Durham E-Theses

The role of foveal and extrafoveal vision in the processing of scene semantics

Gareze, Lynn

How to cite:

Gareze, Lynn (2003) *The role of foveal and extrafoveal vision in the processing of scene semantics*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/4123/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

The Role of Foveal and Extrafoveal Vision in the Processing of Scene Semantics

Lynn Gareze

Thesis submitted for the qualification of Doctor of Philosophy

University of Durham
Department of Psychology

A copyright of this thesis rests with the author. No quotation from it should be published without his prior written consent and information derived from it should be acknowledged.

2003

12 MAR 2004

Lynn Gareze
The role of foveal and extrafoveal vision in the processing of scene semantics
PhD
2003

Abstract

This thesis investigated the ability to process semantic information from foveal and extrafoveal vision during scene viewing. Existing research suggested that object semantics could be detected from extrafoveal vision. This suggestion was investigated using three experimental paradigms.

Semantic inconsistency was defined as a target semantically incompatible with scene gist. In Experiments 1 to 4, fixation position during a brief scene presentation was manipulated relative to a target object. The target's semantic inconsistency, presented foveally or extrafoveally, influenced performance on an object identification task. Extrafoveally presented semantically inconsistent targets were facilitated when simple line drawings were displayed, although this effect was unlikely to be mediated by semantic processing. No similar effect was found with complex line drawings or photographic stimuli.

Experiments 5 and 6 attempted to replicate significant advantages for inconsistent targets in a change detection paradigm. However, no significant difference was found between performance for consistent and inconsistent targets in a two-exposure, forced-choice change detection task or an alternating display change detection task. There was no evidence that changing inconsistent targets were detected more reliably or earlier than changing consistent targets.

Experiment 7 investigated the proposal that the extrafoveal processing of inconsistent objects could influence saccade patterns by attracting earlier fixations. Participants freely scanned both line drawings and photographs of scenes with no task. Again, no evidence was found supporting the earlier fixation of inconsistent objects in scenes.

Therefore, this thesis could not confirm previous evidence of an inconsistent object advantage in either brief scene presentations, change detection or natural scene viewing. The evidence suggested that the preferential processing of inconsistent scene objects could occur under very limited circumstances, but would be unlikely to be mediated by semantic processing. When viewing complex, realistic scenes, there was no evidence of differential processing for consistent and inconsistent objects.

Table of contents

Abstract 2
Table of Contents 3
List of Illustrations..6
List of Tables..9
Acknowledgements..11
Declaration..11
Statement of copyright...11

Chapter 1: General Introduction 12
    1.1 Processing information from different retinal regions 15
    1.2 The processing of scene context or 'gist' 18
    1.3 The study of consistent and inconsistent objects in scenes 21
    1.4 Investigating the perceptual consistency effect 25
    1.5 The perceptual schema hypothesis and the local processing hypothesis 27
    1.6 Evaluating the schema hypothesis 29
    1.7 Evaluating the local processing hypothesis 44
    1.8 The effects of semantic inconsistency on natural scene viewing ...52
    1.9 Conclusions and experimental hypotheses 67

Chapter 2: Brief Presentations of Line Drawing Scene Stimuli 73
    2.1 Introduction to brief presentation experiments 73
    2.2 Experiment 1: Introduction 75
    2.3 Method 78
    2.4 Results 82
    2.5 Discussion 95
    2.6 Experiment 2: Introduction 102
    2.7 Method 103
    2.8 Results 105
    2.9 Discussion 113

Chapter 3: Brief Presentations of Complex Scene Stimuli........................117
3.1 Experiment 3: Introduction 117
3.2 Method 120
3.3 Results 123
3.4 Discussion 135
3.5 Experiment 4: Introduction 138
3.6 Method 140
3.7 Results 141
3.8 Discussion 154
3.9 Summary of brief presentations experiments 1 to 4 155

Chapter 4: Semantic Consistency Effects in Change Detection 162
4.1 Introduction to change detection 162
4.2 Experiment 5: Introduction 173
4.3 Method 174
4.4 Results 178
4.5 Discussion 185
4.6 Experiment 6: Introduction 191
4.7 Method 192
4.8 Results 196
4.9 Discussion 208
4.10 Summary of change detection experiments 5 and 6 213

Chapter 5: Natural Scene Viewing 217
5.1 Experiment 7: Introduction 217
5.2 Method 219
5.3 Results 221
5.4 Discussion 232

Chapter 6: General Discussion 236
6.1 Summary of Experiments 1 to 4 236
6.2 Summary of Experiments 5 and 6 245
6.3 Summary of Experiment 7 249
6.4 Further research 252
6.5 Research questions 260
6.6 Final conclusions 265

Appendix A: Investigation into the Suitability of the Experimental Images 266
A.1 Introduction 266
A.2 Method 267
A.3 Results 270
A.4 Discussion 301

Appendix B: Experimental Scene Images (available on CD inside back cover)

References 306

List of illustrations

Figure 1.1: Example images from Loftus and Mackworth's (1978) stimuli 60
Figure 1.2: Example images from De Graef et al (1990) and Henderson et al (1999) 60
Figure 1.3: Example objects and non-objects used in De Graef et al's (1990) experiment 61

Figure 2.1: Example displaying the sequence of images in a trial 76
Figure 2.2: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12° 84
Figure 2.3: Graph showing the change in accuracy by fixation position and target object consistency 84
Figure 2.4: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes 89
Figure 2.5: Graph showing change in accuracy by fixation position and consistency for small objects only 89
Figure 2.6: Graph showing change in accuracy by fixation position and consistency for medium sized objects only 90
Figure 2.7: Graph showing change in accuracy by fixation position and consistency for large objects only 90
Figure 2.8: Graph showing the change in accuracy by fixation position and target object consistency for the high quality images 94
Figure 2.9: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12° 107
Figure 2.10: Graph showing the change in accuracy by fixation position and target object consistency 107
Figure 2.11: Graph showing the change in accuracy by fixation position and target object consistency for upright and inverted line drawings 109
Figure 2.12: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes 111
Figure 2.13: Graph showing change in accuracy by fixation position and object size 111

Figure 3.1: Examples of scenes used as experimental images 121
Figure 3.2: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12° 125
Figure 3.3: Graph showing the change in accuracy by fixation position and target object consistency 125
Figure 3.4: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes 130
Figure 3.5: Graph showing change in accuracy by fixation position and consistency for small objects only 130
Figure 3.6: Graph showing change in accuracy by fixation position and consistency for medium objects only 131
Figure 3.7: Graph showing change in accuracy by fixation position and consistency for large objects only 131
Figure 3.8: Graph showing the change in accuracy by fixation position and target object consistency for the high quality images 134
Figure 3.9: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12° 143
Figure 3.10: Graph showing the change in accuracy by fixation position and target object consistency 143
Figure 3.11: Graph showing the change in accuracy by fixation position and target object consistency, for photographs and their line drawings 145
Figure 3.12: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes 149
Figure 3.13: Graph showing change in accuracy by fixation position and consistency for small objects only 149
Figure 3.14: Graph showing change in accuracy by fixation position and consistency for medium objects only 150
Figure 3.15: Graph showing change in accuracy by fixation position and consistency for large objects only 150
Figure 3.16: Graph showing the change in accuracy by fixation position and target object consistency for the high quality images 153

Figure 4.1: Procedure for practice and experimental trials 178
Figure 4.2: Scatterplot of hit rate by false alarm rate 179
Figure 4.3: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12° 181
Figure 4.4: Graph showing the change in accuracy by fixation position and target object consistency 181
Figure 4.5: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes 184
Figure 4.6: Graph showing mean fixation times by target object consistency for 'change' trials 205
Figure 4.7: Graph showing mean fixation times by target object consistency for 'no change' trials 205

List of tables

Table 2.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only 87
Table 2.2: Table showing accuracy (in %) by object size and object consistency 88
Table 2.3: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only 110
Table 2.4: Table showing accuracy (in %) by object size and object consistency 110

Table 3.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only 127
Table 3.2: Table showing accuracy (in %) by object size and object consistency 129
Table 3.3: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only 146
Table 3.4: Table showing accuracy (in %) by object size and object consistency 148

Table 4.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only, including mean response times for 'no change' trials 183
Table 4.2: Table showing accuracy (in %) by object size and object consistency 185
Table 4.3: Table showing accuracy (in %) by change condition and target object consistency 197
Table 4.4: Summary table of mean response times (in ms) by change condition and target object consistency 197
Table 4.5: Summary table of mean arrival times (in ms) by change condition and target object consistency 199
Table 4.6: Summary table of mean saccade amplitudes (in °) by change condition and target object consistency 201

Table 4.7: Table showing accuracy (in %) by object size and object consistency 206
Table 4.8: Table showing response times (in ms) by object size and object consistency 207
Table 4.9: Table showing arrival times (in ms) by object size and object consistency 208

Table 5.1: Summary table of mean data relating to measures of saccade behaviour prior to target object fixation 222
Table 5.2: Summary table of mean data relating to measures of saccade behaviour during target object fixation 223
Table 5.3: Summary table of data according to target object size 224
Table 5.4: Summary table of mean data relating to measures of saccade behaviour prior to target object fixation 226
Table 5.5: Summary table of mean data relating to measures of saccade behaviour during target object fixation 227
Table 5.6: Summary table of data according to target object size 228
Table 5.7: Summary of results for consistent and inconsistent line drawings and photographs 231

Table A.1: Table displaying the recognisability results of the line drawing scene images 276-286
Table A.2: Table displaying the recognisability results of the photograph scene images 291-300

Acknowledgements

Primary thanks are due to my supervisor, Prof. John M. Findlay, for his support and encouragement throughout this entire process. Additionally, I am indebted to many other people for their practical help, such as Bob Metcalf for creating the computer programs used in these experiments, Elaine Behan and Ross Devine for assisting me whenever I ran into computing difficulties and also Lora Findlay for converting the photographic images used in Experiment 3 into line drawing stimuli used in Experiment 4. Credit for the manipulation in Experiment 2, the inversion of the line drawing scenes to interfere with semantic processing, must go to Dr. Peter De Graef.

I am also grateful for the financial support I received from the University of Durham, in the form of a studentship. Many thanks to the staff and students at the Department of Psychology for supporting me during my years here, including all members of the Centre for Vision and Visual Cognition, who provided valuable feedback and advice on my research. Finally, I would like to acknowledge the many hundreds of people who assisted me, by volunteering their time and participating in my experiments.

Declaration

The research contained in this thesis was conducted by the author between October 1999 and September 2002. None of the material contained in this thesis has been submitted in candidate for any other degree.

Statement of Copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Chapter 1

General Introduction

The investigations described in this thesis are concerned with the visual processing occurring in foveal and extrafoveal vision during scene viewing. The retina is often differentiated in vision research into three regions, to reflect relative visual acuity and the ability to resolve visual information. The central region of the retina, the *fovea*, encompasses approximately 2° of visual angle and is capable of high resolution visual processing. Visual acuity decreases steadily as distance from the fovea increases, through the *parafovea*, between approximately 2° and 10° from central fixation, and beyond it in the *periphery*. For the purposes of this thesis, *extrafoveal vision* will be defined as vision resulting from the processing of information appearing in the parafovea or periphery. Therefore, a distinction is made between the highly detailed processing resulting from direct foveal fixation and the increasingly degraded visual information available from other regions of the retina.

Much research has been conducted on the visual acuity of different retinal regions. Paradigms such as gaze contingent masks and windows have been applied to investigate whether vision is affected by the removal of foveal or peripheral information. The foveal region is generally found to be of greater importance, with furthest peripheral information being less so. Particularly in reading, much work has been conducted investigating the 'perceptual span', measuring the spread of information used when reading text, beyond which the masking of the stimuli does not affect the reading process.

The perceptual span in reading is not located centrally around the fixation position and fovea, but extends further to the right, when reading text from left to right. This asymmetry is believed to indicate the use of extrafoveal (parafoveal and peripheral) vision when planning a saccade and the pre-processing of text before fixation. If extrafoveal vision were useful in the processing of potential saccade targets, then the type of information available from such processing would affect the selection of a region for subsequent fixation.

Similar research in natural scenes has been hindered by the absence of a predictable saccade pattern. Because of this, it has been difficult to investigate how far a perceptual span in scene viewing extends and to determine how extrafoveal vision is used. An additional difficulty is raised by the nature of the visual stimuli. Pictorial images such as natural scenes contain several types of information, ranging from physical information (such as light/dark) to highly cognitive information (such as scene meaning). While some research has focussed on the *visual* processing of an object viewed in extrafoveal vision, there is also the *semantic* level of processing, at which an object is identified, named and its meaning can be recognised.

To investigate the extraction of semantic information, the semantic relationship between a target object and the scene background in which it is located can be manipulated. Target objects can be semantically related to a scene, in which the object's identity is compatible with the scene's 'gist', or unrelated to the scene, where the object's identity is incompatible with the gist. For example, two objects sharing similar visual features, such as an apple and a ball, would not be semantically associated with the same types of scene background. An apple is semantically compatible, or *consistent*, with a fruit market context for example, but a ball in the same location would be *inconsistent* with the scene's meaning. By comparing performance on a given task between objects categorised as consistent and those categorised as inconsistent with the scene context, it is possible to investigate whether the target object's semantic meaning can be accessed or whether only its visual properties can be processed from extrafoveal vision. The manipulation of this consistency relationship allows for an experimental test of whether semantic information is processed from objects in scenes in foveal and extrafoveal vision.

This manipulation forms the basis of the investigations undertaken in this thesis. The aim was to determine whether information processing from extrafoveal vision is sufficient to allow the semantic identification of an object in a complex visual scene. This issue will be investigated using three different methodologies which have previously been applied to this study. However, before discussing the results of existing research in this specific area, it is important to consider some general issues relating to the wider process of visual perception in scenes.

Three distinct issues will be reviewed, before the specific discussion of the research question outlined above. In the first, the processing of information, not specifically semantic information, from foveal and extrafoveal vision during scene viewing will be discussed, with an attempt made to determine the extent of a perceptual span or the 'useful field of view'. Of particular interest is the ability to resolve detail from extrafoveal vision prior to a saccade, during the viewing of complex scenes as opposed to simplified images. This investigation considers the simultaneous effects of foveal processing of a fixated object with extrafoveal processing for a subsequent fixation target.

The second issue, specific to the viewing of natural scenes, is concerned with the investigation of the extraction of scene meaning or 'gist'. The information obtained from scenes, often from a single fixation, is not restricted to the layout of physical scene properties but includes semantic information such as the identity of component parts and scene context. For a discrepancy to be detected in the semantic relationship between a target object and its scene background, the rapid and accurate perception of the scene's meaning is essential.

The final issue under consideration concerns the effects of semantically inconsistent relationships between objects and their backgrounds. Previous research investigating the effects of inconsistent objects in scenes has considered many different approaches and paradigms. In this section, our current knowledge of the effects of these inconsistent objects on eye movement behaviour and memory will be discussed, before identifying more controversial areas of research within this field.

1.1 Processing information from different retinal regions

The processing of extrafoveally presented information is constrained by the resolution capabilities of the retina. Although the processing of semantic and visual features may be distinct, the ability to identify an object must on some level be dependent on the resolution of its visual features. Research relating to the identification of objects in extrafoveal vision, especially within complex scenes, is of particular interest. Performance on a selected task is affected by the manipulation of target object location, with images presented at greater eccentricities from the fovea exhibiting worse performance than those directly fixated. The degeneration of visual processing across the retina can be mapped relative to performance on a task requiring object identification, to investigate how retinal eccentricity influences the identification of objects.

Nelson and Loftus (1980) investigated the functional visual field using colour slides of complex scenes. Near identical scenes were paired in which the identity of only one critical object differed. Participants viewed one scene of each pair, for 250ms to minimise the probability of saccade initiation, and were required to distinguish, in a subsequent two-alternative forced-choice task, which one of the paired scenes had been presented.

As fixation position was manipulated, accuracy was highest, at almost 80% correct, when the participant had directly fixated the target object. Accuracy decreased as the distance between the participant's fixation and the target object increased, measured in degrees of visual angle, with the greatest decrease between 0° and approximately 1.8°. However, performance for critical objects located over 2.6° away from fixation was still above chance with an accuracy of over 60% suggesting that, under these circumstances, useful information about the objects, for use in a two-alternative forced-choice decision, could be obtained from extrafoveal vision.

Henderson, McClure, Pierce and Schrock (1997) used an artificial scotoma paradigm which could either mask the object fixated or the object to the right of the one fixated. Participants fixated each of four objects in a row and confirmed whether a target named at the end of the trial had been present. As performance was significantly above chance (approximately 85%) even when the foveated object was masked, the authors concluded that the direct fixation of an object was not necessary for its identification.

However, as the objects were approximately 1.5° x 1.5° in size and their centres were 2.4° apart, information about a neighbouring object could be processed in near foveal vision. When fixating the centre of one object, half of each neighbouring object would appear within 2.4° of the current fixation and often the entire object within approximately 3.15°. The evidence indicated that at these eccentricities, objects could be correctly identified, even when inappropriate foveal information was visible at fixation.

A review by van Diepen, Wampers and d'Ydewalle (1998) summarised the findings of Saida and Ikeda (1979) who manipulated the content and availability of peripheral vision during picture viewing. The size of a gaze contingent window, obscuring all visual information outside it, was varied during the viewing of 80 line drawings. Participants later viewed 160 images and identified the ones displayed previously. The results indicated that a window of half the image size produced test performance equal to unrestricted viewing, although image size was confounded with image density. Perceptual span was found to be smaller for complex, photographic scenes (McConkie and Loschky, 1997) than line drawings, an effect which was suggested to be modulated by image density.

Shioiri and Ikeda (1989) investigated the 'useful resolution' of picture viewing, defined as 'the fineness of detail actually required to achieve a normal level of performance', in a similar scene recognition paradigm (van Diepen et al, 1998). They degraded peripheral information outside a square window of variable size and conducted a recognition test like Saida and Ikeda's. Again, the perceptual span at which performance matched unrestricted viewing conditions subtended half the image size. The useful resolution decreased faster than 'available resolution', defined as 'the smallest size of details that could be discriminated', across increasing eccentricity, leading to the conclusion that image memorisation required only low resolution information.

van Diepen, De Graef and d'Ydewalle (1995) introduced an elliptic foveal mask during scene viewing to investigate the use of foveal vision. The mask was centred on fixation, appearing after variable fixation times between 15ms and 120ms, and its size was manipulated. The presence of a mask, regardless of size, increased fixation durations, indicating a disruption in scene exploration caused by the interference to foveal information. Manipulating the mask onset delay indicated that foveal masking disrupted the search task most when presented within 45ms of each fixation, suggesting that fixations included early foveal analysis. The manipulation of mask size, either 1.5° x 1.0° or 2.5° x 1.7°, only affected the saccade amplitude and not the fixation durations or scene inspection times, suggesting that, with larger masks, participants would make larger saccades to regions outside the masked area.

van Diepen et al (1998) described the masking of extrafoveal information attempted by van Diepen, Wampers and d'Ydewalle (1995). They presented only high or low spatial frequency information outside a fixation-centred 6.0° x 4.6° window during an object search task, to investigate the type of visual information used from extrafoveal vision. A clear benefit in scene inspection time and saccade amplitude was found when high frequency information was available, indicating that the high frequency detail assisted in the localisation of scene objects for further processing.

van Diepen and Wampers (1998) used a similar moving window technique to investigate the type of peripheral information used within the first 150ms of fixation. Within the 3.5° wide x 2.6° high elliptic window, the image was always presented without manipulation. However, outside the window, the image could be subject to low-pass, bandpass or high-pass filtering, for the first 150ms of every fixation. Any removal of spatial frequency information slowed scene exploration in a non-object search task but the presence of selective information showed no advantage over the completely masked condition. However, the confound of the suddenly reappearing visual detail after 150ms was found to account for many of the effects, suggesting that the loss of peripheral information may have had little effect after all.

This evidence indicated that foveal information is primarily used during the early part of fixations, with the use of extrafoveal information being possibly delayed and more selective according to the task demands. The perceptual span was summarised by van Diepen et al (1998) as ranging between 20% and 50% of the image size, depending on stimulus complexity. The study of the perceptual span in viewing has remained focussed on the perception of visual detail from extrafoveal processing and the useful resolution from different retinal eccentricities, applicable to the selection of saccade targets from peripheral analysis. However, research related to the processing of *semantic* information from extrafoveal vision is limited.

1.2 The processing of scene context or 'gist'

Sanocki and Epstein (1997) investigated whether gist information from a prime scene background would facilitate a non-gist-related task on a subsequent scene. Although significant facilitation was found for an identical scene background prime (presented without the target objects on which the task depended), the advantage could be affected by knowledge of spatial layout, including the location of the ground plane and other reference objects, rather than gist. A replication by Germeys and d'Ydewalle (2001) attributed a portion of this facilitation to the apparent visual onset of the target objects, when the prime scene displayed the target scene background, but also failed to find evidence of gist facilitating subsequent scene processing. Although gist was not found to significantly facilitate performance in these experiments, other experiments involving scene identification or categorisation have found significant effects.

Much of this research considers visual processing from brief scene presentations, replicating a single fixation on a complex scene, without explicitly considering foveal and extrafoveal vision. Impressively, some specific judgements about briefly presented images can be made. For example, Thorpe, Fize and Marlot (1996) found that participants could categorise images according to whether an animal was present or not, from 20ms presentations, with a mean accuracy of 94%.

McCauley, Parmelee, Sperber and Carr (1980) investigated the ability to extract semantic meaning from a single object using a priming technique. By reducing the prime exposure time until the facilitation on a subsequent related object's identification was extinguished, they concluded that semantic information from a single object could be extracted without conscious awareness in less than 37ms. Although this is not directly applicable to gist extraction from scenes, it indicates that semantic information can be processed very rapidly from certain images.

The identification of semantic scene meaning, sufficient to name it appropriately, is believed to primarily occur from the processing of global contextual information about the scene as a whole, rather than its component objects. Even using complex scenes, it has been determined that the 'gist' of a scene can be extracted from the image within milliseconds. Schyns and Oliva (1994) found that the extraction of scene meaning, in a scene categorisation and a scene identification task, was modulated by the initial (up to approximately 45ms) rapid processing of low spatial frequency 'blobs', providing coarse information on scene layout. This initial processing was followed by the processing of high spatial frequency 'edges' which provided finer details of local boundaries.

Although both low and high spatial frequency information could be processed from brief presentations, the scene identification process in these tasks relied on scene-based information in its earliest stages, before focussing on object-based information. Schyns and Oliva (1997) emphasised that this process was flexible and could be affected by task demands. The pattern of processing described could be altered according to experimental task requirements, by biasing participants towards preferentially processing either low spatial frequency or high spatial frequency information, but presumably would be typical of general scene viewing (no task).

Potter (1975; 1976) used colour photographs of scenes, obtained from magazines, to investigate the speed with which participants could detect an image when given details about its appearance. Half of the participants were shown the target image they were expected to detect and the second half were given a description of the image, outlining the gist of the scene (e.g. ‘a road with cars’ or ‘a girl sitting in bed’). From a sequential presentation of 16 colour pictures presented for variable durations, participants were required to make a manual response when the target image was detected.

The results indicated that accuracy was over 70% even when the images were presented for 125ms each only, regardless of the instructions (Potter, 1975). With a shorter display time of 113ms (1976), accuracy for participants given the scene gist was 64%. Although accuracy at longer presentation times (167ms, 250ms and 333ms) was greater than at 113ms, there was no significant difference between them, indicating that at most 167ms was sufficient to perform this task accurately. From the scene description, participants could rapidly identify the target, implying that the scene gist could be processed within 113ms and 125ms. The author concluded that the results of these experiments and previous studies “support the hypothesis that a preliminary identification of a complex meaningful scene occurs within about 100ms, whether or not the scene is expected” (p521).

Biederman, Glass and Stacey (1973) found that jumbling the regions of a visual scene to destroy global properties made the process of searching for a pre-specified object much more difficult and hypothesised that gist information could be used to direct a search. However, Henderson (1992a) suggested that the facilitation evident for intact images might have been affected by the overall degradation of the jumbled stimuli, rather than a specific inability to detect gist. In a further study, Biederman, Rabinowitz, Glass and Stacey (1974) asked participants to select the appropriate verbal label, from two similar or dissimilar options, to describe a scene viewed briefly. Accuracy decreased when the global components of the scene were jumbled, suggesting that the identification of gist was affected by the coherence of the spatial dimensions.

Participants selected gist from two similar alternatives (e.g. 'bedroom' and 'living room') in a coherent (not jumbled) scene from a 100ms presentation with an accuracy of over 70%. With error rates approximating 30%, it is clear that only some processing of the scene could be completed and the possibility of foveal analysis of useful image components cannot be ruled out. However, this evidence suggested that some detailed processing of the scene meaning could be accessed within 100ms of scene viewing. The converging evidence from all these studies indicates that the semantic meaning of a scene and its global contextual properties can be detected very rapidly and possibly within 100ms of viewing, although the evidence considered above is not intended as a conclusive or exhaustive overview of the relevant research.

1.3 The study of consistent and inconsistent objects in scenes

The effect of an object incompatible with its scene context has been investigated in great detail. For current purposes, the relationship between such objects and the scenes in which they are located will be described as *inconsistent*, with scenes in which all items are compatible with the gist being *consistent*, although other researchers have used different terms, such as 'congruent' and 'incongruent' or 'plausible' and 'implausible'. An object or scene described as inconsistent indicates the presence of an object in a scene which is incompatible with the gist and which would not be expected to be located there.

In 1975, Palmer investigated whether scene context could influence the identification of a target object by using a line drawing of a scene background to prime the identification of a subsequently presented object. Following a two second scene presentation, participants viewed a line drawing of an object for 20, 40, 60 or 120ms and were required to name it. Palmer found that a preceding scene image which was semantically related to the target improved naming accuracy, even when the target was only presented for 20ms. At this shortest exposure, accuracy was just under 80%, compared to approximately 55% when preceded by a blank screen (no context) and 40% when preceded by an unrelated scene context. This significant finding indicated that prior access to relevant semantic information facilitated the subsequent identification of a target object. Although this does not prove that a similar effect occurs for objects presented within scenes, it is evidence that previously acquired semantic information can affect the identification of objects.

Friedman (1979) investigated the effects of inconsistent objects within scenes when the object was fixated. These objects were categorised as unexpected, rather than impossible, objects which would be unlikely to be located in the given scene (e.g. a fireplace in a kitchen). Participants viewed six line drawings of scenes for 30 seconds each while their eye movements were recorded. A label indicating the gist of the scene preceded each trial and instructions were given that a recognition test would follow, in which participants would have to distinguish between the scene displayed and a new scene differing in one small detail only. First fixation durations on objects were found to be strongly correlated with their rated likelihood of appearing in the scene, with longer durations for inconsistent objects. There was a 342ms difference between first fixation durations on consistent and inconsistent objects, falling to approximately 250ms for the second fixation and 78ms for the mean duration of third and later fixations.

Friedman concluded that prior knowledge of scene context allowed participants to access 'global memory structures such as frames' to facilitate the detection and identification of consistent objects in relatively shorter fixation durations. Inconsistent objects required first fixation durations which were approximately twice as long as required for consistent objects, indicating a greater need for foveal analysis in order to memorise in preparation for a test. In the recognition test, inconsistent distractors were more reliably rejected than consistent distractors, perpetuating a facilitation effect for inconsistent objects. The distractor's probability of being located in a given scene influenced the participants' recollection of whether it had been present. These findings of the processing of inconsistent objects in scenes, *once they are fixated*, have been replicated and proved to be reliable and robust.

The evidence that inconsistent objects are fixated for longer than consistent objects and are more reliably rejected in recognition tests suggests that information relating to inconsistent objects may be better retained in memory over time. More naturalistic real-life environments have been used to investigate the effects of semantic consistency over longer time spans. For example, Pezdek, Whetstone, Reynolds, Askari and Dougherty (1989) attempted to replicate Brewer and Treyens' (1981) study, in which participants were asked to wait in an office until called. After 35 seconds, the experimenter arrived and showed them into a different room. At this point, the participants were tested on verbal recall, drawing recall and verbal recognition for the objects in the 'waiting room'. The items considered most consistent with the expectations provided by the context were recalled and verbally recognised more often than inconsistent items.

Pezdek et al (1989) provided two experimental rooms, an office and a preschool classroom, in which the same 16 objects were placed, half of which were consistent with being found in an office but not a classroom and the other half were consistent with a classroom context but not an office. Participants were allowed one minute to intentionally observe the room, before recording a list of items they had seen (recall) and returning to the experimental room to identify which of the 16 experimental objects had been altered (recognition). By this time, an experimenter had replaced half of the objects (four consistent and four inconsistent items) with a similar token (an object with the same name but different visual appearance). Inconsistent objects were recalled better than consistent objects and changes to them were more reliably detected, even after a delay between observation and test of one day.

In a second experiment, Pezdek et al used two offices to investigate whether the results found with the simplified environment of the first office, used in the previous experiment, could be replicated in a more naturalistic, genuine student office. Additionally, participants were divided into an 'intentional learning' group, as in their previous experiment, and an 'incidental learning' group, who were simply asked to wait quietly in the room, as in Brewer and Treyens' (1981) experiment. Recognition improved with intentional instructions and the consistency effect remained in both learning conditions. Inconsistent items were better recognised than consistent items and the false alarm rate was higher for consistent objects than inconsistent objects. These two conclusions seemed to be robust findings despite Brewer and Treyens' results, confirming that memory for inconsistent objects was better than that for consistent objects and that more consistent objects than inconsistent objects were falsely believed to have been present in a scene when they were not.

These effects were further investigated by Lampinen, Copeland and Neuschatz (2001) who replicated the finding that objects inconsistent with the scene schema were better remembered than consistent objects and that participants were more likely to falsely remember consistent items which were not present. These results were found for both the 'incidental learning' and the 'intentional learning' conditions, with participants spending one minute in the experimental room and tested after a substantial training phase. Additionally, the subjective experience of remembering and the recollections on which the 'remember' judgements were made were under investigation. When participants claimed to have 'remembered' an object, they were asked to explore the recollection and state whether it involved a perception, thought, emotion or contextual information. When inconsistent objects were remembered, they were significantly more likely to be associated with an emotion than consistent objects ($p<.05$), with participants recalling amusement or surprise.

The finding that memory for objects can be dependent on their relationship with the scene context has also been investigated over shorter time scales. As inconsistent objects are somehow better represented in long-term memory, it seems plausible to investigate whether any differences exist in the immediate perceptual processing of consistent and inconsistent objects. For example, the increased foveal processing of inconsistent objects could result in such a memory effect.

Using the change blindness flicker paradigm, Hollingworth and Henderson (2000) found evidence that changes to inconsistent objects were more reliably detected than changes to consistent objects in scenes, implying that the representation of an inconsistent target before a change was in some way better than that for a consistent target. Similarly, Hollingworth, Williams and Henderson (2001) found an advantage for the detection of changes to inconsistent objects during a scene viewing task. Participants viewed a scene in anticipation of a memory test and were warned that changes might occur to objects within the scene, which they were to attempt to detect. Participants were better able to detect a change occurring to an inconsistent target during a saccade away from it than during a saccade away from a matched consistent target. An advantage was seen regardless of the gaze duration on the target, which could have modulated the effect, indicating that inconsistent objects were better processed and represented in memory across the course of the trial (up to 20 seconds).

1.4 Investigating the perceptual consistency effect

These intriguing results indicating facilitation for inconsistent objects have been the subject of further research, much of which has concentrated on the effects of fixated inconsistent objects in contextual scenes. The issue of the *foveal* processing of inconsistent objects has been discussed but the influence of an inconsistent object processed extrafoveally, *before it is fixated*, has generated less compatible findings. In terms of the initial processing of an object, experimental hypotheses can be constructed which support different perceptual consistency effects.

The enhanced detection of consistent objects could be predicted, because a compatible scene context would facilitate their identification. Conversely, the enhanced detection of inconsistent objects could be hypothesised, because their contextual incompatibility would render them more visually salient. As most objects in real-life visual scenes would be consistent, the primary experimental concern has been to investigate whether the presence of an inconsistent object could be detected in any way prior to fixation or whether the advantage for inconsistent objects in memory only begins at direct fixation and encoding. In this way, it is of interest to find evidence of differential performance for consistent and inconsistent objects in experimental tasks requiring their detection or identification without direct fixation.

Investigating such differences has required the consideration of two different issues, reflecting the existing research in the field in which a useful distinction is made between different time scales of scene processing. To begin with, it is possible that the presentation of an inconsistent object in a scene is immediately apparent, that it ‘pops out’ of the scene. In this case, the investigation would centre on whether the perceptual processing that enables the detection of scene gist could also identify regions of semantic discontinuity, consciously or otherwise. By presenting scene images to participants only very briefly, to prevent them from being able to initiate a saccade, we can obtain data on the extrafoveal processing of objects in scenes upon the first fixation. Any significant differences in performance between consistent and inconsistent objects must be explained by the detection of semantic inconsistency within a single fixation. Therefore, the rapid identification of gist would need to be combined with the similarly rapid identification of an inconsistent object. This approach reflects the influence of global scene context on local processing of a target object (scene-to-object).

Alternatively, the inconsistency between an object and its background might not be immediately apparent but could have an effect over longer time periods during the scanning of visual scenes. This approach does not rely on the immediate detection of an inconsistent object, as any evidence of differential performance between consistent and inconsistent objects would be explained by the effect of semantic consistency on eye movements. As visual acuity decreases as eccentricity from the fovea increases, objects would become more difficult to identify when viewed further in extrafoveal vision. Without the compatible context to constrain possible interpretations of degraded visual information, inconsistent objects could be more difficult to recognise than consistent objects when viewed extrafoveally.

In this way, it is possible that a single fixation on a scene could not isolate regions of semantic inconsistency if these regions occurred far from fixation. However, when viewing the scenes over extended periods of time, the target object could be processed from nearer extrafoveal vision, enabling the inconsistency to be detected. This information could then influence saccade behaviour. In this case, the effect of semantic consistency need not be influenced by immediate global scene context, but could be modulated through the accumulation of scene information across successive fixations (object-to-object). This approach investigates whether semantic inconsistency could be detected without direct foveation and used to direct saccades.

A similar dichotomy to that applied to research is apparent in the theoretical explanations for a consistency effect. It has been theorised that consistent objects rather than inconsistent objects would be facilitated in most tasks. Two explanations will be discussed, outlining the possibility of scene-to-object influence from the perception of a global consistent scene context or object-to-object influence from previous foveal analyses. Possible explanations for the small but increasing number of studies demonstrating an advantage for inconsistent objects will also be considered, followed by a review of the experimental evidence demonstrating consistent and inconsistent object advantages and the identification of issues suitable for further study.

1.5 The perceptual schema hypothesis and the local processing hypothesis

There are broadly two types of explanation for consistency effects in the viewing of natural scenes. The first approach involves the global processing of context, in the immediate identification of a scene's gist, interacting with the identification of individual objects (scene-to-object). The second type of explanation seeks alternative mechanisms independent of the interaction between global and local processing, which could influence consistency effects (object-to-object). An example of each will be discussed here and associated problems will be considered.

The *perceptual schema hypothesis*, also referred to as the *description enhancement model* (Hollingworth and Henderson, 1999), proposes that scene viewing activates a memory representation of a prototypical scene which allows expectations to develop about the objects likely to be viewed. In this way, the identification of objects consistent with the scene context would be facilitated, compared to the identification of objects inconsistent with the context. This theory predicts that a kettle would be easier to recognise in a kitchen scene than in a farm scene for example, because the perception of the kitchen's gist would facilitate the kettle's identification but the perception of the farm's gist would not. The identification of the scene and object may occur in parallel and provide mutual facilitation.

The schema hypothesis depends on two assumptions, the first being that the activation of a schema requires scene meaning information to be accessible very early in scene processing. The second assumption states that the activation of a schema can interact with the identification of individual objects and produce a top-down facilitation for objects contained within the scene. Henderson (1992a) identified some flaws in these assumptions and provided a critique of the schema hypothesis.

As an alternative explanation independent of the interaction between global and local processing, Henderson (1992a) proposed the *local processing hypothesis*, which suggests that global scene information does not interact with the perception of local objects. Instead, semantically related objects would facilitate each other's identification, resulting in a context effect through object-to-object priming. This proposal is also subject to two assumptions. The first is that the object identification system is 'informationally encapsulated' (Henderson, 1992a) from systems processing global scene information. Context effects would instead occur from interactions at the local object level, involving intra-level rather than inter-level interaction. This 'information encapsulation' is entirely compatible with the first assumption of the schema hypothesis, that scenes can be rapidly categorised according to meaning, but is incompatible with the second assumption, that top-down processing facilitates objects contained in the scene schema.

