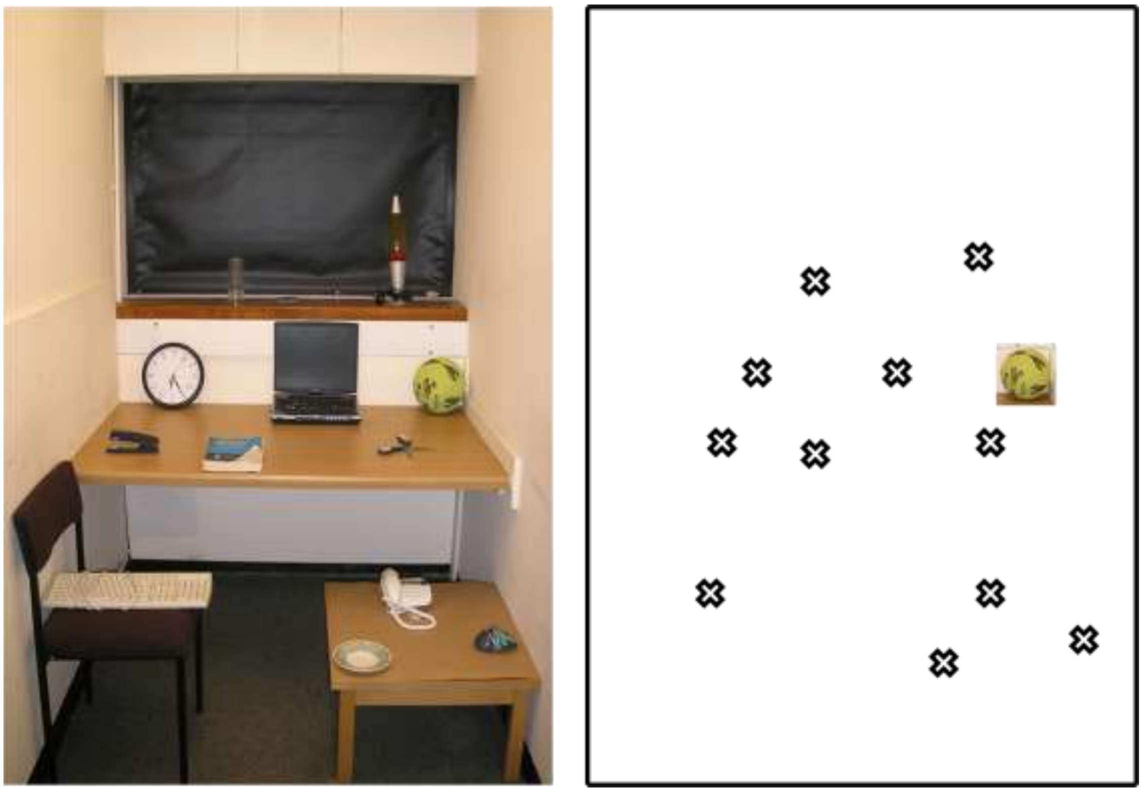

Figure 18. Panel a: The original stimulus photograph. Panel b: The spatial map corresponding to that photograph with one of the slot locations occupied by an object file containing identity detail.

An interesting related issue concerns how this information is stored in memory (i.e., the nature of the representation). In the broadest sense, memory representations can be considered to be abstract and stored in some form of system such as a semantic network (e.g., Collins & Quillian, 1969). Figure 19 illustrates how the actual object may relate to a more abstract representation of the object.

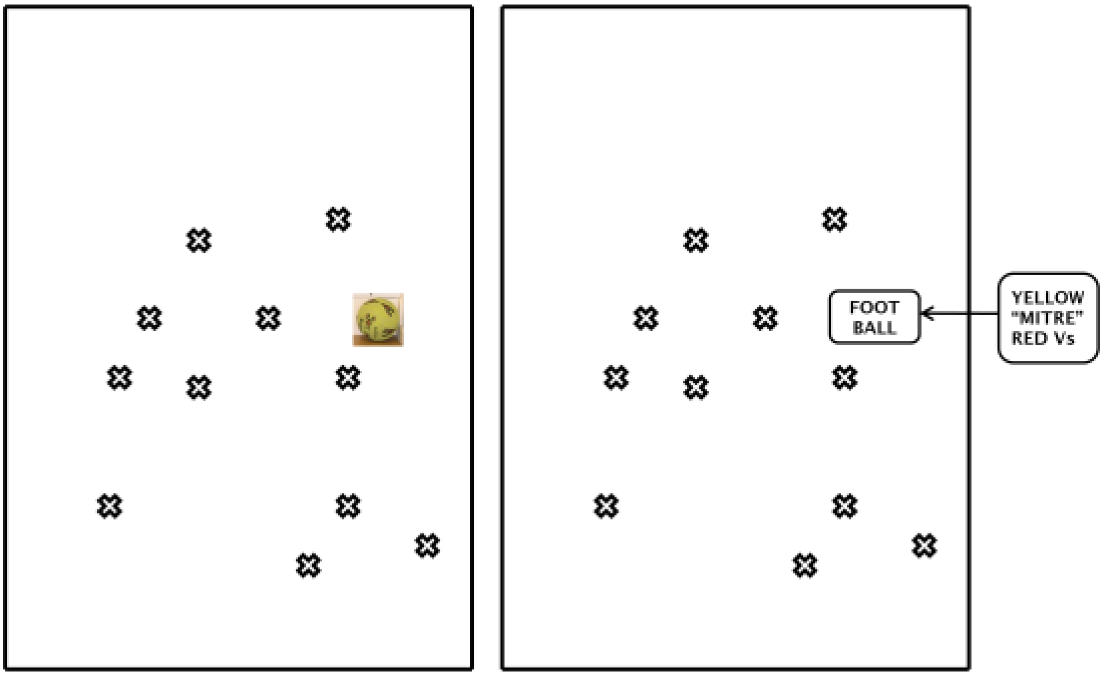

Figure 19. Panel a: The spatial map with an image of an actual object linked to a location. Panel b: The spatial map conceptualised as a semantic network with a node representing an object.

The semantic network perspective states that information is represented in memory in an abstract way, with information stored in an inter-connected structure of nodes. Each node represents a single entity and each node has links connecting it to many other nodes (see Figure 20). Collins and Quillian used the example of the *canary* and suggested that if *canary* could be characterised as *a yellow bird that can sing* then *canary* would be linked to the nodes, amongst others, *bird, yellow,* and *can sing.* They also stated that this information was arranged hierarchically with three levels. According to their hypothetical model, information would be represented in broad categories at the first level (e.g., *animal*), followed then by types of animal at the second level (e.g., *bird, fish*) and then, at the third level, specific animals themselves (e.g., *canary, robin*). If this is true, then there would be certain characteristics associated with canaries (e.g., they fly, they have wings) that would not need to be linked directly to *canary.* Instead, those characteristics would be linked to *bird* as these characteristics are common for most birds.

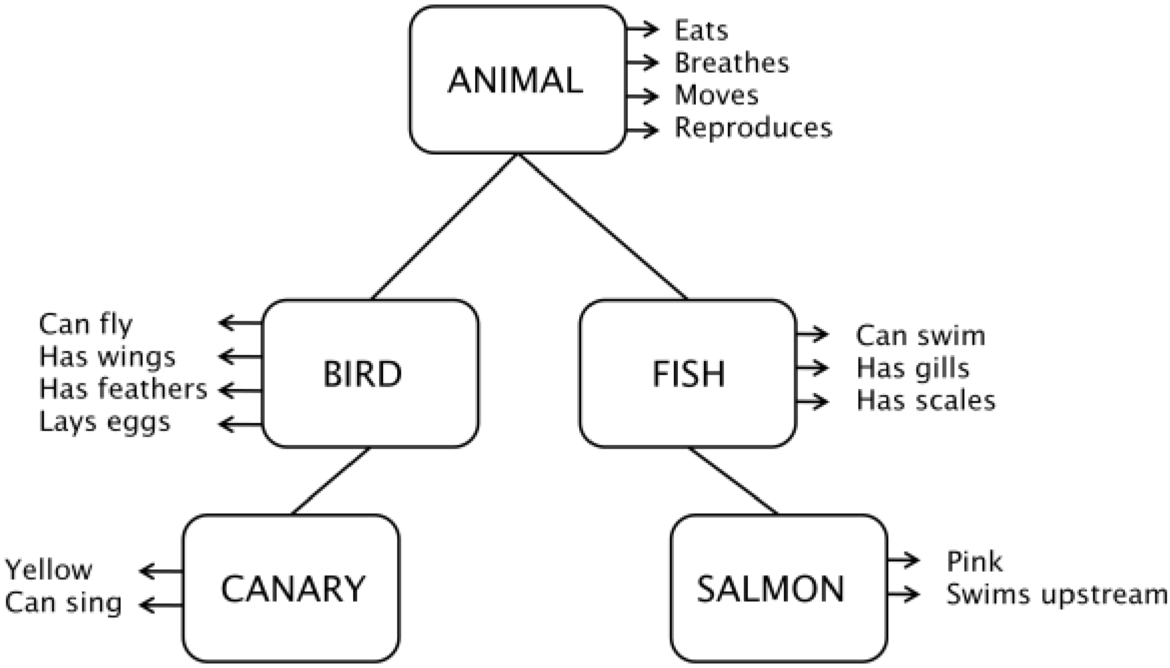

Figure 20. Illustration of the hierarchical semantic network model of memory representation

Regardless of the specifics of the Collins and Quillian's semantic network model, the important point to note is that an object representation is comprised of a set of defining features. In the context of the present research, this semantic network approach would state that each object that is fixated within a scene, and in fact, the specific details associated with that object are encoded over separate fixations. In this way, a more detailed abstract representation of the object could be encoded over fixations. Thus, when an object file is created (e.g., *football*) the defining characteristics of that object in the stimulus photograph are acquired and linked to it via fixations. Such defining features could be represented within some form of semantic network (e.g., see Figure 21). The fact that the ball is round is not linked to the node, as "being round" is not unique to this particular ball. However, what is somewhat unique to this ball, as opposed to other balls, is that it is yellow, manufactured by "Mitre" and has red Vs on its surface. Therefore, these defining characteristics may be linked to the node for that object. The point to take from this discussion is that it seems that theories of how information about objects and their identities is stored in memory fit quite neatly with theorising based on the VMM. Models specifying how information about objects is represented and stored in memory stipulate that they are represented as nodes associated with defining features. In a similar way, the

VMM specifies that object files are instantiated and then embellished through successive acquisition of defining characteristics across fixations. Clearly, the notion of a node within a semantic network and an object file and associated defining characteristics, are conceptually, quite similar.

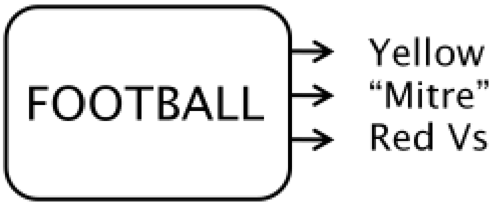

Figure 21. Illustration of a node for "football" with its defining characteristics.

5.1.6 When does consolidation of information from short-term to long-term memory happen?

The final issue raised in relation to Hollingworth and Henderson's Visual Memory Model that the present experimental data speak to concerns the consolidation of information from short-term to long-term memory. The Visual Memory Model states that information is consolidated into long-term memory, but this account does not specify when this takes place. The data from Experiment 1 suggest that consolidation of identity information happens during the second fixation on an object. This is on the basis that the identities of objects fixated only once were remembered accurately only when the test was immediate, but not when the memory test took place after a 24 hour delay. However, the identities of objects fixated twice or more were remembered accurately no matter when the memory test occurred. From a theoretical perspective, this suggests that when overt visual attention is moved to another object within the scene, the initial short-term memory representation for the object decays and what remains in memory is the spatially indexed long-term memory object file, but this only happens if an object has been fixated at least twice. This long-term memory representation is robust enough to support accurate memory for that object even after a delay of 24 hours.

5.1.7 *Summary of the Visual Memory Model and its relationship with the current data*

At this point it seems appropriate to summarise Hollingworth and Henderson's Visual Memory Model in light of the experimental data considered thus far. Whilst the data presented in Experiments 1, 2 and 3 do not speak directly to the spatial map, considerable discussion has taken place as to the nature of the spatial map and the time course of its instantiation. The present data do speak to the way in which object identities and specific object locations are encoded, and also relate to how information is consolidated from short-term to long-term memory. In particular, the data have provided evidence that information about the identities of objects is encoded during multiple fixations, but that information about the specific locations of objects is encoded during only the first fixation on an object. Clearly, the current data allow for extension of the current understanding of how different aspects of the Visual Memory Model might function. This issue will be returned to in the final chapter of the thesis.

An issue that the experimentation reported thus far has not addressed, however, is the way in which information about the overall configuration of objects is encoded over fixations. It has been suggested above that information about the overall configuration of objects (measured via the best fit error, and thought to represent memory for the spatial map) is encoded very early in scene viewing and without the need for each individual object to be overtly attended to via fixation. However, the relationship between encoding and recall for the overall configuration of objects has not yet been formally explored.

This is an important relationship to investigate because there is evidence that suggests that the way in which information about the *specific locations of objects* and information about the *overall configuration of objects* is encoded, may occur differently. Experiment 4 was therefore designed to directly investigate how information about the overall configuration of objects is encoded.

5.1.8 Evidence for a distinction between the overall configuration of objects and the specific locations of objects

The next section of this chapter considers the literature that is relevant to the issue of how the overall configuration of objects is encoded across fixations. Postma and de Haan (1996) investigated whether there were differences in memory for identities of objects, the specific locations of objects, and the overall configuration of objects using a computer-based task. As such, their experiment is very relevant to the current experiment as the present study investigates whether information about the overall configuration of objects and information about the specific locations of objects are encoded differently. Postma and de Haan provided evidence which suggested that memory for the overall configuration of objects (the spatial map in the Visual Memory Model) and memory for the specific location occupied by an object (the process of indexing an object file to a position within the spatial map) are distinct from one another. Postma and de Haan presented participants with computer-based arrays of either identical stimuli (e.g., c c c), different letters (e.g., c d e), or different nonsense letters (e.g., ي ط الله). In their study, Postma and de Haan actually used Hebrew letters with a "c" superimposed to create their nonsense letters. Additionally, participants viewed set sizes of 4, 7, or 10 objects. For half of the trials, participants had to count backwards while looking at the stimuli (articulatory suppression condition). On the other half of the trials, participants were silent.

At test, participants were presented with an empty array and the stimuli listed above the array. Participants had to use the computer mouse to move the stimuli to the locations within the array that they remembered them being positioned. At this point it is important to draw a distinction between requiring participants to place identical letters (c c c) and requiring participants to place different letters (either c d e or ي ط الله). In order for participants to place the identical letters they only had to remember which locations had been filled. In other words, participants only needed to recall the overall spatial configuration of objects. In the context of this thesis, this is analogous to just remembering which locations were filled (i.e. the overall configuration of the objects). However, in order for participants to accurately place the different letters, not only did participants have to remember the overall spatial configuration of objects, but also remember which of the letters belonged in each location. In this thesis, this is analogous to remembering which object was positioned at each of the locations, as well as remembering which locations were filled. Clearly, for Postma and de Haan's participants, relocating the different letters was a more difficult task than relocating the identical letters as it required memory not just for the overall configuration of objects, but also the link between the identity of each object and each location. Based on this logic, the results were unsurprising. Participants demonstrated the smallest average error when placing the identical letters. In the context of the Visual Memory Model, participants only needed to remember the spatial map in its un-indexed form (i.e., without any object identities indexed to the candidate positions). As all of the letters were identical, it did not matter which object was placed in each location; only that the overall configuration was remembered.

Error was higher for the different letters and highest for the nonsense letters. Postma and de Haan suggested that this was because the nonsense letters were more difficult to label and remember than the different letters, leading to confusion about which nonsense letter belonged in each location. Also, the results showed that articulatory suppression led to poorer location memory for the nonsense letters and different letters, but had no effect on location memory for the identical letters.

Based on these results, Postma and de Haan (1996) suggested that encoding the overall spatial configuration is a relatively simple task which requires fewer attentional resources than placing specific objects in specific locations. As a result, the articulatory suppression led to reduced accuracy in the complex task (binding specific objects to specific locations) but had no effect on the more simple task (remembering the overall spatial configuration). Postma and de Haan stated that articulatory suppression was employed in order to disrupt working memory. However, they did not specify whether the effect of articulatory suppression was due to inhibition of potential verbal encoding of information (such as repeating the name of a letter or labelling the nonsense letters) or because it engaged working memory. However, regardless of the mechanism of the articulatory suppression, the findings of Postma and de Haan seem to provide evidence that information about the overall spatial configuration of objects and information about the specific locations of objects seem to be encoded differently. Memory for the overall spatial configuration of objects is very accurate and is not affected by articulatory suppression. Conversely, memory for the specific locations of objects is less accurate and is negatively affected by articulatory suppression. However, Postman and de Haan did not record eye movements in their experiment and therefore did not investigate the mechanisms by which each of the different types of location information were encoded.

The suggestion that memory for specific locations of objects and the overall configuration of objects are different was further developed in a later article by Postma, Kessels and van Asselen (2008). They posited a neuro-cognitive theory which asserted that memory for objects in scenes can be split into at least three separate processes: (a) object identity processing, (b) spatial-location processing and (c) object-to-location binding. However, the main focus of the theory was the brain areas which each process was associated with, rather than the encoding mechanisms which are linked to each type of memory.

5.1.9 Evidence for the overall configuration of objects being encoded automatically

A number of other studies have also demonstrated that memory for the overall spatial configuration of objects seems to be different to memory for the specific locations of objects, and that the overall spatial configuration of objects may be encoded without the need for overt visual attention to be allocated via fixations on each of the objects in the array. All of these studies are relevant to the present study as they investigated whether memory for overall configurations of objects and memory for specific object locations may be different. Simons (1996) had participants view 120 different arrays containing five objects for 2s each. Each array contained five novel shapes that were difficult to label. After a delay of 4.3 s, participants were presented with a test array which was either identical, or was changed in one of the following ways: the spatial configuration was changed (layout-change); two of the objects had switched locations (position-switch); a new object had replaced one of the old objects (identity-switch). Participants had to state whether this test array was the same as the original array, or different in some way and if so, in what way it was different. In terms of the Visual Memory Model, a layout change was analogous to the spatial map (or the overall configuration of objects) whilst a position-switch change was analogous to memory for the binding of objects to their specific location.

The results showed that change detection accuracy for the layout-change test arrays was almost at ceiling. This demonstrates that participants can very accurately remember the overall configuration of objects even if a display was only viewed for 2 s. Accuracy for the position-switch test arrays (which probed memory for the specific locations of objects) was significantly poorer than that for layouts, although it was significantly better than chance (50%). This provides further evidence that memory for the overall configuration of objects and memory for the specific locations of objects are different, and typically participants possess more accurate memory for the overall configuration of objects than for the specific locations of objects.

Simons reported that even though novel and difficult-to-name stimuli were used in the displays, participants reported trying to name the stimuli in order to make the test easier. This was possibly facilitated by each object being presented approximately 9 times, allowing participants to repeatedly view these novel shapes and give them a label. Nonetheless, these results demonstrate that when participants encode a scene containing different objects, accuracy is higher for the overall spatial layout of the scene than it is for the specific locations of objects within that scene, although memory for the specific locations is better than chance.