The second assumption of the local processing hypothesis is that scene processing occurs locally. As attention appears to be allocated to areas as small as one object during scene viewing, the amount of visual information that can be processed semantically can be limited (Henderson, 1992a). Any context effects can be attributed to the integration of local object information across saccades. In this way, the local processing hypothesis predicts that semantic relatedness between a target object and companion objects, rather than the global scene context, gives rise to the facilitatory consistency effect.

1.6 Evaluating the schema hypothesis

Some general criticisms of the schema hypothesis were outlined by Henderson (1992a). They included the lack of detail relating to how the knowledge contained in a schema could influence the object recognition process, as opposed to influencing later processes such as responses. Friedman (1979) suggested that predicted objects in a scene could be identified through resource-free feature matching while resource-intensive feature analysis would be required for unpredicted objects. Alternatively, schema activation may modulate the threshold of information necessary to identify an object, facilitating consistent objects. Distinguishing between different possibilities has proved difficult.

The degree to which top-down information influences the resulting perceptual descriptions of consistent and inconsistent objects also remains unclear. For example, objects presented in isolation can be easily recognised in the absence of a scene schema and object expectations. Schema theories also postulate that a schema's misapplication could result in incorrect object identification, although experience suggests that this happens less often than a misapplication may be expected to occur. Also, world knowledge can be misleading, for example when objects are found in an unexpected position (e.g. a chair on a table). The extent to which the predictions made by a schema can be manipulated is unclear. Therefore, Henderson (1992a) suggested that top-down information might be overridden, ignored or simply unnecessary in the process of object recognition.

Specifically considering the schema hypothesis' assumptions, the claim that a scene's meaning can be accessed very early on is considered robust and valid. However, the second assumption, that the schema's activation would produce top-down processing effects on the identification of individual objects is more controversial. This assumption can be tested by briefly presenting scene stimuli, allowing only global processing of scene gist, and investigating consistent effects at the local object level. The consistency between a target object and its background is manipulated to investigate whether accuracy on an object detection or identification task is affected.

The importance of scene context was first investigated by Antes, Penland and Metzger (1981), who distinguished between local and global information as providers of scene context information. Local information was defined as being carried by specific elements of the scene, such as objects. Global information resulted from viewing the scene as a coherent unit and the perception of scene gist. In their Experiment 1, participants viewed 100ms presentations and decided which one of four objects presented subsequently had been in the image viewed.

In the 'high context' condition, the objects were presented in appropriate locations in a line drawing scene, providing both local and global contextual information. In the 'low context' condition, the same object array was presented without a background, removing the global scene information. In addition, there were two control conditions. In the 'no information' condition (NI), participants did not view an image and simply selected an object at random from the four-alternative forced-choice display. In the 'thematic information' (TI) condition, participants were given a scene name but did not view the image, providing some global contextual information but no local information. Antes et al investigated whether a target object's relationship with either the global context provided by the scene background or the local information provided by the companion objects was important in the task. The relationship between the target object and the three distractors in the four-alternative forced-choice display was also manipulated, with either three consistent or inconsistent objects being presented with the target, introducing a test for possible response bias.

The analysis of most interest compared performance for high context trials, containing a consistent target object, presented with consistent distractors in the four-alternative forced-choice display, with trials containing an inconsistent target object, presented with inconsistent distractors, to determine whether scene-object consistency influenced performance. The target being shown with equally likely selection distractors minimised the effects of response bias. Higher accuracy was found for consistent targets presented with consistent distractors (0.330) than for inconsistent target objects presented with inconsistent distractors (0.159). However, the chance level in this experiment was 0.25 (one in four), so performance was not significantly above chance in either condition. The control conditions indicated that displaying an image produced significantly better performance than that obtained in the NI condition (p<.05) but no higher than accuracy in the TI condition with scene name provided, suggesting that local information was not used. The information processed from the 100ms scene presentations seemed primarily global and provided gist information but there was no evidence that this information influenced performance on consistent or inconsistent objects.

Also investigating the effects of global scene coherence, Biederman, Mezzanote and Rabinowitz (1982) conducted a series of experiments to determine how violations of expected 'relations' within scenes, such as support, interposition, probability, position and size, affected object detection from brief presentations. In 'violating' the probability assumption, objects were located in inconsistent scenes. In Experiment 1, each trial began with an object name presentation, followed by a central fixation cue. Then a scene was displayed for 150ms and finally a mask with a location cue embedded in it appeared. The participants' task was to determine whether an object matching the label presented at the beginning of the trial had appeared in the scene, at the location indicated in the mask.

When a probability violation was present, error rates increased by almost 10% even when other variables were partialled out of the analysis. An advantage for consistent scenes was found in response time, with inconsistent 'violation' trial responses being approximately 20ms slower than consistent 'no violation' trial responses. This result indicated that semantic consistency influenced the target object's detection, suggesting that inconsistent objects were not detected as rapidly or with as high accuracy as consistent objects.

In Experiment 2, Biederman et al investigated the detection of violations, rather than objects. Participants viewed an object name before each trial and were instructed to decide whether this object was undergoing any of the possible violations when the scene was presented. A variable location cue was displayed, marking where the object would appear, followed by the 150ms scene display. Finally, when the mask was presented, participants responded whether the target, which always appeared at the location cue and was therefore fixated, was consistent with the scene context in the probability violation condition. The longest response times were found when no violations were present. A probability violation reduced this response time by over 30ms and the inconsistency was detected with the lowest error rate of all violations, at less than 10%. This finding indicated that the relationship between a fixated object and its scene context could be compared and any inconsistency detected accurately within 150ms.

Boyce, Pollatsek and Rayner (1989) used a similar object detection paradigm to investigate whether the scene gist (global) or additional cohort objects (local) determined whether an object was considered consistent or inconsistent in a scene. Antes et al's (1981) investigation suggested that global information only was processed during brief 100ms presentations. In Boyce et al's experiments, both global and local information were systematically manipulated to determine whether they influenced object detection. An object array was presented against a consistent background, an inconsistent background and in isolation with no background, removing global scene information. The objects were located on a supporting surface in the scene, usually the ground.

Experiment 1 followed the same procedure as Biederman et al (1982). Each trial began with a target object name, followed by a 150ms scene presentation. A mask containing a location cue was then displayed and participants responded whether the cued location had indicated the target named at the beginning of the trial. Significantly better performance was found for consistent than inconsistent objects (66.3% and 58.8% respectively, $p<.05$ for consistency by background interaction).

Accuracy for 'no background' trials was comparable to that measured for consistent scenes (approx 67%), indicating that the consistent background did not facilitate performance but the inconsistent scene background may have inhibited it. This pattern suggested that global gist information did not significantly improve performance for consistent objects above the rate obtained in the 'no background' condition with no scene context. However, gist information incompatible with expectations, generated by the target name displayed before the scene, appeared to inhibit performance.

Experiment 2 investigated the role of local information by manipulating the four non-target objects in the scene, which could be semantically related or unrelated to the target object. Percentage correct results indicated that accuracy across conditions was compatible with the previous experiment but related cohorts did not facilitate performance. This result confirmed Antes et al's (1981) conclusions that global information alone contributed to the consistency effect.

To control for visual complexity, Experiment 3 introduced new control backgrounds which contained the equivalent amount of visual information as the experimental backgrounds, maintaining 3D supporting surfaces for objects, but contained little or no meaning information. The results displayed facilitation for consistent scene backgrounds, compared to nonsense backgrounds (4.8% increase in accuracy), but no decrease in accuracy for objects in inconsistent scene backgrounds compared to nonsense backgrounds. These data suggested that consistent scene backgrounds facilitated performance while inconsistent scene backgrounds did not inhibit it. 'No background' controls in Experiments 1 and 2 artificially improved performance, possibly through the removal of interference from background contours. Boyce et al (1989) concluded that consistency effects could be attributed to the facilitatory consistent global context and not to local object information.

Boyce and Pollatsek (1992) continued this investigation using a different paradigm. They presented participants with line drawing displays of object arrays on consistent, inconsistent or nonsense backgrounds. After fixating the scene centrally for 75ms, a target object would 'wiggle' by moving 0.5° before returning to its original location. Participants were asked to name the wiggling object, invariably fixating it. A significant facilitation of 47ms was found in naming latency for objects in consistent scenes ($p<.025$), compared to the same object in a nonsense background. No difference was found between objects in an inconsistent scene and in a nonsense background. These results supported their earlier findings that a consistent context facilitated object identification but an inconsistent context did not inhibit it.

To summarise, in Biederman et al's (1982) experiment, performance on an object detection task was better for consistent than inconsistent objects. Boyce et al (1989) also concluded that a consistent scene facilitated the detection of consistent objects while an inconsistent context had no inhibitory effects, supported by their 1992 study also finding faster identification for consistent objects. However, their results did not entirely support the schema hypothesis.

In a further investigation, naïve participants rated the target objects on their likelihood of occurring in the scene (Boyce et al, 1989). The authors wanted to investigate whether the consistency effect was influenced by the object's degree of consistency or inconsistency with the scene. This hypothesis was compatible with the schema hypothesis, as objects would need to be predictable from the schema for a consistency effect to be found. A low correlation between the degree of relationship and performance suggested that rated predictability had little influence on the consistency effect and indicated that the important feature was the object's plausibility rather than its probability. Similarly, Boyce and Pollatsek (1992) found a very low correlation of 0.2 between the consistency effect size for each object and its consistency rating. If the participants' ratings were an accurate estimate of how the visual system categorises consistent and inconsistent objects from brief presentations, these results did not support the schema hypothesis. The schema hypothesis predicts that consistent objects would be facilitated through their inclusion in the activated scene schema as predictable objects, but no significant correlation was found between the degree of consistency or likelihood and the size of the consistency effect.

Additionally, several general criticisms of the object detection paradigm implemented by these researchers were identified by Henderson (1992a) and Hollingworth and Henderson (1998). One concern involved the distinction between object detection and object identification. Schema hypotheses predict facilitation for consistent object identification in consistent scenes. However, some of the paradigms outlined above only considered the detection of objects and not necessarily their identification. The presentation of a target name before a scene image generates a memory representation of visual features *before* the scene presentation so the task would be to find an object to match the activated image representation. Object identification in contrast involves identifying a stored memory representation to match the perceived visual representation, as studied by Boyce and Pollatsek's (1992) object naming task. This process would also be ensured by presenting the object name *after* the scene presentation, requiring the search for a stored mental representation to match the visual image (e.g. Hollingworth and Henderson, 1998).

It was also argued that the object detection paradigm did not distinguish between context effects in identification and post-identification processes. Henderson, Pollatsek and Rayner (1987) argued that consistent objects may be facilitated at the time of response, rather than at the time of processing. Objects may be identified equally well regardless of their consistency with the scene context but performance may be affected at a later task, as information relating to consistent or inconsistent objects may be treated differently during memory storage or retrieval.

For example, consistency effects could have been modulated by response bias, rather than facilitation in processing. Participants were more likely to claim a consistent object had been present in the scene, regardless of whether it had been or not, compared to an inconsistent object. It is likely that participants were more reluctant to claim that a television had been present in a farm scene (inconsistent), compared to a horse in a farm scene (consistent), especially during catch trials when the target object was not present.

It was suggested that *d'*, a measure of detection sensitivity, did not adequately control for this bias because participants did not have to detect the same objects in experimental and catch trials (Hollingworth and Henderson, 1998). The appropriate control for an experimental trial depicting a horse in a farm would be a catch trial using the same object label ('horse') and the same scene background in which the target was not present. The catch trials used by Biederman and colleagues and Boyce and colleagues were entirely different to their experimental trials, so the detection sensitivity measure was calculated using the correct detection of a specific object in a scene and the false detection of a different object which was not in the scene. This catch trial design could lead to better performance on consistent target trials than inconsistent target trials.

The presentation of the object name before the scene could also have artificially inflated performance for consistent targets, as they would have had fewer possible scene locations constrained by the object's semantics. Participants could have used this information to identify likely locations to search for the target object, while inconsistent objects would not have had predictable locations. This strategy would have resulted in an advantage for consistent objects which were present in the scene. Additionally, the label would have accurately predicted the type of scene to be viewed in half of all trials. For example, if the object label was 'kettle', participants could have expected to view a kitchen scene, which could also have affected their strategy, again producing an advantage for consistent targets.

The final concern identified by Hollingworth and Henderson (1998) was the use of a location cue after the scene presentation. Participants may not have needed to search for the target object during the scene presentation if they could recall the scene region present at the location cue. For example, if the object label was 'kettle' and a kitchen scene was presented, the location cue could also have assisted in the present/absent decision. Likely locations for a kettle include kitchen surfaces, so recalling whether the cue corresponded to an appropriate surface would facilitate the decision. In this way, the location cue could be a selectively useful source of information primarily for consistent objects, as inconsistent objects would have no predictable location.

To correct these concerns, Hollingworth and Henderson (1998) adapted the paradigm. Experiment 1 attempted to replicate Biederman et al's (1982) results using 200ms scene presentations, rather than the 150ms presentations used by Biederman et al and Boyce et al. The task was again to determine whether the object presented at the location specified by the cue had matched the object label presented. Detection sensitivity was significantly higher (p<.001) in scenes containing a consistent cued object (0.861) than in scenes containing an inconsistent cued object (0.775), resulting in a consistent object advantage, replicating Biederman et al's (1982) conclusions.

Experiment 2 corrected the catch trial design for present and absent targets, so performance was compared when the same object was cued in two trials but only present in one of them. Higher accuracy was again found for consistent objects compared to inconsistent objects (76.6% and 59.2% respectively) but the measure of detection sensitivity (A') displayed no significant difference between them (0.803 and 0.810 respectively, F<1), indicating that performance was no better for consistent objects than for inconsistent objects. This correction replicated hit rates but not false alarm rates, resulting in detection sensitivity measures which contradicted previous studies. This result indicated that the facilitation found for consistent trials could have been generated by the failure to correctly calculate false alarm rates in catch trials.

In Experiment 3, the object label could be presented either before or after the scene. If participants used the target name to identify likely object locations, then performance would be poorer when the label was presented after the scene, particularly for the more predictable consistent objects. Additionally, the location cue was eliminated and participants had to determine whether the named object had been present anywhere in the scene. This change could also decrease performance for consistent objects, if participants used the cue to decide whether the object was likely to have appeared at the location indicated.

The accuracy levels obtained appeared comparable to those found in the previous experiments, indicating that the removal of the location cue did not inhibit performance to any great extent. There was a significant effect of target label presentation with better performance, measured by A', when the label was presented before the scene (0.836) than after (0.755, $p<.001$). There was no effect of consistency in the pre-view condition, replicating Experiment 2, but there was a significant difference in the post-view condition ($p<.05$), with better detection sensitivity for inconsistent objects (0.781) than for consistent objects (0.729). This result contradicted previous research by evidencing a significant advantage for inconsistent objects.

Experiment 4 introduced a two-alternative forced-choice procedure to minimise response bias. Participants were required to discriminate between two object labels (presented after the scene) which were either both consistent or both inconsistent with the scene context, to prevent response biases in favour of the consistent object. The scene was presented for 250ms, rather than the 200ms presentation times in the preceding experiments, and the longer display time may have allowed participants to initiate one saccade during the presentation. The results provided simply indicated that there was no effect of consistency on accuracy, with consistent objects being responded to correctly 70.7% of the time, compared to 71.6% for inconsistent objects.

Hollingworth and Henderson used these data to conclude that the consistent object advantage evidenced in previous object detection experiments may be attributable to the flaws in the paradigm. However, other issues still remained, including the absence of an adequate explanation for the inconsistent object advantage found in Experiment 3. Additional concerns, not considered by Hollingworth and Henderson are identified and discussed below.

Participants viewed 160 experimental trials, consisting of all four possible targets in the same scene background (two consistent and two inconsistent objects). As participants would have been presented with the same object labels (e.g. 'chicken' or 'pig') twice in the forced-choice procedure, on the second occasion they may have been inclined to select whichever label they had not selected previously. Additionally, it appeared that the same background was presented four times with a different object in the same location each time, so participants may have learned where the target object could be found in the scene and directed their efforts towards it. As the trials were randomised for each participant, this was unlikely to have influenced consistency effects but may have resulted in artificially high accuracy rates, especially in later trials.

The 250ms presentation time in Experiment 4 may have allowed participants to initiate or prepare one saccade during the scene presentation and may have confounded two different issues in the effects of semantic consistency, as identified previously. The existence of semantic inconsistency could 'pop out' of the scene during a single fixation or, if semantic information can be processed from extrafoveal vision, this inconsistency could be detected during scene exploration. With longer display times, it is possible that a distinction could exist between very brief image processing from a single fixation and processing in preparation for a saccade.

In a subsequent study, Hollingworth and Henderson (1999) repeated their experiment with a shorter presentation time of 150ms, to prevent participants from initiating saccades. A significant difference was exhibited between performance for consistent and inconsistent objects in scenes, 64.2% and 67.0% respectively ($p<.05$). This significant advantage, for inconsistent objects rather than consistent objects, replicated the inconsistent object advantage found in Experiment 3 (Hollingworth and Henderson, 1998) using the adapted forced-choice paradigm of Experiment 4, although the accuracy rates themselves were quite low, compared to a chance level of 50%. Two further experiments were conducted in which, instead of selecting between two different object types, when participants could be influenced by their relative consistency with the scene (is one object more consistent than the other?), the forced-choice display presented two tokens of the same object type (e.g. two different cars). No evidence of an inconsistent object advantage was found and no clear account was provided for why Experiment 1 (Hollingworth and Henderson, 1999) had found an inconsistent object advantage when Experiments 2 and 3 had not.

The only explanation this thesis can suggest relates to the presentation of all experimental scene-object combinations. In Experiment 3 (Hollingworth and Henderson, 1999), there was no consistency effect evident in the final 4 blocks, but a non-significant trend was reported towards better performance for inconsistent objects than consistent objects in the first half of the experiment (p=.09). This effect was entirely extinguished in the second half, where accuracy for both object types increased, which may support the criticism that performance could improve in later trials if participants viewed all possible consistent and inconsistent objects in each scene background. As the presentation was randomised, there should be no systematic benefits for either consistent or inconsistent trials but it is possible that participants may selectively remember details from different scene-object pairs and the 250ms presentation time may allow them to use this information.

Further evidence of an advantage for inconsistent objects when scenes were presented only briefly comes from the application of the change blindness flicker paradigm (Hollingworth and Henderson, 2000). One scene image was presented for 250ms, followed by a blank mask and then a second image, also for 250ms, which could be identical to the first or differ in one critical detail. From these brief presentations, participants detected orientation changes to inconsistent objects more reliably than to consistent objects (54.7% and 49.5%

respectively, $p<.05$). Additionally, significantly better performance was found for later than earlier blocks of trials ($p<.001$).

To detect a change, participants would need sufficient information about the target object as it appeared in the first image, to compare to the second image and reach a decision. In this way, the advantage for inconsistent objects may have reflected a facilitation in the processing of information from these objects when viewed for 250ms (first scene image) or preferential representation of this information in memory across the trial. The findings of this experiment and further experiments reported in this paper will be considered in more detail later.

From these experiments, it appears that the consistent object advantage in scenes predicted by the schema hypothesis could not be replicated when potential experimental problems were carefully controlled. This difficulty suggested that the consistent object advantage evidenced by Biederman et al (1982) resulted from the inadequate controls implemented in their experiments. There was little reliable evidence that consistent objects in scenes could be better detected than inconsistent objects and, even when an advantage was found, it did not appear to be influenced by an object's predictability. The use of global scene information may not facilitate the detection of individual objects and, in that sense, Henderson's (1992a) proposal of functional isolation between global scene context and local scene objects may be acceptable.

However, an inconsistent object advantage cannot be explained by the schema hypothesis. Contrary to its prediction, recent results have indicated the existence of an inconsistent object advantage from brief scene presentations (Hollingworth and Henderson 1998; 1999). This advantage is unlikely to be anomalous, as Hollingworth and Henderson's (2000) experiments also indicated facilitation in the detection of changing inconsistent objects over 250ms scene presentations.

The authors suggested that the criterion enhancement model could explain an inconsistent object advantage, as the absence of facilitation for inconsistent object identification could lead to a more detailed representation of the visual information. The hypothesis claims that an inconsistent object advantage could be explained by the increased analysis of the object representation during identification, compared to the lesser amount required for a consistent object, as the context could facilitate consistent object recognition. This facilitation however has been difficult to replicate and was not present in any of the experiments cited which controlled for response bias and participant search strategy, calling into question the validity of this explanation. Although longer fixation times on inconsistent objects have been demonstrated, this only supports the increase in processing when objects are directly fixated and not during brief presentations. There is therefore some limited evidence that global scene context can influence the detection and identification of objects located in scenes, but not in the direction predicted by the schema hypothesis.

Alternative hypotheses to explain an inconsistent object advantage include the memory schema hypothesis, the attentional attraction hypothesis and the attentional disengagement hypothesis (Hollingworth and Henderson, 2000). The memory schema hypothesis proposes that the processing of consistent and inconsistent objects proceeds equivalently but information relating to semantically inconsistent objects is preferentially remembered, as is found in real-world memory tests. Information relating to schema-compatible objects could be lost during a normalisation process, while information relating to inconsistent objects would be retained in a more veridical representation.

However, Hollingworth and Henderson (submitted, cited in 2000) manipulated the inter-stimulus interval between the presentations of two images for comparison, predicting that a greater inconsistent object advantage would result from longer delays, as the construction of a veridical representation would be more complete. Their results did not support this hypothesis, as the inconsistent object advantage remained constant despite the manipulation of the retention interval. Additionally, for this proposal to adequately explain the data described above, the memory advantage for inconsistent objects must become apparent at very short delays during brief presentations (e.g. Hollingworth and Henderson 1998; 1999).

The attentional attraction hypothesis suggests that the deployment of covert attention may be influenced by semantic consistency, with inconsistent regions being preferentially attended. The incompatibility between the object's identity and the scene's context would draw covert attention and the effort required to reconcile the issue would result in a more complete visual representation. To explain the inconsistent object advantage in brief presentations, this hypothesis would require the deployment of attention and the localisation of the inconsistent object within 150ms scene presentations.

The final attentional disengagement hypothesis proposes that covert attention is deployed around the scene to regions of interest based on visual features. Although not automatically drawn to regions of semantic or conceptual inconsistency, once located the increased conceptual difficulty captures attention. This capture results in longer fixations on inconsistent objects, producing a more detailed visual representation. This hypothesis is compatible with the finding that inconsistent objects are fixated for longer than consistent objects. Again, the speed at which this process occurs would need to be within the 150ms presentation time used in previous experiments.

All three hypotheses outline ways in which inconsistent objects may be subject to increased visual processing, resulting in a more complete visual representation. The attentional attraction hypothesis suggests that covert attention may be preferentially allocated to regions of semantic inconsistency. However, the memory schema hypothesis and the attentional disengagement hypothesis predict that scene processing proceeds equivalently regardless of semantic consistency, with differences in object processing occurring upon their detection.

According to the memory schema hypothesis, the effect is modulated by the normalisation of consistent information into memory while inconsistent information would be encoded independently. The attentional disengagement hypothesis suggests that semantic inconsistency would require additional attentional resources to process, which may give rise to a more detailed memory representation. These two proposals are similar, although they explain the inconsistent facilitation through different mechanisms, memory and attention, but both would struggle to explain this effect within 150ms presentations. Because of the difficulty in distinguishing between all 3 hypotheses experimentally, it is only possible to conclude that inconsistent objects may be better represented visually, resulting in their facilitation, but the process by which this may occur is undetermined.

1.7 Evaluating the local processing hypothesis

The local processing hypothesis claims that, although the detection of gist can occur rapidly during scene viewing, this global processing cannot influence the identification of local features such as objects. Instead, context effects occur through intra-level priming, with related objects priming each other's identification. Henderson (1992a) reviewed three objections to this theory, the first concerning the possible confounding of semantic and episodic relationships. Objects can be episodically related rather than semantically related when they share the feature of likely co-occurrence in a scene (Boyce, Pollatsek and Rayner, 1989). However, Henderson (1992a) argued that truly semantically related objects are also likely to be episodically related, as objects with similar semantic meaning are likely to co-occur. Additionally, De Graef (1992) investigated priming across object pairs which were episodically but not semantically related and found similar facilitation on fixation duration, indicating that episodic relatedness was sufficient to induce object-to-object priming.

The second objection identified context effects which cannot be attributed to simple semantic priming effects, such as context effects for support violations (Biederman et al, 1982 and De Graef, Christiaens and d'Ydewalle, 1990). Henderson (1992a) suggested that methodological problems with these studies could explain their results and concluded that the evidence for context effects beyond simple probability effects, and possibly positional effects, was "weak".

In addition, the local processing hypothesis was designed to explain context effects in probability violations only, not violations of spatial relationship.

The final objection claimed that object-to-object priming should generate facilitatory effects in both scenes and non-scene object arrays. Boyce et al (1989) claimed that context effects did not appear in non-scene arrays, finding significant facilitation in an object detection task for an object array presented briefly with a consistent scene background compared to when presented in isolation. However, Henderson (1992a) argued that the 150ms presentations were insufficient for intra-level priming to occur, as a prime object would need to be attended prior to the target object. This claim presumably suggests that, had a prime been fixated during a longer presentation allowing scene exploration, a context effect would have been evident.

The investigation of whether the direct fixation of a related prime object can facilitate the identification of a subsequently fixated target object requires the use of experimental techniques which allow free scene exploration but also monitor fixation positions. Eye movement data can provide both a measure of processing difficulty, by measuring fixation time on objects (Rayner, 1978) and a comparison of saccade behaviour when viewing consistent and inconsistent objects in scenes. Henderson (1992a) cautioned that fixation measures must be selected carefully to avoid measures of post-identification processes. Although inconsistent objects in scenes are fixated for longer and more often than consistent objects (e.g. Friedman, 1979), these measures probably include post-identification processing effects, for example, any extra time required to integrate the inconsistent object into a memory representation. Therefore, the best measure of processing difficulty may be the shortest, the first fixation duration (Henderson, 1992a).

An additional concern arises with scene exploration, as the participants' perceived task could influence their viewing strategy. As people direct their eyes to the image regions most informative for the task they are to complete (Yarbus, 1967), it becomes unwise to compare results across studies providing participants with different instructions. For example, when anticipating a memory test, an inconsistent object advantage may be explained by greater efforts to concentrate on diagnostic or distinguishing objects, rather than predictable objects. Therefore, it is necessary to engage participants in a task which will not bias their saccade behaviour.

Further criticism, not discussed by Henderson (1992a), can be made of the local processing hypothesis in terms of its supporting evidence. The evidence for intra-level object-to-object priming has been found primarily in very limited scene arrays, rather than complex scenes resembling real-world stimuli. Henderson, Pollatsek and Rayner (1987) investigated the effects of a foveal prime and extrafoveal preview of the saccade target on target naming. The participant would fixate a related or unrelated prime before saccading to a target object presented extrafoveally, with the extrafoveal preview of the target being available in only half of all trials. The preview was expected to facilitate subsequent encoding, as found by Pollatsek, Rayner and Collins (1984), and shorten naming latency. The prime and target objects were located either 5° or 10° away from each other and the participants' task was to fixate the prime, then saccade to the target object and name it as quickly as possible. The local processing hypothesis would predict that intra-level object priming would result in faster naming latencies for target objects presented with related primes than for targets presented with unrelated primes.

Although the expected effect of related primes facilitating naming latency was found ($p<.05$), this effect was only evident in the analysis of all trials. The effect of a related prime was greatest when no extrafoveal preview of the target was available. Removing trials in which no extrafoveal target preview was allowed, leaving only preview trials which most closely resemble scene viewing, eliminated the significant facilitation in naming latency for related primes (607ms) compared to unrelated primes (606ms) at 5° and produced approximately 15ms advantage at 10°. However, no explanation for the absence of priming at 5°, compared to 10°, was provided. If priming effects were greatest for 10° saccades with no extrafoveal preview of the saccade target, these effects would not be likely candidates to explain context effects in scene viewing.

Inherent problems with this experiment included the fact that it did not approximate natural scene viewing. Although targets were placed at 5° and 10° from the prime fixation position, extrafoveal processing at these eccentricities cannot be compared in simple and more complex images, due to differences in size and stimulus density among other factors. Participants' saccade behaviour within limited displays of individual objects was either carefully controlled or easily predictable (Henderson et al, 1987; Henderson, Pollatsek and Rayner, 1989 and Henderson, 1992b).

The nature of extrafoveal processing prior to a saccade could differ when deciding where to saccade to within complex scenes. In this case, extrafoveal processing is used to determine the most desirable next fixation position, before computing the direction and size of the saccade required. This decision is often removed in experiments, for example when saccade direction is constant over a block of trials and targets appear at one of only two different eccentricities. It is likely that the extent of extrafoveal processing which is then evidenced does not accurately reflect the processing possible or necessary during complex scene viewing. For the intra-level hypothesis to explain consistency effects in scenes, the facilitation for target object identification needs to be evident in more scene-like stimuli, with free exploration and participant-determined, voluntary saccades.

Henderson et al's (1987) Experiment 3 used structured displays, presenting four objects in a square pattern. The use of larger object arrays allowed participants to exercise more control over fixation positions, although predictable patterns often emerged within such limited arrays. An attempt was made to create global context by sometimes presenting one unrelated item in a four-item display, for example presenting a boat, a truck, a car and a shoe, in which the shoe would be the unrelated object as it is not a mode of transport. The same objects could also appear in arrays in which all the items were unrelated. Target fixation times were compared after fixating a related and unrelated prime in these conditions.

The local processing hypothesis would predict shorter fixations on all objects fixated after a related object. However, Biederman (1981) suggested that an unrelated object in an otherwise semantically homogenous group would ‘pop out’ and be detected more rapidly. Investigating the first fixation durations, targets fixated after a related object had a significantly shorter mean time (269ms) than those fixated after an unrelated object (315ms) and those fixated after an unrelated object in a uniformly unrelated display (300ms, $p<.05$). Fixation times on targets following related primes were shorter than those following unrelated primes and did not indicate faster detection of unrelated objects in related arrays. The experimenters investigated whether this effect could be due to global context effects derived from object categorisation, rather than intra-level priming, and concluded that the facilitation evident for objects in related arrays *was* due to the prior fixation of a related priming object. Therefore the data supported the local processing hypothesis.

Similar results were found by Henderson (1992b), who investigated whether related objects presented as lateral flankers could prime the identity of a target object and found significant facilitation only when an extrafoveal preview of the target was unavailable. The greatest evidence of facilitation was found with related flankers when the target was presented further from fixation and when no extrafoveal preview was allowed. These conditions indicated that the semantic relationship between objects in displays can facilitate identification, but only seems to be used when other, possibly more useful, information is removed from the display. The limitation of this priming effect, using both foveated primes and lateral flankers, to trials without extrafoveal preview of the target object weakens the argument that it plays a significant role in scene viewing.

De Graef, De Troy and d'Ydewalle (1992) evaluated whether object-to-object priming could occur in line drawings of contextual scenes. The presentation of objects in scenes introduced a confound between intra-level effects and those arising from global context. However, an attempt was made to exhibit facilitation on target object fixation durations following fixations on a related prime. De Graef et al argued that Boyce et al’s (1989) experiment, which failed to find facilitation from consistent cohort objects, was biased towards global scene processing and therefore was not a true test of the local processing hypothesis. Additionally, Boyce et al's scenes were presented only very briefly, with insufficient time to allow fixation on a prime object prior to a target, which would be necessary for direct object-to-object priming but not for priming provided by flankers as suggested by Henderson (1992b).

De Graef et al's experiment involved a non-object count task. Participants were required to search for non-objects, which are composed of object-like features but which together do not form any semantically recognisable object. In searching for these, participants should have been selectively targeting objects for fixation, improving the likelihood of object-to-object priming occurring. In an attempt to direct fixations to the target object following a fixation on the prime, Boyce and Pollatsek's (1992) 'wiggle' technique was used, in which a target was rapidly moved up and down after 160ms viewing, to attract the next saccade. The first fixation on each trial was on the prime object, achieved by instructing participants to fixate a variable marker prior to each scene presentation. The desired second fixation was on the target, manipulating the semantic relationship between the two objects. The prime could be related or unrelated to the target object and both objects could be consistent or inconsistent with the scene background. Non-objects were also used as control primes, having no relationship with the target object or scene background, but were never target objects.

The desired saccade pattern occurred in only 34% of trials, comprising the experimental data of interest. In 50.7% of trials, participants eventually fixated the target object, with a mean lag of 3.6 fixations or 1443ms. There was a significant 454ms difference between lag time for consistent objects (1670ms) and inconsistent objects (1216ms) (*p*=.023), indicating that inconsistent targets were fixated sooner than consistent targets when intervening saccades occurred. Although this evidence is suggestive of inconsistent object facilitation, the pattern was not reflected in the data of most interest.

First fixation durations are defined as the length of time the eyes rest on an object, from the first saccade into the object region until the first saccade away from this initial position. This measure is distinguished from the gaze duration, also referred to as the first pass fixation duration, which totals the time spent fixating an object for the first time, possibly including more than one fixation, before a saccade away from the object region.

For the correct saccade pattern data, the target's consistency with the scene background failed to affect the first fixation durations on the target (*p*=.18). This result suggested that global context did not facilitate identification, although inconsistent targets did evidence longer gaze durations (*p*=.013) than those in consistent backgrounds, 424ms and 366ms respectively. First fixation durations and gaze durations on the target for related and unrelated prime trials also failed to show a significant difference. Targets fixated after a related prime had mean first fixation durations of 256ms and mean gaze durations of 376ms, compared to unrelated prime trials with first fixation durations of 235ms and gaze durations of 414ms. The first fixation durations did not show the expected pattern of shorter durations for related primes but gaze durations did, although the effect was not significant (*p*=.12). Therefore, no evidence was found of object-to-object priming in complex scenes.

De Graef et al concluded that although the semantic relationship between objects could have an effect on the ease of identification of a target, it could not be explained by simple object-to-object priming as tested in their experiment. Even a consistent global context failed to facilitate target first fixation durations. While gaze durations showed the expected but non-significant increase for unrelated primes, the most appropriate measures of identification difficulty, the first fixation measures, failed to exhibit this effect. In fact, the first fixation measures indicated the opposite effect of a slight increase in first fixation durations when the prime and target were related but both inconsistent with the scene context.

The authors suggested that these data, taken together with Boyce and Pollatsek's (1992) findings of immediate global context effects, indicated that global background had an effect which was most pronounced during the first two or three fixations on a scene and which disappeared after this time. The longer gaze durations for targets in inconsistent backgrounds, which appeared only when the target was fixated immediately after the prime and not after a longer time lag, supported this proposal. They concluded that both global scene background and local object information could affect the ease of object identification in scenes and identified a possible shift from global to local contextual effects as viewing time increased.

The evidence suggests that object-to-object priming, as hypothesised by Henderson (1992a), cannot account for the consistency effects found in studies using complex scenes. Facilitatory effects were found using related objects in simple arrays but even in these cases, the facilitation appeared limited to, or enhanced by, trials in which appropriate target information was not available prior to fixation, removing extrafoveal preview and presenting targets 10° from fixation. De Graef et al's (1992) investigation of object-to-object priming in scenes also failed to produce the predicted facilitation for related objects in more realistic visual stimuli. Additionally, if primes require fixation to produce any strong context effects, although flanking primes produced small facilitatory effects under very specific conditions, the local processing hypothesis cannot explain consistency effects in brief scene presentations.

Particular difficulty arises in explaining an inconsistent object advantage. The memory schema hypothesis and the attentional disengagement hypothesis claimed that consistency effects arose from differential processing of consistent and inconsistent objects only when fixated, which cannot be applied to brief presentations' results. Although the attentional attraction hypothesis predicts that covert attention can be drawn to regions of semantic inconsistency, this attraction would have to occur very rapidly. Unless regions of semantic inconsistency can be detected extrafoveally from 150ms scene presentations, these hypotheses cannot explain the findings of the research discussed so far. Therefore, it is important to analyse existing research for evidence of rapid detection of inconsistent objects in scenes.

1.8 The effects of semantic inconsistency on natural scene viewing

Current theories explaining consistency effects found in brief presentation experiments have needed to hypothesise how regions of semantic inconsistency could be detected for preferential processing, usually without foveal fixation. This approach has influenced the study of whether semantic information can be processed from extrafoveal vision. To begin with, relevant evidence from saccade-contingent displays of simple object arrays will be discussed, before considering research using complex scenes.

Saccade-contingent displays involve the alteration of the visual stimuli, either the entire image or a selected region, according to the monitored eye position. For example, an object can be presented extrafoveally and then changed for another object during a saccade towards it. This procedure allows the relationship between the extrafoveal saccade target and the subsequently fixated foveal object to be manipulated systematically.

Pollatsek, Rayner and Collins (1984) showed that extrafoveal processing of a saccade target occurred prior to its fixation. They manipulated the accuracy of extrafoveal preview by altering an extrafoveally presented target during a saccade towards it, finding that object identification was facilitated by providing an accurate preview. This saccade-contingent display allowed the relationship between the preview and the target to be manipulated, to determine whether a semantically related preview influenced the subsequent naming latency of the target. The shortest latency would be expected when the preview exactly matched the target object and the longest when the two objects were entirely unrelated. Any decrease from this longest naming latency was considered facilitation attributed to the preview (preview benefit).

Experiment 4 investigated visual and semantic relatedness using object pairs which were either visually similar (e.g. tomato - ball), semantically similar (e.g. baseball bat - ball) or neither (e.g. carrot - ball). A naming latency advantage for a semantically related preview would indicate that the extrafoveal processing of the line drawing target was sufficient to detect its semantic meaning, rather than simply its visual form. However, no clear facilitation was found for semantically similar but visually dissimilar objects, suggesting that object semantics, as manipulated in these stimuli, were not processed from extrafoveal vision and did not affect naming latency.

Experiment 5 investigated whether a preview depicting a different image of the same target object (e.g. a different picture of a cow) could exhibit a facilitatory effect. The images were semantically identical but visually slightly different, sharing key features in order to be identifiable as the same object type. No firm conclusions could be reached about the 87ms advantage found, as the control preview in this experiment was a square, rather than an unrelated object, so critical naming latencies could only be compared to trials involving no extrafoveal object processing during saccade planning. Similar transsaccadic studies have established that information relating to visual features can be retained in memory across saccades, including object viewpoint (e.g. Verfaillie and De Graef, 2000) and structural relations between object parts (Carlson-Radvansky and Irwin, 1995). However, no veridical representation of specific spatial location (e.g. Pollatsek, Rayner and Henderson, 1990, Henderson, 1997) or object size (e.g. Pollatsek et al, 1984) could be maintained across a saccade and there was no evidence that semantic information could be recalled accurately transsaccadically.

Scene processing research has included the study of how saccade targets are selected. Mackworth and Morandi (1967) recorded scan paths during photographic scene viewing and determined that a region's 'informativeness' predicted the likelihood with which it would be fixated. One inch square regions were rated on a 10 point 'informativeness' scale, defined in terms of recognisability, with high ratings associated with unusual details and unpredictable contours. Mackworth and Morandi claimed that peripheral processing edited out predictable contours and simple texture to direct fixations to more informative visual regions, concluding that extrafoveal processing of *visual informativeness* could direct fixations.

Loftus and Mackworth (1978) investigated whether *semantic informativeness* could be used to select fixation locations. Semantic inconsistency was defined as the extent to which an object had "a low a priori probability of being in the picture given the rest of the picture and the observer's past history" (p566). The simple line drawing images subtended 20° x 30° of visual angle and contained minimal landscape detail and some foreground objects to establish global context. Informative (inconsistent) and noninformative (consistent) target objects were located in them with some, possibly imperfect, attempt made to control for visual similarity. For example, a farm scene could contain either a tractor (consistent) or an octopus (inconsistent) as the target object. Conversely, an underwater scene could contain either an octopus (consistent) or a tractor (inconsistent).

Participants' scan paths were recorded while viewing the scenes for four seconds, in expectation of a subsequent recognition test which did not occur. The results were analysed for evidence of differential viewing behaviour when inconsistent objects were present in scenes and produced three important conclusions. Participants fixated inconsistent objects faster than their consistent alternatives, with the unconditional probability of having fixated the target after the first fixation on a scene (by the second fixation) being 0.214 for inconsistent objects and 0.147 for consistent objects. This difference was proved to be statistically significant at the $p<.05$ level by a sign test ($z=1.72$, significance level calculated from this value).

The second conclusion claimed that inconsistent objects were fixated more often than consistent objects, which was again proved significant by a sign test ($z=4.43$, $p<.01$). The final conclusion stated that inconsistent objects were fixated for longer than consistent objects ($F(1,11)=8.23$), a difference which was significant at the $p<.025$ level (calculated from Loftus and Mackworth's data). For first fixation durations, consistent objects received a mean fixation of approximately 240ms compared to 270ms for inconsistent objects. The mean of the second and third fixations together showed a greater difference, with consistent objects averaging fixations of 290ms compared to 410ms for inconsistent objects.

These results allow further conclusions to be reached regarding the use of extrafoveal vision in scene viewing. To determine whether the target had been selected as a saccade target from extrafoveal processing, its distance from the previous fixation location was calculated (equivalent to saccade amplitude) and found to be over 7° on average. The large display size and inter-object distance in the sparsely populated scenes may have resulted in such large saccades. There was no significant difference in size between saccades directed to consistent and inconsistent objects but, coupled with the finding that inconsistent objects were fixated sooner than consistent objects, the evidence suggested that semantic inconsistency could be detected in peripheral rather than parafoveal vision.

The faster fixation of inconsistent objects also implies the existence of specific processing abilities very early in scene viewing. It implicitly assumes the rapid detection of the semantic context provided by the scene, which has been discussed earlier and is possible. In addition to this gist detection, the rapid processing of local object information and the integration of this object information with the contextual analysis must also occur, in order to isolate potentially inconsistent objects rapidly. This information would also need to influence saccade control for inconsistent objects to be fixated earlier than consistent objects.

De Graef, Christiaens and d'Ydewalle (1990) investigated whether inconsistent objects were fixated sooner than consistent objects using a different paradigm. They argued against concluding from experiments such as Biederman et al's (1981; 1982) that object detection was affected by scene context. In these experiments, complex scenes were presented very briefly and participants indicated whether a location cue accurately identified the scene region at which a previously named target had appeared. Context effects could alternatively be explained as a result of participants' guessing strategies in the absence of adequate information, such as during brief presentations with extrafoveally presented objects. This procedure would encourage participants to guess and make use of rapidly detected global information. The results indicated that such information *could* be used under appropriate conditions but did not imply that context information *is* habitually used in scene viewing.

In De Graef et al's study, the line drawing stimuli were the same size as Loftus and Mackworth's (20° x 30°) but were considered to be more naturalistic. Target objects were manipulated to confirm or violate the scene-object relationships identified by Biederman et al, namely support, position, size and probability. Objects violating the probability condition were inconsistent with the scene context. Additionally, object-like figures with no recognisable identity (non-objects) were embedded into the images. The non-object search instruction was expected to encourage participants to fixate all discrete object-like features of the scene. As recognisable objects were irrelevant to the task, the conditions were considered appropriate to investigate saccade behaviour when viewing consistent and inconsistent objects in scenes.

Their results indicated that inconsistent objects were subject to significantly longer first fixation durations, compared to consistent objects ($p<.05$). This effect was modulated by whether the object was fixated early or late in the display duration. Significantly longer first fixation durations (and gaze durations) were found for inconsistent objects fixated late compared to late-fixated consistent objects (248ms and 203ms respectively, $p<.05$), indicating that context influenced fixation durations later in the trial. No similar effect was found for objects fixated early, suggesting that context had no immediate effect on the perception of inconsistent objects.

Additionally, a cumulative probability graph of the proportion of targets fixated as a function of ordinal fixation number and violation condition was provided. Comparing the functions for consistent (no violations) and inconsistent (probability violation) objects showed no evidence that inconsistent objects were fixated sooner than consistent objects at any fixation position, unlike Loftus and Mackworth's (1978) results. The only slight difference indicated that later in the trial, after approximately 14 fixations, consistent objects were more likely to have been fixated than inconsistent objects. De Graef et al concluded that early saccade behaviour for inconsistent objects in scenes was no different to that for consistent objects in first fixation duration, gaze duration or probability of being fixated.

Henderson, Weeks and Hollingworth (1999) also investigated saccade behaviour in consistent and inconsistent scenes. In Experiment 1, line drawings were displayed for 15 seconds each and participants viewed them in preparation for a memory test, which was never actually conducted. Henderson et al replicated significant differences in fixation measures, with inconsistent objects producing significantly longer first pass gaze durations and total fixation durations (first fixation durations were not reported) than consistent objects. However, no evidence was found for the earlier fixation of inconsistent objects, with no difference in the probability of fixating the target or the number of saccades taken to fixate it according to consistency. There was also a trend towards smaller saccade amplitudes directed to inconsistent objects than consistent objects (2.86° and 3.21°).

In Experiment 2, participants searched for an object named before each trial. False alarms (reporting the target present when absent) were higher for consistent objects (11.5%) than for inconsistent objects (1.3%), indicating a response bias. Search time was shorter for consistent objects than for inconsistent objects, 1174ms and 1309ms respectively ($p<.05$ by participants, $p<.10$ by items) and the number of fixations before reaching the target was also less for consistent objects (3.11 and 3.46 fixations, $p<.05$). As the mean number of fixations before reaching the target in this search experiment was 3.29, compared to 10.2 fixations in the previous memorisation experiment, the data indicated an alteration in behavioural strategy depending on experimental task. Participants may have selectively fixated objects during their search, rather than scanned the entire scene thoroughly, locating consistent objects faster through their more predictable locations than inconsistent objects.

Henderson et al (1999) proposed the following explanations for the discrepancy between their results and those obtained by Loftus and Mackworth and De Graef et al. Differences in their experimental designs were identified, which can mostly be categorised as either differences in task or stimuli. Loftus and

Mackworth's (1978) experimental task involved participants viewing the scenes in preparation for a memory test which never occurred. This memorisation task could have encouraged participants to focus their efforts on scene regions which would be diagnostic in a subsequent recognition test. An inconsistent advantage could be explained by the motivation to find unusual or highly detailed items for encoding into a memory representation, to supplement gist information, resulting in the increased processing of inconsistent objects.

Henderson et al's (1999) Experiment 1 used the same instructions but failed to replicate Loftus and Mackworth's inconsistent object facilitation, which may be explained by other experimental differences. In De Graef et al's (1990) experiment, the non-object search task should have emphasised discrete object-like figures in the display and participants may have directed their efforts to objects with unusual visual features. Inconsistent objects would have been more unusual than consistent objects when processed semantically but contained equally usual visual features. However, in this case also, no inconsistent advantage was found in fixation patterns.

Henderson et al's Experiment 2 involved a search for a target identified before each scene presentation. Again no evidence was found for the earlier fixation of inconsistent objects although this result could be explained in terms of the task itself. Identifying likely locations in which consistent objects could be found would facilitate search, but there could be no similar facilitation for inconsistent objects. Also, the target label before each trial could generate expectations which might predict the scene type to be presented in half of all trials, possibly facilitating the detection of gist. Inconsistent target labels could induce expectations of a different scene layout which could interfere with search. It therefore seems likely that task demands would influence scene viewing and the least misleading instructions were provided by De Graef et al (1990), who still failed to exhibit an inconsistent object advantage.

Differences in experimental stimuli, such as display size and scene complexity were also identified as possible concerns (Henderson et al, 1999). Loftus and Mackworth's scenes were projected with a size of 20° x 30°, which may have influenced the large 7° saccade amplitudes exhibited in their experiment. Henderson et al argued that this mean saccade amplitude was particularly anomalous compared to other studies of saccadic behaviour during scene viewing, such as van Diepen, De Graef and d'Ydewalle (1995). Larger display sizes require longer saccades to fixate distant objects and presumably include larger scene objects, although detailed information was not provided by Loftus and Mackworth.

De Graef et al (1990) used more naturalistic scenes also subtending 20° x 30°, although they did not report saccade amplitudes in their experiment and the inter-object distance in their images was less than in Loftus and Mackworth's sparse stimuli. Henderson et al's (1999) scenes were modified versions of De Graef et al's stimuli but were displayed at the smaller size of 10° x 14.5°, which would affect average saccade amplitude towards targets (3.04° in Experiment 1 and 3.68° in Experiment 2). These differences in display size could affect saccade behaviour as the extrafoveal processing of scene regions contributes to the selection of subsequent saccade targets.

Scene complexity also varied but is unfortunately difficult to evaluate because full image details were not available in any of the studies and only a limited number of scene examples were provided. Loftus and Mackworth reproduced farm scenes containing a tractor (consistent) and an octopus (inconsistent), with few non-target objects and some large areas containing no visual detail at all, appearing significantly less complex than natural scenes (see Figure 1.1). The presence of fewer contours and larger regions of empty display would have resulted in less lateral masking when processing extrafoveal objects and may explain the larger saccade amplitudes.

De Graef et al and Henderson et al used mostly the same stimuli set, created from photographs, and the examples provided included several non-target objects, unlike those reproduced by Loftus and Mackworth. Examples of the images provided by these researchers are included in Figure 1.2. These images were still rather simplistic however and, as Henderson et al (1999) admitted, they were "a significant visual simplification of natural environments" (page 224).

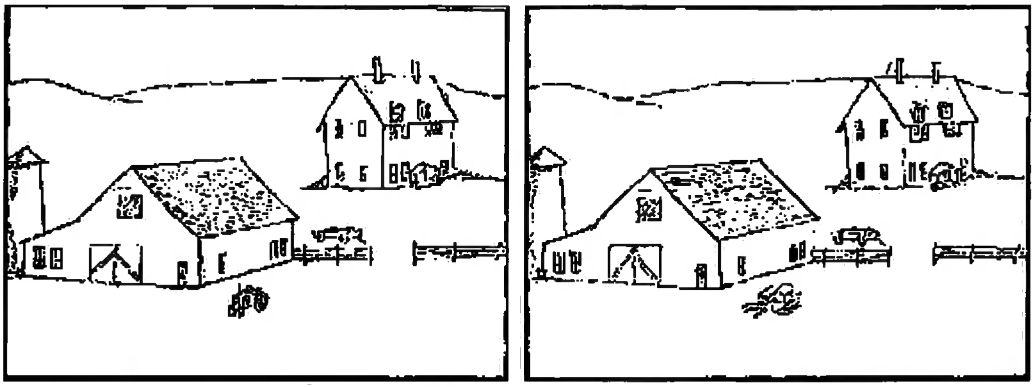


Figure 1.1: Example images from Loftus and Mackworth's (1978) stimuli.
A consistent target (tractor) and an inconsistent target (octopus) located in a farm scene.

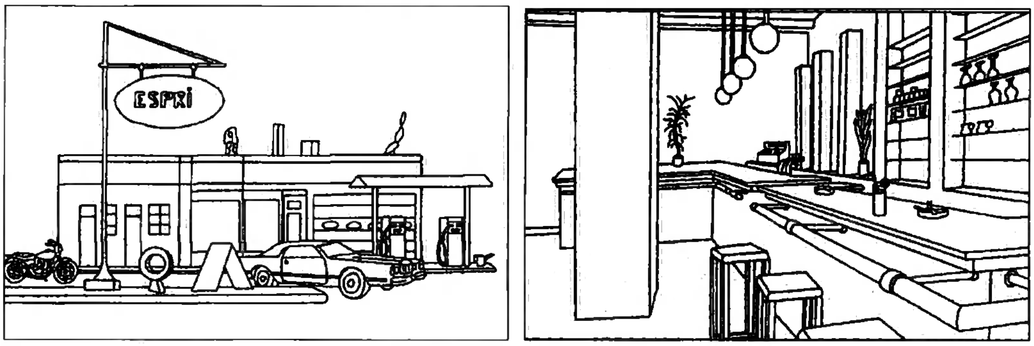


Figure 1.2: Example images from De Graef et al (1990) and Henderson et al (1999).
A 'gas station' scene, containing non-objects, was used by De Graef et al.
A 'bar room' scene example was provided by Henderson et al.

De Graef et al developed images of additional non-objects which, from the examples provided (see Figure 1.3), may have been visually different to both consistent and inconsistent objects, as they contained rather convoluted outlines and few straight lines. Without access to the complete sets of experimental images, it is impossible to conclude whether significant differences could explain discrepancies in results. However, the complexity of the images could affect the ability to process extrafoveal information and select saccade targets.

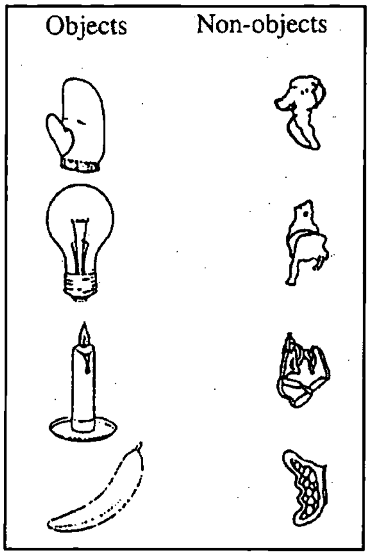


Figure 1.3: Example objects and non-objects used in De Graef et al's (1990) experiment.

Additional concerns were identified by Henderson et al (1999) including the possibility that Loftus and Mackworth's significant result could simply have been due to statistical error resulting from their equipment. The eyetracking equipment available at the time only had a low spatial and temporal resolution which might have provided misleading results. The 12 year delay between the 1978 study and De Graef et al's (1990) experiment emphasised the increased sophistication and accuracy of the available equipment which may explain the differences in results.

The final concern was that visual and semantic consistency could have been confounded by Loftus and Mackworth. Target objects designated as semantically inconsistent could also have been visually inconsistent if, during the creation of the scenes, the consistent objects were drawn into the appropriate images but the inconsistent objects were either drawn independently or within consistent scene backgrounds. The 'swapping' of objects across scenes might have introduced stylistic differences into the completed inconsistent images.

Although not discussed by Henderson et al (1999), visual differences other than style were also indicated in Loftus and Mackworth's example scenes (Figure 1.1), as the tractor and the octopus were not well matched for visual similarity.

The tractor, as a man-made construction, consisted of mostly geometric shapes (e.g. tyres, bodywork, steering wheel) and straight lines intersecting at right angles, but the octopus' outline was more organic and did not include straight lines. Similarly, the backgrounds would be different, as an underwater scene would presumably contain fewer straight lines than a farm. Inconsistent objects could have appeared visually inconsistent in scenes, without any semantic processing needing to occur. However, this possibility is based only on the one example available and not Loftus and Mackworth's entire stimuli set. The images used by De Graef et al and Henderson et al should not have been similarly affected, as there should be less visual variability across scenes of room-like backgrounds (rather than landscapes) containing man-made household items (such as tables and chairs).

Some additional experimental paradigms have contributed to the study of whether semantic inconsistency can attract fixations. The change blindness flicker paradigm involves the presentation of two scene images which are identical except for one small feature. These two images are displayed alternately, with an intervening image (usually blank) acting as a mask, resulting in the appearance of one flashing image. This presentation continues until the participant terminates the trial and the response time indicates the speed with which the participant fixated the target object to detect the change. Hollingworth and Henderson (2000) found that changes to inconsistent objects were detected earlier than changes to consistent objects ($p<.01$), suggesting indirectly that this inconsistent change attracted fixations more successfully than changes to consistent objects.

Further to this, De Graef (1998) used eye movement recording to investigate very early saccade behaviour for evidence of inconsistent object facilitation from extrafoveal vision. He suggested that both a consistent object advantage and an inconsistent object advantage might be valid research findings. A strong conceptual facilitation of consistent objects could simply wash out any inconsistent object effects, resulting in only scenes exhibiting little consistent conceptual facilitation capable of evidencing any possible inconsistent facilitation. De Graef re-analysed a sub-set of De Graef, De Troy and d'Ydewalle's (1992) 'wiggle' study data which investigated object-to-object priming in scenes. Each trial began with a fixation on a prime object and after 160ms, target objects, located on average 7.5° from the prime object, were wiggled to attempt to direct a saccade to them. The primes and targets could be consistent or inconsistent with the scene background, which also affected their consistency with each other. Participants saccaded directly to the target (direct hits) on only approximately 34% of trials, with no difference in accuracy between consistent and inconsistent targets.

To investigate whether consistent or inconsistent wiggling targets attracted participants' saccades better, the first fixation duration on the prime object was recorded, as shorter fixation durations would indicate an earlier saccade to the target. First fixation durations on primes with consistent and inconsistent targets were 392ms and 375ms respectively, which were not significantly different, but indicated slightly faster saccades to inconsistent targets. Significant effects of inconsistent facilitation were only apparent for gaze measures when target skips (target was not fixated at all) and delayed hits (time lag trials) were included in the analysis. The slightly longer prime fixation for consistent targets was explained by consistent objects being easier to process extrafoveally than inconsistent objects. De Graef argued that due to the greater ability to process consistent object information extrafoveally, "the need for foveal target analysis is reduced, resulting in a subsequent delay or even cancellation of target fixation" (p323). However, these predictions were not entirely supported by the data.

No clear reduction was found in direct hits for consistent targets compared to inconsistent targets as a result of saccade cancellation (35.0% and 35.8%). Additionally, a significantly higher target skip rate for consistent targets was predicted due to saccade cancellation but, although a difference was exhibited in the predicted direction (17.5% for consistent targets and 11.1% for inconsistent targets), it was not statistically significant (p=.13). Finally, reducing the need for foveal analysis should result in shorter target fixation durations for consistent objects in direct hit trials, as longer prime fixations would result in increased extrafoveal processing prior to fixation, as found in transsaccadic change studies (e.g. Pollatsek, Rayner and Collins, 1984). The target fixation data were not provided in this paper but the data presented in De Graef et al (1992) provided some indication of whether this pattern occurred.

In De Graef's 1998 paper, the prime was always consistent with the scene, although the 1992 experiment from which the data were derived manipulated the prime-target relationship. To control for this, the data from the original 1992 paper was re-analysed, for the purposes of this thesis, to compare fixation times on consistent targets after a related (consistent) prime with fixation times on inconsistent targets after an unrelated (consistent) prime. A preview advantage for consistent objects should show shorter target fixation times compared to inconsistent objects. Mean first fixation times on consistent and inconsistent targets were 240ms and 238ms respectively, showing no evidence of the expected facilitation for consistent objects due to improved extrafoveal processing. Admittedly, using the 1992 results to investigate the prediction was not ideal, as the 1998 results were only a subset of the earlier data, but the analysis suggested that further evidence is needed to support De Graef's explanation for the slight increase in prime fixation for consistent trials.

De Graef also investigated the time course of saccade initiation, to determine whether consistent or inconsistent wiggling objects attracted saccades better. Very fast gaze shifts, with prime fixation times less than 240ms and considered unlikely to have been triggered by the target wiggle, were rarely directed to the target object. However, when the target was inconsistent, more saccades were initiated, regardless of where they were directed. This evidence suggested that the presence of an inconsistent object triggered more saccades with very short latencies than a consistent object.

With latencies between 240ms and 320ms, the involuntary, reflexive saccades initiated in response to the wiggle were target-directed more often than very fast gaze shifts. In the substantial number of trials when the saccade was not directed at the target, there was a higher proportion of saccades for inconsistent object trials than for consistent object trials, indicating that the presence of an inconsistent object may have elicited more saccades overall. For target-directed saccades, more saccades were initiated towards consistent objects than inconsistent objects at the start of the time band (approximately 240-270ms), an increase which was controversially suggested to be due to the enhanced capture of attention by consistent objects. This effect was followed by an increase in inconsistent target-directed saccades until the end of the time band, possibly explained by an increased salience for wiggling inconsistent objects.

Saccades initiated with 340ms to 400ms latencies were considered voluntary and showed a larger proportion of target-directed saccades than found for saccades with shorter latencies. For both target-directed saccades and those directed elsewhere, a larger proportion of saccades were initiated during inconsistent trials than consistent trials. The increased tendency to fixate inconsistent objects was explained by the author as a result of the greater need for foveal processing.

The final time band contained a further peak in the saccade latency distribution between 420ms and 460ms for consistent object trials only. These delayed, voluntary saccades were explained by De Graef as saccades in which extrafoveal target processing increased prime fixation times and delayed saccade initiation. However, the concerns identified previously with this hypothesis are also relevant in this case and further evidence would be useful.

Together, the data were believed to confirm that inconsistent objects in scenes elicited more rapid, voluntary saccades. Conversely, consistent objects attracted more delayed, voluntary saccades. However, no further supporting evidence for the earlier fixation of inconsistent objects has been found from eyetracking studies. The difference in proportions described in this study may be so small that it would be difficult to evidence from small numbers of experimental trials and participants in eyetracking experiments.

A second unidentified wiggle study was reported in which the distance between the prime and target objects was manipulated (3° or 8° from fixation) and the target was wiggled after 140ms rather than 160ms. Both near and far targets received a higher proportion of direct hits than found previously, attributed to the shorter wiggle latency. The effects of both distance from fixation and object consistency on direct target hits were found to be significant (*p*<.007, *p*<.03) although there was no interaction. Inconsistent targets were fixated directly after the prime more often than consistent targets (62.5% and 56.1% respectively) and the difference was particularly clear for far targets (60.3% and 49.5%). These data were interpreted as evidence that the selection of a saccade target was determined by the need for foveal analysis.

The investigation of saccade latency distributions exhibited a similar pattern for far targets as the previously reported experiment, although no increase in direct hits to consistent targets was found in the reflexive time band, only being evident for near target direct hits and non-target-directed saccades with far targets. Instead, a large increase in saccades to inconsistent targets was seen just after the reflexive time band, suggesting that these were not initiated in response to the wiggle. This increase was found for both near and far targets and for all saccades away from the prime, suggesting that a wiggling inconsistent object elicited more saccades between the reflexive time band and the fast, voluntary time band (280-340ms), 140-200ms after the wiggle. Increases in saccades in consistent object trials were associated with the predicted time bands, retaining a third peak beyond the voluntary time band which was not found for inconsistent objects.

De Graef concluded that the evidence supported inconsistent objects as more salient saccade targets, as there were a larger proportion of voluntary, fast saccades directed to inconsistent targets than consistent targets in both experiments. Also, there was some evidence of a time span in which consistent objects were subject to a greater reflexive orienting response, early in the reflexive saccade time band. Reflexive saccades in consistent trials were also more closely associated to the wiggle than saccades in inconsistent trials. However, the results only compared graphical representations of elicited saccades and did not allow the statistical analysis of these differences.

The experimental results reviewed in this section have confirmed that the evidence supporting the processing of semantic consistency from extrafoveal vision is incomplete. Many studies in which participants were allowed to freely scan visual scenes failed to replicate Loftus and Mackworth's result that inconsistency could be detected from approximately 7° in extrafoveal vision. Although De Graef's conclusion that consistent and inconsistent objects may preferentially elicit saccades at different time periods was highly suggestive, the data only displayed patterns and could not confirm significant differences in saccade behaviour when consistent and inconsistent objects were displayed in scenes. Therefore, further work needs to be conducted before appropriate conclusions can be drawn.

1.9 Conclusions and experimental hypotheses

The evidence described so far has considered the role of semantic information in scene viewing. The theoretical hypotheses proposed to explain consistency effects have been discussed and the evidence relating to these effects has been considered. The detection of semantic inconsistency from brief presentations of scene images has also been investigated and there have been some indications that it may indeed be possible to detect items which are semantically inconsistent with their context, in some way which is as yet unclear.

In order to integrate the results described so far, the existence of both a consistent object advantage and an inconsistent object advantage modulated by experimental design could explain the diversity of findings presented by different researchers. Different experimental paradigms have produced apparently conflicting evidence for both consistent and inconsistent object facilitation. These two positions need to be reconciled to understand the role of semantic information in scene viewing. In this thesis, further evidence of facilitation for inconsistent objects in object detection and identification paradigms will be sought, to determine whether such consistency effects can be replicated with different experimental stimuli and designs.

The primary aim of this thesis is to investigate whether the paradigms providing evidence for consistency effects are replicable or whether the findings could instead be the result of methodological problems. This investigation involves distinguishing between two main issues outlined previously; the first being the ability to detect regions of semantic inconsistency early on in scene viewing and the second being the ability to use any semantic information which can be obtained, either upon first fixation or after extended viewing of a scene, to direct eye movements towards regions containing inconsistent objects. Of additional interest is the investigation of the applicability of consistency effects to real-life scene viewing.

Several problems with existing research have been identified from previous experiments and specific methodological concerns are addressed in the following series of experiments, to determine whether consistency effects can be reliably evidenced. One of these concerns is the use of several similar scenes, including many different target objects located in the same background, which may influence the quality of the data obtained. In the study of early scene viewing, it would be particularly important for each experimental scene to be novel to participants. With familiar scenes, the processing of gist, for example, could be influenced by the prior identification of a similar image. The role of global context on object identification would be particularly affected by familiarity, as context could be primed by a previous image presentation or the remembered location of a diagnostic item from a previous trial. This prior exposure to a similar image may alter the level of processing required in order to perform the required task.

Many experimenters such as De Graef and Henderson and their colleagues reported using stimuli sets which included different target objects embedded into the same or similar scenes, with all possible combinations being shown to each participant. For example, two-alternative forced-choice experiments by Hollingworth and Henderson (1998, 1999) involved the presentation of 160 trials to each participant, created from combinations of 20 background scenes by 2 consistency conditions by 2 target objects of each consistency by 2 label positions in the response screen. This manipulation indicates that each scene background was presented 8 times to each participant, with the target object located in the same place each time (Hollingworth and Henderson, 1999).

Participants might have learned the appropriate region of the scene from repeated exposure to the same background, which would have been a particular concern when the images were presented for longer durations, up to 250ms. In addition, the presentation of a scene background followed by a previously viewed object label may also have influenced participants' responses. For example, if a participant had observed the target in one presentation, the following presentation of the same scene background with the same object label selection screen (e.g. 'chicken' or 'pig') would be likely to prompt the alternate response, regardless of whether the target was viewed or not.

Familiarity with experimental scenes and objects could also be a cause for concern in experiments investigating saccade behaviour. De Graef et al's (1990) study involved participants viewing line drawings containing non-objects and targets possibly undergoing relational violations. Each participant viewed the full complement of 135 experimental scenes, consisting of 5 different versions of each of 27 scene backgrounds.

To begin with, the observation of a scene background would be expected to be less thorough on the fifth presentation than the first. Also, as the 5 versions varied according to the component objects and non-objects included, viewing behaviour may have been affected by the prior observation of either a non-object or an inconsistent object at a specific location, particularly considering that unusual objects (either difficult to identify or inconsistent with the scene context) would be fixated for longer and better represented in memory. Equally, an object previously fixated in another trial could require less processing for identification than a novel object. When investigating scene scanning, the prior presentation of a similar image may direct participants to alter a natural saccade pattern, depending on the outcome of previous trials. Additionally, inconsistent objects in real-life scenes have been found to be better represented in memory over reasonably long periods of time, so it would be unwise to fail to control for possible recall effects of previously viewed inconsistent objects, both for long trial durations and also within shorter time spans.

For the reasons outlined above, the experiments contained in this thesis were designed to prevent unnecessary repetition of experimental stimuli. The experiments limited participants' exposure to the experimental scenes and avoided displaying the same scene to participants more than once, especially when two-alternative forced-choice decisions were required. The same scene context (i.e. its identity), could be used repeatedly across trials but the visual composition of each experimental display was unique within each block of trials, to prevent participants from learning where target objects were likely to be located. For example, more than one bathroom scene could be used in one block of trials but the images would depict the same layout from a different angle, if not a completely different room.

The importance of the participants' fixation position, relative to the target object, was also underestimated in previous research. The study of object detection in complex scenes, especially from brief presentations, must consider the use of both foveal and extrafoveal vision. Most of the studies considered previously used central fixation positions, with targets appearing at random eccentricities in the visual scene. Although pairing consistent and inconsistent objects and presenting them at the same location controlled for eccentricity effects, the data analysis would not consider any variability in accuracy modulated by a possible interaction between eccentricity and consistency.

The ability to identify an object from more highly degraded visual information, such as that available from peripheral vision, may be better for consistent objects than inconsistent objects. The consistent scene context could constrain the possible identity of the item and an 'educated guess' would be possible from crude visual features. Conversely, the identification of an object highly inconsistent with the scene context may be more difficult from degraded visual information and require foveal or parafoveal analysis. In this way, a distinction may occur between the ability to recognise an object, according to its consistency, with respect to fixation position.

This hypothesis would predict that an inconsistent object advantage would be more likely to occur at eccentricities closer to fixation, at which the object could be readily identified. Similarly, an advantage for consistent objects over inconsistent objects may occur further into peripheral vision, where consistent object identification may be facilitated. In the research reported in this thesis, fixation position relative to the target object was manipulated to determine whether semantic inconsistency could be detected from near or far extrafoveal vision. The proposed distinction between processing consistent and inconsistent objects across retinal eccentricity may influence the presence and absence of the consistency effects evidenced across various studies. For these reasons, the participants' fixation position was ensured to be at a known location and distance from the target object, to investigate whether performance on consistent and inconsistent objects varied according to target eccentricity.

A final concern identified that the processing of semantic information has been investigated in many cases with the use of simplistic visual stimuli which assist in the manipulation of specific items or relationships. Limited evidence is available from complex scene viewing, with the stimuli usually used failing to be as realistic as real-world images. Principally, a distinction exists between the possibility of processing semantic information from simple scene-like images and from the naturally complex visual information which we process and interpret during everyday viewing. Attempts need to be made when interpreting experimental results to consider their applicability to real-world viewing.

In the investigations to be described here, the intention has been to apply the improvements described above to diverse experimental paradigms, such as object identification from brief presentations, change detection and obtaining scan paths during scene viewing. Using a number of different techniques, the aim will be to integrate obtained results and attempt to apply the current research to existing work and also our perception of real-world scene viewing. More specifically, a number of research questions were identified from gaps in existing knowledge, which are provided below, and which will be investigated in greater detail.

- Do inconsistent objects in scenes evidence a detection advantage upon first fixation on a scene and is this effect influenced by its location relative to fixation position?
- Is such an effect replicable using more realistic scene stimuli such as photographs?
- Is there any evidence of preferential earlier fixation on inconsistent objects compared to consistent objects, in line drawings or in more naturalistic images like photographs?
- Is there a distinction between whether we *can* process and use semantic information under specific conditions and whether we actually do so during the course of real-world viewing?

Chapter 2

Brief Presentations of Line Drawing Scene Stimuli

2.1 Introduction to brief presentation experiments

Experiments 1 to 4 were designed to investigate extrafoveal processing of semantic information from brief scene presentations. Biederman et al (1982) for example found consistent objects to be detected faster and more accurately than inconsistent objects, although this conclusion was not always reliably replicated. Antes, Penland and Metzger (1981) concluded from their experiment that consistent objects presented in a coherent scene were not identified from a four-alternative forced-choice at a rate better than chance from a 100ms scene presentation. Any discrepancy in performance between consistent and inconsistent trials could have been caused by a response bias, with inconsistent objects resulting in poorer than chance accuracy.

Hollingworth and Henderson's (1998) critique of Biederman et al's object detection paradigm (page 35) concluded that the evidenced facilitation for consistent targets could be explained by inappropriate experimental procedures. They replicated a facilitatory effect for consistent objects when the scenes were presented for 200ms. Participants were more accurate at detecting a consistent object, identified as the target before the scene presentation, at a cued location than a similar inconsistent object. However, introducing a more stringent catch trial design and calculating a measure of detection sensitivity, rather than simply reporting percentage correct trials, eliminated any significant difference between consistent and inconsistent trials.

In a further experiment, the location cue was removed, requiring participants to decide whether the target had appeared anywhere in the scene, and the presentation of the object label identifying the target appeared either prior to or after the scene presentation. When the target was identified before the scene, no difference in performance was found between consistent and inconsistent targets, replicating the previous experiment's results. However, in the post-view condition when the target was identified after the scene presentation, a significant (p<.05) advantage was found for inconsistent objects over consistent objects.

This facilitation for inconsistent objects was also found in their 1999 experiment. Scenes were presented for 150ms only and participants were provided with a two-alternative forced-choice display containing either two consistent objects or two inconsistent objects, to prevent response bias. Again, a significant advantage was found for inconsistent targets (p<.05), which were correctly identified as having appeared in the briefly presented scene more often than consistent targets.

Possible concerns with these experiments were identified (page 39). The repeated presentation of the same scene background with target objects located in the same region each time may have given rise to an advantage for inconsistent objects, which are subject to a memory and recall advantage. Also, the failure to manipulate participants' fixation positions relative to the target objects prevented any conclusions being reached about the processing of semantic information from foveal or extrafoveal vision. Although paired consistent and inconsistent target locations were matched to avoid introducing a bias due to proximity of fixation, the distance between the object and fixation position was not systematically manipulated, to investigate whether facilitatory effects were influenced by the level of visual detail available during the single fixation. It was hypothesised that consistent objects may be easier to identify from less detailed visual information, such as that obtained from extrafoveal vision, because context would constrain the identification possibilities. By manipulating fixation position relative to the target object, it would be possible to determine whether semantic information could be processed from extrafoveal vision.