In Experiment 3 (which tested different participants), Simons ensured that each novel shape was seen no more than 3 times, making the novel shapes more difficult to label. As a result, accuracy for the position-switch trials was no better than chance, but accuracy was at ceiling for the layout trials. This provided evidence that information about the overall spatial configuration may be encoded somewhat automatically and without the need for participants to explicitly label different parts of the scene. It also suggested that making it more difficult to label objects (by reducing the amount of times each object was seen) thereby leading to poorer memory for the specific locations of objects meant that the specific locations of objects are encoded somewhat less automatically than the overall configuration of objects.

Crucially, eye movements were not recorded and therefore the manner in which the information was encoded could not be explored. Despite this, the finding that accuracy for the overall spatial configuration of objects was at ceiling in both Experiment 2 and 3 despite the truncated viewing duration in both experiments (2 s) and the reduction in the amount of trials each display was seen for (from 9 times in Experiment 2 to 3 times in Experiment 3) demonstrates that information about the overall configuration of objects is encoded rapidly and without the need for effortful encoding. However, information about the specific locations of objects seems to require more effortful encoding.

A similar finding was demonstrated by Aginsky and Tarr (2000). They used a flicker paradigm to probe change detection for colour changes and changes of position (location, or appearance/disappearance) of an object. In terms of the Visual Memory Model, a change made to the colour of an object could be considered analogous to information about the identity of an object whereas a change in position for an object (either by virtue of an object moving to a novel location, or a new object appearing) could be considered analogous to the spatial map. On each trial, the original image was presented for 240 ms followed by a blank grey screen for 80 ms. The original image (A) was presented twice in succession, followed by the modified image (B) twice. Thus, the sequence of images in each trial was A, blank, A, blank, B, blank, B, blank etc until the participant responded. In addition, participants were randomly assigned to either the cued or non-cued condition. In the cued condition, participants were informed prior to the start of each trial as to the nature of the change (colour, location, or disappearance/appearance). In the uncued condition, no such information was provided.

The results showed that colour changes were detected faster in the cued condition than in the uncued condition. However, location changes and appearance/disappearance changes were detected equally quickly in both the cued and uncued conditions. These results suggest that information about colour in a scene is encoded less readily than information about the overall configuration of objects. Importantly, this fits with the suggestion of the Visual Memory Model that information about the overall configuration of objects is encoded early and rapidly, whilst information about the specific properties of an object is encoded later and more slowly. This distinction between encoding for overall spatial layout of objects being accurate and rapid, but encoding for object properties (such as colour) being slower and more effortful has also been demonstrated by Jiang, Olson, and Chun (2000), and Alvarez and Oliva (2007).

Hasher and Zacks (1979) proposed a framework which aimed to explain the mechanisms involved in a broad range of memory phenomena. One of their claims was that memory tasks fall somewhere along a continuum. At one end of this continuum are tasks that are automatic; that is, information that can be encoded without the need for direct attention. At the other end of the continuum are tasks which are effortful. These tasks require direct attention. Hasher and Zacks suggested that information about spatial locations is one such aspect of a scene that is encoded automatically. Hasher and Zacks provided two possible explanations for information about spatial layouts being encoded automatically: (i) heredity and (ii) practice. They suggested that the locations of things in the visual environment might be encoded automatically because it is a necessary part of successful function in the world. They also suggested that some types of information about the environment (e.g., the overall configuration of objects) could, with practice, eventually be encoded automatically. However, despite reporting results from several experiments which probed different types of memory tasks, they did not report data from any tasks probing spatial memory. Their claim that spatial information is encoded automatically was not unique; earlier research had provided evidence that supported this. For example, there is evidence that participants can accurately remember the locations of stimuli even if they are not informed prior to the experiment that they will be tested on the locations of those stimuli (Zechmeister, McKillip, Pasko, & Bespalec, 1975; von Wright, Gebhard, & Karttunen, 1975; Mandler, Seegmiller, & Day, 1977).

Therefore, a whole host of evidence suggests that information about the overall configuration of objects (or stimuli) is encoded somewhat automatically and therefore without the need for direct fixations to be made on every portion of the scene. This finding is consistent with the claims of the modified Visual Memory Model that the spatial map is formed early in scene viewing. In addition, other studies have indicated that there is a difference between the specific locations of objects, and the overall configuration of objects (Postma & de Haan, 1996; Postma, Kessels, & van Asselen, 2008). This dissociation has also been observed in Experiment 1 in this thesis. Therefore, it seems possible that information about the overall configuration of objects and the specific locations of objects may be encoded differently. To be clear, the existing literature seems to suggest that information about the overall configuration of objects may be encoded automatically, but information about the specific locations of objects (i.e. binding object identities to available locations within the spatial map) is not encoded automatically and requires effortful processing. To investigate this, the test phase of Experiment 4 was modified to allow investigation into the relationship between eye movements during encoding and memory for both the overall configuration of objects, and the specific locations of objects.

5.1.10 The present experiment

The encoding phase in Experiment 4 was identical to the encoding phase used in Experiment 1, except that in Experiment 4 all participants viewed the stimulus photograph for 10s and all participants were tested immediately. However, the test phase was modified in the following way: participants first placed 12 plastic blocks in the cubicle to demonstrate their memory for the overall locations of objects. This allowed assessment of the degree to which participants remembered which locations were filled by objects in the stimulus photograph without participants needing to remember which objects had been present at each location. Once all 12 blocks had been placed, participants were given the 12 correct objects and told to replace each block with one of the objects. It is important to note that participants were not required to select objects from the pool of 24 objects as in Experiments 1-3. The reason for this was that Experiment 4 was primarily focussed on investigating memory for the overall configuration of objects and the specific locations of objects, rather than the identities of objects.

The predictions were based on the modified Visual Memory Model presented above, the findings of Experiments 1 and 2 in this thesis, and the literature reported earlier. The first prediction was that the overall characteristics of saccades and fixations would be comparable between the block-placing participants in this experiment and the free-viewing control participants taken from Experiment 3 (used as a control group). This is because both groups of participants experienced identical testing phases on an identical eye tracker. Therefore it was expected that the encoding behaviour of Experiment 4 would directly replicate the encoding behaviour of the free-viewing participants in Experiment 3.

The second prediction related to the test phase. It was predicted that the best fit error and the placement error would be very similar for both block-placing participants and the control participants. This is on the basis that both groups of participants experienced identical viewing durations (10s) and neither group were informed as to the nature of the memory test. Instead, both groups were simply told that a memory test on the contents of the photograph would occur immediately after the encoding phase began.

The next prediction related to the way in which locations are encoded. According to the modified Visual Memory Model, the overall configuration of objects is encoded very early during scene viewing, resulting in a spatial map being formed that contains all candidate locations. If this is true, then presumably some (or maybe all) of the locations are encoded without the need for each location to be directly encoded. On that basis, it was predicted that memory for the overall location of an object would be accurate even if that location had not been encoded. In other words, something (although not necessarily the *correct* object) would be placed close to the location of an object even if that object had not been fixated. In addition, it was predicted that this overall location accuracy would not improve with increasing number of fixations. Instead, information about the overall configuration of objects would be encoded very quickly, and would not improve with subsequent fixations.

Memory for the specific locations of objects, however, was predicted to be poor for non-fixated objects, but accurate for fixated objects. However, as in Experiment 1 in this thesis, it was predicted that once the specific location of an object had been encoded, it would not substantially improve with more fixations. This is on the basis that once an object identity has been indexed to a position within the spatial map, that accuracy of that specific location does not improve.

The prediction that information about the overall configuration of objects would have a different pattern of encoding compared to the specific locations of objects was directly based on the framework of the modified Visual Memory Model, on the findings of Postma and de Haan (1996) and Postma, Kessels and van Asselen (2008) who stated that there is a fundamental difference in encoding for the specific locations of objects and the overall configuration of objects. It is also based on the findings that information about the overall spatial layout of objects is encoded somewhat automatically (Zechmeister, McKillip, Pasko, & Bespalec, 1975; von Wright, Gebhard, & Karttunen, 1975; Mandler, Seegmiller, & Day, 1977; Hasher & Zacks, 1979; Simons, 1996, Aginsky & Tarr, 2000, Jiang, Olson, & Chun 2000; Alvarez & Oliva, 2007). Finally, it was predicted that confidence for the overall configuration of objects and the specific locations of objects would closely match the actual memory performance for both. This prediction was based on the findings of Shaffer and Shiffrin (1972) who found that confidence for previously viewed scenes closely matched memory accuracy for the same scenes, and the findings from Experiments 2 and 3 in this thesis that confidence and memory accuracy tend to closely mirror each other.

5.2 Method

Participants and design

Participants were undergraduate and postgraduate students from the University of Southampton who took part in exchange for course credit. There were 19 participants in total of which 7 were male and 12 were female. The average age was 21.21 years (*SD* = 1.69 years) and the ages ranged from 18 to 24 years. The data from 8 participants was removed due to tracker loss during the encoding phase leaving data from 11 participants. A between-participants design was used. Control data was taken from the 10 participants in Experiment 3 who viewed the stimulus photograph for 10 s in the free-viewing condition as both sets of participants were tested on the same eye tracker and experienced identical encoding phases.

Apparatus

The apparatus used in the encoding phase was identical to Experiment 3. The test phase took place in the same research cubicle as in Experiments 1-3. In the test phase, participants were first given 12 identical yellow plastic blocks to place. Only the 12 correct objects were present in the room and were presented to participants in the second half of the test phase. The 12 distractor objects were not used in this experiment.

Stimuli

The stimulus used was identical to that used in Experiment 1 (see Chapter 2).

Procedure

The encoding phase was identical to that of the 10s/0 delay condition of Experiment 1 and the free-viewing condition of Experiment 3. Participants were forewarned that they would see a photograph of a room for a single 10s viewing duration and that a memory test would take place immediately afterwards. A 10 s encoding duration was used to try to yield enough data in the 2 and 3+ fixations bins to allow analysis on the rate at which object identity and location information was encoded across multiple fixations.

As soon as the encoding phase had finished, participants were taken to the test cubicle and the test phase began. This test phase was different to Experiments 1-3. All of the objects had been removed from sight, leaving the room and the four surfaces (chair, small table, desk and shelf) empty. There were two distinct stages to the test phase. First, participants were given 12 numbered plastic blocks and instructed to place the blocks in the locations that they remembered being filled by objects in the photograph. They were instructed to guess locations if they were unsure where some objects had been located. Once all 12 blocks had been placed, participants gave a confidence rating from 0 – 10 for each block which represented their confidence that an object had been located in each of the locations now filled by a block.

Once all 12 confidence ratings had been provided, the second stage to the test phase began. The 12 objects that had been present in the photograph were presented to the participant and they were asked to replace each of the plastic blocks with one of the objects. As participants were replacing blocks with objects, they were allowed to place the objects in a new location if they wished. Once participants had replaced each block with an object, they were asked to give a confidence rating for each object that represented how confident they were that each object was now in the correct location. Participants were not required to select objects, nor were they asked for their confidence regarding the identity of each object.

5.3 Results

As in Experiments 1-3, all fixations shorter than 80 ms or longer than 1200 ms were not included in the analyses.

Comparison of Eye Movement Behaviour between Experiment 4 and the Free-Viewing Condition of Experiment 3

The first analyses compared the basic oculomotor behaviour of participants in Experiment 4 and a control condition taken from the free-

Table 18

Means and Standard Errors (in Parentheses) of Eye Movement Measures for the Block-placing Participants in Experiment 4 and the Control Condition from Experiment 3 in the Encoding Phase

Condition:	Free-viewing condition (from Experiment 3)		Block-placing condition (from Experiment 4)	
Measure	M	SE	M	SE
Total time fixating objects (ms)	7332	(334)	7999	(160)
Total time fixating blank (ms)	818	(198)	603	(94)
Fixations made on blank regions	3.70	(0.84)	2.82	(0.42)
Average blank region fixation duration	203	(29)	219	(17)
Proportion of viewing duration spent fixating objects	0.90	(0.02)	0.93	(0.01)
Fixations made on objects	25.30	(1.35)	27.27	(0.92)
Fixations made in total	29.00	(0.92)	30.09	(0.89)
Number of objects fixated	10.60	(0.54)	10.91	(0.34)
Average fixation duration (ms)	292	(8)	296	(10)
Average gaze duration (ms)	469	(47)	428	(25)
Average number of fixations made during gaze duration	1.54	(0.11)	1.49	(0.10)
Average total time (ms)	701	(34)	733	(39)
Number of fixations made during total time	2.41	(0.12)	2.48	(0.10)

viewing condition in Experiment 3. These participants experienced identical task instructions and encoding phases and were tested on the same eye tracker. Therefore it was expected that the eye movements would be similar between the two groups. If this was true, then Experiment 4 would be shown to replicate the encoding phase of Experiment 3. A series of between-groups *t*-tests on the eye movement variables found no significant differences between the groups on any of the encoding measures (see Table 17 above).

For some of the *t*-tests reported below, the Levene's test for equality of variances was significant. This means that the variances of the two groups were not equal and as a result, the degrees of freedom associated with the test were not N – 2 but instead were decimal numbers. There was no significant difference for total time spent fixating objects, $t(19) = -1.86$, $p > .05$; for total time fixating blank regions, $t(12.88) = .98$, $p > .05$; for the average number of fixations made on blank regions, $t(19) = .96$, $p > .05$; for the average blank region fixation duration, $t(19) = -.49$, $p > .05$; the proportion of viewing duration spent fixating objects, $t(12.08) = -1.24$, $p > .05$; the number of fixations made on objects, $t(19) = -1.23$, $p > .05$; the number of fixations made in total, $t(19) = -.85$, $p > .05$; the number of objects fixated, $t(15.42) = -.48$, $p > .05$; the average fixation duration, $t(19) = .38$, $p > .05$; the average gaze duration, $t(19) = .80$, $p > .05$; the average number of fixations made during gaze duration, $t(19) = .38$, $p > .05$; the average total time, $t(19) = -.60$, $p > .05$; and the average number of fixations made during total time, $(19) = -.46$, $p > .05$.

As both conditions were identical in terms of the task instructions they received, it is entirely unsurprising that the eye movement behaviour was comparable between the two groups. This finding is important as it shows that the overall encoding behaviour in Experiment 4 directly replicates the overall encoding behaviour of Experiment 3.

Comparison of Memory Test Performance between Experiment 4 and the Free-Viewing Condition of Experiment 3

The memory test undertaken by participants in Experiment 4 was slightly modified to that of the original test phase experienced by participants in Experiments 1, 2, and 3. In this experiment, participants were not required to select objects from a pool of 24 objects. Instead, the focus of the test phase was on memory for the overall locations of objects, and the specific locations of objects, rather than the identity of objects. Therefore, the test phase yielded two relevant memory measures, both of which relate to memory for object locations. The measures were best fit error (which represented the overall configuration of objects) and object placement error (which represented the specific location of objects).

The first task that participants completed was to place the 12 plastic blocks in the locations they remembered objects had been located in the stimulus photograph. The second task was to replace each of the plastic blocks with one of the objects. Both the positions of the blocks and the positions of the objects were noted, as were any instances where participants decided to change the location of an object when replacing the block. On these occasions, the location of the block was noted, as was the new location of the object. However, of all the 132 block-object replacements that occurred (12 blocks for each of the 11 participants), only 5 objects were placed in a new location (3.79%).

For the purpose of the analyses, both the best fit error and average placement error were calculated. The best fit error was calculated for all 12 plastic blocks, and average placement error was calculated for all 12 object placements. In line with the predictions, there was no difference in best fit error between the block-placing participants and the control participants, $t(19) = 1.89$, $p > .05$. Additionally, there was a significant negative correlation between the order in which blocks were placed and the confidence for that location, $r(132) = -.52$, $p < .001$ with participants placing blocks in the locations for which they were most confident first, and least confident last. This suggests that participants had some meta-cognitive awareness of the memory they possessed for the overall spatial configuration of objects, and subsequently chose to place blocks in the locations they were most confident about first, leaving the locations they were least confident about until the end of the experiment.

The second finding was also in line with the predictions. There was no difference between the block-placing group and the control participants in the accuracy with which objects were placed to their correct home location, $t(19) = 0.75$, $p > .05$. As stated above, participants were not required to select the 12 objects from the photograph from a pool of 24 objects. Instead, they were given the 12 correct objects and asked to replace each block with an object. Recall that participants were also asked how confident they were that each object was now in the correct location. There was a significant negative correlation between the replacement order and object confidence, $r(132) = -.69$, $p < .001$ with confidence higher for earlier replacements and lower for later replacements. As before, this provides evidence that participants chose to replace blocks with objects in an order which reflected their awareness of their memory.

Table 19

Means and Standard Errors of Recall Measures in the Test Phase for the Block-placing Participants in Experiment 4 and the Control Condition taken from Experiment 3

Condition:	Free-viewing condition (from Experiment 3)		Block-placing condition (from Experiment 4)	
Measure	M	SE	M	SE
Best fit error (cm)	24.06	(2.25)	18.44	(1.96)
Object placement error (cm)	38.28	(4.02)	34.01	(3.98)

Relationship between the Number of Fixations Made on an Object and Memory for the Overall Configuration of Objects

The next analysis considered the relationship between encoding and the best fit error. Best fit error was calculated in the same general way as in Experiments 1-3, except that the average distance that each of the plastic blocks was placed from the 12 home locations was calculated for each participant. In order for the relationship analysis to take place, each of the plastic blocks needed to be assigned to one of the objects. The way in which this was done was to assign each plastic block to the object that the participant chose to replace it with. Recall that participants had to place the blocks in numerical order such that the first block they placed was block number 1 and the last block they placed was block number 12. During the second half of the test phase when participants replaced the plastic blocks with the correct objects, each of the blocks was assigned to the object that replaced it. Therefore, if participants chose to replace block number 1 with the mouse, then any encoding that the mouse received during viewing of the photograph was assigned to block number 1.

Objects were categorised as receiving zero fixations (i.e., being ignored), one fixation, two fixations, or three or more fixations during encoding. As a result of the block-to-object assignment described above, each of the blocks was also categorised in the same way. Then the distance between each of the plastic blocks and its nearest home location was calculated. This gave a measure of how the accuracy with which a block was placed near to any of the candidate locations in the spatial map was influenced by the amount of encoding that object had received.

However, four of the 11 participants fixated all 12 of the objects in the photograph, meaning that they had no data for the non-fixated objects. To fill these blank cells (9.09% of the data set), a linear regression was carried out to obtain an appropriate estimation of error for non-fixated objects. For best fit error, the regression was not significant and yielded an equation of $y = 0.04X + 18.36$. As the regression was not significant it was deemed inappropriate to use the intercept to replace the missing data. It was not possible to use a previous stable mean value because this type of analysis had not been carried out before. Instead, the existing mean value of 17.98 was used to replace the missing values which were all in the "zero fixations" bin. The results of the repeated-measures 2x4 ANOVA are reported below.

Relationship between the Number of Fixations Made on an Object and Memory for the Specific Location of Objects

As in Experiments 1-3, the average distance each object was placed from its own home location was also calculated. All 12 objects were included in this analysis. As before, four participants were missing data in the zero fixation "bin" and a linear regression was significant at the .001 level, yielding an equation of $y = -11.40X + 56.11$. Therefore 56.11 was inserted to the blank portions of the dataset; a value not dissimilar to the previous mean value of 60.54cm. This value is also very similar to the corresponding value from Experiment 1 for participants with a 10s viewing duration and an immediate test of 54.7.