While this possibility of an interaction between consistency and retinal position makes intuitive sense, it may also be supported by some existing data. De Graef (1998) observed the distribution of saccades initiated upon the first fixation on a scene, which indicated more saccades triggered during inconsistent trials than consistent trials, even for saccades with very short latencies less than 220ms. This pattern suggested that the inconsistency in the scene somehow influenced the initiation of the first saccade.

The location of the targets was manipulated relative to the fixation position and the discrepancy in saccade accuracy between consistent and inconsistent targets was larger when they were presented far from fixation (approx 8°). Although there was no significant interaction between consistency and fixation distance, a 10.8% increase in direct hits towards inconsistent objects over consistent objects was found in 'far' locations, compared to a 2.0% increase for 'near' targets (approx 3°). This effect seemed to be caused by a decrease in direct hits for consistent far targets, indicating that, at locations further from fixation, inconsistent objects were equally salient as when located near to fixation but consistent objects were less salient. These results suggested that, when presented extrafoveally, inconsistent objects were more salient saccade targets than consistent objects, contrary to the predictions made above. However, the data did indicate that consistency effects could be influenced by fixation position, justifying its further investigation.

2.2 Experiment 1: Introduction

The paradigm used in Experiment 1 was adapted from those used by Hollingworth and Henderson (1998,1999). In their experiments, a scene image was presented briefly during a central fixation. A response screen was then displayed, depicting two objects, one of which had been present in the scene. Participants were required to identify the target from this two-alternative forced-choice display. This procedure was adapted for the current experiment by setting the scene presentation time at 120ms and manipulating fixation position, relative to the location of the target object (see Figure 2.1). A fixation cross presented before each scene directed the participant to fixate a specific region of the scene.

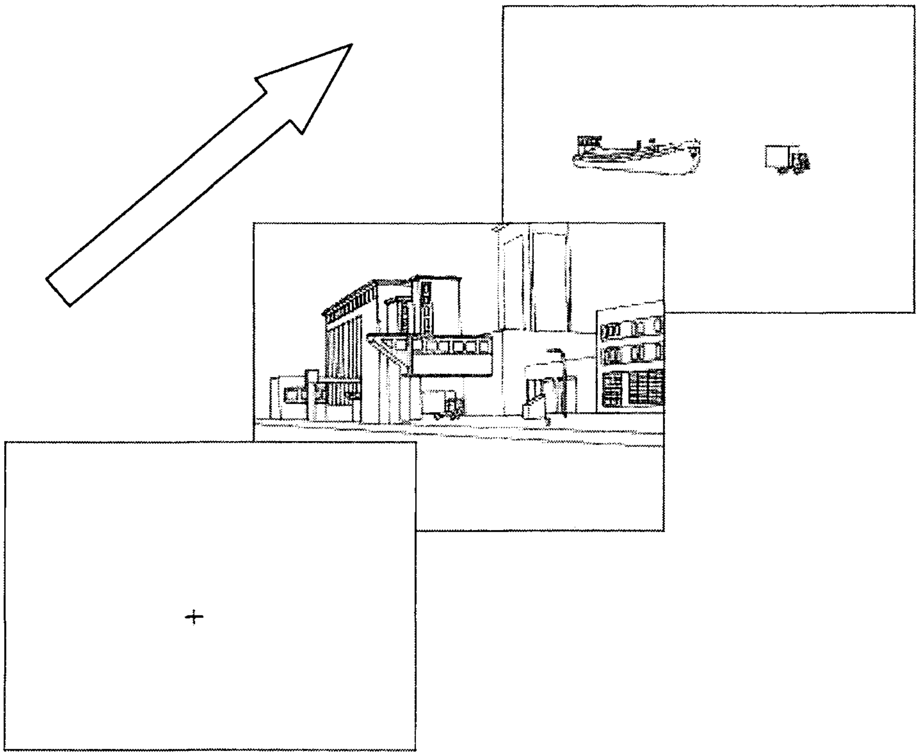


Figure 2.1: Example displaying the sequence of images in a trial.
A variable fixation cross was displayed for 1000ms. A scene image was presented briefly for 120ms. Finally, a display presented two alternative objects, either both consistent or inconsistent, for selection. In this example, the correct response would be pressing the right-hand button.

Target objects were selected which could be embedded into a scene in which they would appear either consistent or inconsistent. This decision was made originally by the experimenter and confirmed by an investigation into the suitability of the scene stimuli. More information can be found in Appendix A, which discusses the investigation of all the scene images used as experimental stimuli in this thesis. Some modifications were made to the objects selected for use on the basis of these data.

For each scene with a consistent target object, an inconsistent object was selected to be embedded in an alternative version of the scene. Every background had at least two paired versions, one containing a consistent target and another with an inconsistent target. In reality, all experimental backgrounds were produced with two consistent and two inconsistent versions. To avoid response bias in the two-alternative forced-choice procedure, with participants selecting a consistent distractor over an inconsistent target, the two alternatives were selected to be either two consistent objects or two inconsistent objects. Backgrounds with only one possible consistent object were used as practice trials, paired with an inconsistent object as the alternative distractor.

Since retinal resolution decreases at increasing retinal eccentricities, higher accuracy should be observed in trials where the target object was presented close to fixation and accuracy should decrease as a function of eccentricity. However, predictions relating to the consistency manipulation are less clear. Previous research suggested that inconsistent object identification was facilitated in brief scene presentations (e.g. Hollingworth and Henderson, 1998, 1999) indicating that an inconsistent target may be detected more reliably from brief presentations in this experiment, facilitating performance for inconsistent objects across all fixation positions. Similarly, De Graef's (1998) data indicated an inconsistent object advantage at further extrafoveal distances, suggesting that the effect could be modulated by foveal or extrafoveal processing.

An alternative proposal suggests that consistent objects could be recognised from fewer or less detailed visual information than would be required to resolve inconsistent objects. At increasing eccentricities, where poorly resolved visual information is available, consistent objects may be facilitated relative to inconsistent objects. Conversely, performance for inconsistent objects would be highest at positions closer to fixation, where sufficient visual detail would be available to identify the target.

The results were analysed for evidence of a significant effect of fixation position, with better performance at fixation positions closest to the target. A significant effect of consistency could be predicted, if performance for consistent or inconsistent objects was facilitated at all fixation positions. In addition, it was possible to investigate how consistency and fixation position interacted to influence accuracy. In this way, any variation in the ease of recognition of consistent or inconsistent objects at different eccentricities could be investigated, perhaps to indicate a consistent advantage at further extrafoveal fixation positions with an inconsistent advantage at closer fixation positions or vice versa.

2.3 Method

Participants

A total of 100 participants took part in this study. 40 male and 60 female undergraduates at the University of Durham volunteered to participate. They were all naïve to the purposes of the experiment.

Apparatus

This experiment was run on a number of Intel Celeron PCs, with 533MHz Processors and 64mb RAM. 15” monitors displayed the experimental images at a resolution of 800 x 600 pixels. The program was written in C++ using Borland C++ Builder. The following measures were taken to present and change the images within a single refresh period (60-75Hz) and display them for the correct durations. The images were presented using DirectDraw, a component of Microsoft DirectX and an accurate hardware timer in the computer was used to measure the presentation periods and response times. Responses were collected using purpose-built button units, which plugged into the parallel port of each computer. All computers had sufficient video memory to simultaneously contain all of the images used in the experiment.

Participants were provided with instructions which included a diagram depicting the trial sequence, a consent form to confirm willingness to participate and a 60 cm measuring ruler, which was used to ensure that all participants were seated at equal distances from the monitors. A debriefing sheet was also provided at the end of the experiment.

Materials

The scene backgrounds and target objects were selected from the Leuven line drawing library. Backgrounds provided with more than one consistent object, of which there were 11, were used as experimental trials, while backgrounds with only one object were used as practice trials. The consistent objects were provided as independent images which could be superimposed onto the backgrounds. Inconsistent targets were also selected from these images but were located in a scene incompatible with their identity.

An attempt was made to match the inconsistent object with its consistent alternative in size and shape where possible. These inconsistent objects were pasted onto the selected backgrounds and some alteration was often necessary. For example, some object images needed rotating or resizing to scale with the background, although it was attempted to maintain line thickness constant throughout the finished scene. Although attempts were made to embed both the consistent and inconsistent objects in the same location, this was not often possible and many inconsistent objects were re-located to a more appropriate region.

Every one of the 11 scene backgrounds was modified to create four alternative versions. The finished scene set for each background consisted of two scene versions, containing one consistent object each, and two scene versions, each containing one inconsistent object. Therefore there were 44 different versions of 11 scene backgrounds used in this study. In addition, a set of 12 practice trials were created, from six backgrounds, with two versions per scene background containing either a consistent or an inconsistent object. Every participant saw the entire set of practice trials, with fixation positions selected to include all fixation distances used in the experimental trials. Response screens were also constructed, containing images of two objects located on a white background. The objects depicted were either the two consistent objects allocated to that particular scene background or the corresponding two inconsistent objects, depending on whether the target object had been consistent or inconsistent.

These images were presented at a resolution of 800 x 600 pixels. The entire scene, which filled the monitor screen, subtended approximately 19° in height and 26° in width at a viewing distance of 60cm. The individual scene targets subtended an average of 3.19° in width and 3.27° in height. An analysis of object sizes found no significant difference between the sizes of consistent and inconsistent objects, measured as the target area in pixels (*t*(42)<1, p=.42). The final images used are provided in Appendix B.

Design

There were two independent variables manipulated in this study, the consistency relationship between the target object and the scene background and the distance between the target object and fixation. The consistency of the target object had two levels, as the target could be either consistent or inconsistent with the scene. The distance of the target object from fixation had 5 levels labelled 0 (for targets presented directly on fixation), 1, 2, 3 and 4, with eccentricity increasing by about 3° with each increasing level. These positions were selected by identifying four points, with equal increment between positions, in a straight line from the target. The direction of this line of fixation positions was usually towards the centre of the image rather than the edges, to allow the fourth, most distant, location to correspond within the scene boundaries. The dependent variables measured were response accuracy and response time.

To balance out any response bias, participants selected between two objects, either both consistent or both inconsistent, after the presentation of each scene. Experimental scenes were created containing each of those objects, so each object depicted in a response display did occur as a target in an alternative experimental scene. However, any participant viewing all 44 possible scene-object combinations may find the second or subsequent presentations of the same scene background easier than the first and would be able to use prior knowledge from previous presentations to perform the task. For example, when on the first presentation one object was selected from the two-alternative forced-choice display, on a second presentation, the other object would be more likely to be selected, particularly if the participant was unsure of the correct response.

This problem could also jeopardise any results comparing consistent and inconsistent objects. For example, memory for inconsistent objects is known to be better than for consistent objects, possibly facilitating inconsistent trials. Conversely, performance for consistent objects could be facilitated through an enhanced search of a repeatedly viewed and processed scene. The subsequent presentations of the same backgrounds may encourage participants to ignore the fixation cross intended to guide their fixation position and attempt to process the image in a different manner, perhaps by focussing on regions of the scene fixated previously. By keeping the backgrounds unfamiliar, participants were considered to be less likely to adopt an independent strategy.

A final concern was the eccentricity manipulation, as a single participant could not view each scene-object pair at different eccentricities. Each participant viewed one consistent and one inconsistent version of each background (i.e. half of the possible 44 scenes) so two participants were required for all the scene images to be viewed. However, considering the five possible eccentricities of each target object, for each of the 44 scene-object combinations there were five possible experimental trials. Therefore ten participants were required for all possibilities to be viewed. To this end, a total of 100 participants were recruited so that there would be ten responses to each possible trial from which to analyse results.

To summarise, each participant viewed 22 of the possible 44 trials which were presented in a random order. The selection of 22 experimental trials ensured that each participant viewed only one consistent and one inconsistent version of each background. As inconsistent objects could rarely be embedded in the same location as consistent objects, viewing one version of each type was considered unlikely to produce practice effects. The distances from fixation were also randomised so that each participant received a distribution of targets at different eccentricities from fixation. Each person viewed a different set of trials out of the 100 which had been constructed to ensure an equal, though random, distribution of trials within the existing constraints.

Procedure

Participants were recruited and voluntarily attended a data collecting session. They were seated at a computer and provided with a set of instructions, a consent form and a 60cm measuring ruler. Participants were allowed to ask questions before proceeding with the experiment. Once they understood the procedure, they followed the instructions to access the appropriate program and continued without needing further intervention from the experimenter unless specifically requested. They were required to provide some personal details in order to subsequently identify their results.

Participants proceeded with the practice trials once they had ensured that they were seated at the appropriate distance. Each participant viewed the same 12 practice trials. After an initial delay of 1000ms, a fixation cross appeared on the screen, directing participants to fixate one of the pre-specified co-ordinates relative to the target object's subsequent location. This fixation cross was displayed for 1000ms before the 120ms scene presentation. Immediately after the scene presentation, the appropriate response screen was displayed and participants were given 5000ms to respond before the next trial began. In practise, participants were mostly responding before this occurred and the button press initiated the next trial after a 1000ms inter-trial interval. The trial sequence was illustrated in Figure 2.1.

At the end of the block of practice trials, participants were again given the opportunity to ask questions or clarify any problems. Then they accessed the specific set of experimental trials assigned to them in their instructions and, after checking that they were seated at the correct distance from the monitor, they completed the 22 experimental trials. The procedure was identical to that of the practice trials but the scenes displayed were entirely different. At the end of the experiment, consent forms were collected from each participant and they were provided with a debriefing sheet which explained the purpose of this study. The entire procedure took less than 15 minutes to complete.

2.4 Results

Results were subjected to a screening process before analysis. Even at chance levels, participants should have performed with at least 50% accuracy. Data relating to any participant failing to score at least 13 out of 22 (60%) were removed and replaced by another participant run on the same set of trials. According to this criterion, 10 participants' data had to be replaced from the original 100.

Distance of the target object from fixation

Figure 2.2 illustrates the decrease in accuracy as the distance between the target object and fixation position increased. From a maximum level of 88% correct at fixation, accuracy decreased to 55% at the furthest distance. Performance was best and significantly above chance level at fixation and gradually decreased to approximate chance level by position 4. This pattern supported the prediction that when participants were fixating a region of the scene closer to the target object, they would respond more accurately than when the target object was further from fixation. It also confirmed that participants were adhering to the instructions and fixating the fixation cross prior to each scene presentation.

To test the prediction that fixation position influenced response accuracy, a binary logistic regression analysis was selected, as the dependent variable was binary and categorical. This analysis investigated whether any number of independent variables (which could also be categorical) or their interactions had a significant effect on the outcome of the binary response variable. The analysis tested the effects of both fixation position and consistency on accuracy and the likelihood ratio test values are reported throughout. There was a significant effect of fixation position on response accuracy ($\chi^2(1)=105.6, p<.001$), with closer fixations to the target more likely to result in accurate responses.

Consistency of the target object

The binary logistic regression analysis indicated no significant main effect of consistency on accuracy ($\chi^2(1)<1, p=.59$). A mean of 69.9% of consistent trials were responded to correctly across all fixation positions with 72.7% correct for inconsistent trials and this 2.8% difference did not reach statistical significance.

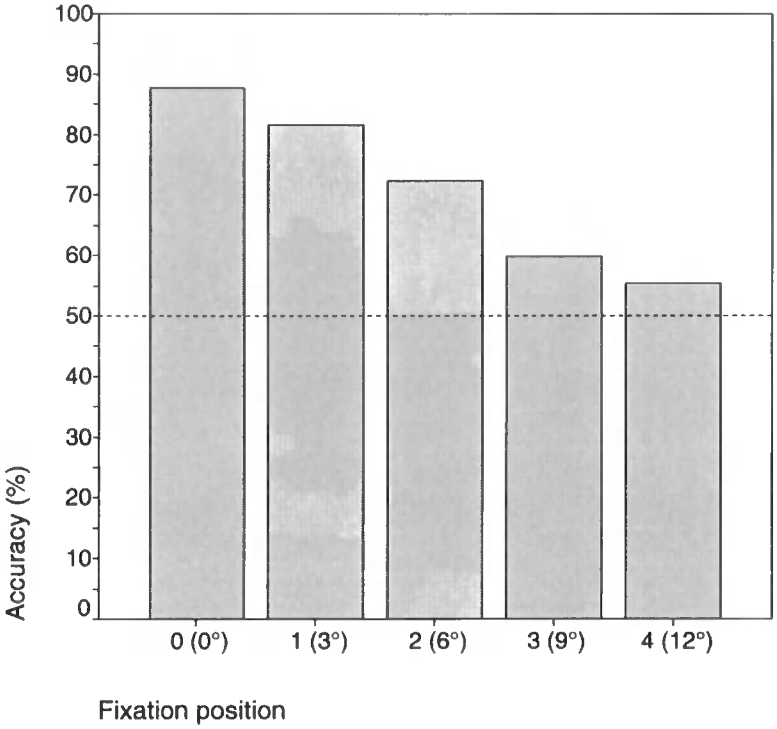


Figure 2.2: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12°. Chance level of 50% is indicated.

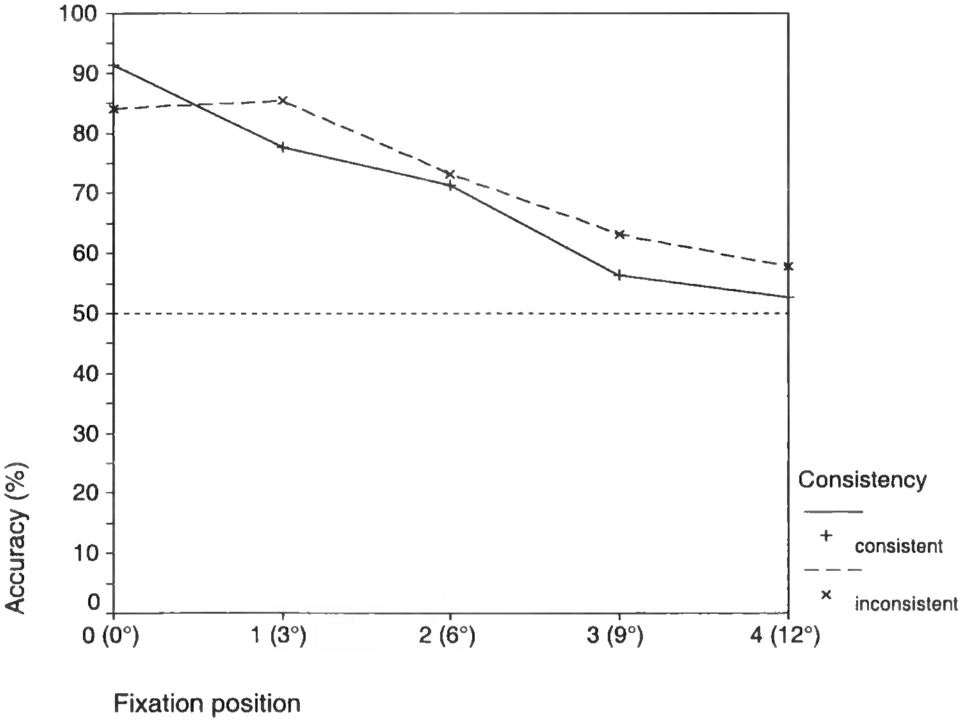


Figure 2.3: Graph showing the change in accuracy by fixation position and target object consistency. Chance level of 50% is indicated.

Interaction of fixation position and consistency

The possible interaction between consistency and fixation position was first investigated by graphically plotting the data. Figure 2.3 illustrates the relationship between accuracy and both fixation position and consistency. Although performance for consistent targets was better than for inconsistent targets when they were presented directly at fixation, inconsistent targets were responded to with more accuracy at all other eccentricities. A significant interaction between fixation position and consistency was confirmed by a binary logistic regression analysis ($\chi^2(1)=4.21, p=.040$).

Participant accuracy was best when the target object was presented at fixation. In these trials, when the fixation position was '0 (0°)', consistent objects were responded to more accurately than inconsistent objects. A binary logistic regression analysis was conducted to determine the effect of consistency on participant accuracy for this fixation position only. The results proved a significant difference between the proportion of correct trials for consistent targets (91%) and inconsistent targets (84%) ($\chi^2(1)=5.48, p=.019$) and indicated that consistent trials were responded to with more accuracy than inconsistent trials.

This pattern was seen to reverse at fixation position 1, when participants were directed to fixate a region approximately 3° from the target object. A binary logistic regression analysis found a significant effect of consistency ($\chi^2(1)=4.40, p=.036$), with higher accuracy for inconsistent targets at this fixation position than for consistent targets (85% and 78% respectively). However, no significant effect was found for data at position 2 ($\chi^2(1)<1, p=.67$) indicating that there was no significant difference between performance on consistent and inconsistent object trials at this fixation position. Similarly, there was no significant difference between performance on consistent and inconsistent trials at fixation position 3 ($\chi^2(1)=2.13, p=.15$) or position 4 ($\chi^2(1)=1.11, p=.29$).

The only significant advantage for consistent objects was evidenced when the participant was directly fixating the target. At all other fixation positions, inconsistent objects produced slightly higher accuracy than consistent objects. To investigate this interaction further, a subsequent analysis was conducted in which data relating to position '0' trials (target object presented at fixation) were removed and a binary logistic regression analysis was conducted on the data relating to positions 1 to 4. As no interaction between the variables was expected at these fixation positions, a main effect of fixation was evidenced ($\chi^2(1)=85.4, p<.001$) and also of consistency ($\chi^2(1)=5.99, p=.014$). This analysis indicated that inconsistent trials were responded to significantly more accurately than consistent trials when the object was not presented at fixation.

Another interesting pattern observed was that performance for consistent object trials appeared to decrease at a faster rate than that for inconsistent object trials. Using a binomial test, it was possible to calculate whether accuracy at each fixation position was significantly different to chance level (50%). Accuracy at position 4, when the target appeared approximately 12° from fixation, was still above chance, with a mean value of 55% correct ($p=.032$). These data were then analysed according to consistency condition, to find that accuracy for consistent objects was not significantly different to chance with a mean of 53% correct ($p=.46$). However, performance for inconsistent targets was significantly above chance, with a mean accuracy of 58% ($p=.026$).

Similarly, at position 3, when the target appeared approximately 9° from fixation, performance for consistent objects was not significantly different to chance at 56% ($p=.069$) although performance for inconsistent targets was, with a mean correct score of 63% ($p<.001$). These results indicated that performance for consistent objects at fixation positions 3 and 4 was not significantly greater than chance but performance for inconsistent objects was still significantly above chance at both positions. However, no significant difference was found between performance for consistent and inconsistent objects at these fixation positions.

Response times

The response times for all correct trials are presented in Table 2.1 and analysed by participants in terms of the independent variables, using a univariate analysis of variance (ANOVA). A significant effect of fixation position was evidenced ($F(4,860)=16.46, p<.001$), with shorter response times when the object was presented closer to fixation than when it was presented further away. Response times increased from a mean of 849ms when the object was presented at fixation to 1296ms when the object was located approximately 12° from fixation. No significant effect of consistency was found ($F(1,860)<1, p=.49$). The mean correct response time for consistent objects was 1035ms compared to 1053ms for inconsistent objects. There was no significant interaction between the two variables ($F(4,860)<1, p=.68$).

Table 2.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only.

	0 (0°)	1 (3°)	2 (6°)	3 (9°)	4 (12°)	Mean
Consistent	830	915	1126	1167	1303	1035
Inconsistent	868	926	1121	1177	1288	1053
Mean	849	921	1124	1172	1296	1044

Object size

Further post-hoc analyses were conducted on performance according to target object size. Object size was calculated as the length and width of the smallest box which would contain the target and was measured in pixels, although these measurements were also converted to degrees of visual angle for inspection. The consistent and inconsistent targets included in each trial were measured independently to consider size differences when the same object was used in more than one scene. The mean object dimensions were 3.55° height and 3.35° width for consistent objects and 3.04° height and 2.96° width for inconsistent objects.

The 44 objects were divided into three groups according to size, discriminating between the smallest objects, the largest objects and those of medium size. Of the 18 'small' objects, defined as a pixel area less than 7,000 (7° square), 8 were consistent and 10 were inconsistent. Of the 16 objects of medium size, with a pixel area between 7,000 and 17,000 (7° to 16° square), 8 were consistent and 8 were inconsistent. Of the 10 objects classed as large, with areas over 17,000 square pixels (16° square), 6 were consistent and 4 were inconsistent. Using these categories, it was possible to investigate whether objects of different sizes modulated differences in results.

Accuracy according to the size of the object was investigated by creating an error plot. Figure 2.4 displays the mean accuracy and 95% confidence intervals around each mean, for the three object sizes. Small objects were responded to with an accuracy of 66%, medium objects had an accuracy of 78% and large objects averaged 71%. Accuracy was worst when the target object was small and higher accuracy was found for medium sized objects, but large objects did not display an increase in performance above this level.

Table 2.2 displays participant accuracy according to object size and consistency condition. Small objects showed no significant difference in performance according to consistency. A slight 2.25% advantage for inconsistent objects was found for medium sized objects and this advantage increased to over 11% for large objects. Accuracy was lower for large consistent objects, compared to large inconsistent objects, contrary to the predictions made earlier. However, the data must be interpreted with caution as the sample sizes were not equal across all groups, so the mean values of the smallest group (large objects) may not be as reliable.

Table 2.2: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	66.08	76.50	66.22	69.91
Inconsistent	65.80	78.75	78.00	72.73
Mean	65.93	77.63	70.94	71.32

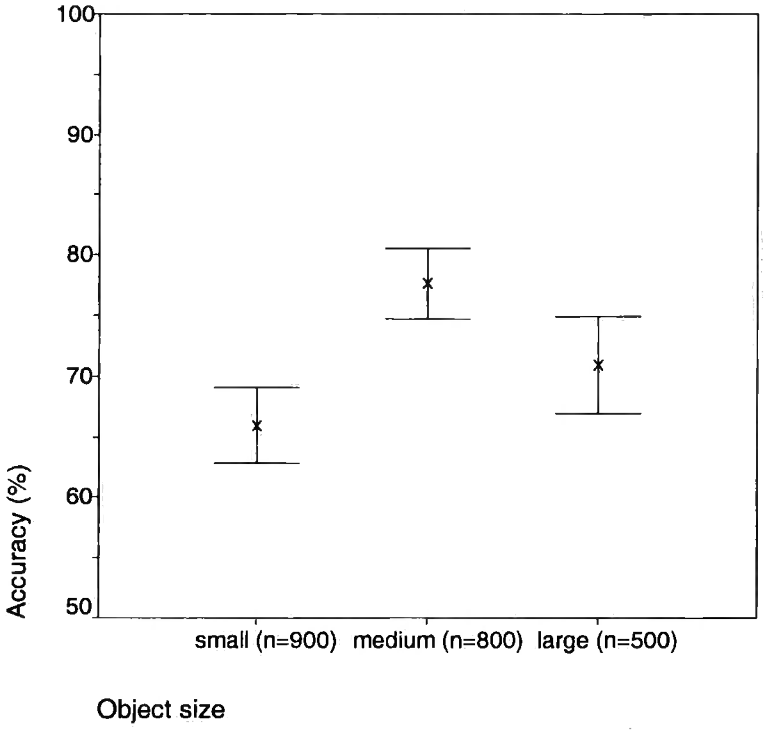


Figure 2.4: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes.

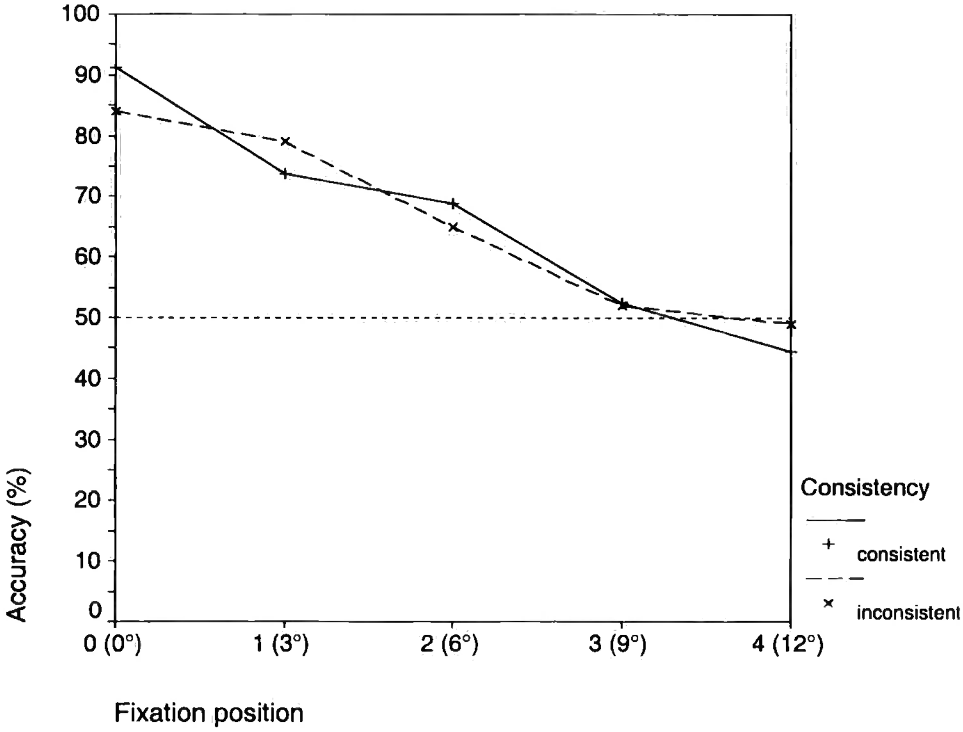


Figure 2.5: Graph showing change in accuracy by fixation position and consistency for small objects only. Chance level of 50% indicated.

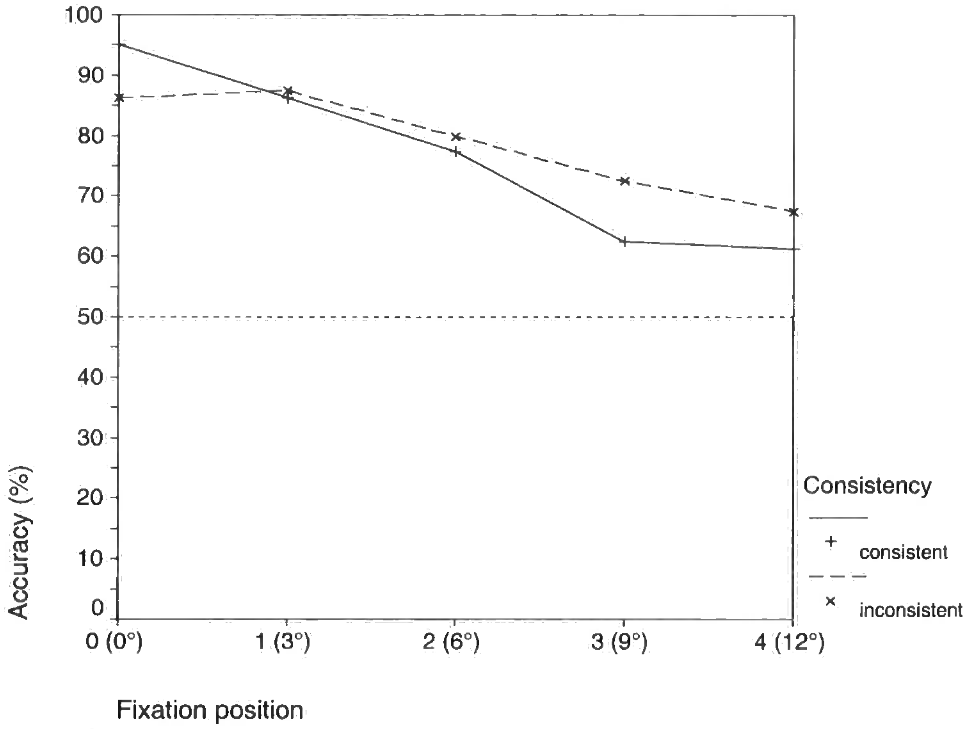


Figure 2.6: Graph showing change in accuracy by fixation position and consistency for medium sized objects only. Chance level of 50% indicated.

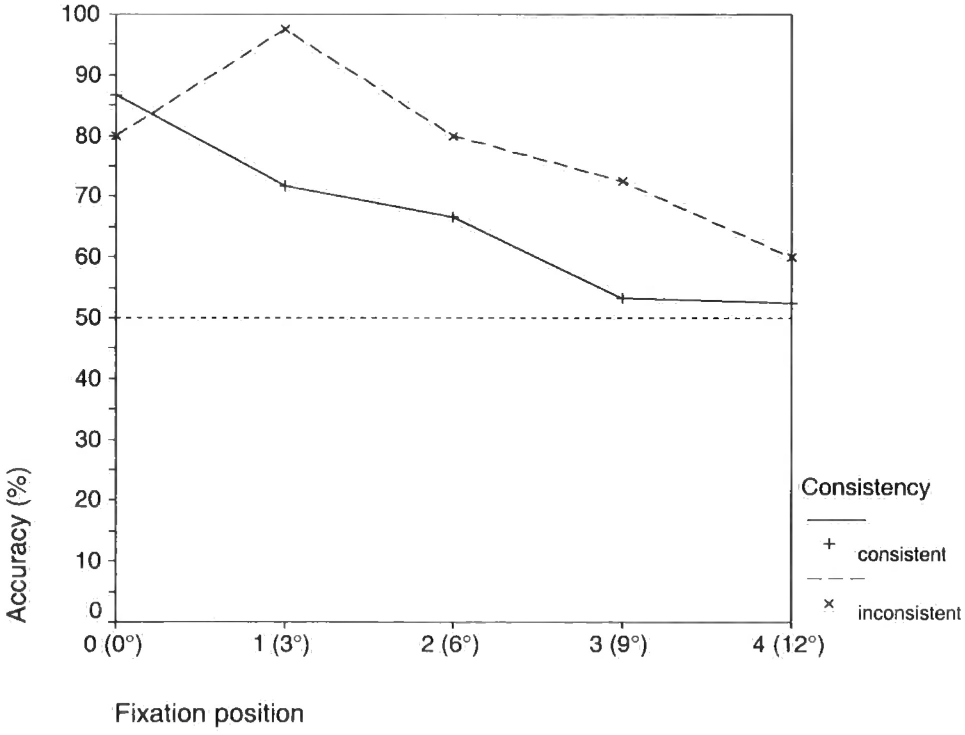


Figure 2.7: Graph showing change in accuracy by fixation position and consistency for large objects only. Chance level of 50% indicated.

A binary logistic regression analysis investigated whether object size, together with the existing independent variables of fixation position and consistency, was a significant predictor of accuracy. There was a main effect of fixation position ($\chi^2(1)=104.6, p<.001$) and also of object size ($\chi^2(2)=32.0, p<.001$) but not of consistency ($\chi^2(1)<1, p=.69$). Additionally, there was a significant interaction between consistency and fixation position ($\chi^2(1)=5.23, p=.022$). Although a consistency effect was found in interaction, together with an effect of object size, the interaction between these two variables was not statistically significant but possibly indicated a trend ($\chi^2(2)=4.51, p=.11$).

Figure 2.5 plots accuracy for small objects according to fixation position and consistency. There were no reliable differences at any fixation position between performance on consistent and inconsistent trials. A non-significant increase in accuracy was found for consistent objects when presented at fixation ($\chi^2(1)=2.16, p=.14$), which was compatible with the findings of the entire data set showing a significant advantage for all consistent objects presented at fixation. Performance for small objects was particularly poor at the furthest eccentricities, as would be expected.

The data relating to medium sized objects is presented in Figure 2.6. An improvement in accuracy was seen at the furthest fixation positions, which may have counteracted poorer performance for small objects. Additionally, there was an advantage for consistent objects presented at fixation which approached statistical significance ($\chi^2(1)=3.74, p=.053$), similar to that evidenced in the entire data set. However, at further eccentricities, performance appeared to be better for inconsistent objects rather than consistent objects, although this indicated a trend towards better performance for inconsistent objects when presented approximately 9° or 12° from fixation, rather than a statistically significant effect ($\chi^2(1)=2.19, p=.14$).

The sample size for large objects was smaller than that for medium or small objects but patterns may still be discerned by displaying the data according to fixation position and consistency in Figure 2.7. These data also evidenced a very slight, non-significant advantage for consistent objects presented at fixation, compared to inconsistent objects. However, performance at all other fixation positions was better for inconsistent objects. The results of a binary logistic regression analysis indicated a significant main effect of fixation position ($\chi^2(1)=30.6$, $p<.001$) and also a significant main effect of consistency ($\chi^2(1)=8.85$, $p=.003$), but no interaction between these two variables ($\chi^2(1)<1$, $p=.96$). Large objects were identified most accurately when they were inconsistent with the scene context, with the largest difference between consistent and inconsistent objects at fixation position 1.