Figure 22 shows both types of location memory (best fit error and placement error) plotted on the same graph. There are two clear distinct patterns of location memory. Memory for the overall configuration of objects (best fit error) is accurate no matter how many fixations an object receives (or indeed whether it is fixated or not). However, memory for the specific locations of objects (placement error) is not accurate for non-fixated objects, but is accurate for fixated objects. In addition, as demonstrated in Experiments 1 and 2, the placement error does not improve substantially with increasing numbers of fixations.

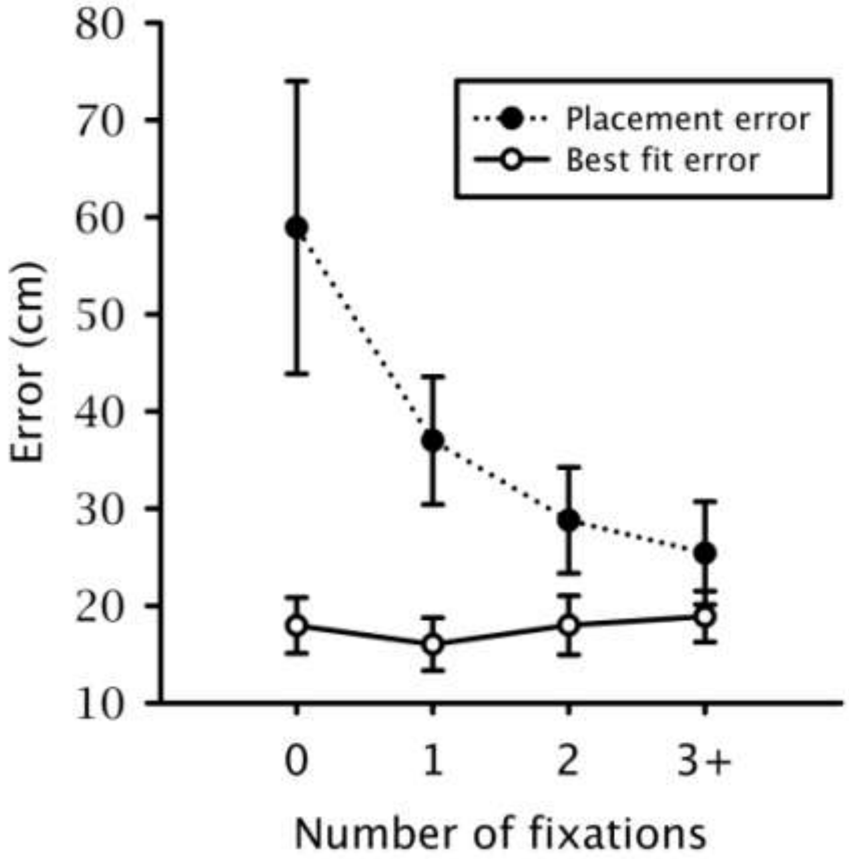

Figure 22. Placement error and best fit error as a function of the number of fixations made on objects during the encoding phase (error bars represent standard error)

To investigate this pattern of results, a 2 (Error type: placement error vs. best fit error) x 4 (Number of fixations: 0, 1, 2, or 3+) ANOVA was carried out. There was a significant main effect of Error type, $F(1, 30) = 37.10$, $p < .001$ with best fit error smaller overall than placement error. In other words, participants' memory was more accurate for the overall configuration of objects than the specific location of objects. This is unsurprising as previous research (e.g., Postma & de Haan, 1996) has demonstrated that it is easier to place objects in the correct overall configuration than to place objects in their correct locations.

There was no main effect of the number of fixations, $F(3, 30) = 2.04$, $p > .05$. This lack of an effect of fixations on location memory was probably driven by the best fit error scores being comparably low across the number of fixations. However, critically, there was a significant interaction between error type and the number of fixations, $F(3, 30) = 3.60$, $p < .05$. Closer inspection of this interaction using Bonferroni-corrected paired *t*-tests with a new alpha level of .05/4 = .013 showed that error scores were significantly different at 0 fixations, $t(10) = -3.14$, $p = .013$ with best fit error lower than placement error; error scores were also significantly different at 1 fixation, $t(10) = -4.25$, $p < .013$ with best fit error lower than placement error; however, error scores were not significantly different at either 2 fixations, $t(10) = -2.26$, $p > .013$ or at 3+ fixations, $t(10) = -1.22$, $p > .013$.

Comparison of the Accumulation of Placement Error with Experiments 1, 2 and 3

To enable comparison with the results of earlier experiments, the placement error for all fixated objects in Experiment 4 was collapsed across all fixation bins, and a paired *t*-test was carried out on placement error for fixated vs. non-fixated objects. However, despite a fairly substantial numerical difference, there was no statistically significant difference in placement error between fixated objects (M = 30.71 cm, SE = 3.86 cm) and non-fixated objects (M = 45.06, SE = 16.38), $t(10) = 0.79$, $p > .05$. This is not what was found in Experiments 1 and 2. However, there are two problems with the non-fixated data in this Experiment.

First, recall that four of the 11 participants fixated all 12 objects, meaning they did not have any data in the “zero fixation” bin. This problem was solved by using the existing mean value of 17.98 but still meant that four participants had artificially generated data. The second problem was that of the remaining seven participants who *did* have data in the “zero fixation” bin, two of them fixated 11 of the 12 objects. This meant that despite these two participants having data in the “zero fixation” bin, the value was for only one object each. For both of these participants, the single non-fixated object was placed extremely far from where it belonged. For one of the participants the object was placed 136.96 cm away, and for the other the object was placed

142.70cm. These two values contributed to the large standard deviation and standard error associated with the non-fixated objects, and it is likely that these two data points contributed to the non-significant result of the *t*-test.

Taken together, these results demonstrate that not only does Experiment 4 directly replicate the findings of Experiments 1 and 2 with regards to the different ways in which the identities and specific locations of objects are encoded, but also that Experiment 4 provides evidence that information about the overall configuration of objects forms without the need for direct fixations to be made on individual objects.

The Relationship between the Number of Fixations made on an Object and Confidence for the Overall Configuration of Objects

The next analysis considered the participants' confidence ratings for the locations in which they placed the plastic blocks, or the overall configuration of the objects. As before, linear regression was used replace missing data in the 0 fixations "bin". The regression analysis was significant,and yielded a regression equation for location confidence of $y = 0.70X + 5.29$ meaning that 5.29 was inserted into the data set; a value comparable to the previous mean value of 4.64. The results of the repeated-measures 2x4 ANOVA are reported below.

The Relationship between the Number of Fixations made on an Object and Confidence for the Specific Locations of Objects

Regarding confidence for the specific locations of objects, recall that once participants had placed the plastic blocks, they had to replace each block with one of the 12 correct objects which had been given to them by the experimenter. Participants then gave their confidence that each object was now located in the correct place (and could move objects if they required). A significant linear regression yielded an equation of $y = 1.45X + 3.83$ meaning that the value of 3.83 replaced the blank cells.

A repeated-measures 2 (Confidence type: location confidence vs. identity-location binding confidence) x 4 (Number of fixations: 0, 1, 2, or 3+) ANOVA was carried out. The ANOVA showed that there was a significant main effect of the number of fixations on confidence, $F(3, 30) = 11.37$, $p < .001$ with a significant linear trend, $F(1, 10) = 29.32$, $p < .001$. Post-hoc Bonferroni-corrected paired *t*-tests showed that confidence scores were higher after 1, 2 and 3+ fixations than 0 fixations (all *p*s < .05). There were no other pair-wise differences.

There was no main effect of confidence type, $F(3, 30) = .11$, $p > .05$ and no interaction between confidence type and the number of fixations, $F(3, 30) = 1.22$, $p > .05$. This finding is interesting as it demonstrates that participants' confidence for the location of non-fixated objects is significantly lower than for fixated objects, yet the behavioural data showed that memory for the overall locations of objects was no less accurate for non-fixated objects than the fixated objects. Therefore, there appears to be somewhat of a dissociation between the memory that participants exhibit for non-fixated objects and their confidence for those same objects. Based on the behavioural data it might be expected that confidence for the plastic blocks would be high regardless of how many fixations an object received; this would closely fit the memory for the overall spatial layout. However, this was not what was found. Possible explanations for this dissociation are explored in the discussion.

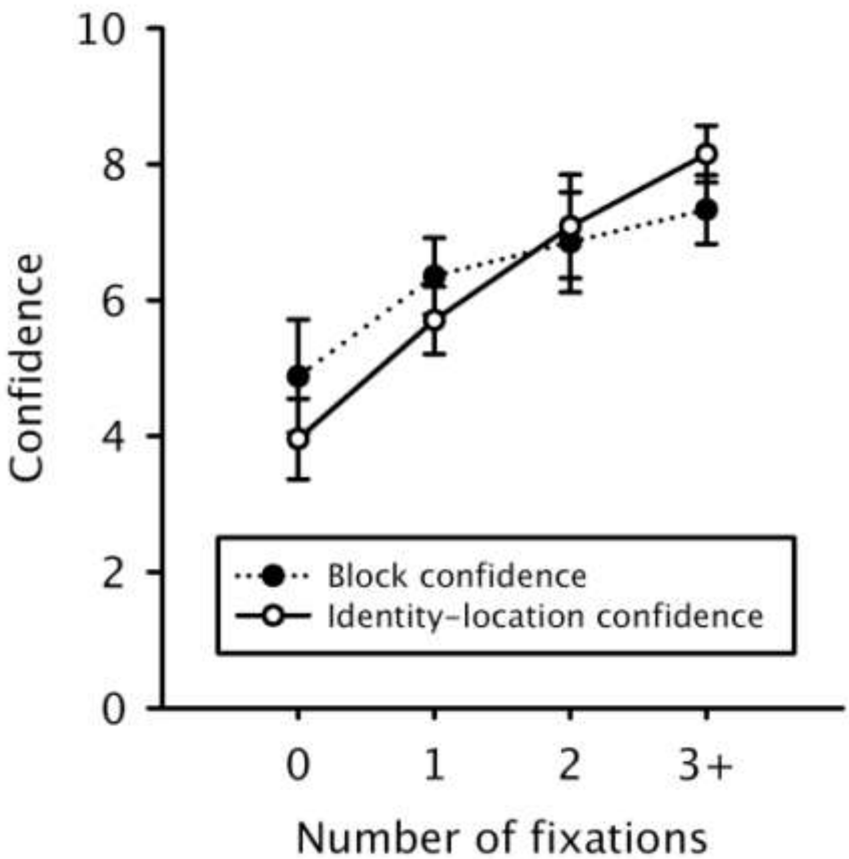

Figure 23. Block confidence and identity-location confidence as a function of the number of fixations made on objects during the encoding phase

5.4 Discussion

The results of Experiment 4 provide a more detailed understanding into the way that memory for the specific locations of objects and particularly the overall spatial configuration of objects are encoded into memory during the viewing of a scene. Each finding will be considered in turn. The first finding was that the basic characteristics of saccades and fixations were very similar between the block-placing participants from Experiment 4 and the control participants from Experiment 3. This is unsurprising because both groups of participants experienced identical encoding phases and task instructions, and therefore there is no reason for them to behave differently from one another. Experiment 4 therefore directly replicated the encoding behaviour reported in Experiment 3.

The second finding was that memory for the overall configuration of objects was very similar for the two groups of participants, and so was the memory for the specific locations of objects. This is unsurprising because both groups experienced identical encoding phases. On the basis that the encoding behaviour was similar, it is unsurprising that the test behaviour was similar.

The third finding is perhaps the most important. Information about the overall configuration of objects (best fit error) and information about the specific locations of objects (placement error) showed two distinct patterns of encoding. Best fit error did not change with increasing number of fixations. Instead, it was accurate (approximately 20cm) no matter how many fixations an object received. Put simply, regardless of the amount of encoding an object received, that object was placed near to one of the available candidate locations within the spatial map. This provides evidence that some information about the overall configuration of objects is formed without the need to directly fixate all of the objects. This finding provides evidence that early in scene viewing a spatial map is formed which contains all of the available candidate locations. This spatial map forms early and presumably without the need for each of the candidate locations to be fixated. If this is true, then it explains why even non-fixated objects were placed close to one of the candidate locations, even if they were not placed close to their actual home location. This finding also supports previous research that suggested memory for spatial locations forms automatically (Zechmeister, McKillip, Pasko, &

Bespalec, 1975; von Wright, Gebhard, & Karttunen, 1975; Mandler, Seegmiller, & Day, 1977; Aginsky & Tarr, 2000, Jiang, Olson, & Chun 2000; Alvarez & Oliva, 2007). It is also in line with the discussion in the Introduction in this chapter.

Memory for the specific location of objects, however, was poor for non-fixated objects, but accurate for fixated objects. In addition, as in Experiment 1, 2, and 3, memory for the specific locations of objects did not substantially improve after one fixation. Therefore the data from Experiment 4 replicate the findings of the previous three experiments. Non-fixated objects were typically placed a long way from their home location, but fixated objects were placed close to their home locations. In addition, once the specific location of an object was encoded, no improvement was observed if subsequent extra fixations were made on that object. This suggests that in order for an object to be indexed to its specific position within the spatial map, it *does* need to be directly fixated. This explains why non-fixated objects are typically placed far from their actual home location. Most critically, the dissociation between the encoding pattern found for overall spatial locations and specific locations supports both the claims made in Experiments 1 and 2 in this thesis, and the findings of Simons (1996), Postma and de Haan (1996) and Postma, Kessels and van Asselen (2008).

The final finding to be discussed relates to the confidence data. Recall that participants gave self-confidence ratings for both their confidence that each location was filled correctly by a plastic block, and, once they had replaced the blocks with objects, also for their confidence that each of the objects was placed in the correct location. The confidence data matched the behavioural data for the specific placement of the objects. That is, non-fixated objects were placed a long way from their own home location, and confidence was low. However, fixated objects were placed nearer to their home location, and accordingly confidence increased. There were small but non-significant increases in accuracy of object placement as the number of fixations increased, and this was reflected in small but non-significant increases in confidence for the specific locations of objects.

However, for placement of the blocks (demonstrating memory for the overall spatial configuration of objects), it was found that memory accuracy was equally good no matter how many fixations an object received. Interestingly, the confidence data did not reflect this. Instead, participants reported low levels of confidence for the overall configuration of non-fixated objects even though accuracy was high (i.e. error was low). As fixations on an object increased, the confidence ratings also increased, despite the finding that memory for the overall configuration of objects did not change. This finding is interesting as it suggests that there may be a dissociation between participants' beliefs about their memory for the overall configuration of objects, and the memory they actually exhibit for the overall configuration. Specifically, even though participants may not be confident about the locations they did not fixate, they still demonstrate accurate memory for those overall locations. The most likely explanation is that the spatial map forms early in scene viewing somewhat automatically as discussed above. However, participants do not seem to possess meta-cognitive awareness about the accuracy of their memory for the overall configuration of objects. As a result, participants seem able to place objects in locations that they did not fixate, and are not confident about. It seems as though participants possess some overall configuration memory for objects they have not fixated, even if they may not be aware of having that memory.

Summary

The findings of Experiments 1 and 2 extended the framework of the Visual Memory Model and suggested that once an object is directly fixated, information about the identity of an object continues to be encoded across multiple fixations, but information about the specific location of an object is encoded on the first fixation on an object (and there appears to be no improvement on subsequent fixations). The findings of Experiment 4 seem to add two key details to the Visual Memory Model.

First, the findings provide robust evidence that information about the overall configuration of objects is encoded somewhat automatically and without the need for each location to be directly fixated. Even non-fixated objects can be placed close to one of the locations in the spatial map. However, for objects to be placed close to their specific home location, objects need to be directly fixated.

Second, memory for the overall configuration of objects seems to form somewhat automatically without participants being aware of it, as shown by the low confidence ratings for the overall locations of non-fixated objects and the finding that memory for the overall locations of objects was accurate even for non-fixated objects. These findings are integrated into a modified Visual Memory Model which is presented in Chapter Seven.

6 The relationship between the number of fixations on an object and memory for both the specific location and correct surface location of an object

6.1 Introduction

Experiments 1-4 in this thesis added detail to the framework proposed in Hollingworth and Henderson's Visual Memory Model. Specifically, the results of Experiments 1-4 have provided evidence that (a) there is a very close relationship between encoding and recall for the identities and locations of objects in scenes; (b) the encoding processes associated with the encoding of identities and specific locations of objects appear to operate differently; and (c) the encoding processes associated with the encoding of the overall configuration of objects and the specific locations of objects also seem to operate differently. One important finding is that information about the specific location of an object seems to be encoded during the first fixation on that object and subsequently does not appear to improve with more fixations. That is, memory for specific object locations seems to be somewhat categorical. The lack of improvement for object location memory suggests that there are no degrees of specificity associated with accuracy of memory for the specific location of objects. That is, there is no sense of knowing the general area in which an object is located, without a sense of knowing an exact location that the object belongs in.

As presented in Chapter 1, to date, there has been very little research that has directly investigated the relationship between the encoding and recall for the identities and locations of objects in scenes. Most research has either focused on the encoding of objects without considering how that encoding might affect subsequent recall, or focused on the recall of objects without investigating how that information has been encoded. However, one relevant study was carried out by Tatler, Gilchrist and Land (2005) which directly investigated the relationship between encoding and recall for objects in scenes.

6.1.1 Evidence for accumulation of specific object location information across multiple fixations

Tatler, Gilchrist and Land (2005) used a head-mounted eye-tracker to record participants' eye movements as they viewed real-world environments for 5 s. Participants stood in the doorway of 6 rooms and looked at them in preparation for an immediate memory test. The test involved completing a multiple choice pencil-and-paper test which probed memory for a variety of different aspects of the scene including memory for the specific locations of objects. To demonstrate their object location memory, participants had to draw a cross on a line drawing of the scene indicating where they remembered a particular target object had been located.

As in the experiments reported in this thesis, Tatler et al. found that memory for the specific locations of non-fixated objects was no better than expected by chance, thereby providing further evidence that in order for information about the specific location of an object to be encoded, that object needs to be directly fixated. However, they also found that the accuracy of the specific location seemed to accumulate across fixations. They argued that this was unsurprising, as in the real world, objects frequently change position (e.g., the position of a car on the road) and therefore it is appropriate for the visual system to regularly encode and update information about the specific location of an object. Note, however, the objects were completely stationary in the rooms in their study. Also, it might be more appropriate to characterise repeated encoding of the location of a moving object as updating the new location associated with that object over time (rather than developing a more accurate representation of the same location).

6.1.2 Comparison between Tatler et al. and the present experiments

Importantly, the type of objects, and scenes selected as stimulus environments by Tatler et al. were somewhat different to the stimulus photograph used in the current experiments. All of the scenes used by Tatler et al. featured scene-congruent objects in appropriate (expected) locations within the scene (indeed, it was stated that the rooms were set up to look exactly like a laboratory, an office, a waiting room, a seminar room, a dining room, and a kitchen). This means that if participants did not fixate one of the objects they were questioned about at test, they could presumably use their prior knowledge about the most likely location of that object within that environment to influence their performance in the memory test. As a result, participants might have been able to correctly state the general location at which an object was located based on their previous experience, despite possessing no memory for the specific location of that object.

Tatler et al.'s data suggested that memory for the specific location of objects *did* improve as the number of fixations on those objects increased, thereby suggesting that there are degrees of specificity associated with the accuracy of object location memory. In other words, Tatler et al.'s data suggest that contrary to the claims of the Visual Memory Model, information about the specific location of an object may accumulate in memory across separate fixations.

In the present experiment, there were several objects which could be considered somewhat unusual in the scene. Notably, informal post-hoc discussions revealed that several participants expressed surprise that there had been a football, a lava lamp and a glass in the photograph. They based their surprise on their assumption that the stimulus photograph was meant to look like an office. As these objects could be judged not to belong in the scene, it is likely that participants could not use their previous experience to place these objects in the absence of memory for their specific location. For example, where would a ball typically be located in an office? In addition, some of the objects in the stimulus photograph were in unusual locations. The keyboard was on the chair (not on the desk); the mouse was on the small table (rather than the desk); the clock was on the desk (rather than the wall). As these objects were in unusual locations within the scene, it is likely that participants, again, could not use previous experience to place the objects if they did not have memory for their specific locations.

The critical point to make here is that the objects and their particular locations within the stimulus photograph used in the current experiments meant that the specific locations of some of the objects (that had not been directly fixated and encoded) could perhaps not easily be inferred on the basis of the context within which the objects appeared. This difference between Tatler et al.'s study and the present experiments may explain why Tatler et al. found that memory for the specific locations of objects accumulated across separate fixations, whereas in the current experiments no such accumulation has been observed.

It seems as though there is a contradiction regarding the way in which specific object location information accumulates across separate fixations. Tatler et al. provided evidence which seems to suggest that memory for the specific location of an object might improve as the number of fixations made on that object improves. However, the experimental data reported thus far in this thesis has provided evidence which demonstrates that memory for the specific locations of an object does *not* improve as the number of fixations made on that object improves. Instead, the experiments reported in this thesis support the view that specific location information is encoded rapidly after only a single fixation and does not improve with further fixations.

As a direct result of this contradiction, the present study set out to investigate whether modifying the *test phase* of the standard object location paradigm used so far might enable further investigation into the nature of accumulation of object location memory across separate fixations. Recall that Tatler et al. provided evidence that improvement is observed across multiple fixations whereas the experiments reported in this thesis have provided evidence that object location memory is encoded accurately during the first fixation on an object and does not become any more accurate with increased fixations. To directly probe this issue in more detail, the test phase was modified slightly. .

6.1.3 Modification of the test phase of the present paradigm

The modification of the test phase of the present experiment was as follows: a recognition component was introduced in the test phase for the locations of objects. Prior to the test phase, a marker was placed at each of the 12 object locations in the cubicle. With the locations of all the objects in the cubicle pre-marked, participants did not have to recall the locations of objects in the stimulus photograph. Instead, they simply had to recognise which object belonged on each of the markers. Recall that in Experiments 1-4 participants had to (a) *recognise* which objects they had seen in the photograph; (b) *recall* which locations were occupied by objects in the photograph; and (c) *recall* which object belonged at each of these locations. This meant that there was presumably a three-fold demand on their memory for the objects and their locations in the scene. In contrast, participants in Experiment 5 only had to (a) *recognise* which objects they had seen in the photograph; and (b) *recognise* which object belonged at each location. They were not required to recall from memory the locations that the objects occupied.

Simplifying the test phase in this way meant that participants were likely to experience a reduction in the cognitive demands that the test phase placed upon them. Previous research has provided evidence that recall tasks tend to be more difficult than recognition tasks (e.g., Schonfield & Robertson, 1966; Hasher & Zacks, 1979; Burke & Light, 1981) It has been suggested that this difference in difficulty occurs due to recall tasks requiring more self-initiated processing than retrieval tasks. In recognition tasks, the stimuli that participants have viewed during the trial are typically presented to participants again at test (often alongside some unseen distractor objects). This means that participants can use the stimuli to guide their recognition. In contrast, recall tasks typically do not provide any retrieval cues. Instead, participants are required to effortfully retrieve detailed information about objects that is stored in memory (Craik & McDowd, 1987). Consequently, recognition tasks typically demand a higher level of cognitive processing than recall tasks.
It was expected that this simplification would mean that participants invested all of their cognitive resources during the test phase into recognising which of the objects belonged on each of the markers (i.e. the binding between identities and their locations). Note that participants were not forewarned about the nature of the memory test prior to the encoding phase of the experiment, and therefore, did not modify their encoding behaviour accordingly. It was anticipated that the simplification of the memory task in this experiment might allow for the observation of improved object location memory across fixations.

There was also a second change in methodology. Unlike in Experiments 1-4, participants were not allowed to place distractor objects in the cubicle. During the test phase, they were allowed to select from both the objects that had appeared in the photograph, along with distractor objects, but before participants could place any of the objects, any distractor objects that they had erroneously selected were removed by the experimenter and replaced by objects that had been present in the stimulus photograph. This meant that once participants began placing objects they had all 12 correct objects available to them for placement, of which, some had been provided by the experimenter.

This change in methodology was introduced to maximise the possibility that improvement in accuracy of object location memory might be observed across fixations. In Experiments 1-4 participants were required to place all the objects that they selected regardless of whether they were correct or they were distractor objects. However, as distractor objects were not present in the stimulus photograph, and therefore did not receive any visual encoding, it would not have been possible to have included them in any of the analyses that investigated the relationship between encoding and recall. Giving participants the correct 12 objects to place meant that, potentially, more data would be available for analysis which would add more power to any potential findings. One might expect that participants would be less likely to place experimenter-provided objects on the correct surface than objects they had selected themselves. This is because it is likely that experimenter-provided objects had received fewer fixations during encoding than objects that the participants had selected. As Experiments 1-4 have demonstrated that the likelihood of selecting an object at test was closely related to the number of fixations made on that object at test.

To reiterate, the modifications made to the test phase of the present experiment allowed examination of how accurately participants could *recognise* the correct location for each object rather than having to *recall* the correct location for each object. On this basis it was expected that if object location memory is indeed categorical, then the accuracy with which objects were placed on the correct marker would be well above chance after a single fixation and would not improve substantially with subsequent fixations. In this way, such a finding would demonstrate that the identity of an object is bound to one of the markers in the spatial map during the first fixation on an object. In addition, it was expected that non-fixated objects would be placed on the correct marker no better than chance, providing further evidence that objects need to be directly fixated if their specific location is to be remembered at test.

6.1.4 The present experiment

Several predictions were made on the basis of the findings of earlier experiments in this thesis. First, it was expected that the eye movement behaviour exhibited during the encoding phase would be comparable for the participants in this experiment and the control participants (taken from Experiment 3). This is on the basis that both groups of participants were

tested on the same eye tracker and experienced identical stimuli and testing conditions (10s viewing duration and immediate memory test).
In addition it was expected that the number of objects correctly selected at test would be comparable to the number selected by the experimental participants from Experiment 5 and the control participants from Experiment 3, and the 10s/0 delay participants in Experiment 1. Again, this would be due to the identical testing conditions experienced by all participants. Again, these analyses were conducted to ensure replication.

The third prediction was based on the findings of Experiments 1, 2 and 3 in this thesis. The results of these experiments suggested that memory for the identity of objects accumulates across separate fixations. Therefore, this prediction was that object identity would accumulate across multiple fixations in a similar way to Experiments 1, 2 and 3 in this thesis. A fourth prediction was that experimenter-provided objects would be less likely to be placed on the correct marker than objects that the participants had selected themselves. This prediction was made on the basis that experimenter-provided objects were likely to have received less visual attention via fixation during the encoding phase than objects selected by the participants (since objects that were fixated were much more likely to be selected than objects that were not).

The fifth prediction related to the probability of placing an object on the correct marker. As this was considered to be analogous to placing an object in the correct location in Experiments 1 – 4 it was expected that the probability of placing a non-fixated object on the correct marker would be no better than chance, but that fixated objects would be very likely to be placed on the correct marker. In addition it was expected that there would be no clear improvement in the probability of placing objects on the correct marker with increased numbers of fixations on those objects. That is, one fixation on an object should be sufficient to allow for accurate placement. If this pattern of results were found, it would provide further evidence that information about the specific location of an object is encoded during the first fixation on an object and does not substantially improve with subsequent fixations. In addition, it would provide further evidence that in order for the specific location of an object to be encoded, then that object must be directly fixated. However, if improvement in placing objects on the correct marker did occur as the number of fixations made on that object increased, this would undermine the claim that an object's location is encoded accurately during the first fixation on it, and that accuracy does not improve with subsequent fixations. Such a finding would suggest that there are degrees of increasing specificity across fixations that are associated with memory for the locations of objects. The final prediction related to the confidence ratings. As in Experiment 4 it was expected that selection and placement confidence would gradually increase linearly across fixations, and if it did, this would provide a further replication of the data from Experiment 4.

6.2 Method

Participants and design

There were 14 participants all recruited from the University of Southampton undergraduate community. All participants took part in exchange for course credit and were naive to the aims of the experiment. None of the participants had taken part in any of the other studies reported in this thesis. There were eight males and six females. The average age was 20.36 years (*SD* = 0.74 years) and ages ranged from 19 – 22 years. Data from three participants was discarded due to tracker loss leaving data from 11 participants. A between-participants design was used, with control data taken from the 10 free-viewing participants in Experiment 3.

Apparatus

The apparatus used was identical to that used in Experiment 4. In addition, small pieces of Blu-Tack were used to mark the correct object locations.

Stimuli

The stimulus photograph was identical to that used in Experiments 1 and 4 and the free-viewing condition of Experiment 3.

Procedure

The encoding phase was identical to that of Experiment 3, Experiment 4 and the 10 s/0 delay condition of Experiment 1 reported in this thesis. As in Experiment 4 a 10s encoding duration was used to try to generate enough eye movement data in the 2 and 3+ fixations bins to allow for analyses to be carried out the rate of encoding for object identity and location information

across multiple fixations. However, there was a modified test phase. At test, participants entered the cubicle shown in the stimulus photograph. They were required to perform two tasks. The first task was identical to the first part of the test phase in Experiments 1-3. Participants were presented with 24 objects arranged randomly on the floor of the cubicle. Of these, 12 had been present in the photograph, and 12 were distractor objects that had not been present in the stimulus photograph. Participants had to select up to 12 objects. If participants selected fewer than 12 objects, they were told to keep selecting objects until they had picked 12, guessing if needed. Participants then had to give confidence ratings for the identity of each object from 0 – 10 where 0 represented a guess, and 10 represented full confidence.
Next, any incorrect object selections (i.e. any distractor objects that had been erroneously selected) were taken away by the experimenter and replaced by the correct objects that the participant had not selected. This left the participant with all 12 correct objects, some of which they had selected, and some of which may have been provided to them by the experimenter.

The second part of the test phase probed memory for the locations of the objects. The 12 correct locations had been pre-marked in the cubicle using small pieces of Blu-Tack and participants were required to place each of the 12 objects on to one of the markers, guessing if needed. Once all 12 objects had been placed on a marker, participants were required to give a confidence rating that represented their confidence that each object was on the correct marker.

6.3 Results

As in the earlier experiments reported in this thesis, any fixations shorter than 80 ms or longer than 1200 ms were excluded from the analyses.

Between-Groups Analysis of Eye Movement Behaviour

As predicted, the eye movement measures were comparable between the experimental condition and the control condition (taken from Experiment 3). A series of unpaired *t*-tests found no significant differences between the groups. For one of the *t*-tests reported in Table 19, the Levene's test for equality of variances was significant and as a result the degrees of freedom associated with the *t*-test was not (N-2) but a decimal value instead. Table 1 displays the values of the t-tests. The finding that eye movements were comparable is unsurprising as participants experienced identical task instructions and were tested on the same eye tracker. In addition, it demonstrates that Experiment 5 replicates the basic encoding behaviour demonstrated in Experiment 3.

Between-Groups Analysis of Memory Behaviour

As the locations of objects were pre-marked, best fit error was not an informative measure to record as participants were not given the opportunity to place objects in incorrect locations. Given this, it is not possible to directly compare placement data from the current experiment with preceding experiments. However, the number of correct object selections could be compared between the experimental and control conditions of the present experiment. There was no significant difference between the experimental condition (M = 9.55, SE = 0.37) and the control condition (M = 8.90, SE = 0.38) in the number of objects correctly selected, $t(19) = 1.23$, $p > .05$. This is unsurprising as both the experimental and control groups had experienced identical testing conditions up to this point.

Table 20

Means and Standard Errors (in Parentheses) and t-test results for the Eye movement measures for the experimental and control groups

Condition:	Control condition (from Experiment 3)		Marker condition (from Experiment 5)		Between-groups *t*-test results		
Measure	M	SE	M	SE	df	*t*	*p*
Total time fixating objects (ms)	7332	(334)	8091	(278)	19.00	1.76	0.09
Total time fixating blank (ms)	818	(198)	691	(188)	19.00	0.46	0.65
Fixations made on blank regions	3.70	(0.84)	2.82	(0.70)	19.00	0.81	0.43
Average blank region fixation duration	203	(29)	198	(35)	19.00	0.10	0.92
Proportion of viewing duration spent fixating objects	0.90	(0.02)	0.92	(0.02)	19.00	-0.72	0.48
Fixations made on objects	25.30	(1.35)	25.91	(1.00)	19.00	-0.37	0.72
Fixations made in total	29.00	(0.92)	28.73	(1.02)	19.00	0.20	0.85
Number of objects fixated	10.60	(0.54)	10.18	(0.58)	19.00	0.52	0.61
Average fixation duration (ms)	292	(8)	316	(16)	19.00	-1.30	0.21
Average gaze duration (ms)	469	(47)	549	(51)	19.00	-1.14	0.27
Average number of fixations made during gaze duration	1.54	(0.11)	1.67	(0.10)	19.00	-0.88	0.39
Average total time (ms)	701	(34)	825	(61)	15.67	-1.77	0.10
Number of fixations made during total time	2.41	(0.12)	2.61	(0.14)	19.00	-1.06	0.30

Placement of Experimenter-Provided Objects vs. Participant Selected Objects

Next, the likelihood of placing objects on the correct marker was considered as a function of whether the participant had chosen the object or whether the object was provided to them by the experimenter. This is an interesting analysis to carry out because it is likely that objects selected by the participants would be placed more accurately than objects provided to the participant by the experimenter, since experimenter-provided objects are very likely to have received fewer, if any, direct fixations during encoding than objects selected by the participants.

Accordingly, experimenter-provided objects were looked at for shorter total times during the encoding phase (M = 403 ms, SE = 114 ms) than objects chosen by participants (M = 770 ms, SE = 49 ms), $t(10) = -2.52$, $p < .05$. In addition, experimenter-provided objects received fewer fixations in the encoding phase (M = 1.32, SE = 0.34) than objects chosen by participants (M = 2.43, SE = 0.11), $t(10) = -3.05$, $p < .05$.

In line with these findings, a paired t-test showed that the probability of placing an object on the correct marker was significantly higher for objects selected by the participant (M = 0.55, SE = 0.07) than for objects provided by the experimenter (M = 0.13, SE = 0.06), $t(10) = -3.96$, $p < .01$. This finding is consistent with the data above that show that experimenter-provided objects were fixated for less time during encoding than the objects selected by the participants.

The Effect of Fixation Behaviour on Probability of Correct Object Selection

The next analysis considered the encoding of object identities in a similar way as in Experiments 1-3. To analyse the relationship between encoding and memory for object identity, each object was categorised as having received zero, one, two, or three or more fixations during the encoding phase. The reason that fixations were categorised in this way (0, 1, 2, and 3+) rather than as 0, 1, and 2+ (as in Experiments 1 and 2) was that in this experiment, all participants had 10 s to view the photograph. This meant that there was enough data in the “two” and “three or more” bins to justify analysing these categories separately.

The probability that each object was selected at test was calculated. It is important to note that, due to the 10 s viewing duration, five of the 11 participants fixated all 12 of the objects in the photograph. This meant that those five participants had no data in the “zero” fixations bin. Therefore, a linear regression analysis was carried out on the object selection probability using the number of fixations as a predictor variable. A significant regression was observed, and the regression equation obtained was $y = 0.13X + 0.55$. As a result, all of the blank cells were replaced by the value of 0.55 (which was comparable to the previous mean value of 0.42).

To enable a direct comparison between Experiment 5 and Experiment 1, the data from Experiment 5 were analysed alongside the 10s/0 delay condition from Experiment 1. This is because both groups of participants experienced identical encoding phases and identical first parts of the test phase i.e. selecting 12 objects from the pool of 24 objects. Figure 24 shows the respective values for Experiment 1 (10s/0 delay participants) and Experiment 5.

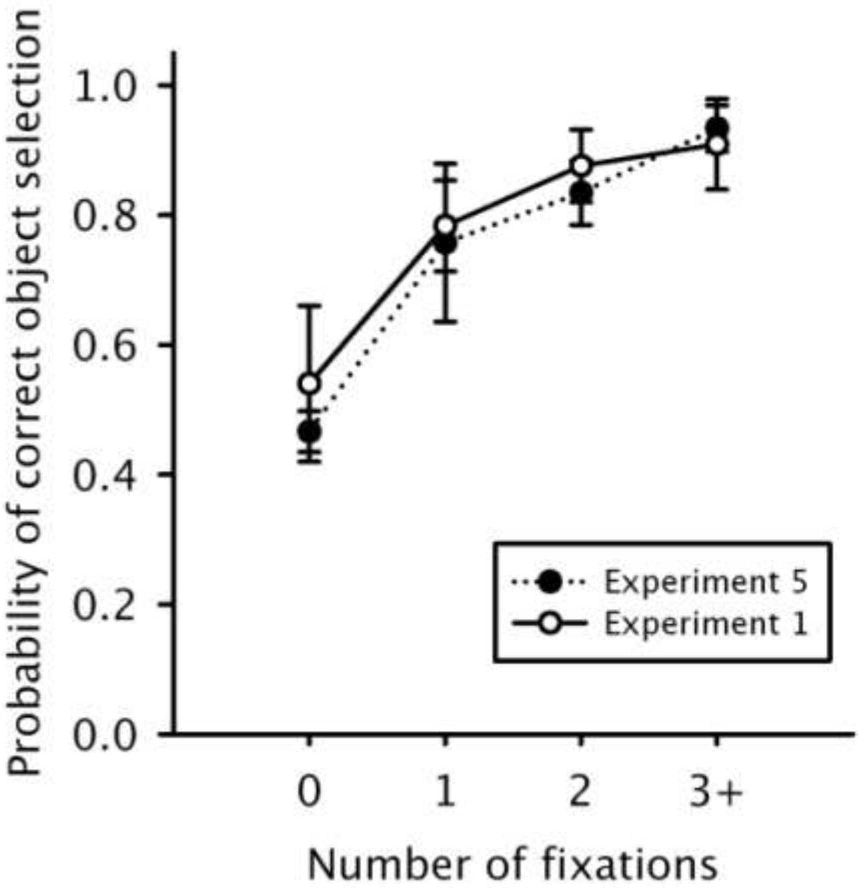

Figure 24. Probability of correct object selection as a function of the number of fixations made on objects during the encoding phase for both Experiment 1 and Experiment 5

A 2 (Experiment: 1 vs. 5) x 4 (Number of fixations: 0, 1, 2, or 3+) mixed ANOVA showed that there was no overall difference between Experiment 1 and Experiment 5, $F(1, 18) = .13$, $p > .05$, demonstrating that Experiment 5 directly replicated the relationship between encoding and recall observed in Experiment 1. There was no significant interaction between the number of fixations and the Experiment, $F(3, 54) = .23$, $p > .05$. However, there was a significant main effect of the number of fixations, $F(3, 54) = 9.32$, $p < .001$ and planned post-hoc Bonferroni-corrected *t*-tests showed that objects fixated twice, or three times or more were selected more often than non-fixated objects (*p*s < .01).