The results of the object size analyses suggested that the significant inconsistent object advantage evidenced for the entire data set at positions 1 to 4 was caused by the medium and large target objects, which produced higher accuracy at further eccentricities for inconsistent objects. No clear difference in performance was seen when target objects were small but larger objects exhibited significant differences in performance between consistent and inconsistent targets. This indicated that the size of the target object, as well as its consistency and its distance from fixation may influence the processing of semantic detail.

High quality image subset

During experimental debriefing, participants indicated that the images presented were not always recognisable (see Appendix A for further details), suggesting that the images might not be suitable for use in this experiment. As a post hoc analysis, the possible influence of the difficulty in identifying the inconsistent targets was assessed. For semantic inconsistency to be determined, both the scene and the target would need to be identified. Some participants indicated that the objects presented in the two-alternative forced-choice were sometimes unrecognisable, suggesting that the inconsistency between the target and the scene could not be determined in these trials. If targets presented within consistent scenes were facilitated in identification, then inconsistent targets could have been distinguishable from consistent targets in extrafoveal vision, not due to their semantic consistency, but due to an inability to identify them. To investigate this further, an analysis of the recognisabilty of the images was conducted by asking participants to name targets and scene backgrounds presented in a questionnaire and also rate them on the probability of finding the target in the scene. This investigation is described in Appendix A and the data collected in this study enabled the selection of a subset of the stimuli which were most reliably identified by participants. From the 44 images, 10 consistent scenes and 10 inconsistent scenes were selected for further analysis. These images were chosen because both the scene background and the target object were identified correctly by most participants and appropriately rated, on a scale of 1 to 5, as consistent or inconsistent scenes.

Of the 10 consistent scenes, six were recognised appropriately by all participants and four were identified by all but one participant. The mean rating given to these scenes was 4.83 (*SD*=.15), where a rating of 5 indicated a target which was very likely to be located in the scene. Two of the targets were categorised as small, five were medium sized and three were large. Only four of the inconsistent scenes were suitably identified by all participants, one scene was identified by all but one participant and five by all but two. The mean consistency rating was 1.16 (*SD*=.21), where a rating of 1 was assigned to targets very unlikely to be found in the scene. Four of the inconsistent targets were small, four were medium sized and two were large.

If the differential performance evidenced between consistent and inconsistent trials when analysing the entire data set was caused by the perception of semantic inconsistency, then the effect should remain and possibly be enhanced when analysing a high quality subset of the data containing the most reliably recognisable consistent and inconsistent trials. However, if the advantage found for inconsistent trials in the analysis of the entire data set cannot be replicated when investigating these higher quality scenes, it would suggest that the detection of semantic inconsistency did not modulate the significant effect found.

The data from the high quality image subset is presented in Figure 2.8. A binary logistic regression analysis found a significant effect of fixation position on accuracy ($\chi^2(1)=99.9, p<.001$) but no effect of consistency ($\chi^2(1)=2.42, p=.12$) and no interaction between the two variables ($\chi^2(1)<1, p=.41$). Unlike the results of the entire data set, which showed a significant interaction between consistency and fixation position (*p*=.040), no similar effect was found when analysing the scene subset.

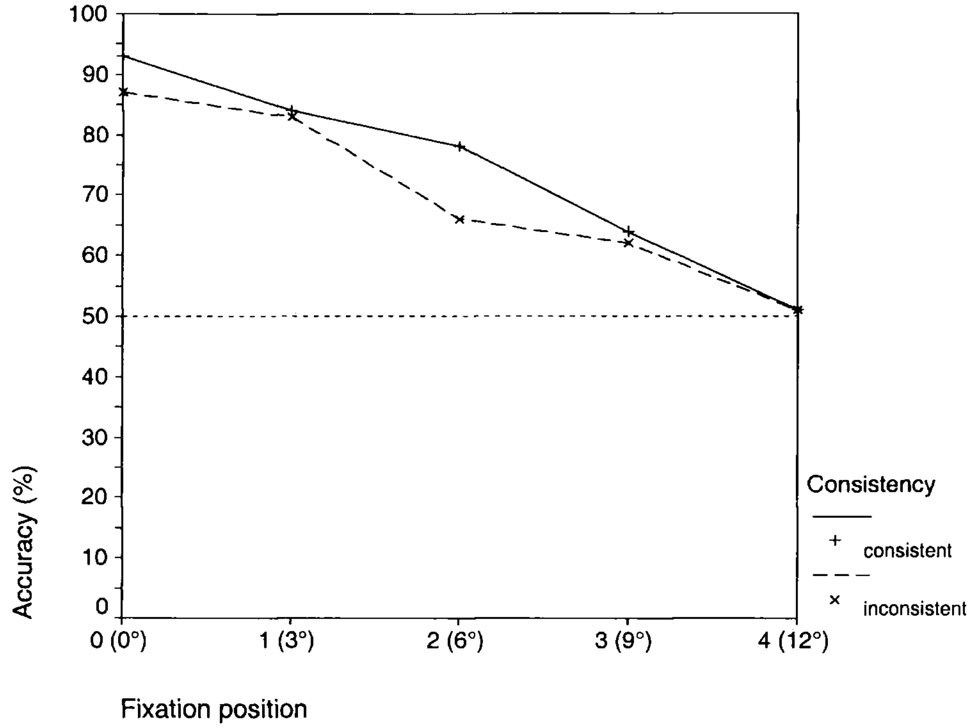


Figure 2.8: Graph showing the change in accuracy by fixation position and target object consistency for the high quality images. Chance level of 50% is indicated.

Accuracy was approximately equal for consistent and inconsistent scenes at all fixation positions except positions 0 and 2. At these fixation positions, accuracy was higher for consistent trials than inconsistent trials, which was contrary to the prediction made. The 6% difference in accuracy when the target was presented directly at fixation was not statistically significant ($\chi^2(1)=2.03, p=.15$) and the 12% difference at position 2 just failed to reach statistical significance ($\chi^2(1)=3.59, p=.058$).

The absence of any evidence indicating an inconsistent object advantage when analysing the data from the most reliably recognised scene-object pairs did not support the conclusion that semantic inconsistency caused the differential performance. Instead, the data suggested that the consistency effect could have been caused by the difficulty in identifying objects without a consistent scene presentation. In this way, the facilitation appeared to be evidenced for targets which were difficult to identify, rather than those which were readily identifiable and reliably considered to be inconsistent with the scene context.

2.5 Discussion

As expected, fixation position was found to influence performance, with significantly higher accuracy when participants were directed to the correct region in which the target object would appear (p<.001). Response times were also significantly influenced by fixation position, exhibiting faster responses for targets presented closer to fixation (p<.001). No overall effect of consistency was evidenced on either accuracy or response time, but a significant interaction was found between the two variables for accuracy only (p=.040).

When the target appeared at fixation, there was a significant advantage in performance accuracy for consistent objects (p=.019) but when fixations were directed at position 1, approximately 3° from the target object, there was a significant advantage for inconsistent objects (p=.036). Although accuracy for inconsistent targets was higher than for consistent targets at further fixation positions, the differences were not statistically significant. An analysis of the data from fixation positions 1 to 4, removing trials in which the target object appeared at fixation, found a significant advantage for inconsistent targets (p=.014).

A consistent object advantage was found when the target object was presented at the participant's fixation position. This result indicated that when directly fixating the target object, performance was better when the target was consistent with the scene context. The object's identity may have been primed by the activation of the scene schema during the image presentation, which would be consistent with the schema hypothesis of object processing in scenes. The prior identification of the scene could have facilitated the identification of the correct consistent target object, compared to selecting the correct target when the scene was inconsistent.

This consistent object advantage was not found at other target positions. When the target object was presented in parafoveal vision, the advantage was reversed and consistent objects were responded to less accurately than inconsistent objects. This did not support the schema hypothesis or the intra-level priming approach which suggested object-to-object processing in scenes. Existing theoretical explanations for this inconsistent object advantage would need to account for such an effect during brief presentations. The attentional attraction hypothesis predicted that covert attention would be drawn to inconsistent objects faster than consistent objects but whether this can occur during 120ms is still unclear.

Additionally, the pattern of consistent and inconsistent object facilitation did not support the hypothesis suggested previously, that inconsistent objects would be easier to identify at regions closest to fixation and more difficult at further eccentricities. The predicted consistent object facilitation at greater extrafoveal distances did not materialise, instead exhibiting a consistent object advantage at fixation and an inconsistent object advantage most pronounced at fixation position 1, approximately 3° from fixation, but still evident at further eccentricities. This pattern suggested that some difference between consistent and inconsistent objects could be detected extrafoveally, but was either not occurring during foveal fixation or was overridden by a consistent scene facilitation, possibly through priming, when the target was presented at fixation. This exhibition of both a consistent and an inconsistent object advantage at different fixation positions also justified the manipulation of fixation position relative to the target's location and may explain how both effects had been found in previous research. These results indicated that future work should consider the effect of target eccentricity when investigating consistency effects.

From post-hoc analyses of object size, the inconsistent object advantage at extrafoveal locations appeared to be caused by medium and large sized targets in this study. Small objects showed no consistency effects, possibly because at further fixation positions, the targets were not large enough to enable the processing of their semantic identity. However, with medium and large sized objects, an advantage was found for inconsistent objects presented at further fixation positions, coupled with a consistent object advantage when the target was presented at fixation.

These data need to be evaluated cautiously because this manipulation was not designed prior to conducting the experiment and, therefore, object size was not varied systematically. Target sizes were calculated subsequently and assigned to groups of unequal size for further data analysis. For clearer conclusions, target size would need to be included as an additional independent variable and objects should be selected which can be clearly assigned to different groups. As object size was only included as a post-hoc analysis in this case, target area was a continuous variable, with no pre-determined boundaries between groups. However, the discovery that object size may influence whether consistency effects were exhibited or not indicated that this variable must also be considered when investigating consistency effects in scenes and provides an additional explanation for divergent results in previous research.

The inconsistent object advantage evidenced at extrafoveal positions appeared to be compatible with Loftus and Mackworth's suggestion that inconsistent objects could be distinguished from consistent objects in extrafoveal vision and supported Hollingworth and Henderson's findings of an inconsistent object advantage (1998, 1999, 2000). However, when inconsistent objects were presented at fixation, this advantage was not seen. This could have been because the effect does not occur at fixation and is specific to the extrafoveal processing of inconsistent objects. Alternatively, the consistent object advantage may have been even greater than the facilitation for inconsistent objects and sufficient to negate any advantage for inconsistent objects. The results suggested that the processing of consistent and inconsistent objects in scenes may not be equivalent at different retinal locations as previously assumed.

However, the assumption that the consistency effect was caused by the detection of semantic inconsistency was questioned by the analysis of data relating to the high quality image subset. The selected scenes were the most reliably identified images and therefore prime candidates to express semantic consistency effects. Despite their improved quality, the analysis found no evidence of differential performance for consistent and inconsistent trials within this group. This result indicated that the scene differences influencing the inconsistent object advantage were unlikely to have been caused by the detection of semantic inconsistency.

Alternative explanations could be proposed to explain the discrepancy in performance for consistent and inconsistent objects, without resorting to semantic differences. The improvement in accuracy for inconsistent objects could have resulted from an unintentional bias in the two-alternative forced-choice response procedure. The task of selecting the target object would be easiest when the two objects were large, because a larger target should have been easier to detect from a brief presentation. If inconsistent objects were unintentionally larger than consistent objects, better accuracy for inconsistent objects at all eccentricities would be predicted. As object size could not be counterbalanced to include each object as both a consistent and an inconsistent target, further investigation into possible differences in object sizes was required.

The height and width of each object was measured in pixels and their areas were calculated. These measurements were analysed using an independent samples t test to compare sizes for consistent and inconsistent targets. No significant difference was found between the groups, indicating that consistent and inconsistent objects did not vary in size ($t(42)<1$, p=.42). The mean height of a consistent object, at a viewing distance of 60cm, was 3.55°, with a width of 3.35°, while the mean height and width of an inconsistent object were 3.04° and 2.96°. Inconsistent targets were found to be slightly, but not significantly, smaller than consistent objects. Contrary to the above proposal, consistent targets were slightly larger than the inconsistent targets, so inconsistent targets should not have been easier to identify in the brief presentations of scenes on the basis of their size. Therefore, this suggestion cannot explain the inconsistent object advantage found.

Alternatively, the task would have been affected by the two-alternative forced-choice displays. If the two objects presented were of obviously different sizes, accuracy could be enhanced, resulting in an advantage for either consistent or inconsistent trials. Assuming that the participant was not aware of having viewed the target in any of the following scenarios, a response bias could result according to the relative sizes of the target and the distractor in the two-alternative forced-choice display.

If both alternatives were small, the participant would be equally likely to select the correct target as the incorrect distractor and there would be no response bias. Conversely, if two large objects were presented in the two-alternative forced-choice display and the participant did not know which had been present in the scene, the probability of selecting the target would still be equal to the probability of selecting the distractor. However, if the two alternatives had been obviously different in size, the participant could make an educated guess about which was the target, based on the probability of failing to detect each object in the scene presentation. The most likely target would be the smaller of the two objects, if none had been perceived, as the participant would have been more likely detect the larger object if it had been present. In this way, the similarity in size between the two objects in the two-alternative forced-choice display could influence accuracy and provide an explanation for the inconsistent object advantage in extrafoveal vision, without assuming the processing of semantic consistency.

The relative sizes of the object pairs presented in the two-alternative forced-choice displays were investigated. The height and width (in pixels) of each object were multiplied to calculate the object's area and the difference in area between each pair was calculated individually for consistent and inconsistent objects. Although there were only 11 object pairs in each group, an independent samples t test analysis was conducted to investigate whether any significant differences in object size disparity existed between consistent and inconsistent object pairs.

The mean difference in size for consistent objects was 12,856 square pixels (approx 14.47° squared) compared to a mean of only 7,817 square pixels (approx 7.49° squared) for inconsistent objects, but these size differences were not statistically different from each other (*t*(20)=1.43, p=.17). Contrary to the above proposal, inconsistent object pairs in the two-alternative forced-choice displays were better matched for size than consistent object pairs, which should have facilitated consistent object trials rather than inconsistent object trials. Therefore the inconsistent object advantage evidenced in extrafoveal trials could not be explained by the size of the object pairs in the selection screen.

The inconsistent object advantage could alternatively be explained by visual differences introduced during the creation of the experimental scenes. The images used were from the Leuven line drawing library and consisted of background images and individual object images which could be located within the scenes. Consistent object images fit perfectly into consistent scene backgrounds, including all relevant background contours. Inconsistent objects were the same object images placed in any other inconsistent background.

As these objects were consistent targets in different scenes, they were often adjusted before insertion into an inconsistent background. For example, for an inconsistent object to be placed on a table top, the background detail suited to the consistent background would need to be removed from the object image and the object would need to fit into and occlude any inconsistent background contour behind it (e.g. part of the rear edge of the table). In practice, inappropriate background details were removed from most object images and relevant background contour for the inconsistent scene needed to be included. Therefore there may have been some low level visual differences inadvertently introduced between consistent and inconsistent objects in scenes, irrespective of the scene semantics, which may have resulted in differential performance for these trials.

In order to investigate this possibility, a further experiment was conducted in which the experimental scene images were inverted. This was expected to interfere with the processing of both local and global semantics but not affect the processing of visual features. The object identification task would not require participants to interpret the semantics of the image, only match the visual features of a target with the two presented (also inverted) objects in the two-alternative forced-choice display. As inverting the scene images should not interfere with the processing of visual features, if visual differences caused the inconsistent object advantage in Experiment 1, the facilitation would still be evident. However, if semantic differences between consistent and inconsistent scenes modulated the consistent and inconsistent object advantages, the effects should not be present. This proposal was investigated in the following experiment.

However, Experiment 1 could also be subject to further methodological criticism. One concern raised by the debriefing of participants was the difficulty in identifying the objects in the two-alternative forced-choice display. Several participants claimed that they had responded randomly in many trials because they could not confidently identify the objects. If the objects were difficult to identify from the response display, 120ms scene presentations could have been insufficient for semantic processing. Any suggestion that the inconsistency between object and scene could influence performance would require participants to be able to identify both the scenes and the individual objects. If the scenes were not identifiable, they would not activate a scene schema specific enough to generate facilitation for predicted objects. Again, consistency effects could not be attributed to semantic inconsistency between an object and a scene background if the scene gist could not be rapidly and accurately identified.

This concern was investigated by obtaining ratings on the consistency of the scene-object images, which required participants to identify both the scene and the target located within it. This analysis can be found in Appendix A, together with an evaluation of the Leuven stimuli as experimental images. Many of the target objects and scene backgrounds obtained from the Leuven line drawing library were difficult to identify. In some cases, few diagnostic objects were present to aid scene identification, with the target object being the only movable object present. Although the scenes could all be broadly identified as indoor or outdoor scenes, it may have been difficult to generate any predictions about likely object components in specific terms. The lack of companion objects in some circumstances could also make the target object particularly salient.

Therefore, it was possible that the nature of the scenes themselves had influenced the processing of individual objects in them. Certain criticisms relating to the complexity and realism of both the objects and the scene backgrounds may be levelled at the conclusions of this experiment. However, it was decided to begin by investigating whether the inconsistent object advantage evidenced in Experiment 1 could be attributed to the introduction of visual differences between consistent and inconsistent scenes. This was achieved by inverting the images used in order to inhibit the processing of semantic information.

2.6 Experiment 2: Introduction

The purpose of this experiment was to determine whether the significant facilitation found for inconsistent objects in Experiment 1 was due to visual or semantic differences between consistent and inconsistent targets in scenes. The consistent scene line drawings had been created by selecting the relevant consistent target, which had been designed to fit into the scene, and locating it in the correct area. However, to create scenes with inconsistent targets, objects were placed into a scene they were not designed to be located in. In this way, visual differences may have been introduced between an inconsistent object and the scene in which it was located, which would not be the case for consistent scenes. It was important to determine whether this scene creation technique resulted in the facilitation evidenced for inconsistent objects or whether the facilitation was driven by the processing of semantic information.

In order to test the hypothesis that the inconsistent object advantage was caused by visual discrepancies which were evident in inconsistent scenes but not in consistent scenes, the same images were inverted and used in the same experimental procedure, in an attempt to inhibit the processing of semantic features. This should maintain any visual discrepancies between inconsistent targets and their scenes but disrupt the processing of semantic information. If an inconsistent object advantage similar to that evidenced in Experiment 1 were to be exhibited, then the consistency effect could be attributed to the discrepancy in visual features. However, if no such facilitation were observed, the disruption of semantic processing could have abolished the effect, suggesting that the facilitation evidenced in Experiment 1 was due at least in part to the semantic processing of inconsistency between the scene-object pair.

2.7 Method

Participants

Participants were recruited from the undergraduate population of the University of Durham to participate in this experiment and in Experiment 4. There were 15 males and 85 females participating in the study and all were naïve to the purposes of the experiment.

Apparatus

The same apparatus were used as in Experiment 1. Participants were again provided with instructions, a consent form and a 60cm ruler for measuring viewing distance. A debriefing sheet was provided at the end of the experiments.

Materials

The images used were in every way identical to those in Experiment 1 except for their orientation. They were inverted, flipped horizontally, and presented upside down. An object located in the bottom left hand corner of the screen in Experiment 1 would now be found, upside down, in the top left hand corner of the screen in Experiment 2.

Other necessary alterations included modifying the fixation positions prior to each scene display, to account for the new location of all embedded targets. The fixation crosses still appeared at the same distances from the target object as in the previous experiment, but the actual coordinates were recalculated to correspond to the same positions in the inverted images. The response screens were also inverted so the task would still involve the direct matching of visual features present in an (inverted) object in the two-alternative forced-choice display, to the features present in the scene presentation, without requiring the processing of semantic information.

Design

As in Experiment 1, the independent variables were the proximity of the fixation position to the target object, which had five levels, and the consistency of the relationship between the scene and the object, which had two levels, consistent or inconsistent. The dependent variables were primarily response accuracy and to a lesser extent, response time. As in Experiment 1, each participant viewed 22 trials in an order which was randomised separately for each.

Procedure

Participants were recruited from undergraduate lectures. They were seated at a computer and provided with an information sheet, a consent form and a 60cm measuring ruler. They were instructed to read the information provided and complete the consent form if they agreed to participate.

The same 12 practice trials were used in this experiment as in Experiment 1, but the images were inverted. All participants viewed the same practice trials. A fixation cross was presented on the screen which participants were instructed to fixate. This presentation was visible for 1000ms, after which the inverted scene image was presented for 120ms. Subsequently, a final screen illustrating two inverted objects, either both consistent or both inconsistent with the scene, was presented and remained visible until a response was recorded or 5000ms had elapsed. There was an inter-trial interval of 1000ms before the next trial began.

At the end of the practice trials, participants continued with the experimental block of 22 trials. The experimenter was available to answer questions at this point. After this first block of experimental trials, participants were reminded to participate in the second set of practice and experimental trials (for Experiment 4) or were allowed to collect a debriefing sheet if they had already done so. The images used in each experiment were completely different and debriefing did not occur until both had been completed. The procedure for both experiments lasted approximately 20 minutes.

2.8 Results

Due to participant shortage, the first 100 data sets completed were used in the analysis and participants with low accuracy rates were not replaced. According to the previous criterion (60% accuracy), 12 data sets would have been replaced for failing to achieve at least 13 correct trials out of 22, which was comparable to the 10 data sets replaced in Experiment 1. The data were analysed in the same way as in Experiment 1.

Distance of the target object from fixation

Figure 2.9 displays performance according to the distance between the participants' fixation position and the target object's location. When the target object was presented directly at fixation, accuracy was highest at 86% and dropped to chance level at 51% at the furthest distance investigated. Performance was highest and significantly above chance when the target object was presented at fixation, again supporting the hypothesis that participants were fixating the cross and were better able to identify objects which appeared closest to their fixation position.

The decrease evident in the data indicated a main effect of fixation position ($\chi^2(1)=167.4$, $p<.001$) when analysed using a likelihood ratio test in a binary logistic regression analysis. The greatest decrease in accuracy occurred between positions 1 and 2. These locations represented a retinal distance of approximately 3° to 6° which may reflect a change from near foveal to extrafoveal processing.

Consistency of the target object

There was no significant main effect of consistency on participant accuracy ($\chi^2(1)=1.40$, p=.24). Consistent trials had a mean accuracy of 68.8% across all fixation positions while inconsistent trials had a mean accuracy of 66.5%.

Interaction of fixation position and consistency

Figure 2.10 demonstrates accuracy for consistent and inconsistent trials at each fixation position. Inconsistent targets were responded to more accurately at fixation position 1 only, approximately 3° from fixation. Consistent targets showed slightly higher accuracy than inconsistent targets at all other fixation positions. There was no interaction between the two manipulations ($\chi^2(1)=1.14$, $p=.29$).

As expected, accuracy was highest when the target was closest to fixation. When the target was presented directly on fixation, there was also a slight advantage towards improved accuracy for consistent objects compared to inconsistent objects. However, no significant difference was found using a binary logistic regression analysis ($\chi^2(1)=1.51, p=.22$).

Unlike Experiment 1, there was little evidence of an advantage for inconsistent objects at extrafoveal eccentricities. There was a slight though non-significant advantage at fixation position 1 when the target was presented approximately 3° away from fixation ($\chi^2(1)=1.74, p=.19$). However, at all other fixation positions, accuracy was marginally higher for consistent objects than for inconsistent objects although not significantly so. Therefore, there was no evidence for an inconsistent object advantage as found in Experiment 1.

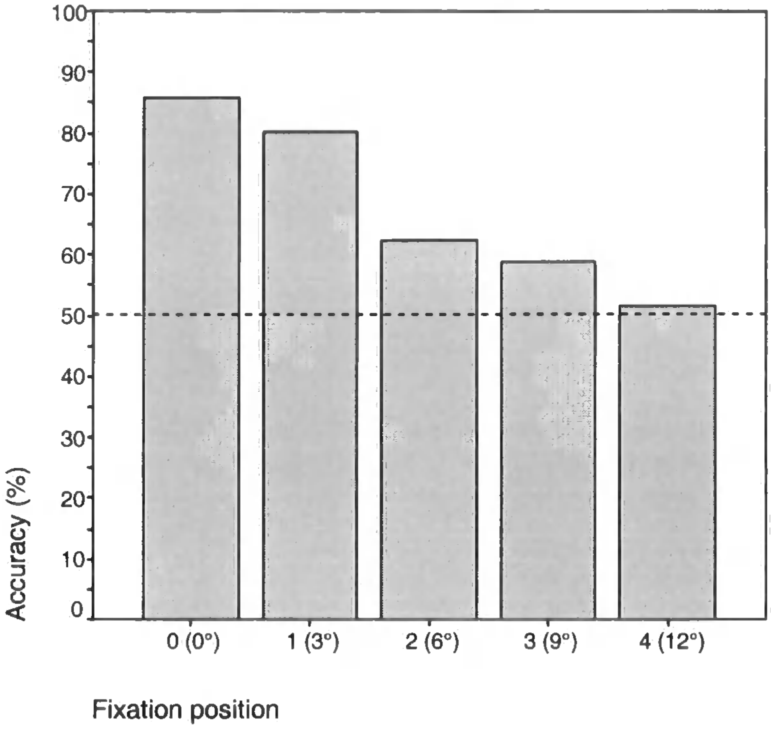


Figure 2.9: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12°. Chance level of 50% is indicated.

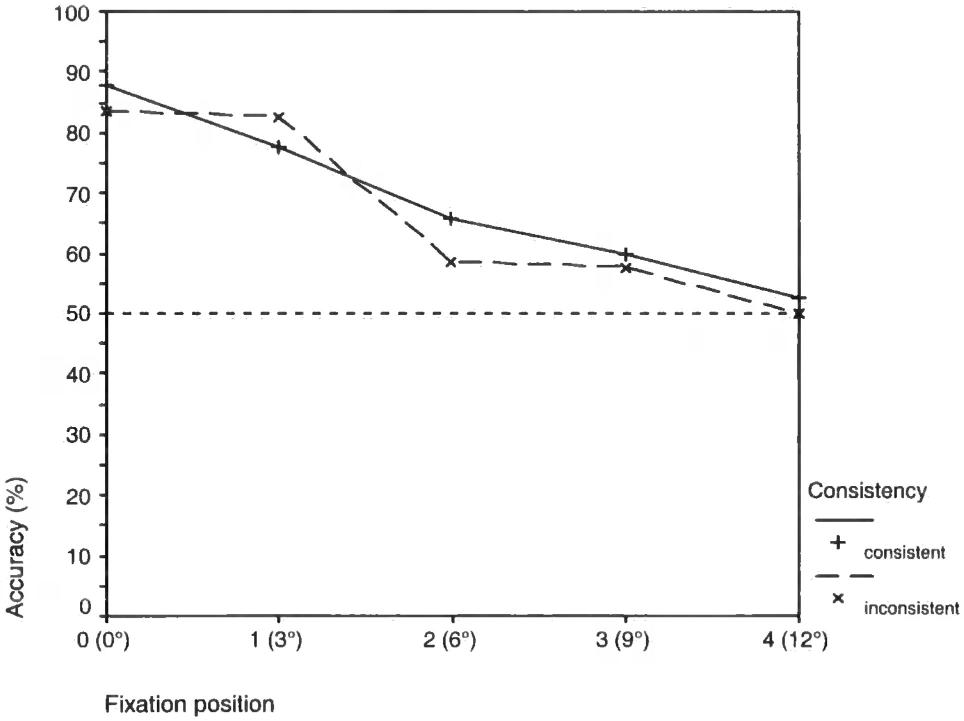


Figure 2.10: Graph showing the change in accuracy by fixation position and target object consistency. Chance level of 50% is indicated.

The data also indicated slightly poorer levels of accuracy than in Experiment 1, which may have reflected the inability to replace participants performing below chance level. Accuracy for both consistency conditions decreased rapidly, particularly for inconsistent targets at fixation position 2. A binomial test indicated that performance for both consistent targets (*p*=.46) and inconsistent targets (*p*=1.00) was not significantly higher than chance at the furthest fixation position. At fixation position 3 (9°), accuracy for consistent targets was significantly above chance at 60% (*p*=.004) and also for inconsistent objects at 58% (*p*=.026). This indicated that performance was only at chance level for both consistency conditions at the furthest fixation position (12°). In Experiment 1, performance for inconsistent objects remained above chance at this 12° fixation position.

Comparing upright and inverted images

A comparison of the two sets of experimental data was made. Figure 2.11 shows the decrease in accuracy as fixation position increases both when the images were presented upright (blue) and inverted (green). The solid lines represent accuracy for consistent trials and the dashed lines represent accuracy for inconsistent trials.

Performance for consistent objects (solid lines) was slightly better than for inconsistent objects when the target was presented at fixation in both experiments. At further fixation positions, the inversion of the images appeared to decrease accuracy selectively for inconsistent trials. Performance for consistent trials in both experiments (solid lines) showed a comparable decrease in accuracy across fixation positions regardless of the orientation of the images. However, performance for inconsistent objects (dashed lines) showed a considerable decrease, from the upright images (blue) to the inverted images (green), at distances greater than approximately 3°.

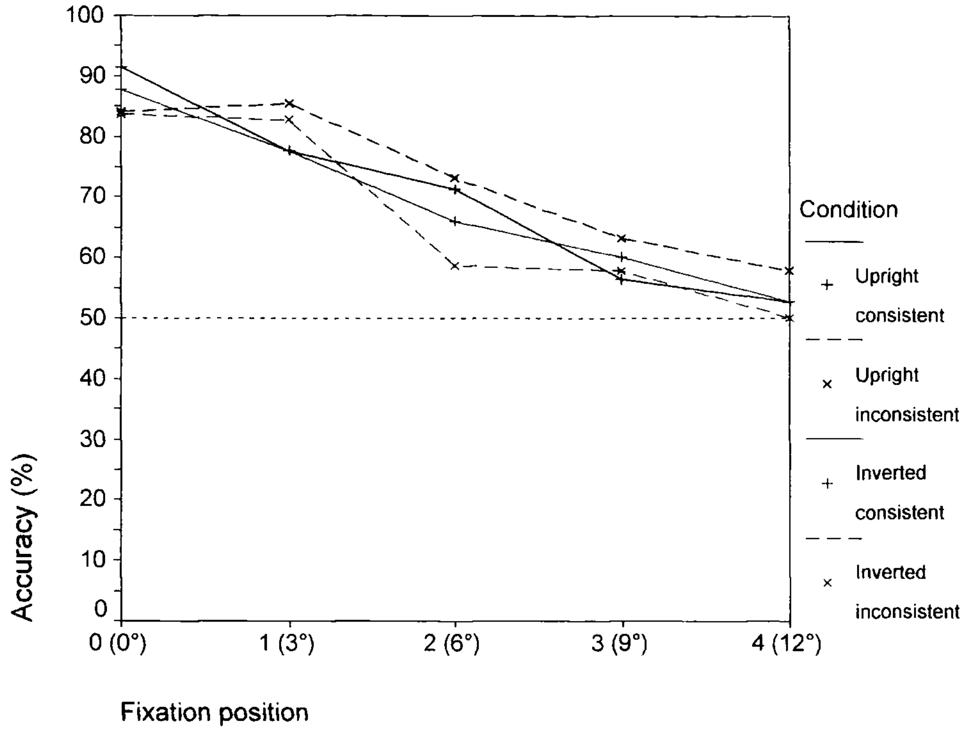


Figure 2.11: Graph showing the change in accuracy by fixation position and target object consistency for upright and inverted line drawings. Chance level of 50% is indicated.

At position 1, the difference was 2.8%, followed by a maximum of 14.6% at position 2, 5.5% at position 3 and 8.8% at position 4. The mean difference in accuracy for consistent trials between the two experiments was only 2.3% with no reliable direction. However, the mean difference for inconsistent trials was 7.9% with inverted images always being responded to less accurately than upright images. This evidence suggested that inverting the scene images affected performance for inconsistent trials to a greater extent than performance for consistent trials.

Response times

The response times for correct trials are presented in Table 2.3 and were analysed using a univariate ANOVA by participants. The results indicated that there was a significant effect of fixation position ($F(4,816)=21.1, p<.001$), replicating the effect evidenced in Experiment 1. Response times increased from a mean of 905ms for objects presented at fixation to a mean of 1295ms for objects presented at fixation position 4 approximately 12° from fixation. No significant effect of consistency was found ($F(1,816)<1, p=.41$) as the mean response time for consistent trials across all fixation positions was 1084ms compared to 1034ms for inconsistent trials. There was no interaction between the two variables ($F(4,816)=1.60, p=.17$). These data were comparable to those evidenced in Experiment 1.

Table 2.3: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only.

	0 (0°)	1 (3°)	2 (6°)	3 (9°)	4 (12°)	Mean
Consistent	922	1039	1128	1185	1251	1084
Inconsistent	887	887	1057	1170	1341	1034
Mean	905	961	1095	1178	1295	1060

Object size

Additional post hoc analyses were conducted on object size. The same criteria were used to measure object size as in Experiment 1 and as the images were identical, simply inverted, no alterations were made to the analysis. Figure 2.12 displays accuracy according to object size. As in Experiment 1, lowest accuracy was found for small objects as expected, with medium sized objects being responded to significantly better but no further advantage for large objects.

Table 2.4 presents mean accuracy by consistency and object size. There was a slight difference in accuracy for consistent and inconsistent small objects, although this was not statistically significant ($\chi^2(1)=1.50, p=.22$). There was no difference in accuracy for medium or large objects according to consistency.

Table 2.4: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	64.75	72.75	69.00	68.82
Inconsistent	61.00	72.50	68.50	66.55
Mean	62.67	72.63	68.80	67.68

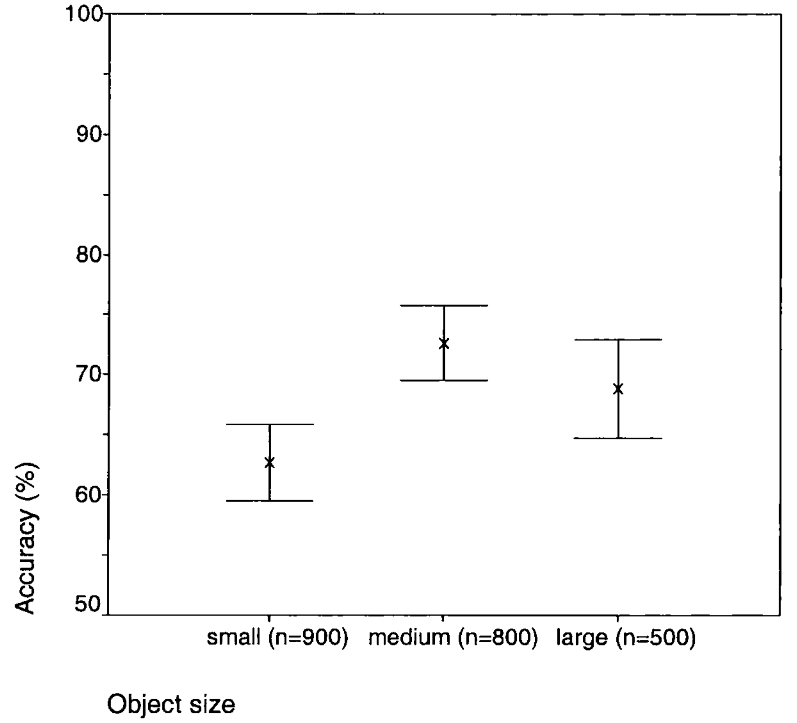


Figure 2.12: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes.

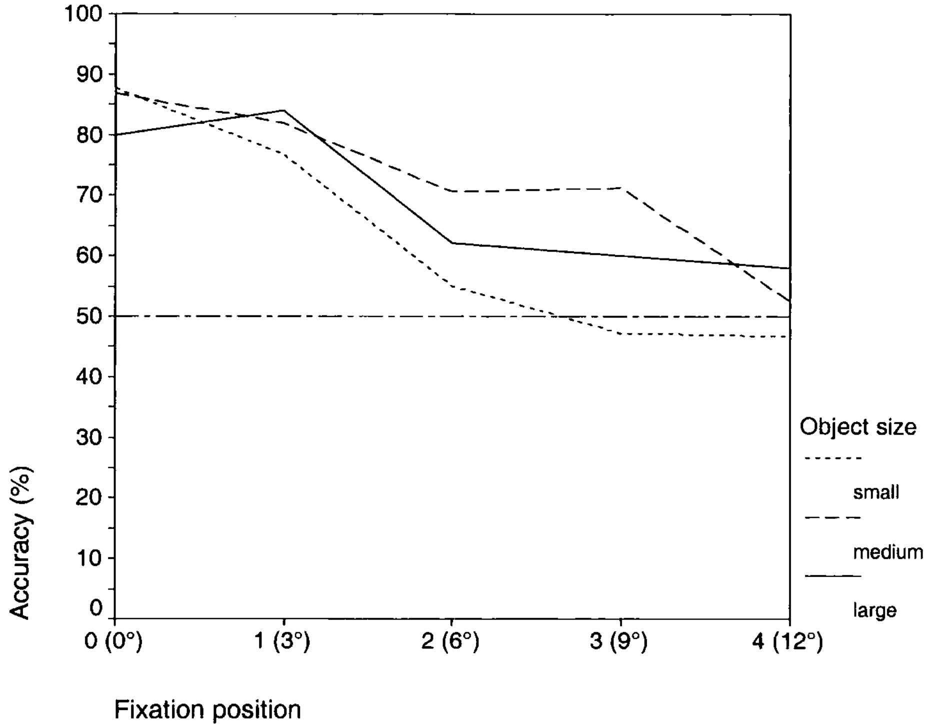


Figure 2.13: Graph showing change in accuracy by fixation position and object size. Chance level of 50% indicated.