Finally, four one-sample *t*-tests were carried out for each fixation bin in Experiment 5 to compare the identity data with chance (0.5). Non-fixated objects were selected no more often than chance, $t(10) = -1.07$, $p > .05$; objects fixated once were selected marginally more often than chance, $t(10) = 2.11$, $p = .06$; objects fixated twice were selected significantly more often than chance, $t(10) = 6.72$, $p < .001$ as were objects fixated three times or more, $t(10) = 12.14$, $p < .001$.

These findings are important because they show that data from Experiment 5 directly replicate the effects reported in Experiment 1 in terms of the relationship between encoding and recall for object identity. This constitutes one of the primary experimental findings in this thesis, namely, that object identity memory accumulates over fixations. Also, as before, non-fixated objects were selected no more often than chance, but fixated objects (and especially objects fixated twice or three times) are selected more often than chance, indicating that to remember the identity of an object it is important to directly fixate it.

Further Comparison of Experiment 5 with Experiment 1 using the same types of analyses

To closely match the style of the data analyses carried out in Experiments 1-3, a paired *t*-test was also carried out on the probability of correct object selection between non-fixated and fixated objects for the data from Experiment 5. Note, however, that no statistical comparison was carried out at this point between the data from Experiment 1 and Experiment 5. Instead, the types of analysis used in Experiment 1 were used on the data from Experiment 5 i.e. a paired *t*-test between fixated and non-fixated objects.

As in Experiments 1, 2 and 3, fixated objects ($M = 0.87$, $SE = 0.03$) were significantly more likely to be selected than non-fixated objects ($M = 0.47$, $SE = 0.03$), $t(10) = -8.85$, $p < .001$. However, as demonstrated by the ANOVA above, although some improvement in memory for object identities did occur from one fixation ($M = 0.76$, $SE = 0.12$) to two or more fixations ($M = 0.88$, SE

= 0.03), this effect was not statistically reliable (unlike Experiments 1 and 2), $t(10) = -1.00$, $p > .05$. A key question, therefore, is why this numerical pattern of results (improvement in identity memory from one to two fixations) is not statistically reliable in Experiment 5 as it was in Experiments 1 and 2?

Closer inspection of the data from Experiment 1, where the improvement in memory for object identity from one to two or more fixations was initially observed, provides a possible explanation. In Experiment 1, the improvement in memory for object identity from one to two or more fixations seemed to be mainly driven by the participants with 5 s to view the data. Recall that in Experiment 1, half of the participants had a 5 s viewing duration. As a result, the mean probability of correct object selection for objects fixated only once by participants with a 5 s viewing duration in Experiment 1 was 0.66 (*SE* = 0.06) (see Table 3 in Chapter 2) whereas in the present experiment, where all participants had a 10 s viewing duration, the mean probability for objects fixated once was 0.76 (*SE* = 0.12). This difference demonstrates that objects fixated only once in Experiment 1 were not remembered as often as objects fixated only once in Experiment 5 and this is likely to be because average fixation durations for participants with a 5 s viewing duration in Experiment 1 (*M* = 269 ms, *SE* = 10 ms) were shorter than for participants with a 10 s viewing duration in Experiment 5 (*M* = 316 ms, *SE* = 16 ms) and this difference was statistically significant, $t(31) = -2.62$, $p < .05$.

In turn, this suggests that the amount of information that can be encoded during one fixation for participants with a 5 s viewing duration may not be equivalent to the amount of information that can be encoded during one fixation for participants with a 10 s viewing duration as there is a 47 ms difference in the average fixation duration. This raises an interesting point concerning the analysis procedures. In all experiments reported in this thesis, the number of fixations made on an object has been used as the unit by which the amount of visual encoding has been measured. In Experiments 1, 2, 4 and 5 the number of fixations made in total was used, and in Experiment 3 the number of fixations made during gaze duration was used. This use of the number of fixations to segment visual encoding means that, at a basic level, an assumption has been made that one fixation made on an object in a 5 s viewing condition is equivalent to one fixation made on an object in a 10 s viewing condition. However, as explained above, fixation durations vary somewhat depending on the time available to view the photograph. Therefore, it is important to acknowledge that the amount of time an object is fixated for, as well as the number of fixations made on an object, may be an additional source of information about the amount of visual encoding an object has received. This is clearly an interesting and relevant issue, and further research could empirically control both the number of fixations made on an object and the fixation durations made on objects to investigate whether differential relationships between encoding and recall are observed using each method. In summary, what is clear to see is that when the mean values from Experiment 5 are analysed alongside the identical condition from Experiment 1 (10s/0 delay) the pattern of data for Experiment 5 very closely replicates the pattern of data for Experiment 1. This serves to reinforce a key conclusion formed on the basis of the findings so far that object identity memory accumulates across fixations.

The Effect of Fixation Behaviour on Confidence for Object Identity and Object Location

The next analysis investigated the effect of the number of fixations made on objects on confidence ratings. Confidence for the identity of objects and the locations of objects was analysed. As the 10 s encoding period allowed participants to fixate most of the 12 objects, five of the participants had missing data in the 0 fixations bin and two participants had missing data in the 1 fixation bin. A linear regression was carried out on the identity confidence ratings using the number of fixations as the predictor variable. The regression was significant at the $p < .001$ level and yielded the equation $y = 1.21X + 5.49$. This equation was used to replace the missing data.

A second linear regression was carried out on the location confidence ratings as five participants had missing data in the 0 fixations bin and this regression was also significant at the $p < .001$. This yielded the equation $y = 1.32X + 3.10$ and this equation was used to replace the missing data for location confidence.

A 2 (Confidence type: Identity vs. Location) x 4 (Number of fixations: 0, 1, 2 or 3+) repeated-measures ANOVA showed that there was a significant effect of the number of fixations made on an object on confidence, $F(3, 30) = 9.72$, $p < .001$. There was also a significant linear contrast, $F(1, 10) = 37.27$, $p < .001$ showing that confidence increased as the number of fixations increased. Post-hoc Bonferroni-corrected t-tests showed that confidence was higher for objects fixated twice than non-fixated objects, and for objects fixated three times or more than non-fixated objects (both *p*s < .05). There were no other pair-wise differences. The pattern of data is shown in Figure 25.

There was also a significant effect of confidence type (identity vs. location), $F(1, 10) = 33.87$, $p < .001$, with confidence higher for the identities of objects than for the locations of objects. There was no significant interaction, $F(3, 30) = .29$, $p > .05$. This demonstrates that, as in Experiment 4, confidence for the identities and locations of objects increases as the number of fixations made on an object increases. This pattern of data is similar to the memory data and suggests that, unlike memory for the overall configuration of objects (Experiment 4, Chapter 5), participants are aware of the memory that they possess for the identity and specific locations of objects.

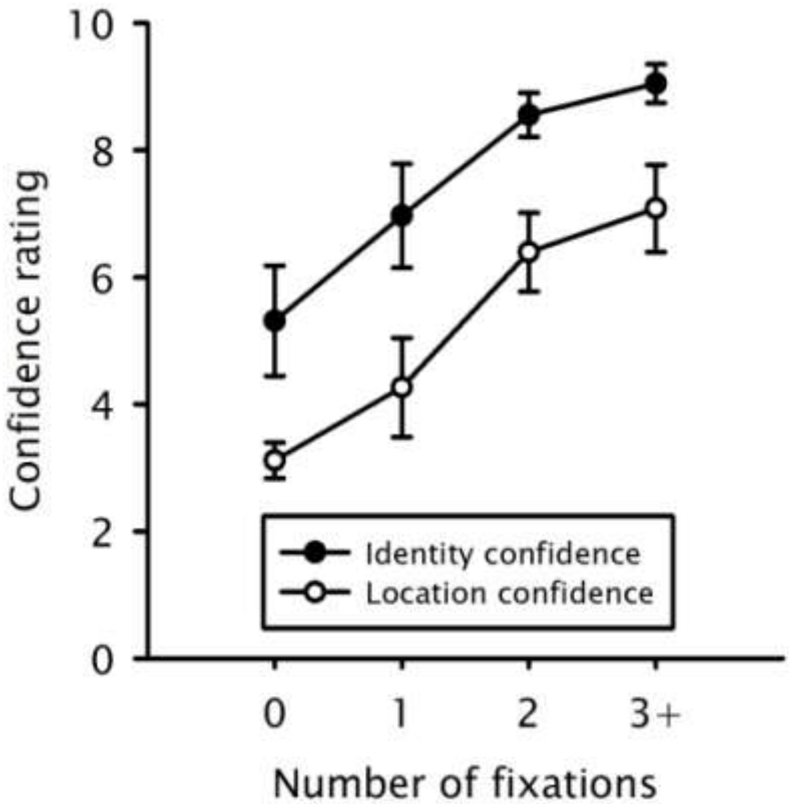

Figure 25. Identity and location confidence ratings as a function of the number of fixations made on an object

The Effect of the Number of Fixations on Probability of Placement on Correct Marker

The final analysis investigated whether information about the specific locations of objects was encoded in a categorical way as found in earlier experiments in this thesis. The probability of an object being placed on the correct marker was calculated. There were 12 markers in the cubicle; one for each object and as each object was placed in the cubicle, the marker onto which it was placed was noted. To replace blank cells in the "0 fixation" bin, a linear regression was carried out. This was significant and yielded the following formula: $y = 0.12X + 0.26$ meaning that 0.26 replaced all blank cells.

A repeated-measures ANOVA with number of fixations (0, 1, 2, or 3+) as the independent variable showed that there was a significant main effect of the number of fixations, $F(3, 30) = 4.46$, $p < .05$ and a significant linear contrast, $F(1, 10) = 19.30$, $p < .01$. Planned post-hoc Bonferroni-corrected paired *t*-tests showed that objects fixated twice, or three times or more were placed on the correct marker more often than non-fixated objects (*p*s < .05). The pattern of data is shown in Figure 26 below. Importantly, there was no evidence for accumulation of location information across separate fixations. This conclusion is consistent with the earlier claim that the specific location of an object is encoded in detail during the first fixation of an object and subsequent fixations on the same object do not improve the accuracy of the location information associated with that object that is stored in memory.

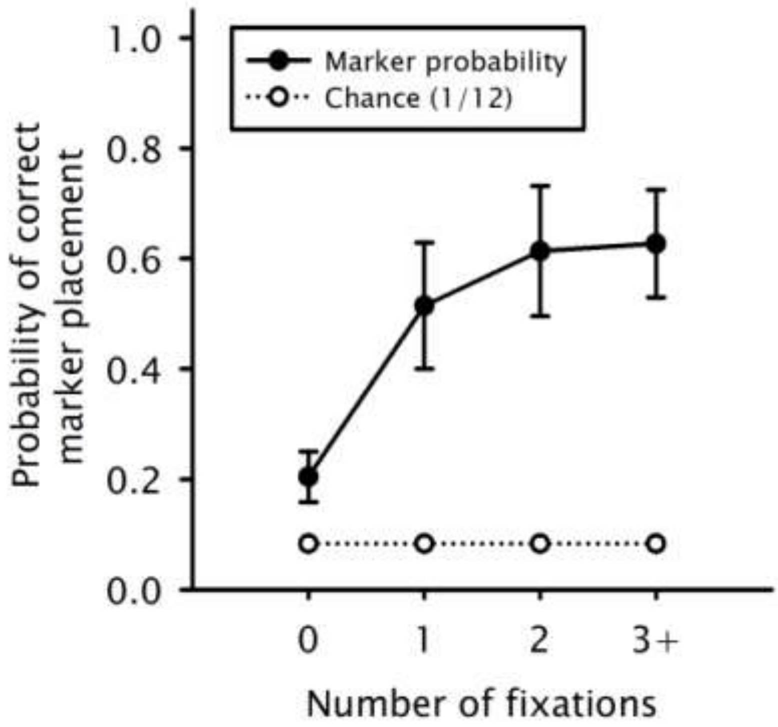

Figure 26. Probability of correct marker placement as a function of the number of fixations made on an object

To compare performance to chance, one-sample Bonferroni-corrected *t*-tests were carried out. The new alpha level was 0.05/4 = 0.013. As there were 12 potential markers, the probability of placing an object on the correct marker by chance was calculated as 1/12 = 0.083. In line with the experimental predictions, non-fixated objects were placed on the correct marker no more often than expected by chance, $t(10) = 2.65$, $p > .013$. However, fixated objects were always placed on the correct marker more often than chance; this was true for objects fixated once, $t(10) = 3.14$, $p < .013$; objects fixated twice, $t(10) = 6.75$, $p < .001$; and objects fixated three times or more, $t(10) = 7.23$, $p < .001$. This suggests that for the specific location of an object to be encoded, that object needs to be directly fixated.

6.4 Discussion

The findings of Experiment 5 serve two clear purposes. First, they directly replicate the findings of earlier experiments in this thesis: the basic encoding behaviour directly replicated the encoding behaviour observed in

Experiment 3; the number of objects correctly selected directly replicated both Experiment 3 and Experiment 1; the relationship between encoding and recall for the identities of objects is identical to that observed in Experiment 1; and the confidence ratings replicate the confidence ratings from Experiment 4. Thus Experiment 5 clearly validates the findings reported in earlier experiments.

Second, the findings of Experiment 5 allow for elaboration of Hollingworth and Henderson's Visual Memory Model. Non-fixated objects were placed on the correct marker no more often than would be expected by chance, but fixated objects were always placed on the correct marker more often than expected by chance. Clearly objects need to be fixated in order for their specific location to be encoded and remembered i.e. for the object to be bound to its correct position within the spatial map. Crucially, the data are consistent with earlier experiments showing that when an object has been fixated once there is no substantial improvement for its location with subsequent fixations. Note that this consistent finding was obtained using a recognition task, rather than a recall task, in the test phase. This provides further evidence to support the claim that information about the specific location of objects is categorical in that when an object's location has been encoded in detail after the first fixation, further fixations do not lead to improvement in accuracy of the memory for the specific location of that object.

Recall that throughout the discussion of the experiments in this thesis it has been argued that there is a dissociation between the encoding processes associated with the identities of objects and the specific locations of objects. The nature of the dissociation is that the identity of an object is typically comprised of a wide range of different characteristics (e.g., name, shape, size, colour, etc) which enables encoding of object identity to continue across multiple fixations. In essence, the identity of an object is complex and has many different details associated with it. Furthermore, these details are encoded across fixations such that the identity of an object is refined over time. In contrast, it has been argued that the object file for an object is indexed to a position in the spatial map during the first fixation on that object and once this binding has occurred there is limited scope for substantial improvement in object location memory. The acquisition of this location information is not the same as the continued encoding of object identity properties over fixations. Put simply, once an observer forms a representation of where an object is located within the scene (e.g., on the back left of the desk) refinement of that location memory does not occur regardless of whether the object is re-fixated.

At this point, it is perhaps worth reconsidering this claim. This is particularly the case in light of the data from Experiments 4 and 5 (see Figure 22, Chapter 5 and Figure 26, in the present chapter). In Experiments 4 and 5, the experimental conditions were optimal to allow for potential observation of an increase in the accuracy of object location memory across fixations. Specifically, all participants had a 10 s viewing duration, which increased the likelihood that some objects would be fixated twice, and there was no manipulation imposed on the encoding phase, meaning participants could freely view the photograph. Close scrutiny of the object location data from these experiments (Figure 22, Chapter 5 and Figure 26, in the present chapter) reveals that, in fact, there was a consistent, but non-significant linear trend for increased accuracy of object location memory over the second, third and subsequent fixations. Importantly, this trend was observed using two different methods of measuring object location memory. Experiment 4 required participants to recall particular object locations, whilst Experiment 5 required participants to recognise that a particular object was associated with a particular given location. Given that the data from Experiments 4 and 5 both seem to show that object location memory accumulates across separate fixations, albeit modestly, at this point it seems appropriate to discuss whether the very firm theoretical conclusion that has been formed regarding the encoding of object location memory is appropriate. To reiterate, this claim assumes that object location memory is encoded very accurately during the very first fixation on an object; so accurately that no improvement occurs if the same object is fixated more than once.

If this claim is correct, then one might expect that participants in the test phases of Experiments 1-4 should have been able to place all fixated objects extremely close to the exact location that those objects belonged to. However, this was not the case. Although in Experiments 1-4 participants consistently placed fixated objects more accurately than non-fixated objects, they did not place fixated objects perfectly. The average placement error for fixated objects in Experiments 1, 2, 3 and 4 was 40.69 cm, 42.38 cm, 40.28 cm and 30.71 cm respectively; an average of 38.52 cm. This demonstrates that participants typically placed fixated objects with some degree of error presumably indicating some uncertainty as to the precise locations of the objects. Whilst the total area of the surfaces in the cubicle was 21,684 cm^2 and therefore an average placement error of about 40 cm clearly reflects accurate memory for the approximate location of objects, it does not reflect perfect memory for the location of objects.

Therefore, a slightly modified theoretical account might state that instead of participants encoding the exact location of an object during the first fixation on it, they instead encoded a small region of the scene to which they assigned the object file. Such a region would constitute a “vicinity of error”. That is, having made a single fixation on an object, participants may have encoded a location such that they could place an object within the vicinity of its location (approximately 40 cm based on the data from the current experiments). If this suggestion is correct, and participants encode a location with a vicinity of error for each object that they fixate (rather than the exact location), then it follows that they might be able to refine this vicinity of error (i.e. reduce the amount of uncertainty associated with the location) over different fixations made on the same object. Thus, the spatial map that forms for the visual scene might consist of 12 vicinities of error, rather than 12 placeholders that mark the exact position of the 12 objects. Figure 27 illustrates the exact position of an object being encoded (left) and the corresponding vicinities of error in which the objects may be located (right).

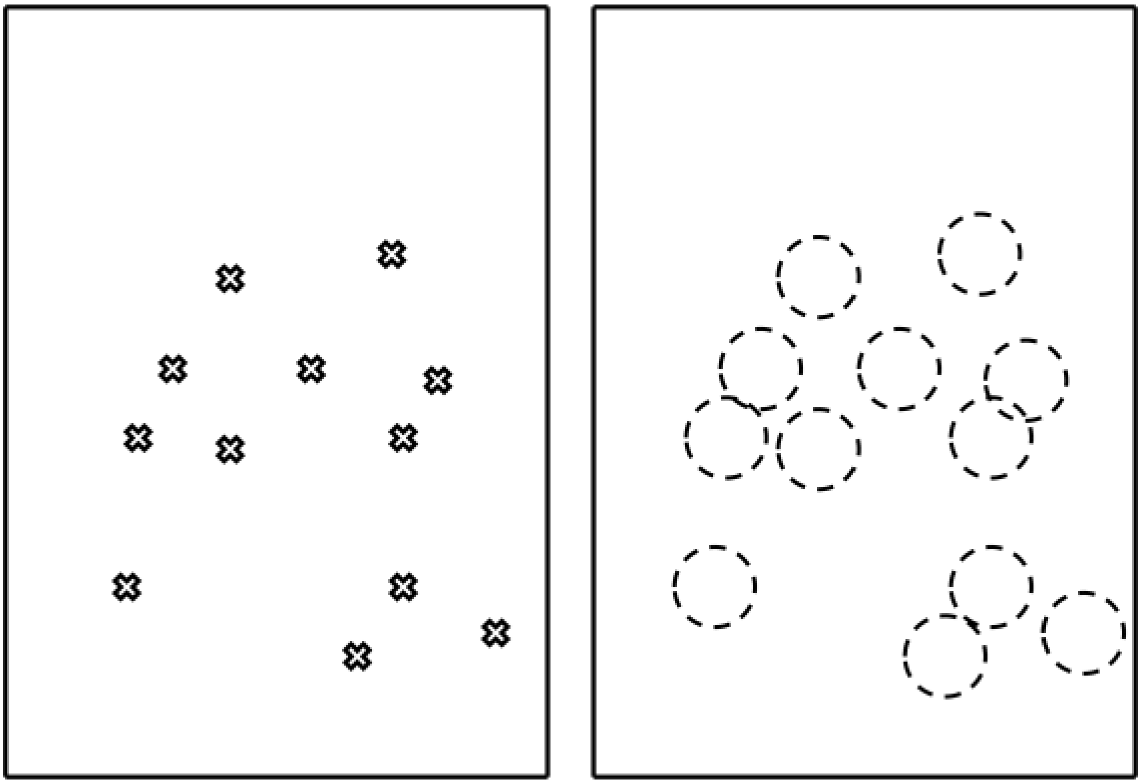

Figure 27. Illustration of the exact positions of objects being encoded (left panel) and the corresponding vicinities of error (right panel)

In Figure 27 (above) all of the vicinities of error are the same size. This is intended to reflect the status of the spatial map early in scene viewing. It is also possible that as the number of fixations made on an object increases, the vicinity of error for that object might decrease. As a result, for example, the vicinity of error for an object fixated, say, three times might be smaller than the vicinity of error for an object fixated only once. Figure 28 illustrates this possibility. In Figure 28, some of the vicinities of error are the same size as in Figure 27 indicating non-fixated locations in the spatial map. Note that non-fixated locations do not have an object file associated with them. Recall that the assignment of an object file to a location occurs upon a fixation on that object. Other vicinities of error within the spatial map are smaller, due to these objects having been fixated more than once. It seems plausible to suggest that participants might be able to place objects with smaller vicinities of error more accurately. However, note that the reduction in size for the vicinities of error is only modest, and this is consistent with the non-significant linear trends observed over two and more fixations on an object in the present experiments.

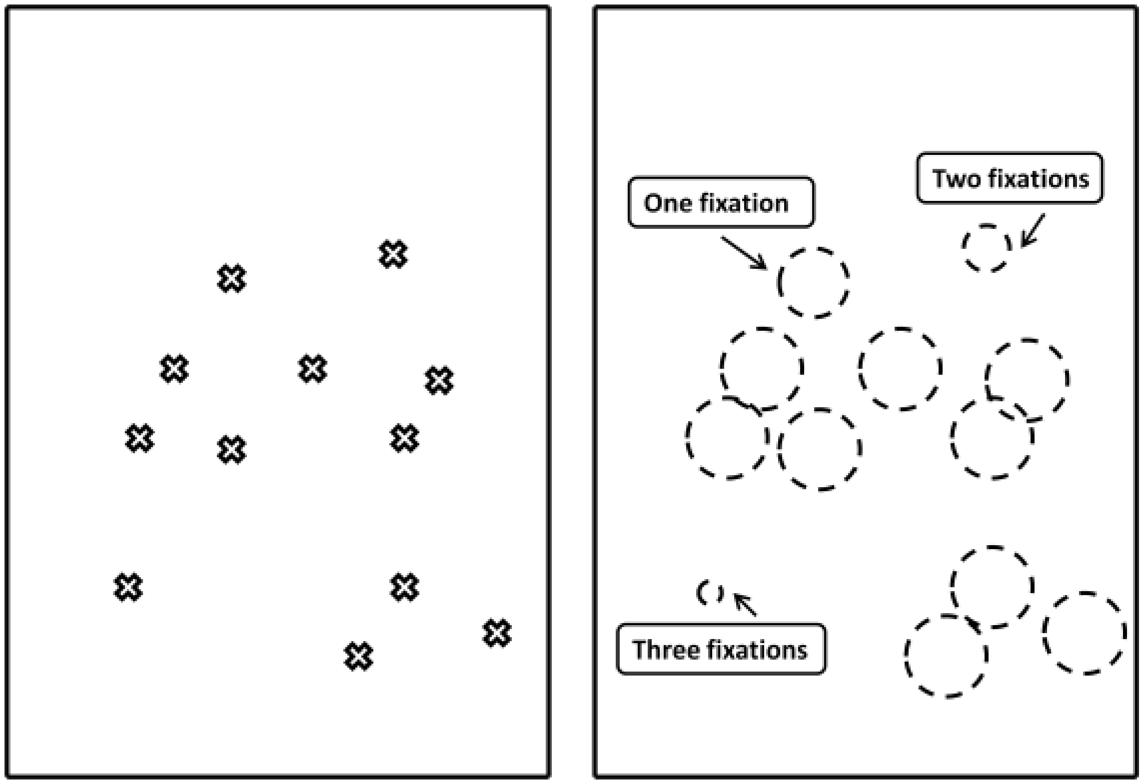

Figure 28. Illustration of the exact positions of objects being encoded (left panel) and a range of differently sized vicinities of error corresponding to the number of fixations each object has received (right). Smaller regions result from more fixations on that object

A Possible Future Experiment

Recall that in the present experiment there were only 12 markers in the cubicle: one for each of the 12 objects. This meant that if participants knew the general region in which an object was located (i.e. they had a vicinity of error for that object) then it was presumably easy for them to choose which of the 12 markers the object should be placed on. This is because in the present experiment the markers were not located close to one other. Instead, the markers were distributed quite evenly across the surfaces in the cubicle. As a result, it is unlikely that the 12 markers substantially competed with one another (in the sense that their vicinities of error would be unlikely to overlap). However, if the test phase of Experiment 5 was modified such that, instead of 12 markers being positioned in the cubicle, there were, say, 156 markers (13 clustered closely around the location of each object) then participants would probably find it difficult to state which of the 13 markers each object belonged on. However, presumably participants would still be able to select which *group* of 13 markers each object belonged to (i.e. the general region that the object was located in), but would not possess the granularity of object location memory to allow them to specify which of 13 closely spaced markers an object was associated with.

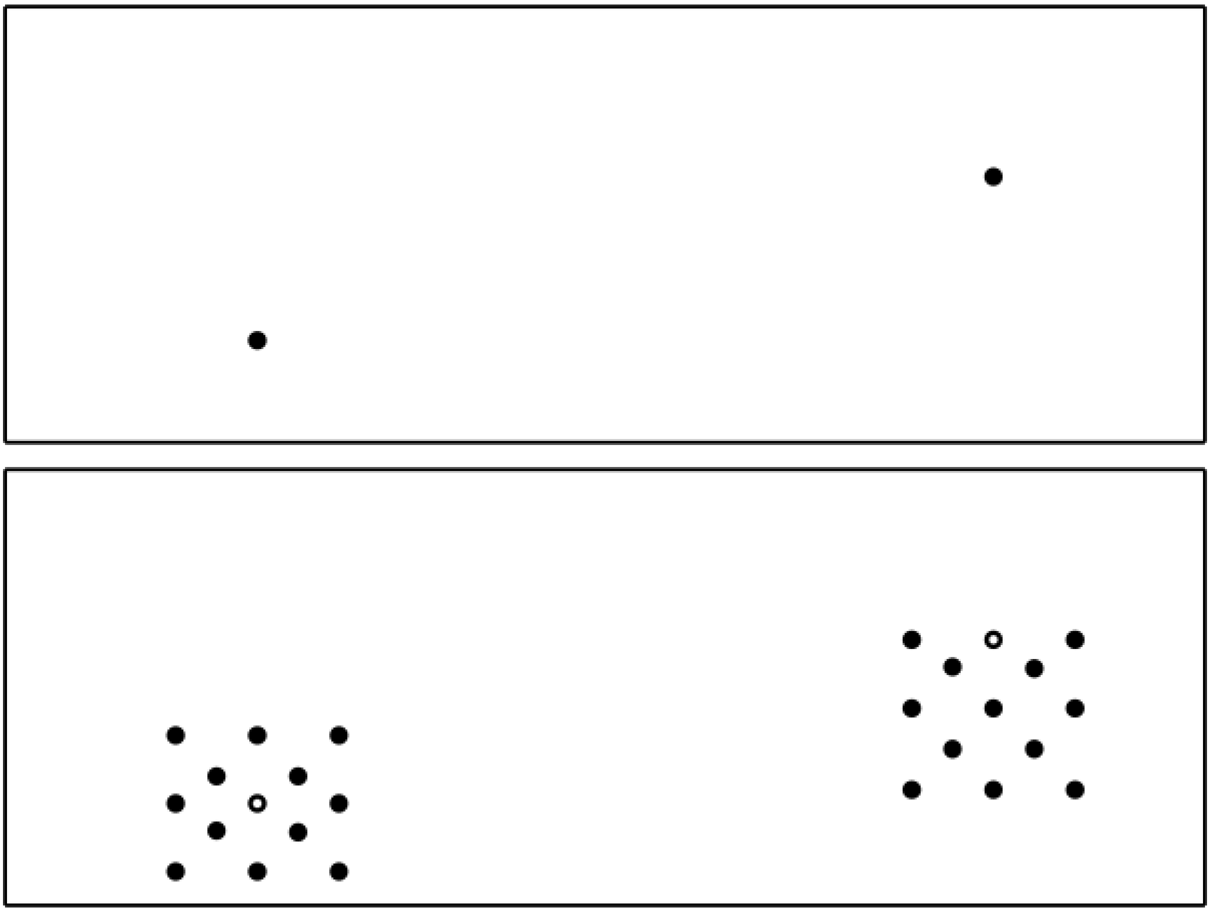

Figure 29. Illustration of no competitor markers in the present study (top panel), and 13 competitor markers in a potential future experiment (bottom panel). In the bottom panel the unfilled circle denotes the actual position of the object.

To illustrate this, Panel A in Figure 29 shows how the markers were distributed on the sill in Experiment 5 and Panel B in Figure 29 shows how the markers might be distributed in a future experiment. Note that Figure 29 is shown from a top-down perspective of the sill surface. This modification to the test phase would allow examination of whether the specific location of an object is encoded perfectly (i.e. participants could easily identify which of the 13 markers an object belonged on) or whether participants simply encode the general region in which the object is located and do not fine-tune this information with subsequent fixations. If the results showed that the accuracy with which participants could select the correct marker from the choice of 13 markers in the region increased as the number of fixations made on an object increased, it would demonstrate that information about the specific location of an object *does* accumulate over separate fixations. Alternatively, the results might show that no matter how many fixations participants make on an object, participants cannot distinguish between the 13 competitor markers. If this was found, it would provide evidence that the information that is encoded about the specific location of an object is quite accurate, but does not improve with subsequent fixations. Put simply, this modification to the test phase would directly investigate whether or not object location information accumulates across the second and subsequent fixations on an object.

To further develop this future experiment, the value of placement error exhibited by participants in earlier experiments in this thesis could be used to dictate how close the competitor markers were placed to one another. For example, if the competitor markers in the new experiment were placed at a distance which exceeded the average placement error taken from earlier experiments (38.52cm) then one might expect that there would be no confusion between markers. Had participants formed vicinities of error each with a diameter of 38.52cm then the markers would not overlap the existing vicinities of error and compete with one another. However, if the competitor markers were positioned at distances smaller than the average placement error then one might expect there to be confusion between the markers. This might be because the markers would fall within the vicinities of error for the objects.

Clearly, this issue could be easily investigated in future research and this would allow more detailed investigation of the way in which information about the specific location of an object is encoded. Specifically, further modification of the test phase as described above would allow detailed examination into the possibility that specific location information is encoded across multiple fixations. The important point to note here is that, contrary to the claims made up to this point in this thesis, it is possible that there might be some granularity associated with accuracy of memory for the specific locations of objects. Note that even if this is the case, the fact remains that there is a highly pronounced and significant difference in object location memory for those objects that were, and those that were not fixated at all. Furthermore, it is also without question the case that there is still a very clear dissociation between the way in which the identities and the locations of objects are encoded.

Analysis of Density of Surface Markers

Whilst the data from Experiment 5 do not allow the issues discussed above to be directly explored, they do allow some tentative investigations to be made into the possibility that competition between object location markers may influence the accuracy with which object locations are remembered. If, as suggested throughout the experiments in this thesis, the specific location of an object is encoded in very precise detail, then it would be expected that fixated objects would be placed on the correct marker more often than expected by chance regardless of how densely populated each surface was. However, if the specific location of an object is encoded in relation to a vicinity of error then one might expect that object location memory would be poor for densely populated surfaces, but accurate for sparsely populated surfaces. This pattern of results would be expected on the basis that as the distance between the markers decreased (i.e., the surface became more densely populated) the degree to which the markers competed with one another increased.

To investigate whether the degree to which a surface was densely populated influenced accuracy of object location memory, the total area (cm^2) of each surface in the cubicle was divided by the number of objects located on each surface in the stimulus photograph. This gave a measure of how densely populated each surface was. The most densely populated surface was the small table (3249 cm^2/3 = 1083 cm^2 per object), the least densely populated surface was the sill (5320 cm^2/2 = 2660 cm^2 per object), and the desk was somewhat densely populated (11475 cm^2/6 = 1912.50 cm^2 per object)[4]. Each surface was then considered in turn, except for the chair which only had one marker and therefore did not afford the potential for competition between location markers.

The probability that an object would be placed on the correct marker by chance on the sill was calculated to be 0.50 as there were only two candidate markers on that surface. A one-sample *t*-test showed that fixated objects that belonged on the sill were placed on the correct marker more often than would be expected by chance (M = 0.75, SE = 0.11), $t(10)$ = 2.24, p = .05. This

[4] The surface dimensions were as follows: desk 135 cm x 85 cm; sill 133 cm x 40 cm; small table 57 cm x 57 cm.

suggests that, as the sill was a sparsely populated surface, there was very little competition between the markers and therefore participants found it easy to recognise which object belonged on each marker.

Next, the same analysis was carried out on the small table (which had three objects located on it in the stimulus photograph). Chance was calculated to be 0.33. However, surprisingly, fixated objects (M = 0.26, SE = 0.12) were placed no better than expected by chance, $t(8) = -.59$, $p > .05$. Unlike the sill data above, participants did not place objects on the correct marker more often than chance, though arguably, this may not be surprising as the small table was the most densely populated surface.

Finally, the same analysis was carried out on the desk (which had six objects on it) and therefore chance was calculated to be 0.17. Fixated objects (M = 0.56, SE = 0.11) were placed on the correct marker more often than expected by chance, $t(10) = 3.60$, $p < .01$. As the desk was less densely populated than the small table, it is perhaps unsurprising that participants demonstrated accurate memory for the correct markers.

Although these analyses are clearly exploratory, and caution must be exerted when interpreting the findings, these analyses are at least suggestive that when competitor markers are more densely populated, participants find it more difficult to accurately remember object locations (presumably due to location marker competition). However, when competitor markers are more dispersed on a surface, they do not compete with one another to the same degree, and participants can recognise which object belongs on which marker comparatively easily. Thus, the present findings offer tentative support for the claim that participants may not encode the precise location of an object perfectly after a single fixation, but may instead encode a location as existing within a vicinity of error that is reduced over subsequent fixations.

How is location information encoded?

An interesting theoretical question concerns the way in which information about the specific locations of objects is encoded. There seem to be at least two possibilities. First, it may be that all locations are equally likely to be encoded. That is, none of the locations are more likely than any of the others to be looked at during the encoding phase. Instead, all locations are as likely as each other to be encoded. Alternatively, the locations might be encoded in a more hierarchical way. In other words, some of the locations might be more likely to be looked during encoding. Clearly this is a complicated issue and the present data do not speak directly to this question.

However, the present data can provide some clues as to the answer to this question. First, memory performance in Experiment 5 demonstrates that fixated objects were placed on the correct surface more often than would be expected by chance (which is 0.25 as there are four surfaces) regardless of the number of fixations (all *p*s < .013 due to the Bonferroni correction). However, non-fixated objects were placed on the correct surface no often than expected by chance, $t(5) = 2.26$, $p > .05$, though note that 45.45% of the data were missing from the "zero fixation" bin. If the missing values are replaced by the previous mean of 0.52 then the one-sample *t*-test with chance is significant for non-fixated objects, $t(10) = 4.32$, $p < .013$. This suggests that once an object has been fixated then the surface on which it belongs has also been encoded. In turn this suggests that the surfaces themselves might be equivalently encoded.

A second piece of evidence which speaks to the theoretical question is found in Table 17. This table provides evidence that the locations of objects might *not* be encoded equivalently. Instead, there is substantial variation of average placement error between the 12 objects. This variation ranges from 28.13 cm for the glass to 60.23 cm for the phone. On this basis it seems as though the locations of some objects might be remembered more accurately than the locations of other objects. However, the likelihood that an object is remembered is also important to consider, and in the example of the glass and the phone it is important to note that the phone was correctly selected 90% of the time whereas the glass was only selected 60% of the time. This is clearly a complicated issue and requires directly examination in future research.

A third analysis which might be relevant relates to the accuracy with which participants placed objects on the correct surfaces and markers in Experiment 5. When participants placed an object on a marker that object could be categorised as either (a) on the correct marker (and therefore on the correct surface by definition); (b) on the correct surface but on the incorrect marker; or (c) on the incorrect surface (and by definition on the incorrect marker). It was found that on average participants placed 5.73 objects on the correct marker. In other words, participants placed about half of the objects on exactly the right marker. What is interesting to examine is what they did with the other objects – which of the possible two errors (right surface but wrong marker or wrong marker) was more prevalent?

It was found that on average participants placed more objects on the incorrect surface (3.82) than on the correct surface but the incorrect marker (2,45) although this difference was not statistically significant, $t(10) = 1.36$, $p > .05$. This is an interesting finding as it demonstrates that when participants did not know the exact location of an object, they were more likely to place an object on the wrong surface than to place an object on the correct surface but the wrong marker. In turn this suggests that memory for the location of an object might be somewhat categorical – it is either known or not known. Participants do not seem to know the general area within which an object belongs i.e. its correct surface but not the correct marker. However, this issue cannot be directly answered using the current data and needs to be directly examined in future experiments.

Summary

The results of Experiment 5 provide further evidence that not only is there a close relationship between encoding and recall for the identities and locations of objects in scenes, but there is a fundamental dissociation between the way in which identities and locations are encoded. However, what is still yet to be conclusively demonstrated is whether object location memory accumulates across fixations or is encoded accurately after the first fixation on an object, not improving over subsequent fixations. The data from Experiment 5 cannot directly address this issue, however, exploratory analyses presented in the Discussion tentatively suggest that there is a possibility that memory for the location of objects improves over fixations subsequent to the first fixation on an object.