A binary logistic regression analysis, including object size as an independent variable, found that fixation position had a significant main effect ($\chi^2(1)=63.8$, $p<.001$) but neither consistency ($\chi^2(1)<1$, $p=.38$) nor object size ($\chi^2(2)=3.06$, $p=.22$) were significant predictors of accuracy. A significant interaction was found between fixation position and object size, as found in Experiment 1 ($\chi^2(2)=22.5$, $p<.001$). Investigating results by object size also failed to exhibit any systematic variation by object size or consistency.

Figure 2.13 displays accuracy by fixation position and object size. For small objects, performance was less accurate overall, with a steady decrease in accuracy to the approximate chance level of 55% at fixation position 2 (6°) ($p=.21$) and beyond. Accuracy for medium sized objects was the highest overall. Accuracy remained high until the furthest fixation position, decreasing to chance levels only at position 4 (12°). At position 3 (9°), accuracy was still significantly above chance at 71.25%. Large objects were responded to slightly less accurately than small and medium objects when presented at fixation, with a mean accuracy of 80% compared to approximately 87%. However, a binary logistic regression analysis on data from targets presented at fixation only failed to find a significant effect of object size ($\chi^2(2)=3.25$, $p=.20$), suggesting that the difference was not significant. Performance for large objects did not decrease to chance level even at the furthest fixation position (12°), remaining at 58% ($p<.001$).

The results from the object size analyses indicated that small objects were responded to least accurately and only above chance at positions closest to fixation. Medium objects appeared to be the most accurately responded to, although large objects also displayed high accuracy at fixation positions furthest from the target. Unlike Experiment 1, no significant effects of consistency were found, indicating that the inversion of the experimental images did not result in the same effects found previously.

2.9 Discussion

The results obtained by inverting the line drawings evidenced a significant main effect of fixation position (*p*<.001), indicating that target objects which were presented directly at fixation were more likely to be responded to correctly than objects presented further away. There was a visible decrease in performance as the distance between the target object and the participants' fixation position increased. However, the data showed no significant difference between performance on consistent and inconsistent objects, either as a main effect or as an interaction with another independent variable.

The aim of this experiment was to determine whether the processing of semantic information was involved in the discrepancy in performance evidenced between consistent and inconsistent trials in Experiment 1. Accuracy was seen to be highest when the target object was presented at fixation but no significant difference between performance for consistent and inconsistent trials was found at this fixation position (*p*=.22). This result indicated that the facilitation evidenced for consistent objects in Experiment 1 may have been the result of more detailed processing of the upright scenes and objects and could not occur when the images were inverted, possibly due to interference with semantic processing.

The inconsistent object advantage in Experiment 1 was found at all extrafoveal fixation positions and appeared to be affected by target object size, with larger targets exhibiting a larger facilitatory effect for inconsistent objects at further fixations. This effect was again not found when the images were inverted. Inverted small objects produced the only slight effect of consistency, with a non-significant increase in accuracy for consistent objects. However, evidence of a facilitatory effect for medium and large objects was entirely extinguished when the images were inverted.

Inverting the images was suggested to interfere with the processing of both scene and object semantics, without affecting the matching of patterns of visual features, hypothesised to be the minimum level of processing necessary to carry out the task. If inverting the images did affect processing in this way, it must be concluded that the lack of an advantage for inconsistent objects when the images were inverted indicated the importance of semantic information processing in modulating this consistency effect. As seen in Figure 2.10, there was no evidence of facilitation for inconsistent objects at the expected fixation positions. The only position at which performance for inconsistent objects was higher than that for consistent objects was position 1 and the difference in this case was only 5.0%.

Figure 2.11 compared performance for Experiments 1 and 2 and found that accuracy for consistent trials remained approximately equal when the images were inverted. However, accuracy for inconsistent trials decreased by a mean of 7.9% at each fixation position between 1 and 4, with a maximum decrease of over 14% at position 2. This result suggested that the inversion of the images affected scene processing, by selectively inhibiting performance on inconsistent objects but not consistent objects. This result suggested that the facilitation for inconsistent objects evidenced in Experiment 1 was influenced at least in part by the processing of semantic information from extrafoveal vision.

Therefore, the suggestion that there were unintentional visual differences introduced between consistent and inconsistent scenes, through the creation of inconsistent scenes, was not supported by the data. These visual differences, if present, were hypothesised to have facilitated the identification of objects which were in some way different to the scene background in which they were located, for reasons other than their semantic congruency. However, these intact visual differences would have remained in the inverted images and could have facilitated the detection of inconsistent objects, which would still have been rendered more salient than consistent objects by these differences. The lack of evidence for an inconsistent object advantage with inverted images did not support this theory and suggested that any visual differences between the images did not generate the inconsistent object advantage found in Experiment 1.

However, it is also important to consider the criticisms of the line drawing stimuli identified previously and discussed in Appendix A. Although it appeared that the inconsistent advantage evidenced in Experiment 1 was genuinely due to the processing of semantic information facilitating inconsistent objects, closer analysis of the stimuli indicated that they were unsuitable and unlikely to result in the detection of semantic inconsistency. The possibility that there were some visual differences, as yet unidentified, between consistent and inconsistent targets in scenes still remained.

Possible object differences between consistent and inconsistent targets were not properly controlled for by including each object as both a consistent target and an inconsistent target in different scenes. Each object would then act as its own control, allowing the comparison of performance for the same target presented in a consistent and an inconsistent scene. Otherwise, it would be possible to argue that the inconsistent targets were different to the consistent targets in some way not considered in this investigation.

The consistent and inconsistent targets in the same scene background were also not closely matched. The shortage of potential targets meant that it was not always possible to locate the inconsistent target at the same spatial location in the scene as the consistent target, without violating any rules of positioning. For example, if a consistent object was located on a wall or on a table, it was not always possible to find a suitably inconsistent object which could have replaced it in the same location. The salience of the targets in terms of their size, shape and also their specific location was not adequately controlled for in these experiments.

Additionally, the recognisability of the line drawing stimuli was also questioned by participants who claimed that objects, especially when presented in isolation with no background, as occurred on the two-alternative forced-choice display, were particularly difficult to identify. A further investigation of whether the stimuli were identifiable or not was conducted by asking naïve participants to name both the object and the scene background and attempt to rate the likelihood of finding the object in the specific scene. More detailed analysis is given in Appendix A. The results indicated that only one quarter of all experimental images were reliably identified by participants. Therefore any conclusions reached about performance for consistent and inconsistent trials would be subject to the criticism that the effect may have been modulated not by semantically consistent and inconsistent objects but by recognisable and unrecognisable objects.

For this reason, further investigations were conducted to investigate whether the inconsistent object advantage found in Experiment 1 could still be evidenced using more naturalistic stimuli. The stimuli were created by photographing a specific set of target objects which were carefully controlled to be of approximately the same size. Because the objects were all highly recognisable household objects, the consistent scenes were equally familiar indoor household scenes. In this way, it was intended to investigate whether differences in the processing of consistent and inconsistent objects could occur under realistic viewing conditions, with easily identifiable scene stimuli.

Chapter 3

Brief Presentations of Complex Scene Stimuli

3.1 Experiment 3: Introduction

Experiment 3 was designed to evaluate whether the inconsistent object advantage evidenced in Experiment 1 could also be produced using photographs of natural scenes and, additionally, to challenge the criticisms of the previous experiments. Experiment 2 investigated whether visual differences between consistent and inconsistent objects, possibly introduced in the construction of inconsistent scenes, could explain the inconsistent object advantage found in Experiment 1. This hypothesis was not supported by the results, in which no evidence of an inconsistent object advantage was found when the images were inverted. However, it was still possible that other visual differences between consistent and inconsistent targets may have affected performance.

As suggested in the debriefing of participants after Experiment 1 and confirmed by the analysis of scene images (Appendix A), the line drawings of both scenes and target objects used in Experiment 1 were often difficult to identify, even under free viewing conditions. This finding implied that many of the scenes and objects were likely to have been unrecognisable during the brief 120ms scene presentation allowed in the experimental procedure. Therefore, any attempt to explain differences in performance using arguments relating to semantic consistency would be subject to criticism, as detecting semantic inconsistency would require the identification of both the scene and the target object to determine the relationship between the two.

If participants viewing the images at leisure could not identify some of the scenes or objects, the categorisation of these object-scene pairs as consistent or inconsistent was arbitrary and cannot be used to explain differences in performance across conditions. The objects and the scenes would need to be recognisable by the general public and the objects would also need to be considered consistent and inconsistent with the scene context, as assigned by the experimenter. The use of photographs in Experiment 3 was believed to enhance the recognisability of both the scene images and the individual objects and this assumption was confirmed by the analysis described in Appendix A.

A final criticism of the previous experiments could suggest that the scenes used in Experiments 1 and 2 were rather simplistic and did not include many non-target objects. Photographic scenes were used in Experiment 3 to investigate whether an inconsistent object advantage would also occur when viewing scenes of sufficient complexity to approximate natural scenes, both in the composition of the scenes (i.e. number and types of objects included) and the nature of the visual stimuli. Realistic scenes would contain many more items than simplified line drawings and photographs contain more visual information relating to depth, shape, contour and other subtle cues missing from line drawings. Therefore, if an inconsistent object advantage could also be evidenced with photographs of natural scenes, this would indicate that the phenomenon may occur during real-life viewing of the visual environment.

Experiment 3 attempted to address these issues by using photographs rather than line drawing stimuli. Photographs were taken of real-world scenes, placing first a consistent and then an inconsistent object in the same location and photographing the scene containing each object from the same viewpoint. This was done to avoid any stylistic differences between images of consistent and inconsistent object trials. In order to minimise luminance variation between the consistent and inconsistent versions of a scene, mostly indoor scenes were used.

The aim was to obtain two photographs of the same scene, differing only in the target object, located in the same place, which could be either consistent or inconsistent. This would allow the comparison of performance, on a two-alternative forced-choice task identical to the one used in Experiments 1 and 2, for consistent and inconsistent objects in the same scene. The counterbalancing of targets, in which each object served as both a consistent and an inconsistent target in different scenes, would also allow the comparison of performance for the same object in a consistent and inconsistent scene.

The scenes and objects displayed in the photographs would be readily identifiable because they were genuine examples of their category rather than a simplified artistic representation of a stereotypical scene. In this way, the photographs accurately depicted the nature of visual information which we are accustomed to viewing during the course of everyday life. These images should have provided an accurate assessment of whether an inconsistent object advantage could occur with more realistic scenes.

These controls ensured that any differences in performance evidenced between consistent and inconsistent objects could be less easily dismissed as due to low-level image differences. If no inconsistent object advantage should be found, it would have repercussions on the interpretation of previous experimental data. The failure to replicate an inconsistent object advantage using naturalistic images would suggest that the effect may at best be a laboratory phenomenon which cannot be reproduced in real life. It is possible for example that the nature of the photographs would make them more difficult to process, particularly in extrafoveal vision, either because of their composition or the increased image complexity. Inconsistent object facilitation might only occur in limited situations, such as when viewing line drawings of scenes, and might not be applicable to real-life visual processing.

Finally, evidence of any advantage for inconsistent targets would suggest that inconsistent object facilitation could be a real-life phenomenon, occurring when viewing realistic scenes, even when only presented briefly. The semantic information relating to both the scene and target object would need to be processed from the 120ms presentation and any inconsistency detected with sufficient time to preferentially process inconsistent target objects compared to consistent targets. This result would imply that semantic processing could occur early during a fixation and could be directed to regions outside the fovea, such as inconsistent objects presented extrafoveally.

3.2 Method

Participants

100 participants were recruited from the undergraduate population of the University of Durham. There were 27 males and 73 females, who were all naïve to the purposes of the experiment.

Apparatus

The same apparatus were used as in Experiment 1. Participants were provided with instructions appropriate to this experiment, a consent form and a 60cm measuring ruler. A debriefing sheet was provided upon completion.

Materials

Colour photographs of household scenes were obtained using a digital camera. These photographs had an original resolution of 640 x 420 pixels. The colour photographs were converted into grey scale photographs with a resolution of 800 x 600 pixels using PaintShop Pro, for presentation in this experiment. The decision was taken to convert the original photographs to grey scale images to prevent colour from affecting scene processing, for example when targets or distractors were brightly coloured. This was particularly important in the garden scene, as the predominant background colours were green and brown and often targets would have been distinguishable and more salient by their colour alone. The entire set of experimental stimuli can be found in Appendix B.

The photographs were of real-life scenes in colleagues' homes. Target objects were selected which would be consistent in a certain scene or room of the house but inconsistent in another. The targets were counterbalanced, with each object appearing as both a consistent and an inconsistent target in different scenes. This manipulation was included to control for the salience of an object's visual features by presenting it as both a consistent and an inconsistent target. For example, the toaster was a consistent target in the kitchen and an inconsistent target in the child's playroom, as depicted in Figure 3.1.

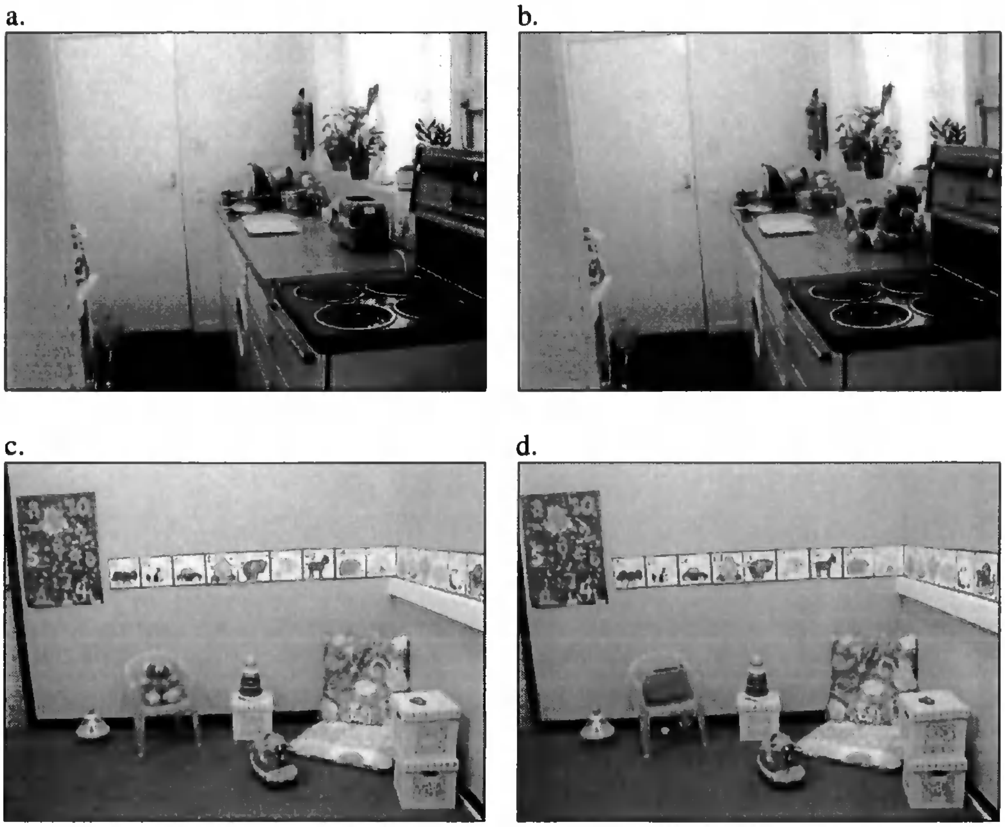


Figure 3.1: Examples of scenes used as experimental images.
a. Kitchen (consistent target toaster), b. Kitchen (inconsistent target teddy bear),
c. Playroom (consistent target teddy bear), d. Playroom (inconsistent target toaster)
A complete set of 64 photographs consisting of 9 different scene types (7 scenes with 4 consistent and 4 inconsistent targets and 2 scenes with 2 consistent and 2 inconsistent targets) with 32 different target objects, can be found in Appendix B.

Each consistent target object was paired with an inconsistent target object which was matched for actual size and shape where possible. For example, the consistent toaster in a kitchen scene was paired with a teddy bear, which was an inconsistent target in the kitchen but a consistent target in the child's playroom. In the playroom, the toaster acted as an inconsistent target. This counterbalancing of objects allowed the comparison of performance on the same object when presented in a consistent or inconsistent scene context or performance on different objects in the same location in the scene.

In total, there were 9 differently named scene backgrounds, for example, bathroom, kitchen and living room. However, to ensure that participants did not view the same scene backgrounds containing different target objects repeatedly, photographs were taken of more than one scene of each type. For example, although the toaster was presented in a kitchen with an inconsistent teddy bear located in an identical version, an alternative consistent target, a kettle, was placed in a different kitchen scene with a matched inconsistent target, a football, in the alternative inconsistent version. This allowed the presentation of both a consistent object and an inconsistent object in each named scene background to the same participant, without repeating the same specific background image twice. As an example, a participant could view a consistent toaster in a kitchen scene and also an inconsistent football in another kitchen scene, without viewing the same background twice.

Where possible, several different backgrounds were obtained to allow participants to view a greater number of scenes, as participants could view two versions of each scene type when two different backgrounds were used. Some scene types contained four different backgrounds, allowing participants to view four images of this scene type. Two consistent objects and two inconsistent objects could be viewed in different backgrounds. Scenes such as the living room and dining room did not produce as many possible inconsistent objects and so fewer scene versions were constructed.

The completed images, experimental scenes and two-alternative forced-choice displays, were presented at a resolution of 800 x 600 pixels. The forced-choice displays contained images of two individual objects located against a white background. The alternatives were either the two consistent objects for a particular scene (e.g. a toaster and a kettle, for a kitchen scene) or the two matching inconsistent objects (e.g. a teddy bear and a football). Only one of the images was ever the target in the given scene background. This design would minimise response bias according to perceived probability of an object being located in a scene.

A set of photographic practice trials was also created. Eight scenes which were not used as experimental trials were photographed. Four of these contained an inconsistent object which was selected as a target object and the other four did not, so a consistent object was selected to be the target. For each of these images, the target object's coordinates were calculated and participants' fixation positions were manipulated relative to this location.

Design

The independent and dependent variables were the same as in the two previous experiments. The distance between the participant's fixation position and the target object was manipulated, as was the consistency between the target object and the scene background. Each participant viewed 32 trials in this experiment, two consistent and two inconsistent trials for each of the seven scene types made up of four different backgrounds and one consistent and one inconsistent trial for the remaining two scene types with only two different backgrounds. The distance between the fixation position and target object was randomised and each participant viewed a unique set of trials.

Procedure

The procedure in this experiment was identical to that used in the previous experiments. Participants received instructions and were required to complete a consent form if they agreed to participate. A set of practice trials was conducted before the experimental trials. In each trial, a fixation cross was presented for 1000ms to direct participants' fixation to a specific region of the scene. Then a rapid 120ms presentation of a scene image was visible, in this experiment a photograph, and was followed by a two-alternative forced-choice display in which two objects were presented. When participants made a response, the next trial would begin automatically after an inter-trial interval of 1000ms. The entire procedure of practice and experimental trials took no longer than 15 minutes.

3.3 Results

Participant accuracy appeared higher than in the previous experiments. The approximate chance level was again set at 60%, at least 19 correct responses out of 32 trials. Only 5 participants did not achieve this level of accuracy and their data were replaced by other naïve participants on the same set of trials.

Distance of the target object from fixation

Figure 3.2 indicates the decrease in performance as the distance between the target object and the participants' fixation position increased. From a maximum of over 90% correct when the target was presented at fixation, accuracy dropped to a minimum of 66% at positions 3 and 4, when the eyes were fixated over 9° away from the target object. The data indicated that performance was best when the target object was presented at the participants' fixation position and decreased as the target object was presented further away, suggesting that participants were fixating the fixation cross as instructed.

A binary logistic regression analysis confirmed that fixation distance had a significant main effect on accuracy ($\chi^2(1)=182.9$, $p<.001$), as closer fixation positions resulted in higher accuracy. However, the decrease in accuracy appeared to plateau between 9° and 12° at approximately 66%, with no obvious further decrease in performance beyond 9°. Even at 12° from fixation, performance did not fall to chance level ($p<.001$). The use of photographs therefore seemed to improve performance, compared to when line drawing images were presented.

Consistency of the target object

There was no main effect of consistency ($\chi^2(1)<1$, $p=.89$) on response accuracy. Mean accuracy for consistent trials was 78.3%, compared to 78.1% for inconsistent objects. This could be compared to Experiments 1 and 2, where accuracy was approximately only 70% or less. Performance was improved when viewing photographs compared to when viewing line drawings, indicating that the target objects were easier to process in photographs where more detailed information was available compared to in simple line drawings.

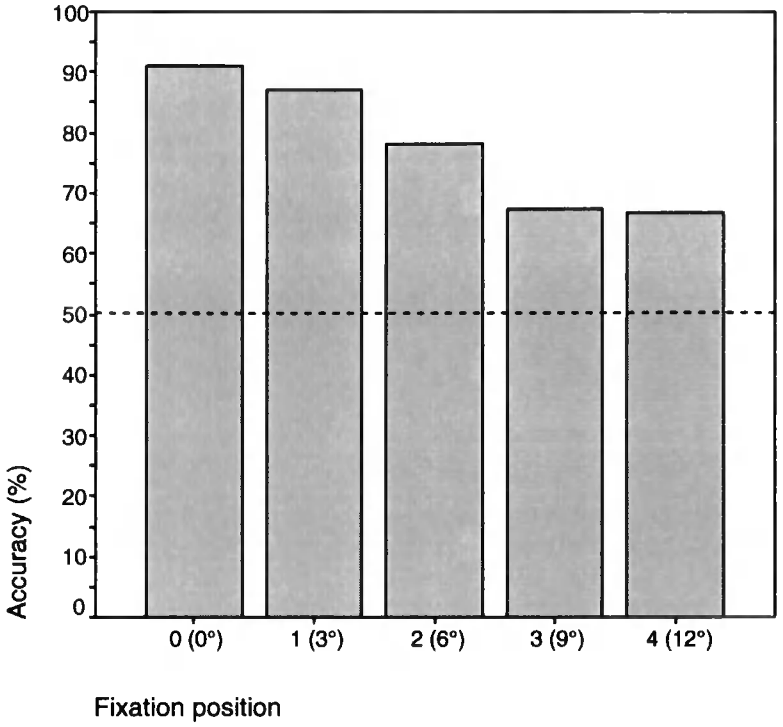


Figure 3.2: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12°. Chance level of 50% is indicated.

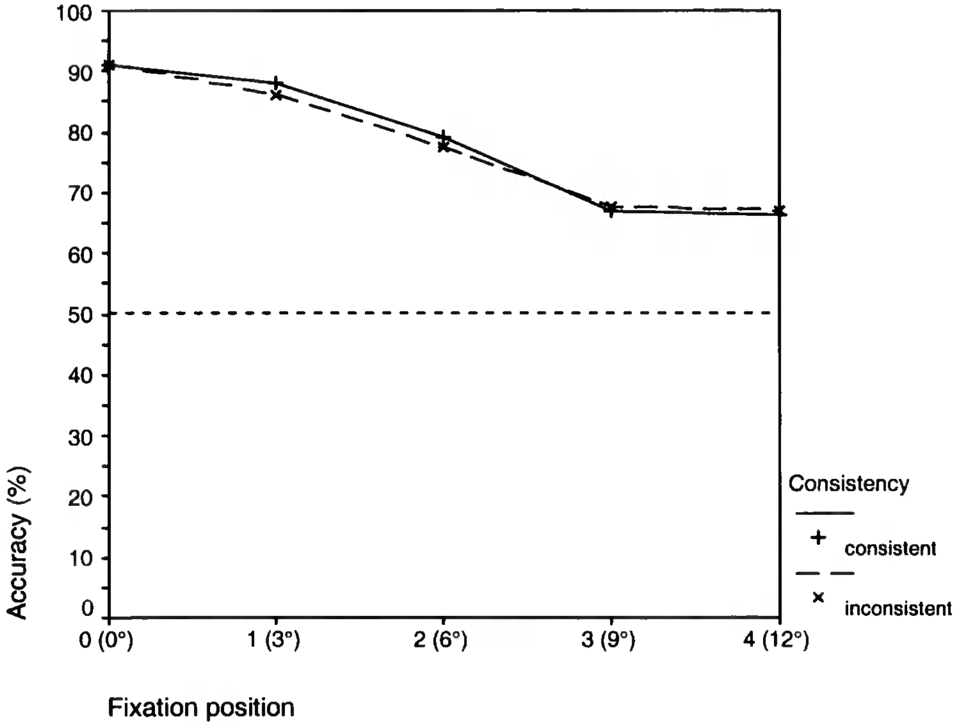


Figure 3.3: Graph showing the change in accuracy by fixation position and target object consistency. Chance level of 50% is indicated.

Interaction of fixation position and consistency

Figure 3.3 illustrates the relationship between fixation position and consistency, allowing the comparison of performance at each fixation position for consistent and inconsistent targets. No significant interaction was found for consistency and fixation position ($\chi^2(1)<1$, p=.94). Unlike the results of the previous experiments, there was no significant difference between performance for consistent and inconsistent targets at any fixation position. Even when the target object was presented at fixation, there was no advantage for consistent objects. Similarly, there was no advantage for inconsistent objects at any other fixation position.

Counterbalanced objects

The experimental design allowed the comparison of performance for the same object in different scenes, with consistent or inconsistent backgrounds, and also for different objects in the same scene, with consistent or inconsistent targets. The accuracy for each object in each scene was calculated and the accuracy for the same object was compared across consistent and inconsistent scenes. The assumptions for a parametric test were explored and the distribution of accuracy scores for inconsistent scenes was found to be skewed containing one outlier, as was the distribution of difference scores. However, variances were found to be approximately equal so a matched pairs t test was conducted.

This test found no significant difference in accuracy between the same object placed in a consistent and an inconsistent scene background ($t(31)<1$, p=.94). Closer observation indicated that out of the 32 targets, 13 objects showed higher accuracy when presented in the consistent scene, 18 showed an advantage in inconsistent scenes and one object showed no difference. Therefore it became clear that there was no reliable difference between performance for an object in a consistent or an inconsistent scene.

It was also possible to compare accuracy on consistent and inconsistent objects in the same scene background, as the objects were approximately matched for size and shape. Again, the data were explored and the distribution of accuracy scores for inconsistent objects was slightly skewed with one outlier, but the distribution of the mean differences was normally distributed and variances were approximately equal so a parametric analysis was conducted. The matched pairs t test indicated that there was no significant difference in accuracy between consistent and inconsistent objects placed in the same scene background ($t(31)<1, p=.95$). Of the 32 consistent-inconsistent object pairs, 15 showed higher accuracy for the consistent object, 16 evidenced higher accuracy for the inconsistent object and one pair showed no difference at all. These data provided no evidence of any difference between performance for consistent and inconsistent object trials with the same scene background.

Response times

Response times for correct trials, according to consistency and fixation position, are presented in Table 3.1 The response times were considerably shorter than those found in Experiments 1 and 2, again indicating that the task was easier when the images were photographs. A univariate ANOVA of the data by participants indicated a significant main effect of fixation position, with faster response times when the target was presented at fixation (631ms), increasing to 980ms when the target was presented 12° from fixation ($F(4,938)=28.5$, $p<.001$). There was no significant main effect of consistency ($F(1,938)<1$, $p=.62$), with consistent objects being responded to with a mean time of 795ms compared to 783ms for inconsistent objects. There was also no significant interaction between fixation position and consistency ($F(4,938)<1, p=.82$). Response times for inconsistent targets were shorter than for consistent targets at positions closest to fixation but not significantly so. At greater distances from fixation, the relationship between response time and consistency was less clear.

Table 3.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only.

	0 (0°)	1 (3°)	2 (6°)	3 (9°)	4 (12°)	Mean
Consistent	645	722	803	940	943	795
Inconsistent	618	671	818	880	1017	783
Mean	631	697	811	910	980	789

Object size

Further analyses were conducted to determine why no difference was found between performance for consistent and inconsistent objects in scenes, including considering the relationship between performance and target object size. To investigate whether target size had any effect on performance, each object was labelled as small, medium or large. As in Experiment 1, object size was calculated by determining the pixel area of each object in each scene. Small objects had a pixel area less than 4,000 (under 4° square), medium objects measured between 4,000 and 8,000 pixels (4° to 8° square) and large objects were greater than 8,000 pixels in area (over 8° square).

24 objects were small in size, of which 11 were consistent and 13 were inconsistent. Of the 23 medium sized objects, 12 were consistent and 11 were inconsistent. The 17 large objects consisted of 9 consistent objects and 8 inconsistent objects. The toaster and teddy bear targets in the kitchen scene, depicted in Figure 3.1, were of medium size but were classed as small when located in the playroom.

Figure 3.4 displays accuracy according to object size, with 95% confidence intervals around each mean. The graph indicated that performance for small and medium sized objects was approximately equivalent, with large objects responded to with significantly higher accuracy. The absence of the predicted monotonic increase in accuracy as object size increased could have been due to the fact that the sizes were distributed over a large range and could not be separated into three discrete groups, without some objects' areas corresponding to the borderline regions for classification.

The interactions between the variables of interest were investigated by tabulating accuracy scores according to object size and consistency condition (Table 3.2). This table indicated that there was no difference in performance according to consistency overall and this was reflected in the accuracy rates for small objects. However, there was a 6.42% advantage for consistent targets of medium size, which was reversed to show a 9.75% inconsistent object advantage for large targets.

Table 3.2: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	77.27	79.33	78.00	78.27
Inconsistent	76.46	72.91	87.75	78.06
Mean	76.83	76.26	82.59	78.16

The improved performance for consistent objects of medium size over inconsistent objects of medium size was proved significant by a binary logistic regression analysis indicating a main effect of consistency ($\chi^2(1)=7.24, p=.007$). From the data presented in Table 3.2, the difference in accuracy appeared to be caused by a decrease in accuracy for inconsistent objects rather than an increase in accuracy for consistent objects. Comparing accuracy for inconsistent small and medium objects, the expected increase due to the target being larger and more salient did not occur. Accuracy for medium sized consistent objects remained comparable to both small and large consistent objects, not reflecting the expected increase in accuracy by object size.

Similarly, the difference evident for large consistent and inconsistent objects was also significant ($\chi^2(1)=14.6, p<.001$), indicating higher accuracy for inconsistent objects. This effect seemed to be caused by a significant increase in accuracy for large inconsistent objects, compared to both large consistent objects and medium sized targets. The evidence indicated that, in some way, large inconsistent objects were more salient than large consistent objects.

To investigate these interactions further, a binary logistic regression analysis was conducted including all the independent variables and their interactions as possible predictors. The results indicated a main effect of fixation position ($\chi^2(1)=26.1, p<.001$) and a main effect of object size ($\chi^2(2)=7.47, p=.024$), with higher accuracy for larger objects than small or medium sized objects. Additionally, significant interactions between object size and fixation position ($\chi^2(2)=8.56, p=.014$) and between object size and consistency ($\chi^2(2)=22.0$, $p<.001$) were evidenced. To observe the patterns in the data more closely, accuracy rates for small, medium and large objects were investigated individually.

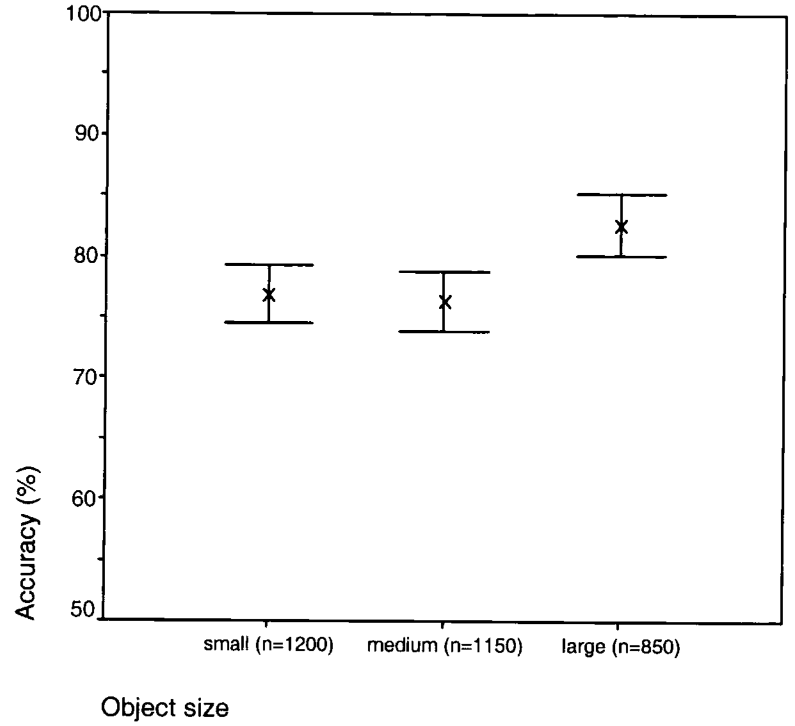


Figure 3.4: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes.

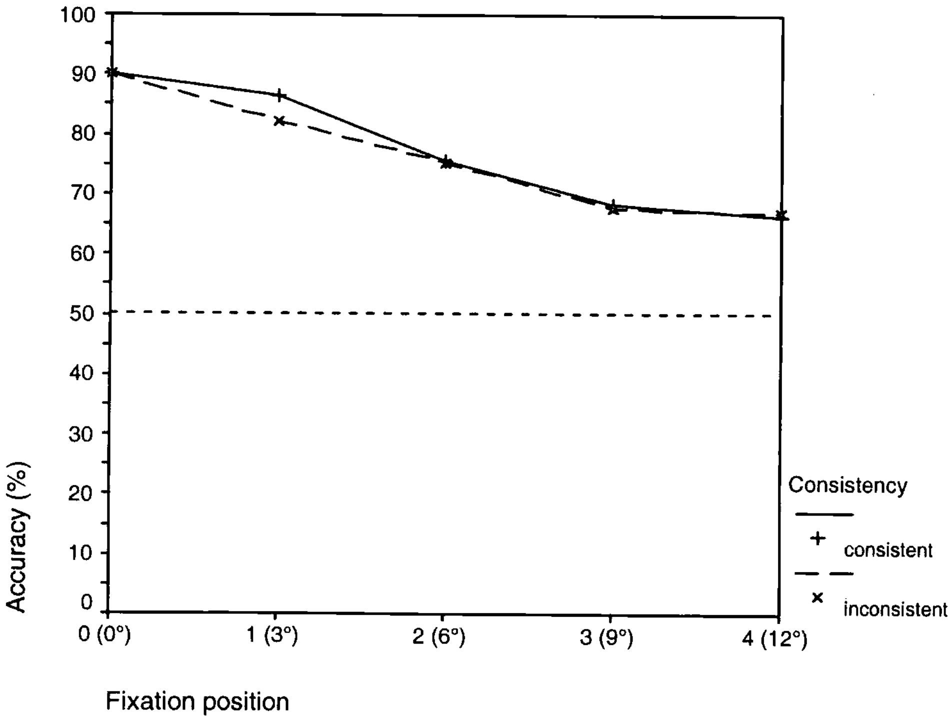


Figure 3.5: Graph showing change in accuracy by fixation position and

consistency for small objects only. Chance level of 50% indicated.

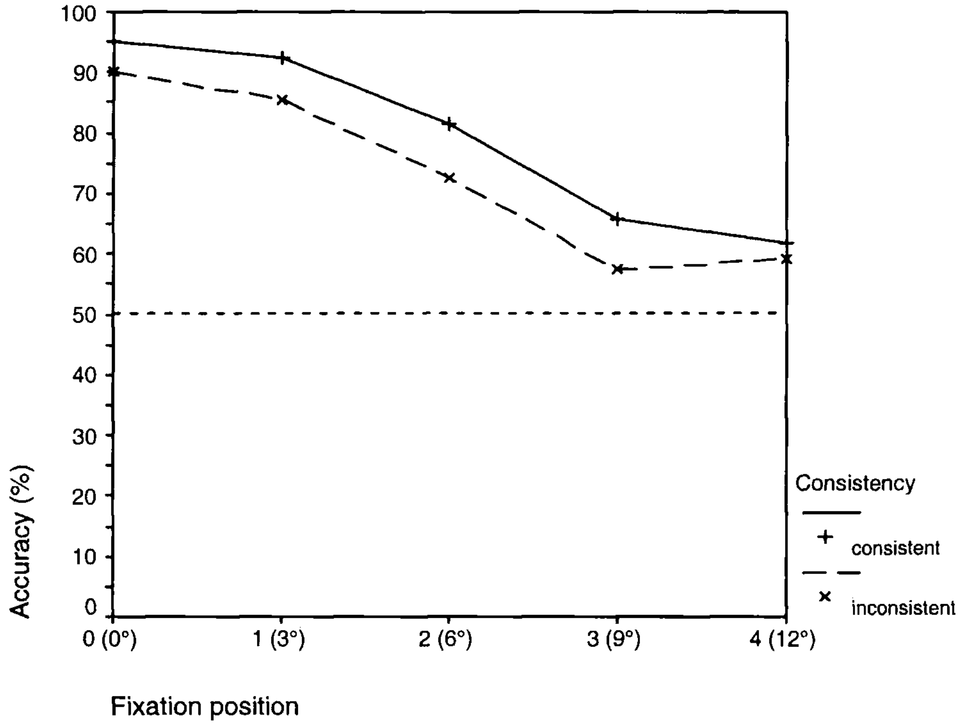


Figure 3.6: Graph showing change in accuracy by fixation position and consistency for medium objects only. Chance level of 50% indicated.

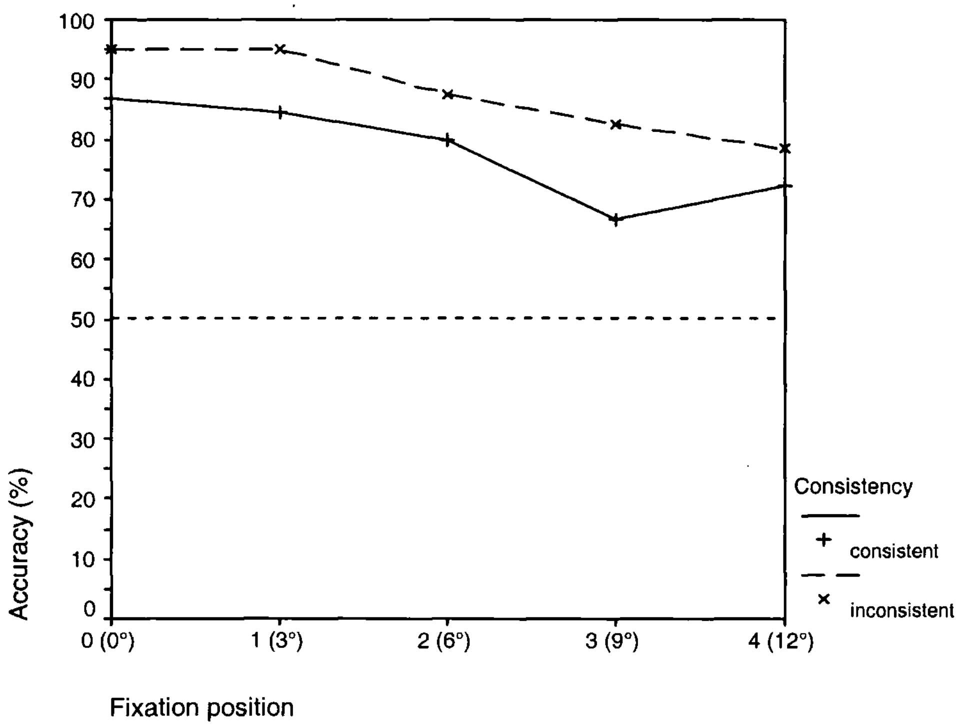


Figure 3.7: Graph showing change in accuracy by fixation position and consistency for large objects only. Chance level of 50% indicated.

The data relating to small objects showed no difference in performance according to consistency but the distribution of results was investigated to provide a comparison for medium and large objects. Figure 3.5 displays accuracy for trials in which the target objects were small, across consistency and fixation conditions, indicating no difference in performance between consistent and inconsistent objects. As expected, performance was best for small objects when they were presented closest to fixation and decreased as the distance between the participants' fixation position and object location increased.

Figure 3.6 displays the relationship between accuracy and the experimental variables for medium sized objects only. Consistent objects were responded to more accurately than inconsistent objects at all fixation positions, resulting in a significant main effect of consistency (*p*=.007). The advantage seen in Table 3.2 was not affected by fixation position, indicating that consistent objects were responded to more accurately, regardless of where they were presented in the scene.

The opposite effect was exhibited by large objects, with a significant advantage for inconsistent objects at all fixation positions (*p*<.001). Figure 3.7 indicates that performance for inconsistent objects was significantly better than that for consistent objects at all fixation positions and the advantage was not influenced by the target's location. However, these results were based on less experimental data as there were fewer large objects than small or medium sized objects so the results could be less reliable.

The analysis of small, medium and large objects individually indicated that there were significant differences between performance for consistent and inconsistent objects within these size categories. The advantage for consistent medium sized objects and the advantage for inconsistent large objects could have cancelled each other out in the analysis of the entire data set, resulting in no clear consistency effects. Again, even with more complex visual images, object size was seen to modulate consistent and inconsistent advantages, indicating that this variable needs to be considered in the investigation of consistency effects.

High quality image subset

In order to investigate whether the failure to evidence a consistency effect from the entire data set could be attributed to the consistency manipulation being insufficiently strong, a subset of the stimuli was selected for use in a post-hoc analysis. It was considered that the consistency manipulation for the line drawing stimuli was stronger than for the photographic stimuli, because line drawings could contain objects located in impossible locations, such as a swing in a laboratory. As a result of obtaining readily recognisable photographic stimuli depicting household objects and scenes, the consistency manipulation was constrained to likely and unlikely scene-object combinations.

As many of the targets objects in the photographic scenes were not considered extremely unlikely to occur in the scene, a selection of 20 images was made from the most reliably recognised scenes, isolating the most reliably rated consistent and inconsistent images. The analysis described in Appendix A was used to identify the scene-object pairs which were correctly identified by all participants and which were the most reliably classified as consistent and inconsistent. To ensure that no effects were attributable to object size, the stimuli were selected to ensure an equal number of small, medium and large targets in each set of images. The mean likelihood rating for the 10 selected consistent scenes was 4.95 (*SD*=.07), where a rating of 5 indicated an object which was very likely to appear in the scene. The mean rating for the 10 selected inconsistent scenes was 1.17 (*SD*=.12), where a rating of 1 was considered very unlikely to appear in a scene.

The data are displayed in Figure 3.8. A binary logistic regression analysis indicated a significant main effect of fixation position on accuracy ($\chi^2(1)=79.8$, $p<.001$) but no significant effects of consistency ($\chi^2(1)<1, p=.51$) or the interaction between the two variables ($\chi^2(1)<1, p=.96$). Even with the strongest consistency manipulation possible with the available realistic scene images, no significant effects of consistency were found. Accuracy for these 10 consistent trials was 71.4% compared to 73.2% for the 10 inconsistent trials, indicating that there was no difference in accuracy according to the semantic consistency of the target object.

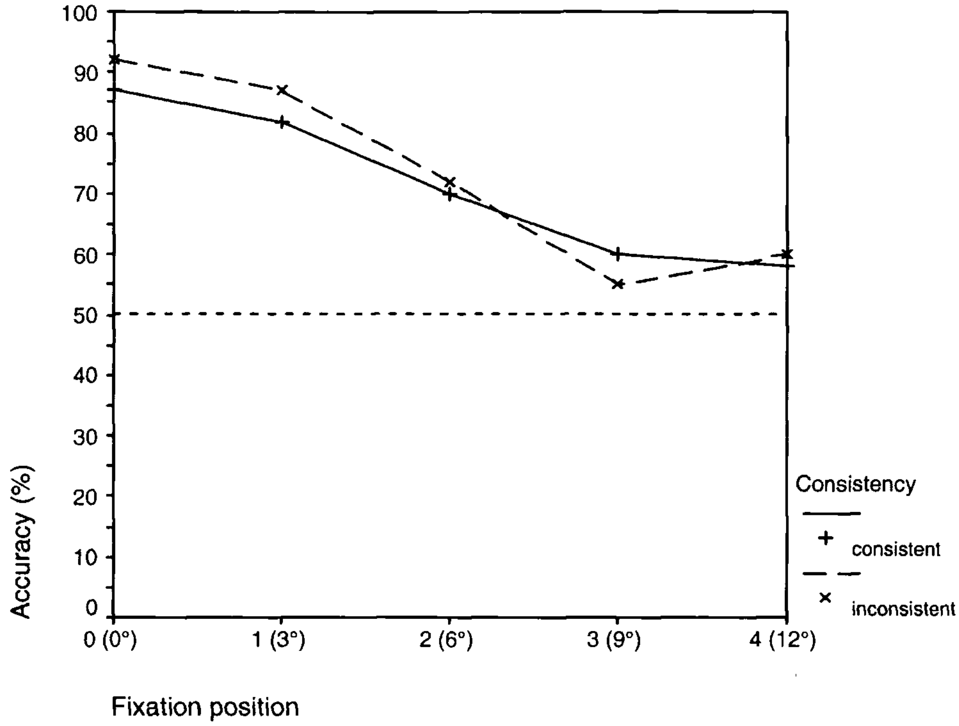


Figure 3.8: Graph showing the change in accuracy by fixation position and target object consistency for high quality images. Chance level of 50% is indicated.

Unlike the results obtained from analysing the entire data set, this restricted data set indicated a slight inconsistent object advantage in accuracy at the closest fixation positions. However, performance at further positions showed no clear effects of consistency. The data were analysed separately at each fixation position to investigate possible differences in performance. When the target was presented directly at fixation, a non-significant 5% difference in accuracy was found between consistent and inconsistent targets ($\chi^2(1)=1.34$, $p=.25$). A similar 5% difference was found at fixation position 1 but this also failed to reach statistical significance ($\chi^2(1)<1$, $p=.33$).

These analyses indicated that the slight advantage found for inconsistent targets when presented closer to fixation was not statistically significant. There was no reliable difference between performance for consistent and inconsistent scenes, even within this stimuli set containing the most reliably identified consistent and inconsistent images. However, the slight differences found were in the direction predicted by Hollingworth and Henderson so this argument could remain a possible explanation for the absence of a consistency effect with photographs of household scenes.

3.4 Discussion

Performance was found to be better in this experiment than in Experiment 1 which used line drawings of scenes, suggesting that the processing of objects, at least to the extent necessary to perform this task, could be performed more successfully from photographs than from line drawings. As expected, a significant main effect of fixation position was found, with significantly better performance when the target object was presented closer to the participants' fixation position. The consistency manipulation between the object and the scene context did not have a significant main effect and no significant interaction effects were found.

There were no significant differences between performance for consistent and inconsistent objects when viewed at any eccentricity, including at fixation. Experiment 1 evidenced a significant consistent object advantage when the target object was presented directly at the participant's fixation position. A similar though non-significant advantage was found in Experiment 2 when the scenes were inverted. This suggested that when the images were line drawings, the contextual advantage of viewing a related scene prior to the object identification task assisted in the selection of the target object. This advantage was extinguished when the scenes and objects were inverted, which was hypothesised to interfere with the processing of semantic information without influencing the matching of visual features. It was possible that the increased volume of information in the photographs made the additional contextual information unimportant in the selection of a target when it was foveated.

Experiment 1 also evidenced a significant inconsistent object advantage for objects presented extrafoveally, which was not seen when the images were inverted. This inconsistent object advantage was not replicated with the use of photographs, indicating that the effect evidenced in Experiment 1 could not be replicated using more complex and naturalistic images. The analysis of the high quality image subset also failed to evidence a reliable effect of consistency. However, the non-significant differences found between performance for consistent and inconsistent targets suggested slightly higher accuracy for inconsistent than consistent targets, so the possibility of evidencing a reliable inconsistent object advantage, using a strong enough consistency manipulation, could not be discounted.

In Experiment 3, unlike Experiment 1, target objects were more carefully controlled to ensure that consistent and inconsistent target pairs appeared at the same location and were as closely matched for size and shape as possible. As the chosen targets were ordinary household objects and the scenes were real-world locations, the context of the scene background and the semantic information relating to the target objects should have been easy to determine. The photographs also retained additional visual information relating to contour and texture which could be lost in simple line drawings. Any effects found with the photographic stimuli could therefore be attributed to the detection of inconsistency between the scene and the target object when both the objects and the scenes could be readily identified.

A consequence of the increased level of visual detail present in the photographic scene images was that any single fixation contained more visual information for foveal processing during the brief presentation of a photograph than a line drawing. This could have interfered with the extrafoveal processing of target objects when they did not appear foveally. However, greater accuracy for photographs of scenes than for the line drawings used in Experiment 1 was found, even when the target was presented extrafoveally, indicating that, contrary to the suggestion outlined, the detection of a target was easier in photographic stimuli than in line drawings, despite the increase in visual detail.

The data were also analysed according to individual object sizes in order to investigate possible explanations for the absence of a consistency effect. Accuracy improved as object size increased and a significant interaction between consistency and object size was found (*p*<.001). Accuracy was highest for large objects, due predominantly to an increase in accuracy for large inconsistent objects to over 87%, while large consistent objects showed no significant increase in accuracy above that found for medium sized consistent objects (approx 79%). This significant advantage for large inconsistent objects was mirrored by an advantage for medium sized consistent objects (*p*=.007) which indicated a decrease in accuracy for inconsistent targets, rather than a facilitation for consistent targets.

It was hypothesised that these two effects could have cancelled each other out to produce no significant effect overall. It would therefore be desirable to control approximate object size when investigating consistency effects in scenes, as including a large range of object sizes may introduce conflicting evidence and occlude any effects specific to larger or smaller objects. This finding of the modulation of consistency effects by object size could also help to explain incompatibility in the results obtained by previous studies in this field.

As each object acted as a consistent and an inconsistent object in different scenes, each large inconsistent object replaced a large consistent object so the large object could not have appeared ‘out of place’ or unusually large in its location in the scene. The advantage must have been caused by an increase in target salience at all fixation positions but the exact nature of this salience remains unknown. Therefore, it can be concluded that an inconsistent object advantage was found but only in favour of large inconsistent objects compared to large consistent objects. The evidence of an opposite advantage for consistent medium sized objects or an inhibition for inconsistent medium sized objects may have cancelled out any main effects.

Several possible alternative explanations for the absence of a main consistency effect in this experiment, compared to Experiment 1, can be considered. As described above, the objects used in consistent and inconsistent scenes were carefully controlled to ensure that there was no advantage for consistent or inconsistent trials due to obvious object differences. This may indicate that the objects used in Experiment 1 produced an advantage for inconsistent objects in a way which was not investigated or controlled for. Alternatively, the effect evidenced in Experiment 1 could have been a valid and genuine effect and the inability to replicate it in Experiment 3 could be due to the nature of the materials. It was possible that the level of detail included in photographs interfered with the extrafoveal processing of semantic information from the entire scene during a brief presentation. In this way, an inconsistent object advantage may have been a genuine effect which could not be reproduced under realistic viewing conditions. In order to determine whether the use of photographs rather than line drawings prevented the expression of a consistency effect, a further experiment was designed which incorporated the stricter control of target objects in this stimuli set with the use of line drawings.

3.5 Experiment 4: Introduction

The purpose of this experiment was to investigate the discrepancy in results between Experiment 1, using simple line drawings of scenes, and Experiment 3, using photographs of complex scenes. This difference could have been due to the nature of the two stimuli types, with photographs being responded to more accurately than line drawings, possibly influenced by the improved recognisability of the photographed scenes. The line drawing stimuli were visually much simpler than the photographs, which contained more detailed visual information about perspective, depth and shading for example. The line drawings were simplified images, created by selecting only relevant major vertices, boundaries and edges from a more complex scene image, suggesting that line drawings could be easier to process. However, the more detailed visual information available in a photograph could allow easier and faster recognition of objects, because their representation in a photograph would more closely resemble their real life appearance than the approximation resulting from a line drawing.

In addition, the photographs used in Experiment 3 resembled real life scenes more than the line drawings used in Experiment 1. The photographs depicted rooms in genuine homes and therefore must have illustrated familiar types of indoor scenes and household objects. In comparison, the line drawings used in Experiment 1 depicted a larger proportion of less familiar outdoor scenes, which may have been more difficult to identify. For example, a theatre or a waterfront would be encountered less frequently than a household scene and could subsequently have fewer rigid constraints on which objects could be found located in them, even if identified correctly. These outdoor scenes also contained correspondingly larger objects which were not necessarily portable, such as a target barge in a waterfront scene, so the size range of the line drawing targets in real life would be much larger than for household targets in photographs.

A final consideration was that the naturalistic photographs were also more complex than the line drawings, not only in the quality of visual information present but also in the quantity of information, as they contained a greater number of objects than most line drawings. Within the Leuven stimuli set, many items usually present in naturalistic, photographic images had not been included in the line drawings. This image simplification may have contributed to the difficulty in identifying some scenes, as there could have been insufficient non-target objects to assist in the identification. In some cases, the target object was one of very few items depicted in the scene and this may have affected the processing difficulty of both the scene and the object.

For these reasons, it was decided to investigate whether the nature of the photographs as photographs was responsible for the absence of any inconsistent object facilitation in Experiment 3. To this end, the photographs were converted into line drawings whilst maintaining both the familiarity of the scenes and the complexity of their composition, estimated by the proportion of non-target objects included. The same experimental procedure was applied to this new set of line drawing stimuli.

A direct comparison of performance for photographs and comparable line drawings would indicate whether performance for photographs was indeed improved over the line drawings used in Experiment 1 due to the nature of the visual images or the improved recognisability of the stimuli. If, using this set of line drawing images, there should be no evidence of inconsistent object facilitation in extrafoveal vision, this result would suggest that it was not the use of photographs as such which extinguished the effect but possibly the complexity of the scene. However, evidencing an inconsistent object advantage comparable to that in Experiment 1 would indicate that it was the nature of the photographic stimuli, with enhanced visual detail, which extinguished the effect and that the complexity of the scene, in terms of its components, did not affect the processing of semantic information from extrafoveal vision.

3.6 Method

Participants

100 participants were recruited from the undergraduate population of the University of Durham to participate in this experiment and in Experiment 2. As described previously, 15 males and 85 females who were naïve to the hypotheses tested by this series of experiments participated in this study.

Apparatus

The same apparatus were used as in Experiments 1 to 3. Appropriate instructions, a consent form and a debriefing sheet were provided along with a 60cm measuring ruler.

Materials

The images used in this experiment were line drawings created from the photographs used in Experiment 3. These line drawings were designed to match the photographs as closely as possible and maintain an appropriate level of detail. Two-alternative forced-choice displays also depicted line drawings of the targets. The images were all presented at a resolution of 800 x 600 pixels and they are provided in Appendix B. Practice trial line drawings from the Leuven set used in Experiment 1 were used as practice trials in this experiment.

Design

The independent variables remained the distance between the fixation position and the target object and the semantic consistency between the scene and the target object. The dependent variables measured were response accuracy and response time, measured in ms. The order of trials was randomised for each participant.

Procedure

This experiment used the same procedure as the previous 3 experiments. A block of practice trials was presented to each participant before the experimental trials. At the start of each trial, a fixation cross was presented for 1000ms which would direct the participant to fixate a specific region. The scene was then presented for 120ms and was followed by a two-alternative forced-choice display in which line drawing representations of two consistent or two inconsistent objects were presented for the participant to select from. At the end of this experiment, participants were instructed to proceed to Experiment 2.

3.7 Results

As in Experiment 2, there were insufficient participants to replace data sets in which accuracy was not above the selected level of 60%. Therefore, the first 100 data sets collected were used in this analysis. Under the criteria implemented in Experiment 3, there were 31 data sets in which participants did not perform better than the chosen level, indicating that performance was much worse for line drawings of photographs than for the original photographs, with only 5 participants failing to achieve the appropriate accuracy rate in Experiment 3.

Distance of the target object from fixation

Figure 3.9 illustrates how accuracy decreased as the distance between the participants' fixation position and the target object increased. At fixation, accuracy was 81%, which was lower than found in previous experiments, and fell to 54% when the object was presented approximately 12° from fixation. As expected, performance was most accurate when the target object was presented at fixation or close to fixation. Accuracy then decreased to approximately chance levels by position 3, when the target object was presented at 9°. The effect of fixation position was confirmed to be statistically significant by a binary logistic regression analysis ($\chi^2(1)=137.9, p<.001$). Unlike performance in Experiment 3, when photographs were displayed, performance on the line drawings of the photographs did decrease to chance levels which suggested that the task was easier to perform when more visual information was available, even during such brief presentations.

Consistency of the target object

No significant effect of consistency was found although a trend of higher accuracy for consistent objects than inconsistent objects was indicated ($\chi^2(1)= 3.52, p=.061$). Consistent objects were responded to correctly in 64.25% of trials, compared to the 61.13% accuracy rate for inconsistent objects. Accuracy was again less than was evidenced when photographs were used as stimuli, when approximately 78% accuracy was obtained for both consistent and inconsistent objects.

Interaction of fixation position and consistency

Figure 3.10 displays accuracy for consistent and inconsistent targets across fixation positions and found no evidence of a reliable advantage for either consistent or inconsistent objects, although consistent objects produced higher accuracy at three of the five fixation positions. An investigation of the interaction between consistency and fixation position found no significant effect on accuracy ($\chi^2(1)=1.95, p=.16$).

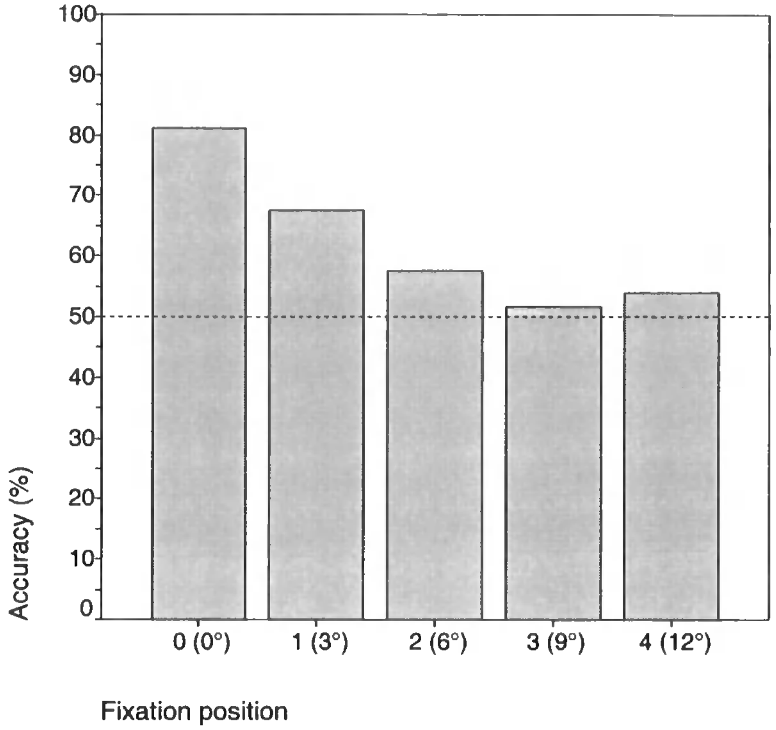


Figure 3.9: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12°. Chance level of 50% is indicated.

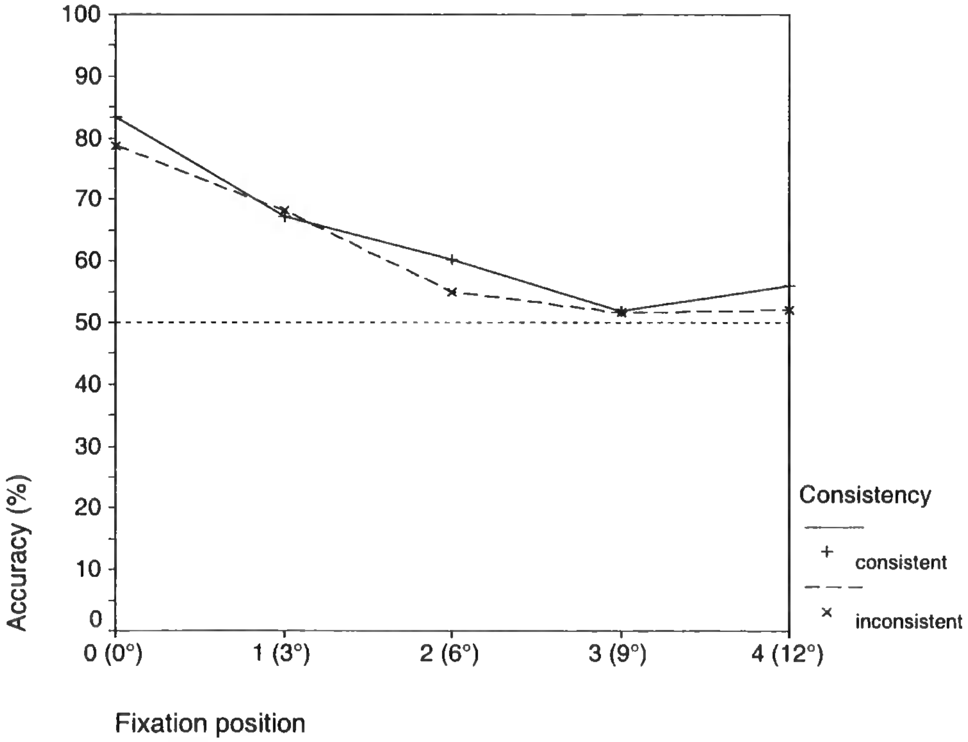


Figure 3.10: Graph showing the change in accuracy by fixation position and target object consistency. Chance level of 50% is indicated.

Accuracy decreased rapidly but the mean value at position 4 (54%) was still significantly different to chance (*p*=.040). Although there was no significant effect of consistency at this position, indicating no significant difference between accuracy for consistent and inconsistent trials ($\chi^2(1)=1.14$, *p*=.29), accuracy for consistent trials at this position was significantly different to chance at 56% (*p*=.029) but accuracy for inconsistent trials at 52% was not (*p*=.50). Similar analyses for targets appearing at fixation position 3 indicated that accuracy was not significantly different to chance (*p*=.40) with accuracy rates for both consistent and inconsistent trials under 52%. Accuracy for inconsistent trials was less than accuracy for consistent trials at fixation position 2 also and not significantly different to chance at 55% correct (*p*=.083), but performance for consistent trials was different to chance at 61% (*p*<.001). Accuracy levels at closer fixation positions were also significantly above chance level.

The observation that accuracy decreased more rapidly when the photographs were converted into line drawings was investigated further by plotting the data for both stimuli types together (Figure 3.11). The discrepancy between performance for photographs (blue) and their line drawing equivalents (green) was illustrated clearly. There was an obvious decrease in performance equally for consistent and inconsistent trials across fixation positions. As expected, performance was highest when the target object was displayed at fixation and showed the least difference in accuracy between the different stimuli types at just 10%. However, performance for line drawings was up to 20% worse than accuracy for the equivalent photographs at greater fixation positions.

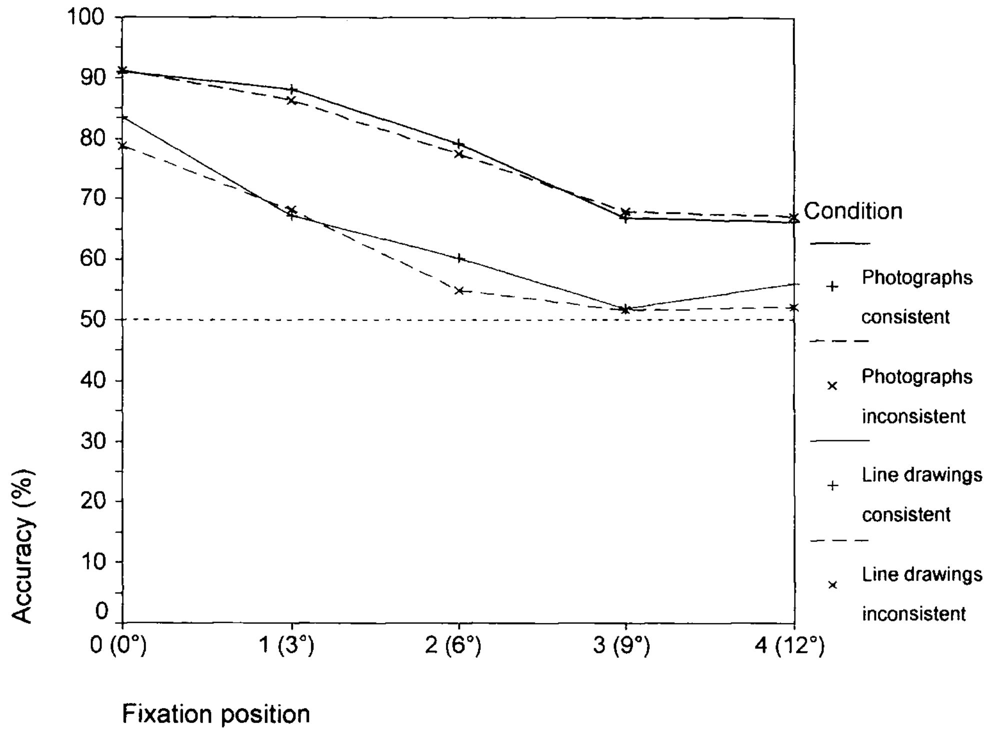


Figure 3.11: Graph showing the change in accuracy by fixation position and target object consistency, for photographs and their line drawings. Chance level of 50% is indicated.

Counterbalanced objects

Accuracy for each target object was calculated, according to whether it was located in a consistent scene or an inconsistent scene. The assumptions of parametric testing were explored. The distributions of accuracy for consistent and inconsistent scenes were approximately normal, with only one outlier, the distribution of mean accuracy differences were normally distributed and the variances were equal so a matched pairs t test was conducted. The results indicated that accuracy did not vary reliably according to whether an object was located in a consistent or an inconsistent scene ($t(31)<1, p=.38$). Of the 32 targets used, 3 objects showed no difference in scores between the consistent and inconsistent scenes, 17 had higher accuracy in consistent scenes than in inconsistent scenes and the remaining 12 displayed better performance in inconsistent scenes than consistent scenes. This analysis concluded that there was no reliable difference between accuracy for each object according to the consistency of the scene in which it was located.

The relationship between accuracy for the consistent and inconsistent objects placed in the same location in the scene was also evaluated. Again, accuracy distributions were approximately normal with only one outlier, the distribution of mean accuracy differences were slightly skewed with no outliers and the group variances were equal. A matched pairs t test indicated no significant difference between accuracy for matched consistent and inconsistent objects in the same scene backgrounds ($t(31)=1.05, p=.30$). 19 object pairs showed higher accuracy for the consistent object than the inconsistent object in the same scene background, 12 pairs showed higher accuracy for the inconsistent object than the consistent object and one pair evidenced no difference at all between consistent and inconsistent object accuracy. Therefore the data indicated that there was no difference between performance for consistent and inconsistent objects in scenes when line drawings of photographs were presented.

Response times

Table 3.3 provides mean response times for correct trials, according to fixation position and consistency. From a mean time of 947ms when the target was presented at fixation, response times increased to 1371ms at 12° eccentricity. Overall, these response times were considerably longer than those found for the photographic stimuli which ranged from 618ms to 1017ms, suggesting that the task was more difficult when line drawings were presented. The response times recorded in this experiment were more comparable to those found in Experiment 1 with Leuven line drawing stimuli, which ranged from 830ms to 1303ms.

Table 3.3: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only.

	0 (0°)	1 (3°)	2 (6°)	3 (9°)	4 (12°)	Mean
Consistent	910	1129	1241	1340	1323	1162
Inconsistent	985	1166	1378	1426	1423	1245
Mean	947	1147	1306	1383	1371	1202

A univariate ANOVA by participants of the response times for correct trials indicated a significant effect of fixation position ($F(4,889)=26.5, p<.001$) and a significant effect of consistency ($F(1,889)=11.7, p=.001$). No significant effect was found for the interaction between fixation position and consistency ($F(4,889)<1$, $p=.95$). The data in Table 3.3 indicated an increase in response time when target objects were presented at further fixation positions. The main effect of consistency reflected a significant mean delay of 83ms when responding correctly to inconsistent object trials. Response times for consistent trials were reliably shorter than those for inconsistent trials across all fixation positions and there was no interaction. The increased difficulty of this task, illustrated by the decrease in accuracy, could have resulted in the consistency effect evident in response times.

Object size

As the scenes underwent some modification in being converted to line drawings, the size of the targets objects were re-calculated using the same criteria, to ensure accuracy. Of the 20 small objects, defined as having a pixel area less than 4,000 (4° square), 11 were consistent targets and 9 were inconsistent targets. There were 23 medium sized objects with pixel areas between 4,000 and 8,000 (between 4° and 8° square) consisting of 10 consistent targets and 13 inconsistent targets. 21 of the target objects were classified as large, with areas exceeding 8,000 pixels square (over 8° square), of which 11 were consistent and 10 were inconsistent.

Figure 3.12 displays accuracy by object size with 95% confidence intervals around each mean. Performance improved with larger object size and the difference between medium and large objects appeared significant. Accuracy overall was again lower than in previous experiments but the expected trend of increasing accuracy with larger objects was found. To investigate whether consistency influenced this effect, the data were tabulated according to this variable (Table 3.4). Inconsistent targets produced lower accuracy than consistent targets for all object sizes, although the difference was not always large. Unlike previous data, although an advantage was found for consistent objects over inconsistent objects of medium size, this was not matched by an inconsistent advantage over consistent objects when the targets were large.

Table 3.4: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	58.00	64.60	70.18	64.25
Inconsistent	56.44	58.46	68.80	61.13
Mean	57.30	61.13	69.52	62.69

These effects were confirmed using binary logistic regression analyses. A significant main effect was found for fixation position ($\chi^2(1)=139.3, p<.001$) and also for object size ($\chi^2(2)=36.4, p<.001$). However, consistency was only a marginally significant variable ($\chi^2(1)=3.55, p=.060$) and no significant interactions were found. The data were further analysed according to individual object size categories.

Figure 3.13 plots the data for small objects only. No reliable effects of consistency on accuracy were displayed. Performance decreased to approximately chance levels by fixation position 2 and remained low at further eccentricities. This pattern suggested that any differences between accuracy for consistent and inconsistent objects at these eccentricities would be unreliable, as participants were not performing better than chance.

Figure 3.14 depicts accuracy for medium sized objects and indicated a pronounced effect of consistency at closer fixation positions, with consistent objects displaying higher accuracy than inconsistent objects. At the furthest fixation positions, performance decreased to chance levels and no difference was found according to consistency. A binary logistic regression analysis on these data indicated a significant effect of fixation position ($\chi^2(1)=65.4, p<.001$) and also of consistency ($\chi^2(1)=4.76, p=.029$), confirming that consistent targets were responded to with more accuracy than inconsistent targets. No significant interaction was found between the two variables ($\chi^2(1)=2.03, p=.16$).

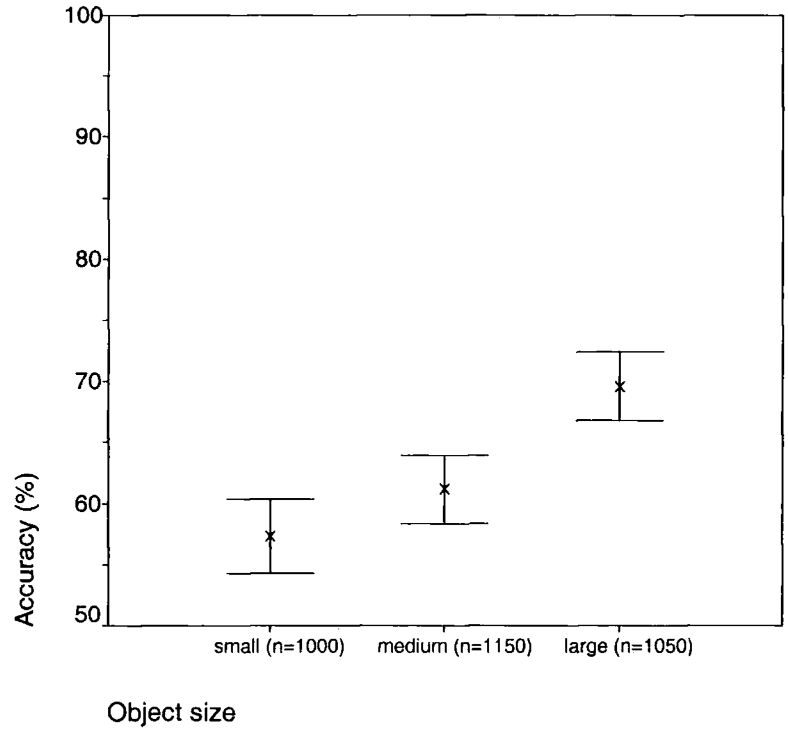


Figure 3.12: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes.