7 General discussion

7.1 Chapter outline

The structure of this chapter is as follows: first, the motivation for this thesis is described. Next, the key findings are summarised and a modified version of the Visual Memory Model is proposed. Lastly, the closing remarks are made.

7.2 Motivation for the thesis

Research on picture memory from the 1960s and 1970s demonstrated that participants possess accurate memory for previously viewed scenes even if a large number of scenes were viewed; memory was accurate for 200 scenes (Nickerson, 1965); 612 scenes (Shepard, 1967); or even over 2000 photographs (Standing, Conezio & Haber, 1970). In addition, participants were able to accurately detect changes made to spatial configurations even if the test was delayed by four months (Mandler & Ritchey, 1977) or up to a year (Nickerson, 1968; Fajnstzein-Pollack, 1973).

Taken together, these findings provided evidence that memory for visual environments is typically detailed and robust. As a result, a dominant perspective on memory representations was that a complete and veridical representation of the visual environment is stored in memory. However, research in the 1990s on change blindness showed that participants failed to notice large changes to visual scenes under a number of different circumstances: if the change was made during a saccade (Grimes, 1996); if an actor changed identity across a cut in a movie (Levin & Simons, 1997); if a conversation partner changed identity in a real-world scenario (Simons & Levin, 1998); or if changes were made gradually during scene viewing (Simons, Franconeri & Reimer, 2000). Changes were also introduced using flicker paradigms (Rensink, O'Regan & Clark, 1997). Collectively this body of research concluded that, contrary to the picture memory literature of the 1960s and 1970s, observers do *not* have detailed and robust memory for visual environments. As a result, the traditional view of visual memory for visual scenes was challenged, and alternative explanations were suggested.

Rensink (2000) proposed a *Coherence Theory* to explain how visual information is encoded and stored. This perspective asserts that only very abstract information is stored in memory once a scene is no longer available to view. This explains why participants tend to find it difficult to detect changes made to objects in the environment. A more extreme perspective was suggested by O'Regan (1992) who claimed that observers use the world as an "outside memory" and do not store any detailed information in memory.

The introduction of eye tracking methodology led to a shift in perspective regarding memory for visual scenes. When eye movements were recorded during flicker paradigms it was found that changes *could* be easily detected, but were typically noticed only when the change was made after the changed region had been fixated both in its pre- and post-change state (Henderson & Hollingworth, 1999; Hollingworth, Williams & Henderson, 2001; Hollingworth & Henderson, 2002). In other words, it demonstrated that the entire visual scene is not overtly attended to at the same time, and instead visual attention is typically allocated to the object that is being fixated at that time. In addition, it suggested that unless an object was fixated both before and after the change, the change was unlikely to be noticed. Furthermore, the finding that participants could easily detect changes once the changed region had been fixated demonstrated that detailed information about objects in scenes *is* stored in detail in memory. Consequently, Hollingworth and Henderson (2002) proposed a *Visual Memory Model* to describe how information about objects in scenes is acquired and stored. The model states that, contrary to the claims of Coherence Theory, detailed information about objects is encoded during scene viewing and stored in memory. This theoretical framework formed the basis for the five experiments reported in this thesis.

Subsequent empirical research investigated the degree to which detailed information about objects in scenes is encoded and stored in memory. This research either investigated how encoding processes influenced memory for the identities of objects (Melcher, 2001; Melcher & Kowler, 2001; Hollingworth, 2004; Liu & Jiang, 2005; Hollingworth, 2006; Melcher, 2006; Hollingworth, 2009) or investigated the extent to which information about the locations of objects is stored in memory whilst neglecting to investigate associated encoding processes (Uttl and Graf, 1993; Postma & de Haan, 1996; James & Kimura, 1997; Postma, Izendoorn & De Haan, 1998; Kohler, Moscovitch & Melo, 2001; Levy, Astur & Frick, 2005; Iachini, Sergi, Ruggiero & Gnisci, 2005).

Collectively, this research provided convincing evidence that detailed information about both identities and locations of objects is stored in memory. However, critically, the precise nature of the relationship between visual encoding and recall for both the identities of objects and the locations of objects was not directly investigated and consequently is not well understood. As a result, a novel paradigm was proposed to directly investigate how the identities and locations of objects are encoded during scene viewing, and how memory exhibited for the objects is related to the encoding behaviour.

7.3 Key Findings

The experiments reported in this thesis have generated several key findings and each one will be described in turn.

i. There is a very close relationship between encoding and recall for objects and furthermore, information about object identities and specific object locations is encoded differently

The first key finding was initially observed in Experiment 1, and then replicated in Experiments 2, 3, 4 and 5. The results of these experiments showed that there is a very close relationship between encoding and recall for objects. The relationship for object identity memory is considered first. Memory for object identities was no better than chance for non-fixated objects, but was better than chance for fixated objects. This is very important as it suggests that in order for the identity of an object to be encoded that object needs to be directly fixated. However, if an object is not fixated, it is no more likely than chance to be remembered. In addition, the results showed that memory for the identity of objects was better for objects fixated twice (or more than twice) than for objects fixated only once. This suggests that information about the identity of objects is encoded during multiple fixations on an object.

Importantly, this finding suggests that memory representations for the identity of objects accumulate detail the more times an object is looked at. Chapter two provided a plausible explanation for this accumulation of object identity information. It is possible that information about object identities accumulates across separate fixations because the identity of an object is comprised of a range of different inherent object properties. Initially, the object file that is formed for an object may contain basic information about its form and identity (e.g., a pen). However, if more fixations are made on the same object, more information about that object's identity can be encoded, embellishing the detail associated with the object file and leading to a more detailed memory representation. Thus, over multiple fixations, the memory representation might progress from "a pen" to "a small thin blue Bic pen", meaning that memory is more detailed for objects fixated twice or more than for objects fixated only once. In this way, accumulation of detail in relation to object identities can be explained.

There is also a very close relationship between encoding and recall for the specific locations of objects. As for object identities, memory for the specific locations of non-fixated objects was no better than chance. However, memory for the specific locations of fixated objects was better than chance. This demonstrates that if the specific location of an object is to be remembered at test, that object must be directly fixated. If it is not fixated, memory for its specific location is no better than chance. Critically, information about the specific location of objects seems to be encoded in a different way to information about object identities. The results from Experiments 1, 2 and 3 suggest that memory for object location does not improve across separate fixations. Instead, the results suggested that the specific locations of objects are encoded accurately after only one fixation on an object and do not improve with subsequent fixations on the same object. If this is correct, this means that if the specific location of an object is to be remembered then an object only needs to be looked at once.

However, the results of Experiments 4 and 5 provided some tentative evidence that memory for object locations might actually accumulate across separate fixations, albeit in a different way to object identities. Whereas memory for object identity seems to improve gradually across fixations, the data from Experiments 4 and 5 suggest that information about the specific locations of objects initially improves very rapidly (from zero fixations to one fixation), but then improves much more slowly (from one to two, or two to three fixations). Note that the data presented in this thesis in relation to whether information about specific locations of objects accumulates across the second and third fixations on an object were not statistically reliable. However, consistent numerical trends were observed in both Experiments 4 and 5. Furthermore, these effects were demonstrated independently and using two different test methodologies. It seems likely that with more power, future experiments could produce similar statistically reliable effects. Clearly, further research is required to directly investigate this possibility and a potential future experiment has been proposed in Chapter Six. However, what can be firmly stated is that there is a very close relationship between encoding and recall for the identities and locations of objects in scenes, and furthermore the nature of the relationship is quite different for identities and specific locations.

ii. The relationship between encoding and recall can be modulated using a lower-level attentional capture technique

Experiment 2 demonstrated that the fundamental relationship between encoding and recall for the identities and locations of objects in scenes (as described in (i) above) can be influenced by using a lower-level attentional capture technique during the encoding phase. Green boxes surrounded half of the objects in the stimulus photograph and this rendered those objects highly visually salient. As a result, the boxed objects were fixated more often and for longer than non-boxed objects during the encoding phase and consequently were selected more often, earlier and more confidently during the test phase. Crucially, however, the green boxes did not change the basic oculomotor characteristics as objects were encoded. That is, the basic relationship between encoding and recall was no different for boxed and non-boxed objects - the boxes did not make the boxed objects themselves more memorable. Instead, the boxed objects were more visually salient and were more likely to be fixated. Because they were more likely to be fixated, they were more likely to be remembered at test. Therefore, it was demonstrated that visual attention can be driven to certain objects in the scene at the expense of other objects, and these differences in encoding behaviour lead directly to differences in memory performance at test. This finding is important as it demonstrates that the memory that participants form for a visual scene can be fundamentally influenced in the encoding phase by the use of simple attentional capture techniques to render some objects more visually salient than others. In this way, it further emphasises the close relationship between the visual system and memory as described in (i) above.

iii. Clear primacy and recency effects are not observed in memory for the identities and locations of objects

The results of Experiment 3 provided no clear evidence for primacy and recency effects in memory for the identities and locations of objects in scenes. Recall that previous research on word lists (e.g., Postman & Phillips, 1965) found that words presented at the beginning and ends of sequences of words were remembered more often than words presented in the middle of the sequence. However, as this pattern of results was not found in Experiment 3, it seems likely that the way in which words in lists are encoded is different to the way objects in scenes are encoded. As Experiments 1 and 2 in this thesis demonstrated that, under free-viewing conditions, participants demonstrate variation in the amount of time that objects are fixated during encoding, it seems likely that imposing a restriction on the amount of time each object could be looked at in Experiment 3 meant that some of the objects were processed to a greater degree than they might have been during free-viewing, and some were processed to a lesser degree. This therefore meant that some objects were remembered accurately (the objects that received more visual processing than during free-viewing) and some objects were not remembered accurately (objects that received less visual processing than during free viewing). In support of this claim, a negative correlation was obtained between average fixation duration made on each object during the encoding phase of Experiment 1 and the probability that an object was selected in Experiment 3 (see Chapter 4). The failure to obtain robust primacy and recency effects, therefore, suggests that the order in which objects are looked at in a scene does not influence the accuracy with which the identities and locations of those objects are remembered. Instead, as demonstrated in Experiments 1, 2, 4 and 5, what seems to influence memory accuracy is the amount of fixations that are made on an object. In other words, objects fixated three times in the middle of a sequence are likely to be remembered more accurately than objects fixated once at the beginning or end of a sequence.

iv. Information about the overall configuration of objects is encoded early and without the need to fixate each object

The next important finding is that information about the overall configuration of objects (the spatial map according to the Visual Memory Model) seems to be encoded early in scene viewing and without the need for each of the objects in the scene to be directly fixated. The results of Experiment 1 showed that best fit error, which is considered to reflect memory for the overall configuration of objects, was no better after a 10 s viewing duration than after a 5 s viewing duration. This suggests that information about the overall configuration of objects is largely, or completely, encoded during the first 5 s spent viewing a scene.

The results from Experiment 4 provided further understanding about the way in which the overall configuration of objects seems to be encoded. It was found that at test, one of the objects was placed close to each of the 12 possible locations in the cubicle, regardless of the number of fixations that had been made on the object at that location during the encoding phase. That is, even if an object was not fixated during encoding, its location seemed to have been encoded in the spatial map. This was demonstrated by the finding that participants could place one of the objects close to all of the 12 locations even if a location had not been fixated.

This finding is interesting because it shows that there is a markedly different relationship between encoding and recall for the overall configuration of objects compared to the relationship between encoding and recall described above for object identities and specific object locations. It appears that the spatial map is formed early in scene viewing. Although the present data cannot demonstrate that it forms any quicker than 5 s, it seems highly likely that it is encoded substantially earlier. From a theoretical perspective, it seems reasonable to suggest that the spatial map is fully encoded before an object file is created for the first object that is fixated in the scene. This is because it is proposed that each individual object file is indexed to one of the positions in the spatial map once the object file has been instantiated. As average fixation durations during scene viewing are approximately 330ms (Henderson, 2003) it seems reasonable to suggest that the spatial map probably forms within 300-400 ms of scene viewing.

v. A 24 hour test delay affects memory for the specific locations of objects, but not object identities or the overall configuration of objects

The next important finding concerns the influence of the 24 hour test delay (which was implemented in Experiment 1). Participants who experienced a 24 hour test delay remembered the same number of objects as participants with an immediate test and their memory for the overall configuration of objects was comparable. However, memory for the specific location of objects was significantly poorer when the test was delayed by 24 hours rather than when the test was immediate. Thus, the binding between identities and locations decays somewhat after a delay of 24 hours. Note that memory for the identities of objects and the overall configuration of objects remains accurate after a delay, but the link between the two does not appear to be so resilient. This raises the question of why memory for the specific locations of some objects was poorer after a 24 hour delay than at immediate test (whilst memory for the overall configuration of objects as no poorer). Presumably, some objects are consolidated into long-term memory whilst others are not. The objects that are consolidated are remembered accurately after a delay and their specific locations are also remembered accurately, explaining why memory for object locations after a delay remains better than chance. However, presumably memory for the objects that are not consolidated decays during the delay, resulting in less accurate recall at test.

Further analyses demonstrated that if an object was to be consolidated into long-term memory then it needed to be fixated at least twice. Objects that were fixated at least twice were remembered accurately regardless of when the memory test occurred. However, objects fixated once were remembered accurately if the test was immediate, but were remembered no better than chance if the test was delayed. This finding indicates that consolidation of information from short-term to long-term memory happens during the second fixation on an object. Therefore, at a delayed test, the objects that are fixated twice are likely to be remembered and placed accurately, but objects fixated only one are less likely to be remembered and placed accurately.

vi. Confidence ratings typically mirror memory performance, although not for the overall configuration of objects

The final important finding was that confidence for object identities and specific locations tended to closely mirror memory for object identities and specific object locations. Confidence ratings were not recorded in Experiment 1, but in Experiment 2 participants gave higher confidence ratings for the identities of the boxed objects than the non-boxed objects. This closely matches the finding that participants remembered more boxed objects than non-boxed objects, and is probably because boxed objects received more fixations and longer fixations during encoding than non-boxed objects.

In Experiment 3 the confidence ratings were closely related to the order in which objects were placed at test, with confidence highest for objects placed early during the test phase and lowest for objects placed late in the test phase. It was also demonstrated that memory for object locations was most accurate for objects placed early, and least accurate for objects placed late. This suggests that participants chose to place objects that they were most sure about first, and least sure about last, and both the placement error values and confidence values reflected this.

In Experiment 4, however, there was an interesting dissociation between confidence and memory for the overall configuration of objects. Whilst memory for the overall configuration of objects was accurate regardless of the number of fixations made on each object, confidence for those locations increased with fixations. This suggests that participants possessed better memory for the spatial map than they were aware of. Perhaps, the spatial map forms somewhat automatically but participants possess poor meta-cognitive awareness as to the degree to which they have encoded the spatial map in memory. They therefore rate their confidence for the overall configuration of objects lower than their memory performance indicates they should. In Experiment 4 participants' confidence for the specific locations of objects increased linearly as the number of fixations increased. This mirrored the data for memory of the specific locations of objects which also seemed to improve with fixations, albeit gradually (see (i) above). This suggests that participants were aware of their memory for the specific locations of objects. This was also found in Experiment 5 where confidence for identities and locations improved linearly as the number of fixations made on an object increased.

Taken together, these findings suggest that participants generally possess some degree of meta-cognitive awareness with regards to the accuracy of their memory for the identities and specific locations of objects. Participants tend to rate their confidence as being high for the identities of objects that they select and place accurately and low for objects that they select and place less accurately. However, similar meta-cognitive awareness in relation to memory for overall spatial configuration of objects did not seem to occur. As a result, in Experiment 4 participants seemed to under-rate their accuracy for the overall configuration of objects. Confidence ratings improved as the number of fixations increased, but memory accuracy for the overall configuration of objects was accurate regardless of the number of fixations that were made on an object. In turn, this is consistent with the idea that the spatial map forms in memory somewhat automatically and without participants directly allocating overt attention to each of the objects in the scene.

7.4 Modified Visual Memory Model

As Hollingworth and Henderson's Visual Memory Model falls short of explaining all of the findings described above, it seems appropriate to re-consider it in light of the key findings reported in this thesis. As such, the following section describes a modified Visual Memory Model that is proposed on the basis of the experimental data reported in this thesis. The modified model is intended to augment the existing framework of Henderson and Hollingworth.

1. Formation of the Spatial Map

According to this modified version of the model, the first thing to happen upon encountering a visual scene is that a spatial map is formed. This spatial map is hypothesised to act as a frame within which the locations occupied by objects are marked. It forms rapidly and without the need for visual attention (via fixations) to be allocated to every object in the scene (e.g., Zechmeister, McKillip, Pasko, & Bespalec, 1975; von Wright, Gebhard, & Karttunen, 1975; Mandler, Seegmiller, & Day, 1977; Hasher & Zacks, 1979; Simons, 1996).
It is clear from current data that the spatial map forms within 5 s, but in reality, it is likely that it actually forms within approximately 400 ms of scene

onset, that is, by the end of the first fixation on an object in the scene. The reason that the spatial map is proposed to form so quickly is that once an object file has been created for the first fixated object, according to the Visual Memory Model, that object must be allocated to a position within the spatial map. If the spatial map has not been formed once an object file has been created, then there is nothing to index the object file to. Therefore, the modified model states that the spatial map is fully instantiated before the first object file has been formed. As the spatial map is believed to form early in scene viewing parallels were drawn between the spatial map and the notion of scene gist. The spatial map and scene gist may be similar in at least two ways. First, both seem to form early in scene viewing. Second, both seem to be formed on the basis of visually salient information. Research has indicated that scene gist might be extracted on the basis of colour (Oliva & Schyns, 2000; Goffaux, Jacques, Mouraux, Oliva, Schyns & Rossion, 2005; Castelhano & Henderson, 2008) or on the basis of the spatial layout of a scene (Biederman, Mezzanotte & Rabinowitz, 1982). In a similar way, it is proposed that the spatial map marks placeholders at the locations in a scene which contain visually salient information – such as the objects.

Parallels were also drawn in this thesis between the spatial map and saliency maps. Saliency maps are proposed to guide oculomotor behaviour in terms of decisions as to where visual attention is allocated within a scene, with saccades being directed to the most visually salient regions. It could be stated that the spatial map is somewhat analogous to the notion of a saliency map. Potentially, the spatial map might play a role in determining which visual attention is allocated during scene viewing in that the placeholders within the spatial map typically correspond with visually salient regions of the scene. However, it would be premature and somewhat dangerous to claim that the spatial map and the saliency map are the same thing. Instead, what can be stated is that they seem to be similar and may share some functional properties.

In addition, it is proposed that this spatial map is robust and does not decay easily. Accordingly, Experiment 1 demonstrated that best fit error was no poorer when tested after a 24 hour delay than after an immediate test.

2. Instantiation of Object Files for Fixated Objects

The next stage of this modified Visual Memory Model states that overt visual attention is allocated via fixations to one of the objects in the scene at a time. During the first fixation on an object, an object file is created for that object. At first, the object file contains abstract information about the object, such as the identity of the object. As the number of fixations made on an object increases, the degree of detail associated with its identity also increases. This means that over multiple fixations on the same object, the object file can become more and more detailed, as additional information pertaining to the identity of the object accumulates (e.g., the shape, size, colour etc.) The experiments reported in this thesis have provided evidence that information about object identity is encoded during at least two separate fixations on an object. However, it is possible that information about the identity of an object is encoded during more than two fixations on the same object. The present data do not inform whether encoding of object identity information continues indefinitely, or whether it ceases after a certain number of fixations or a certain amount of time. What is clear is that object information can be encoded during multiple fixations on the same object. In addition, it does not seem to matter whether these multiple fixations are successive or interrupted by fixations made to other objects or regions in the scene.

3. Indexing of Object Files to the Spatial Map

A second important event that seems to occur during the first fixation on an object is that the object file, which contains the information about the identity of the object, is linked to one of the vacant placeholders, or slots, within the spatial map. The results of Experiment 1, 2, 3 and 5 in this thesis provided convincing evidence that this binding of object identities to locations in the spatial map happens during the first fixation made on an object. The results consistently showed that non-fixated objects were placed no better than chance, suggesting that those non-fixated objects had not been indexed to a position in the spatial map (though recall that non-fixated objects could be placed close to one of the locations in the spatial map). However, fixated objects were consistently placed better than chance and better than non-fixated objects. This substantial improvement in accuracy for the specific locations of objects from zero fixations to one fixation was observed in Experiments 1, 2, 3 and 5. Clearly, a single fixation on an object allows for the identity of an object to be encoded very accurately.

What is still unclear, however, is whether this information about the specific location of an object is refined across subsequent fixations, or whether it plateaus after one fixation and does not improve with subsequent fixations. The results of Experiments 1 and 2 seemed to suggest that no improvement in object location memory was observed after one fixation, but the results of Experiments 4 and 5 provided tentative evidence that this information does accumulate in memory across separate fixations, albeit gradually. From a theoretical perspective it seems more plausible that object location information does improve (slightly) across subsequent fixations, and in Chapter Six a potential mechanism for this improvement was proposed. Therefore, in relation to the modified Visual Memory Model, what is proposed is that during the first fixation on an object, the exact position of that object is not encoded. Instead, an object location with a vicinity of error is encoded for the location of an object. That is, a small region of the scene is encoded within which lies the exact location of an object. Having made only a single fixation on an object, participants can assert that an object belongs somewhere within that vicinity of error, but are unlikely to be able to remember the exact position within the vicinity of error. The current data suggest that, for the scene used in the present experiments, initially this vicinity of error has a diameter of approximately 40cm.

This theoretical account, therefore, stipulates that if multiple fixations are made on an object, the size of the vicinity of error associated with that object might decrease. In other words, participants may be able to refine the accuracy of object location memory over fixations. This suggestion has implications for the nature of the spatial map. As stated above, it is proposed that the spatial map consists of placeholders that represent each of the objects in the scene. What seems most plausible is that these placeholders are in fact vicinities of error which mark out a region of the scene within which each object is located, rather than marking the exact position of each object. A possible future experiment was proposed in Chapter Six which would allow direct investigation into the possibility that memory for object locations is refined across separate fixations.

4. Consolidation of Information from Short-term to Long-term Memory

The data from Experiment 1 also provided detail regarding the consolidation of information from short-term memory to long-term memory. This was by virtue of the experimental manipulation of the 24 hour delay for half of the participants. The data showed that consolidation of identity information from short-term memory to long-term memory seems to happen during the *second* fixation on an object. This is on the basis that the identities of objects fixated only once were remembered accurately only when the test was immediate, but not when the memory test took place after a 24 hour delay. However, the identities of objects fixated twice or more were remembered accurately no matter when the memory test occurred. Note that this does not necessarily mean that two separate fixations are required if an objects is to be consolidated into long-term memory. Potentially, this consolidation might happen after the object has been fixated for a certain amount of time. However, as the present experiments used the number of fixations as the measure of visual encoding, it is not possible to know for sure whether or not this is the case. What can be firmly concluded, however, is that consolidation from short-term memory to long-term memory does not happen immediately upon initial fixation of an object. In addition, if an object is not consolidated into long-term memory, it will be remembered accurately at an immediate test, but remembered poorly at a delayed test.

7.5 Closing Remarks

In summary, the five experiments presented in this thesis have used a novel paradigm to provide significant insight into the way in which information about the identities and locations of objects is both encoded during inspection of a scene, and stored in memory. In order for the identities and specific locations of objects to be encoded, objects need to be directly fixated. Furthermore, the results of the present experiments have shown that there is a very close relationship between encoding and recall for identities and locations of objects. Importantly, the relationship is different for identities and specific locations. Memory for object identity is encoded over multiple fixations, thereby accumulating, whereas the memory for specific locations of objects is primarily encoded during the first fixation on an object. It remains unclear whether the specific locations of objects continue to be encoded over multiple fixations, or whether encoding ceases after a single fixation. The evidence seems to suggest that object locations may accumulate in memory, albeit quickly at first and then more gradually. In contrast, information about the overall configuration of objects is encoded in detail early in scene viewing and without individual objects needing to be fixated. There are a large number of possible manipulations that can be made to the novel paradigm to allow further investigation into the nature of memory representations and the time course of encoding for different object properties. The present thesis represents an initial investigation into the important relationship between encoding processes during inspection of a visual scene, and memory for aspects of that scene. Future research is required to further develop understanding of this aspect of human cognitive processing.

8 References

Aginsky, V. & Tarr, M. (2000). How are different properties of a scene encoded in visual memory? *Visual Cognition, 7 (1),* 147-162.

Agostinelli, G., Sherman, S. J., Fazio, R. H., & Hearst, E. S. (1986). Detecting and identifying change: Additions versus deletions. *Journal of Experimental Psychology: Human Perception and Performance, 12(4),* 445-454.

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm. *Cognition, 93*, 79-87.

Alvarez, G. A., & Oliva, A. (2007). The role of global layout in visual short-term memory. *Visual Cognition, 15,* 70-73.

Beck, M. R., Levin, D. T., & Angelone, B. (2007). Change blindness blindness: Beliefs about the roles of intention and scene complexity in change detection. *Consciousness and Cognition, 16,* 31-51.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14(2),* 143-177.

Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90,* 513-546.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology of perception in art.* Chicago: University of Chicago Press.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12(6),* 1017-1040.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 33 (4),* 753–763

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34(3),* 660–675.

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

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24 (4),* 403–416.

Chun, M. M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240-247.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(3),* 474-479.

DeAngelus, M., & Pelz, J. (2009). Top-down control of eye movements: Yarbus revisited. *Visual Cognition, 17(6/7),* 790-811.

De Goede, M., & Postma, A. (2008). Gender differences in memory for objects and their locations: A study on automatic versus controlled encoding and retrieval contexts. *Brain and Cognition, 66*, 232–242.

Eals, M., & Silverman, I. 1994. The Hunter-Gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology, 15,* 95–105.

Fajnstzejn-Pollack, G. (1973). A developmental study of decay rate in long-term memory. *Journal of Experimental Child Psychology, 16,* 225-235.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral And Brain Sciences, 22,* 661–721.

Flin, R. H., & Shepherd, J. W. (1986). Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning, 5,* 29-38.

Franconeri, S. L. & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65 (7),* 999-1010.

Glanzer, M. & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5 (4),* 351-360.

Gibson, B. S. & Jiang, Y. (1998). Surprise! An unexpected colour singleton does not capture attention in visual search. *Psychological Science, 9 (3),* 176-182.

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12,* 878–892.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. In K. Akins (Ed.), *Perception (pp.* 89-110). New York: Oxford University Press.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108 (3),* 356-388.

Hayhoe, M. (2000). Vision using routines: A functional account of vision. *Visual Cognition, 7 (1/2/3),* 43-64.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*, 51–55.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Science, 7(11),* 498-504.

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 10,* 438-443.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review, 16 (5),* 850-856.

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.

Hollingworth, A. (2005a). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31 (3),* 396-411.

Hollingworth, A. (2005b). Memory for object position in natural scenes. *Visual Cognition, 12(6),* 1003-1016.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 58-69.

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.*

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

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory and Cognition, 29 (2),* 296-304.

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

Iachini, T., Sergi, I., Ruggiero, G., & Gnisci, A. (2005). Gender differences in object location memory in a real three-dimensional environment. *Brain and Cognition, 59,* 52-59.

Intraub, H. (1997). The representation of visual scenes. *Trends in Cognitive Sciences, 1(6),* 217-222.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology, 23*, 420–456.

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 307-317.

Irwin, D. E., Colcombe, A. M., Kramer, A. F., & Hahn, S. (2000). Attentional and oculomotor capture by onset, luminance and color singletons. *Vision Research, 40,* 1443-1458.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489-1506.

Itti, L. & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2,* 194-203.

James, T. W., & Kimura, D. (1997). Sex differences in remembering the locations of objects in an array: Location-shifts versus location-exchanges. *Evolution and Human Behavior, 18*, 155-163.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26,* 683-702.

Johansson, R., Holsanova, J., & Holmqvist, K. (2006). Pictures and spoken descriptions elicit similar eye movements during mental imagery, both in light and in complete darkness. *Cognitive Science, 30,* 1053-1079.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 21 (4), 1008-1018.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception and Psychophysics, 43 (4),* 346-354.

Kessels, R. P. C., Feijen, J., & Postma, A. (2005). Implicit and explicit memory for spatial information in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 20(2-3),* 184-191.

Kessels, R. P. C., Hobbel, D., & Postma, A. (2007). Aging, context memory and binding: A comparison of "what, where and when" in young and older adults. *International Journal of Neuroscience, 117,* 795-810.

Kohler, S., Moscovitch, M., & Melo, B. (2001). Episodic memory for object location versus episodic memory for object identity: Do they rely on distinct encoding processes? *Memory & Cognition, 29 (7),* 948-959.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: An investigation with higher order statistics. *Spatial Vision, 13*, 201–214.

Kuhn, G., Tatler, B. W., Findlay, J. M. & Cole, G. G. (2007) Misdirection in magic: Implications for the relationship between eye gaze and attention. *Visual Cognition, 16(2),* 391-405.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B, 352*, 1231–1239.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559-3565.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3 (12),* 1340-1345.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311–1328.

Levin, D. T., Drivdahl, S. B., Momen, N., & Beck, M. R. (2002). False predictions about the detectability of unexpected visual changes: The role of metamemory and beliefs about attention in causing change blindness blindness. *Consciousness and Cognition, 11,* 507–527.

Levin, D.T., Momen, N., Drivdahl, S.B., & Simons, D.J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition, 7,* 397–412.

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin and Review, 4(4),* 501-506.

Levin, D. T., Simons, D. J., Angelone, B., & Chabris, C. F. (2002). Memory for centrally attended changing objects in an incidental real-world change detection paradigm. *British Journal of Psychology, 93,* 289-302.

Levy, L. J., Astur, R. S., & Frick, K. M. (2005). Men and women differ in object memory but not performance of a virtual radial maze. *Behavioral Neuroscience, 119*, 853-862.

Liu, K., & Jiang, Y. (2005). Visual working memory for briefly presented scenes. *Journal of Vision, 5(7),* 650-658.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*, 6-14.

Liversedge, S. P., Rayner, K., White, S. J., Findlay, J. M., & McSorley, E. (2006). Binocular coordination of the eyes during reading. *Current Biology, 16,* 1726-1729.

Liversedge, S. P., Rayner, K., White, S. J., Vergilino-Perez, D., Findlay, J. M., & Kentridge, R. W. (2004). Eye movements when reading disappearing text: is there a gap effect in reading? *Vision Research, 44,* 1013–1024.

Mandler, J. M., & Ritchey, G. H. (1977). Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 3(4),* 386-396.

Mandler, J. M., Seegmiller, D., & Day. J. (1977). On the coding of spatial information. *Memory & Cognition, 5,* 10-16.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10 (3),* 165-188.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 563-581.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics, 25*, 221-224.

Melcher, D. (2001). The persistence of memory for scenes. *Nature, 412,* 401.

Melcher, D. (2006). Accumulation and persistence of memory for natural scenes *Journal of Vision, 6*, 8–17.

Melcher, D. & Kowler, E. (2001). Visual scene memory and the guidance of saccadic eye movements. *Vision Research, 41,* 3597–3611.

Metzger, R. L., & Antes, J. R. (1983). The nature of processing early in picture perception. *Psychological Research, 45,* 267–274.

Miller, L. K., & Santoni, V. (1986). Sex differences in spatial abilities: Strategic and experiential correlates. *Acta Psychologica, 62,* 225–235.

Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D'Esposito M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging, 15 (3),* 527-541.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66(8),* 1268-1281.

Mounts, J. R. W. (2000). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception & Psychophysics, 62*, 969−983.

Munkres, J. (1957). Algorithms for the assignment and transportation problems. *Journal of the Society for Industrial and Applied Mathematics, 5 (1),* 32-38.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology, 19,* 155–160.

Nickerson, R. S. (1968). A note on long-term recognition memory for pictorial material. *Psychonomic Science, 11,* 58.

Nickerson, R. S., & Adams, J. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11*, 287-307.

Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7 (1/2/3),* 93-106.

Northdurft, H. (1993). Saliency effects across dimensions in visual search. *Vision Research, 33 (5-6),* 839-844.

Oliva, A., & Schyns, P. G. (1997). Course blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46,* 461–488.

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: looking without seeing and seeing without looking. *Visual Cognition, 7*, 191–211.

Orsini, A., Chiacchio, L., Cinque, M., Cocchiaro, C., Schiappa, O., & Grossi, D. (1986). Effects of age, education, and sex on two tests of immediate memory: A study of normal subjects from 20 to 99 years of age. *Perceptual and Motor Skills,* 63, 727–732.

Parker, R.E. (1978). Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4 (2),* 284-293.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42 (1),* 107-123.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics, 44 (4),* 369-378.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15 (4),* 587-595.

Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32 (1),* 3-25.

Postma, A. 1996. Reconstructing object locations in a 7x7 matrix. *Psychologische Beitrage, 38,* 383-392.

Postma, A., & De Haan, E.H.F. (1996). What was where? Memory for object locations. *Quarterly Journal of Experimental Psychology, 49 A,* 178-199.

Postma, A., Izendoorn, R., & de Haan, E. H. F. (1998). Sex differences in object location memory. *Brain and Cognition, 36,* 334-345.

Postma, A., Jager, G., Kessels, R. P. C., Koppeschaar, H. P. F., & van Honk, J. (2004). Sex differences for selective forms of spatial memory. *Brain and Cognition, 54 (1),* 24-34.

Postma, A., Kessels, R. P. C., & van Asselen, M. (2008). How the brain remembers and forgets where things are: The neurocognition of object–location memory. *Neuroscience and Biobehavioral Reviews, 32*, 1339–1345

Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *The Quarterly Journal of Experimental Psychology, 17 (2),* 132-138.

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

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin, 124(3),* 372-422.

Rayner. K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 7(1),* 167-179.

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive free-viewing of eye movements. *Psychological Science, 14,* 385–388.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20(1),* 6-10.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems, 10,* 1-10.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7,* 17-42.

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.

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

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

Rowland, M. L. (1990). Self-reported weight and height. *American Journal of Clinical Nutrition, 52,* 1125-1133.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200.

Schofield, A. J., Bishop, N. J., & Allan, J. (2006). Oscillatory motion induces change blindness. *Acta Psychologica, 121*, 249–274.

Schonfield, D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20,* 228-236.

Schumann-Hengsteler, R. (1992). The development of visuo-spatial memory: How to remember location. *International Journal of Behavioral Development, 15(4),* 455-471.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependant scene recognition. *Psychological Science, 5,* 195-200.

Shaffer, W. O., & Shiffrin, R. M. (1972). Rehearsal and storage of visual information. *Journal of Experimental Psychology, 92(2),* 292-296.

Shepard, N. (1967). Recognition memory for words, sentences and pictures. Journal of *Verbal Learning and Verbal Behavior, 6,* 156-63.

Shoquierat, M. A., & Mayes, A. R. (1991). Disproportionate incidental spatial-memory and recall deficits in amnesia. *Neuropsychologia, 29(8),* 749–769.

Shore, D. I., Burack, J. A., Miller, D., Shari, J., & Enns, J. T. (2006). The development of change detection. *Developmental Science, 9 (5),* 490-497.

Silva, M. M., Groeger, J. A., & Bradshaw, M. F. (2006). Attention–memory interactions in scene perception. *Spatial Vision, 19 (1),* 9–19.

Simons, D. J. (1996). In sight, out of mind: When object representations fail. *Psychological Science, 7,* 301-305.

Simons, D. J., Chabris, C. F., Schnur, T. T., & Levin, D. T. (2002). Evidence for preserved representations in change blindness. *Consciousness and Cognition, 11,* 78-97.

Simons, D. J., Franconeri, S. L., & Reimer, R. L. (2000). Change blindness in the absence of a visual disruption, *Perception, 29,* 1143-1154.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people in a real-world interaction. *Psychonomic Bulletin and Review, 5(4),* 644-649.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9(1),* 16-20.

Smyth, M. M., & Scholey, K. A. (1996). Serial order in spatial immediate memory. *The Quarterly Journal of Experimental Psychology, 49A (1),* 159-177.

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

Standing, L. (1973). Learning 10, 000 pictures. *Quarterly Journal of Experimental Psychology, 25,* 207-222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19 (2),* 73-74.

Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition, 17 (6-7),* 777-789.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*, 643-659.

Tatler, B. W., Gilchrist, I. D., & Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files. *Quarterly Journal of Experimental Psychology Section A, 58(5),* 931-960.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50 (2),* 184-193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51(6),* 599-606.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception & Performance, 20,* 799-806.

Theeuwes, J., Kramer, A. F., & Irwin, D. E. (2011). Attention on our mind: The role of spatial attention in visual working memory. *Acta Psychologica, 137,* 248-251.

Todd, S., & Kramer, A. F. (1994). Attentional misguidance in visual search. *Perception and Psychophysics, 56 (2),* 198-210.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113*, 766–786.

Turatto, M., & Galfano, G. (2000). Color, form and luminance capture attention in visual search. *Vision Research, 40*, 1639-1643.

Uecker, A., & Obrzut, J. E. 1993. Hemisphere and gender differences in mental rotation. *Brain and Cognition, 22,* 42–50.

Uttl, B., & Graf, P. (1993). Episodic spatial memory in adulthood. *Psychology and Aging, 8*, 257-273.

Uttl, B., Graf, P., & Siegenthaler, A. L. (2007). Influence of object size on baseline identification, priming, and explicit memory. *Scandinavian Journal of Psychology, 48,* 281–288

von Wright, J. M., Gebhard, P., & Karttunen, M. (1975). A developmental study of the recall of spatial location. *Journal of Experimental Child Psychology, 20,* 181-190.

Voyer, D., & Bryden, M. P. 1990. Gender, level of spatial ability, and lateralization of mental rotation. *Brain and Cognition, 13,* 18–29.Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

Zechmeister, E. B., McKillip, J., Pasko, S., & Bespalec, D. (1975). Visual memory for place on the page. *Journal of General Psychology, 92,* 43-52.

Zelinsky, G. J., & Loschky, L. C. (2005). Eye movements serialize memory for objects in scenes. P*erception & Psychophysics, 67 (4),* 676-690.