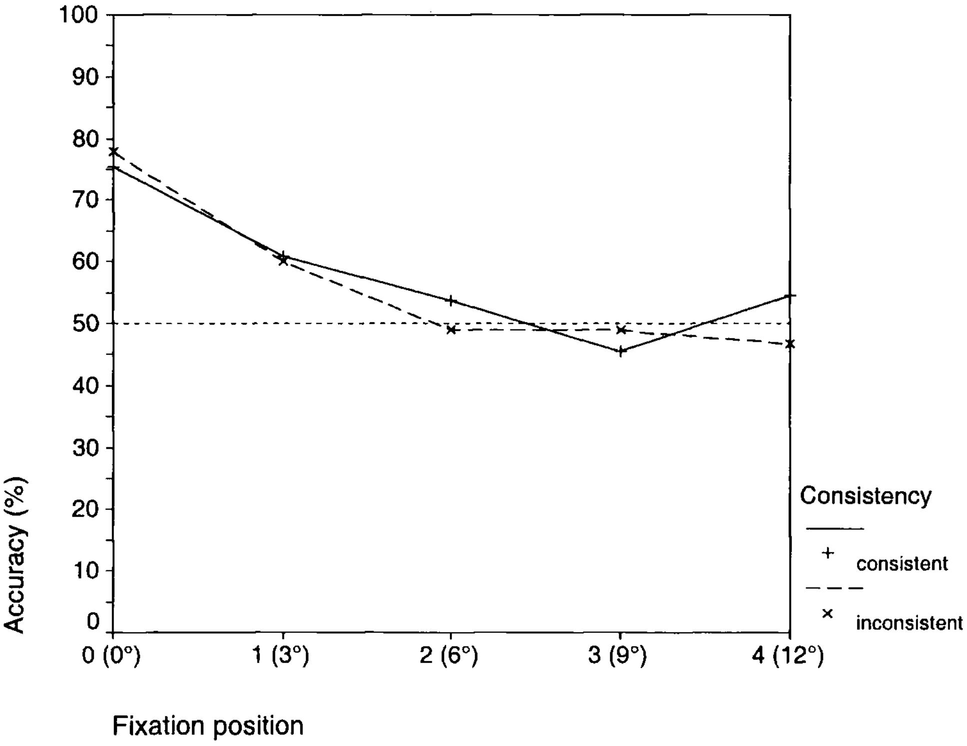


Figure 3.13: Graph showing change in accuracy by fixation position and consistency for small objects only. Chance level of 50% indicated.

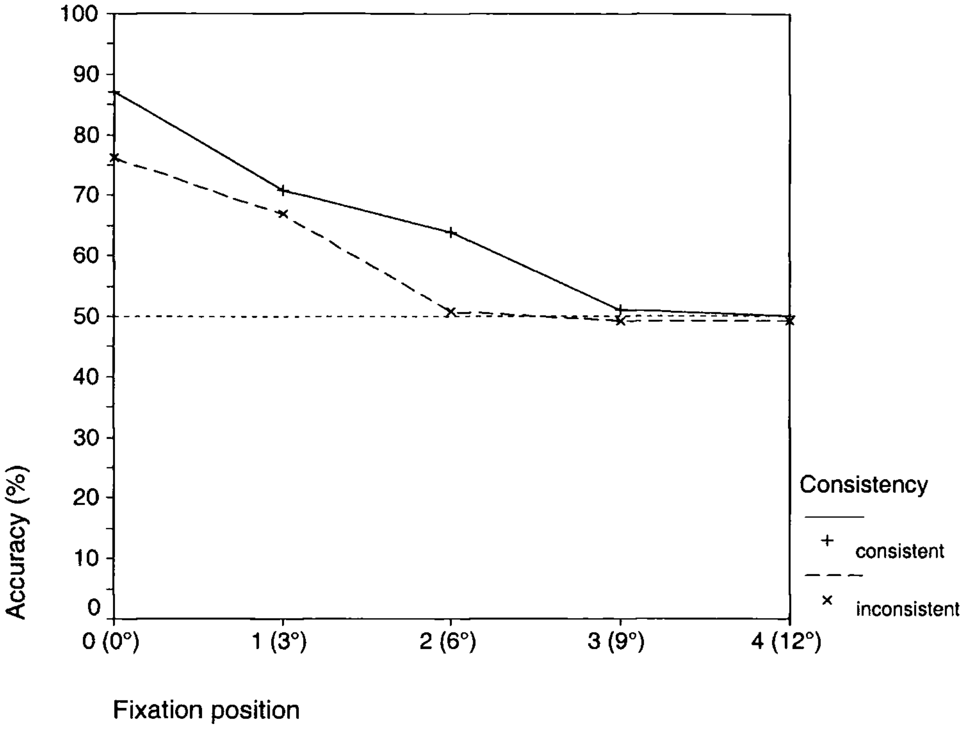


Figure 3.14: Graph showing change in accuracy by fixation position and consistency for medium objects only. Chance level of 50% indicated.

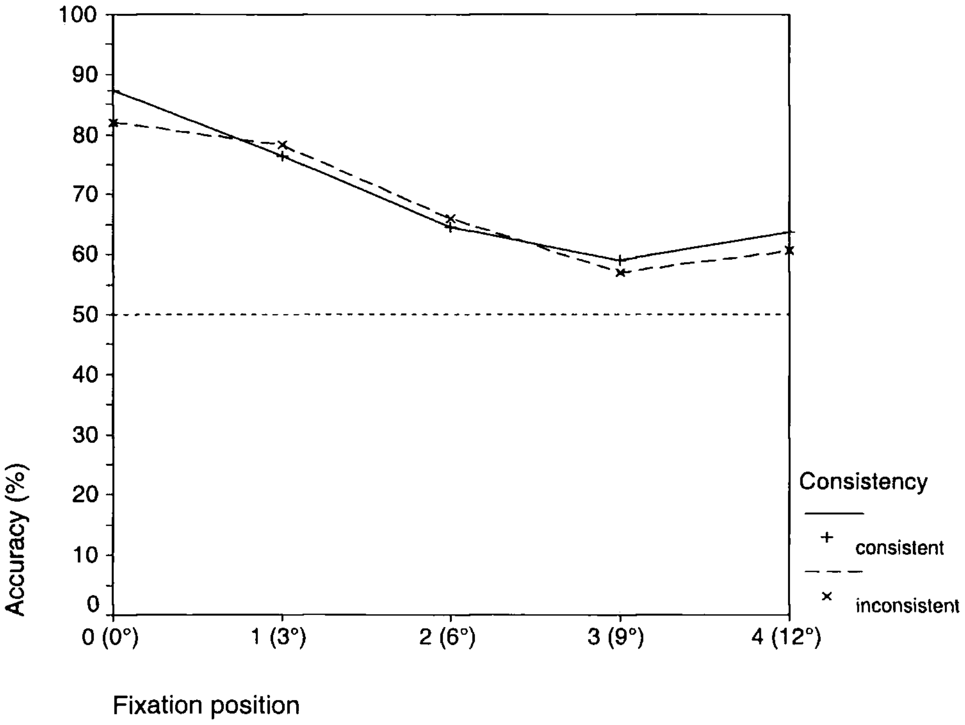


Figure 3.15: Graph showing change in accuracy by fixation position and consistency for large objects only. Chance level of 50% indicated.

The main effect of consistency was apparent when performance was above chance, up to fixation position 2 (6°), but disappeared when performance decreased to chance level. Performance for inconsistent objects decreased to chance level at position 2, at a closer eccentricity than consistent objects (position 3). These results can be compared to those of Experiment 3 in which performance for medium sized objects also provided an advantage for medium sized consistent objects over medium sized inconsistent objects.

Figure 3.15 presents data for large objects only. From this graph, it was apparent that consistency did not influence accuracy in any way. Unlike performance in Experiment 3, which indicated a pronounced advantage for inconsistent targets at all fixation positions, no advantage for inconsistent objects was found with the line drawings. This comparison suggested that the conversion of the photographs to line drawings had somehow interfered with the detection of semantic inconsistency. Accuracy remained above chance level even at furthest fixation positions however, indicating that these objects were responded to more accurately than smaller objects and that the absence of an effect could not be attributed to poor accuracy.

The analysis of object size indicated that the marginal consistency effect, suggesting that consistent objects were responded to more accurately than inconsistent objects, was solely generated by medium sized objects. Neither small nor large objects indicated any reliable effect of consistency. The advantage for consistent medium sized objects over inconsistent medium sized objects was found at positions closer to fixation, possibly as performance was above chance level at these positions. The same consistent object advantage in Experiment 3 was pronounced at all fixation positions but accuracy at all positions was significantly above chance when photographs were displayed.

High quality image subset

The data were analysed according to the most reliably rated consistent and inconsistent scenes, as in Experiment 3. Although no significant evidence was found for an inconsistent object advantage in Experiment 3, the pattern of results was in the direction predicted by Hollingworth and Henderson. To determine whether this possible effect was enhanced, possibly to significant levels with the use of line drawings, further investigation of the data from Experiment 4 was considered appropriate.

A binary logistic regression analysis was conducted to investigate possible effects of consistency on accuracy when analysing only data relating to the 10 most consistent and the 10 most inconsistent target trials. A significant main effect of fixation position was found ($\chi^2(1)=43.1, p<.001$) but there was no significant main effect of consistency ($\chi^2(1)<1, p=.75$). However a marginally significant interaction between fixation position and consistency was evidenced ($\chi^2(1)=3.94, p=.047$). To observe this effect, the data are presented in Figure 3.16.

The significant interaction appeared to be caused by higher accuracy for consistent objects than inconsistent objects at three fixation positions, with the largest difference found at the furthest fixation position. A slight 5% difference in accuracy was found at fixation, indicating that consistent targets were detected more accurately than inconsistent targets when they were directly fixated. This difference was compatible with the results of the entire data set but was not found to be statistically significant ($\chi^2(1)<1, p=.42$).

There was no difference in accuracy at fixation position 1 but an 8% difference was found at position 2, again indicating higher accuracy for consistent targets. However, this difference also failed to reach statistical significance ($\chi^2(1)=1.28$, $p=.26$). Accuracy at position 3 was equal for both consistent and inconsistent targets and the largest difference of 15% was found at the furthest fixation position. This difference was statistically significant ($\chi^2(1)=4.52, p=.033$) and indicated that consistent targets were responded to with higher accuracy than inconsistent targets.

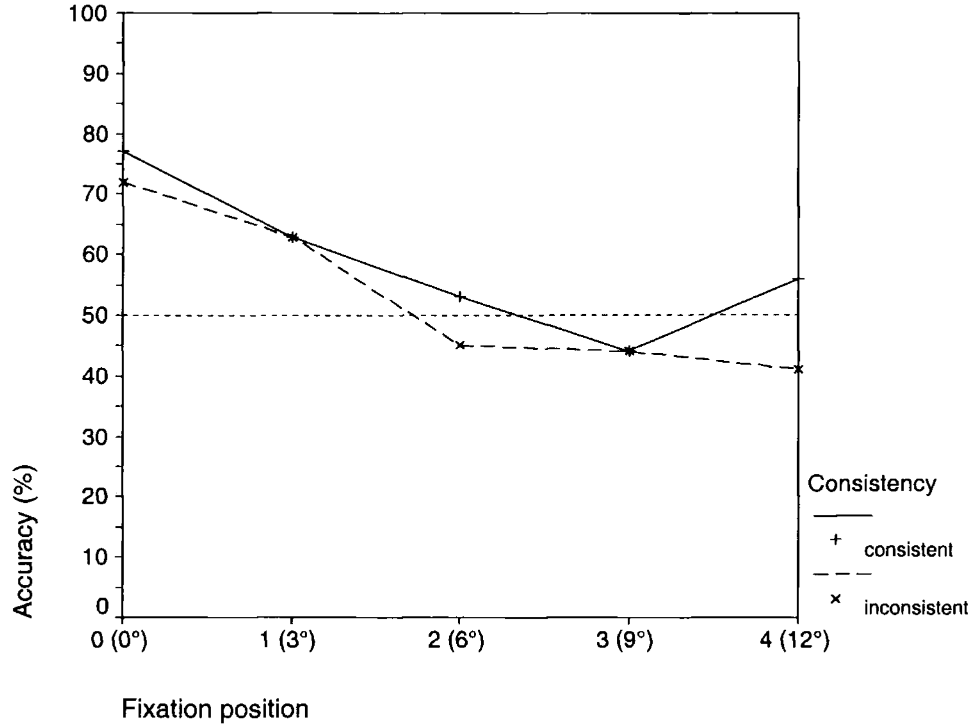


Figure 3.16: Graph showing the change in accuracy by fixation position and target object consistency for high quality images. Chance level of 50% is indicated.

This result runs contrary to the predictions made by Hollingworth and Henderson that inconsistent objects were facilitated in their brief presentation experiments. A consistent object advantage could be explained by the facilitation generated by the relevant contextual information provided by the scene presentation. However, the inconsistent object advantage detected reliably by Hollingworth and Henderson was not replicated in this analysis.

This analysis investigated the most reliably consistent and inconsistent scenes from the realistic photographic images, when the images had been converted into line drawings to mimic the stimuli types used by Hollingworth and Henderson. This investigation still failed to evidence a significant inconsistent object advantage in this task and also failed to support the proposal that the consistency manipulation in Experiments 3 and 4 was insufficiently robust to generate the consistency effect found by Hollingworth and Henderson in their experiments. The absence of a reliable inconsistent object advantage using high quality images in both Experiments 3 and 4 did not support Hollingworth and Henderson's conclusions and suggested that the inconsistent object advantage detected by these researchers could not be reliably replicated with naturalistic scene images, either photographs or line drawings.

3.8 Discussion

Accuracy was found to be affected by the participants' fixation position, with higher accuracy at closer fixation positions. A trend indicated a marginal effect of the consistency manipulation, favouring performance for consistent objects, but this was not influenced by fixation position. Specific investigation of performance for the same object in different scenes failed to evidence any effect of consistency. Additionally, there was no reliable pattern of facilitation for either consistent or inconsistent objects in the same scene background, indicating that the consistency manipulation failed to evidence any differential processing for consistent and inconsistent objects. The analysis of the high quality images confirmed this conclusion, failing to find a reliable advantage for inconsistent targets over consistent targets, when the most reliably rated scenes were selected for further analysis.

A comparison of performance with photographs (Experiment 3) and their equivalent line drawings (Experiment 4) indicated that performance with the line drawings was much less accurate at all fixation positions. The conversion of the images not only influenced extrafoveal processing but also foveal processing and resulted in poorer performance, proving that the increased visual detail in photographs did not have an inhibitory effect on accuracy. Instead performance was enhanced by the inclusion of more detailed visual information in the images, suggesting that the absence of a consistency effect was not caused by an inherent difficulty in processing extrafoveal information from more complex images.

The effects of object size were investigated by analysing the data according to the size of the target object presented. As expected, performance improved as object size increased, with larger objects being easier to detect in brief presentations of scenes. However, no significant interaction was found between consistency and object size as was evidenced in Experiment 3, which displayed an inconsistent object advantage for large targets and a consistent object advantage for medium sized targets. The advantage for large inconsistent objects compared to large consistent objects in photographs must have been generated by features which were not present in the line drawing images of the same objects and scenes, as the effect was not found with these stimuli. These results suggested that the additional visual detail present in photographic stimuli may have affected the level of processing achieved during a brief presentation and that using line drawings of stimuli could have detrimentally affected performance.

Similarly, Experiment 3 provided evidence for a consistent object advantage for medium sized objects at all fixation positions, while Experiment 4 found an advantage only when the targets were presented closer to fixation, at positions less than 6° eccentricity. It was hypothesised that the absence of a consistency effect was influenced by the decrease in accuracy to chance levels beyond this eccentricity, so a consistency effect was only found when the visual processing of extrafoveal objects was sufficient to complete the task. These parallel results did not indicate a dissociation between the eccentricity at which semantic processing and general object identification processing became impossible due to increasingly degraded visual information.

3.9 Summary of brief presentations experiments 1 to 4

In Experiment 1, line drawings from the Leuven library were used to investigate how fixation position influenced the processing of semantic consistency in a briefly presented visual scene. The results indicated an advantage for inconsistent objects when they were presented in locations not subject to foveal processing, up to approximately 12° into extrafoveal vision. The inconsistent object advantage was modulated by the size of the target object. Small objects showed no variation in accuracy according to consistency but both medium and large objects displayed higher accuracy for inconsistent objects presented at extrafoveal fixation positions, an effect particularly significant for large targets.

However, the analysis of the high quality image subset indicated that semantic consistency was unlikely to generate the consistency effect found, as the most recognisable scene images failed to evidence any effect.

It was hypothesised that the process of creating inconsistent scenes might have introduced visual differences between scenes containing an inconsistent target and those containing a consistent target, giving rise to a 'consistency effect' unrelated to semantic processing. A further experiment was conducted in which the original scene images were inverted. This was proposed to interfere with the processing of semantic information from the scenes and objects, without affecting the process of pattern matching which would be sufficient to perform the task.

Experiment 2, using the inverted images, did not replicate the inconsistent object facilitation found in Experiment 1. Overall, no reliable effects of consistency were found and analyses by object size failed to evidence any differences modulated by consistency. A comparison of results for Experiment 1 and Experiment 2 indicated that accuracy decreased overall when the images were inverted and closer inspection suggested that the cause of the decrease was the selective reduction in accuracy for inconsistent targets presented at extrafoveal locations.

Performance for consistent objects was not significantly affected by the image inversion but accuracy for inconsistent trials decreased significantly. This pattern suggested that, if the inversion had affected processing as hypothesised, the interference in the processing of semantic information removed the facilitation for inconsistent objects found in Experiment 1. Any visual differences present between consistent and inconsistent scenes were therefore considered unlikely to have caused the inconsistent object advantage, or Experiment 2 would also have evidenced an advantage as hypothesised. Alternative explanations for the inconsistent object advantage included the inadequate control of target objects, which may have contributed to the consistency effect by making the task easier when the target was inconsistent. However, statistical investigation of size differences between consistent and inconsistent objects failed to provide evidence to support these alternative hypothesis.

The inconsistent object advantage was called into question by the fact that participants claimed to find both the objects and scenes difficult to identify and the most recognisable images did not display a similar consistency effect. It was therefore decided to produce a more appropriate and recognisable stimuli set by using photographic images of household objects and scenes, which would also allow more control over the selection of target objects. Even using these stimuli and the high quality subset derived from them, Experiment 3 failed to evidence any difference between performance for consistent and inconsistent objects at any fixation position. However, although small objects showed no difference in accuracy by consistency, a reliable consistent object advantage was found for medium sized objects at all fixation positions and large objects displayed a reliable inconsistent object advantage. These two effects appeared to cancel each other out when analysing the entire data set but indicated that consistency effects could also be modulated by the size of the target object.

To determine whether the use of photographs affected results detrimentally, Experiment 4 used line drawing equivalents of the photographic experimental images and also failed to find a reliable advantage for either consistent and inconsistent objects in the entire data set. Analysing the high quality images, a marginally significant interaction (p=.047) was found, supporting higher accuracy for consistent targets at fixation positions 0, 2 and 4, although accuracy at positions 2 and 4 were approximately at chance levels. An analysis by object size indicated no significant effects of consistency on small or large objects but an advantage was found for consistent medium sized objects over inconsistent medium sized objects, when they were presented within approximately 6° of fixation.

It was concluded that although the inconsistent object advantage for large objects was not replicated with line drawings, the consistent object advantage for medium sized objects was replicated at fixation positions where the target could be adequately processed and identified. This specific conclusion suggested that line drawings of medium sized objects in scenes could not be processed sufficiently for identification in this task beyond approximately 6°, although photographs of the same objects in scenes were processed sufficiently for performance to remain significantly above chance level even at the furthest position, 12° from fixation. The comparison of accuracy for Experiments 3 and 4 indicated that the additional visual information inherent in photographs did not adversely affect accuracy and, on the contrary, accuracy was higher when photographic stimuli were used.

Possible explanations for these patterns of results need to be considered. The advantage evidenced for inconsistent objects in Experiment 1 must be interpreted with caution, as identification difficulties made it appear unlikely that this effect was caused by semantic inconsistency. According to the schema hypothesis, the rapid categorisation of a scene schema also activates representations of objects likely to be found in it. Unless it could be determined that a scene was sufficiently detailed to activate a specific schema, resulting in the activation of objects likely to be found in it, it would be impossible to conclude that participants could in some way distinguish between objects consistent and inconsistent with that scene.

The scenes used in Experiments 1 and 2 often contained very few diagnostic objects which would be used to generate expectancies, implying that some of the scenes may not have been categorised specifically enough within a 120ms presentation to determine the consistency of component objects. Even with unlimited, self-paced viewing times, participants often expressed difficulty in deciding whether an object was likely to appear in a scene, usually due to an inability to conclusively identify the scene or object (Appendix A). Therefore, it seemed unlikely that the processing of semantic information could be generating the difference in performance for consistent and inconsistent objects in Experiment 1 and this conclusion was confirmed in the analysis of the high quality images. However, none of the investigated visual differences were found to evidence the consistency effect either.

In the same way as the scenes were considered difficult to recognise, it was also claimed by participants that they had difficulty selecting between the two alternative objects when they were both equally unrecognisable. They claimed that the objects which were likely to be found in the scene were easier to recognise than unlikely objects when they were presented in the two-alternative forced-choice display. The consistent scene background presumably assisted in the identification of the objects displayed subsequently. When two objects were presented which seemed unrelated to the previous presentation, some participants claimed to have randomly selected between the two, neither of which could be identified. If this were true, the effect evidenced by objects classified as being inconsistent could have been generated not by the perceived inconsistency between the object and the scene, but by the increased processing difficulty in identifying the object.

The fact that photographs of natural scenes failed to produce any effect of consistency may have provided support for this idea, as the photographs of household objects were very easy to recognise. It was possible however that the selection of objects for use in this experiment failed to provide a strong enough manipulation of semantic consistency. As each object had to serve as a consistent target in one household scene and an inconsistent target in another, the targets in inconsistent scenes were often considered unlikely, rather than impossible, to be found there. Although consistent objects were selected to be diagnostic items in the relevant scene where possible, they could have been considered simply misplaced in another room. An additional concern was that the participants were all undergraduate students who may have had less rigid ideas about acceptable locations for specific objects. While the participants appropriately claimed that inconsistent objects were unlikely to be present in specific scenes under free viewing conditions, this relationship may not have been obvious from brief presentations.

There was also a substantial difference between the Leuven line drawings used in Experiment 1 and the line drawings used in Experiment 4. As the line drawings in Experiment 4 were created from household scenes, they were easier to identify and often contained more non-target objects than the Leuven line drawings. It seemed likely that the line drawings in Experiment 4 contained more visual information and appeared more complex than those in Experiment 1. It is possible that the factor inhibiting the inconsistent object advantage in Experiment 4 compared to Experiment 1 was the additional visual information present in the line drawings. With more component objects resulting in more 'lines' included in the scene image, there would be more information to process which may have interfered with the increased salience of inconsistent objects in extrafoveal vision.

From the available data, it seems possible to draw one of two conclusions. The inconsistent object advantage found in Experiment 1 might have been the result of the detection of semantic inconsistency, even when targets were presented at 12° eccentricity, although this would be unlikely if not replicated with the high quality images. The failure to evidence a similar advantage with stimuli of or derived from photographs of household scenes could suggest that the effect does not occur with realistic scene stimuli or plausible consistency manipulations. Perhaps naturalistic scene stimuli cannot provide a strong enough manipulation of consistency, by including likely and unlikely objects rather than the stronger manipulation of possible and impossible objects.

This hypothesis was not supported by Boyce et al's (1989, 1992) conclusion that the consistency effect evidenced was not modulated by object probability, concluding that it was the object's plausibility rather than its possibility which was important when categorising consistent and inconsistent objects. However, this analysis was conducted on data displaying a consistent object advantage rather than an inconsistent object advantage, so the conclusions may not be entirely applicable. It may be possible for realistic images to fail to evidence differential processing according to consistency as more simplistic images might, limiting the value of an inconsistent object advantage to a curious laboratory phenomenon. While inconsistent object advantages have been evidenced in several experiments, the conditions under which they occur could be highly selective and of no realistic application.

Conversely, it could be argued that the inconsistent object advantage found in Experiment 1 was not caused by the processing of semantic information but instead by the difficulty in identifying the objects, especially without congruent scene information. This may explain why the inversion condition failed to maintain the consistency effect. When presented upside down, all target objects would have been equally difficult to recognise and any advantage for less recognisable inconsistent objects relative to more recognisable consistent objects, facilitated by a related scene context, would have been removed.

The data indicated that congruent semantic information such as that originating from context did not influence performance on this task, with simplified line drawings, as seen by the comparable performance for consistent objects in upright and inverted scenes. However, objects presented in extrafoveal vision must have been processed sufficiently to distinguish between recognisable and unrecognisable objects, on a semantic level, for an advantage to be found for objects considered to be inconsistent with the scene in Experiment 1. Therefore the consistency effect may not be caused by the processing of semantic consistency but by processing difficulty.

However, accepting either of these conclusions is subject to the incorporation of previous research findings supporting the differential processing of consistent and inconsistent objects in scenes. Several studies by Hollingworth and Henderson (1998, 1999, 2000) have evidenced a reliable inconsistent object advantage, most recently by using change detection techniques. These conclusions will be considered and an attempt will be made to integrate the findings of brief presentations and change detection studies.

Chapter 4

Semantic Consistency Effects in Change Detection

4.1 Introduction to change detection

This chapter considers the evidence obtained from change detection studies. Contrary to subjective experience, recent research suggests that we do not accumulate and store all detailed information about visually perceived stimuli. Although we perceive our viewing of the world to be a coherent and cumulative experience, in actual fact we do not construct a veridical representation of the environment during viewing. Rather than storing information obtained through successive fixations, each fixation is largely independent of prior and subsequent fixations. Because of this, large changes to visual stimuli can be made without conscious perception, such as a change in size, colour, presence or identity of an object, under specific circumstances.

Changes occurring during an interruption of the visual process, such as during a blink or saccade, often result in 'change blindness'. Experimental manipulations of these interruptions have included the design of the 'flicker paradigm' in which the two critical images depicting the change are presented alternately, introducing a masking field between the images to prevent apparent motion effects. For a change to be detected in these circumstances, the information relating to the critical object undergoing a change, as viewed in one image, would need to be retained in memory across the mask until the presentation of the second image, at which point the representation of the first image could be compared to the visible representation of the second.

Several researchers have suggested that focussed attention is required in order to detect changes occurring to a visual stimulus across fixations. Rensink, O'Regan and Clark (1997) used the flicker paradigm to determine the stimulus properties which influenced the retention of information across views. Changes to regions of the scene categorised as areas of 'high interest' were detected more quickly than changes to areas regarded of 'low interest'. They interpreted this in terms of attention, assuming that regions of high interest were preferentially attended to, which implied that a participant would need to be directly attending a change to detect it.

The role of eye movements has been investigated in this process. Hollingworth, Schrock and Henderson (2001) monitored eye movements during a flicker paradigm experiment and reported that responses indicating change detection occurred when the participant was directly fixating the target object in 74.5% of trials. Participants responded that they had detected the change when the target object was located in peripheral, rather than foveal or parafoveal, vision in only 5.9% of trials.

This research can be applied to the processing of semantic information from scenes by comparing performance on a change detection task when manipulating target object consistency. Any difference in performance between consistent and inconsistent targets could be attributed to the detection of semantic inconsistency. Hollingworth and Henderson (2000) conducted three change detection experiments using consistent and inconsistent target objects. In all three experiments they found evidence for the faster and more accurate detection of changes to inconsistent compared to consistent target objects, proposing that the inconsistent regions of a scene were preferentially represented across views.

In a subsequent study, Hollingworth, Williams and Henderson (2001) introduced a saccade-contingent change during scene viewing in which the target object, which could be consistent or inconsistent, was replaced during a saccade away from it by a visually different object, within the same category (a token change). Therefore, the target would have been fixated prior to the change and a saccade away from the region would trigger the substitution. Token changes were detected in 35.2% of trials when the target was inconsistent, compared to 18.1% when the target was consistent ($p<.05$). They concluded that inconsistent objects in scenes were better represented over the course of a trial than consistent objects, which is compatible with the long-term memory advantage found for inconsistent objects.

Hollingworth and Henderson's (2000) change detection experiments provided equally robust evidence supporting the facilitation of change detection for inconsistent targets compared to consistent targets for shorter image presentations. This finding implied the detection of semantic inconsistency from brief presentations of scenes. For the purposes of this thesis, an attempt was made to reproduce their findings including an additional manipulation of fixation position, to determine whether the effects of semantic inconsistency observed in the results were influenced by the detection of changes in near foveal or peripheral vision.

Hollingworth and Henderson's series of experiments investigated whether semantic 'informativeness' or consistency could affect the process of change detection. In Experiment 1, the flicker paradigm was used to introduce changes to objects in scene images which were simple line drawings, identical to those used in their previous experiments. Two change conditions were used, the deletion-addition change condition, where the object would appear and disappear across successive views of the scene, and the orientation change condition, where objects were 'flipped' vertically left and right across views, presenting an original and mirror image of the target.

Selected target objects would change as described above in 'change' trials and the same scene image was presented twice in 'no change' trials. The target object in each scene could be an object consistent or inconsistent with the scene context. Each scene image was presented for 250ms, with a 80ms masking period between each image presentation, when a blank screen was presented. The alternating images were presented until the participant terminated the trial by pressing a response button to indicate whether they had detected a change or not. Participants were not required to describe or identify the change if they detected one.

Accuracy was between 85% and 95% correct in the 'change' conditions and over 95% correct in the 'no change' condition, suggesting that responses indicating change detection in 'change' trials were accurate. The mean response time for 'no change' trials was 2257ms, with significantly faster mean responses to 'change' trials, ranging between less than 1200ms and 1500ms (*p*<.001). Although the change conditions were not matched for difficulty, a significant advantage was found for inconsistent targets in both the deletion-addition change condition and the more difficult orientation change condition. Responses were reliably faster to changing inconsistent targets than to changing consistent targets, with response times to consistent trials averaging 1676ms, compared to 1622ms for inconsistent trials (*p*<.05). However, there was only a 23ms difference in response times between consistent and inconsistent trials when no change was occurring, with consistent trial responses taking slightly longer (2268ms and 2245ms). The percentage correct rates also indicated slightly better performance for consistent objects (98.3%) than for inconsistent objects (97.7%), perhaps accounting for the slight increase in response time.

In the orientation change condition, a similar pattern was observed with a 73ms response time difference between consistent and inconsistent trials. Change detection in consistent trials took slightly longer than in inconsistent trials (1500ms and 1427ms). These findings were again reflected in the accuracy rates with 85.4% correct responses for consistent trials and 84.6% for inconsistent trials. The slight increase in response times may have generated the increase in accuracy for consistent trials.

A slightly different pattern was produced in the deletion-addition change condition. Again consistent trials had longer mean response times than inconsistent trials, 1261ms and 1190ms respectively, producing 71ms facilitation for inconsistent trials. However, inconsistent trials showed slightly higher response accuracy than consistent trials (95.0% and 93.3% respectively). The data indicated that inconsistent objects changing in this condition could be detected more rapidly and more accurately than consistent objects. Analysing the two 'change' conditions independently from the 'no change' condition, a reliable effect of semantic consistency was found for the faster detection of changes to inconsistent objects compared to consistent objects (1309ms and 1380ms respectively, $p<.01$) but no reliable effect was found for accuracy.

However, certain criticisms of the experimental design can be raised, the first of which concerns the change conditions. In order for the orientation change condition to be detected reliably, the target objects would need to be clearly asymmetrical across their vertical axis. However, the only example of this condition depicted a coat hanging on a coat rack. The entire object was of a substantial size so a change in the deletion-addition condition would be reasonably salient, but a left-right orientation change would be more difficult to detect. The coat rack itself contained no clear asymmetries and the drawing of the coat included one sleeve slightly longer than the other. The orientation change in this example appeared much less salient than a deletion-addition change, to the point of being largely undetectable. If this object were representative of other targets, the left-right orientation change would result in lower accuracy than expected, for objects which were ambiguously drawn.

A further criticism which has been described previously and is relevant to all the experiments described here remains that participants viewed both a consistent and inconsistent target in each scene background. Each object-scene pair was presented four times, twice in the 'no change' condition, once in the deletion-addition change condition and once in the orientation change condition. Although this design controlled for object location and eccentricity by presenting both consistent and inconsistent objects in the same position in the scene, presenting similar displays to the same participant repeatedly could result in faster responses to subsequent presentations of the same background and object. Participants could perhaps use knowledge obtained during a previous presentation, whether there was a change occurring or not, to direct their search for a change in subsequent presentations.

This hypothesis was supported by analyses conducted by Hollingworth and Henderson to investigate whether performance improved as the experiment progressed. Response times decreased as the experiment progressed and the largest advantage for inconsistent object trials over consistent object trials occurred in the first block out of four (143ms). They acknowledged that the presentation of the same object-scene pairs may have enabled the participants to identify the likely objects changing in each scene, resulting in shorter response times to later trials and reducing the advantage for inconsistent over consistent target trials.

The authors recognised that the inconsistent object advantage evidenced in this experiment could have been explained by participants actively searching for inconsistent objects and fixating them sooner. The contingency relationship between inconsistency and change condition could have been identified and used by participants. When an inconsistent object was viewed in a scene, it would be changing in 50% of trials. Although consistent targets also only changed in 50% of trials, the presence of additional non-target consistent distractors which never changed, in both consistent and inconsistent trials, made the probability of viewing a consistent changing object much less than 50%. This fact would suggest that responses to 'no change' inconsistent trials would be faster and more accurate than responses to 'no change' consistent trials.

If participants viewed an inconsistent object which was not changing, they could prematurely terminate the trial. Viewing a non-changing consistent object would not allow a participant to make any judgement about the correct response and would require continued search. However, the data did not support this hypothesis as the difference in response time between consistent and inconsistent 'no change' trials was only 23ms and accuracy was marginally higher for consistent object trials. If a change was occurring to a consistent or inconsistent object, the trial would be terminated when the participant detected the change, so no consistency effect would be predicted unless consistent or inconsistent objects were fixated sooner, a hypothesis not supported by existing research (e.g. De Graef et al, 1990; Henderson et al, 1999). Therefore the hypothesis that the inconsistent object advantage in change detection was generated by participants searching for inconsistent objects was not entirely supported by the available evidence.

An alternative hypothesis is related to the robust finding that participants fixate inconsistent objects for longer than consistent objects. In order to detect a change, participants would need to attend to the critical object during one scene presentation and again after the intervening mask and during the presentation of the alternate image. Therefore, a natural tendency to fixate inconsistent objects for longer might increase the probability of maintaining fixation across a mask and modulate the faster change detection for these objects compared to consistent objects in scenes.

Hollingworth and Henderson's Experiment 2 presented each alternative image only once, rather than cycling them until the participant terminated the trial, to control for potential differences in the allocation of overt attention to consistent and inconsistent object regions. The deletion-addition change condition was removed and only the orientation change condition was implemented. The first scene image was presented for 250ms, followed by a 30ms mask and then the second scene image, which would be identical to the first in 'no change' trials or contain a left-right orientation change of the target object in 'change' trials. This second scene remained visible until the participant responded.

In order for the participant to detect a change in this change condition, information relating to the target region would need to be encoded to include the object's orientation during the first 250ms image presentation, retained across the mask and compared to the object's orientation in the second image. If the two representations were identical, the participant could respond that no change had occurred. However, if the two representations illustrated an orientation change, the participant could respond that a change had occurred. Therefore, this paradigm investigated any discrepancy between the processing of consistent and inconsistent object information during a brief 250ms presentation and the ability to retain the relevant information across a mask.

Accuracy for 'no change' trials in this experiment was approximately 78% for both consistent and inconsistent trials. This level of accuracy indicated that a substantial number of false alarms were reported, with participants believing they had detected a change when none had occurred. The decrease in accuracy compared to the previous experiment could be explained, as the change was originally difficult to detect and would be even more so from only two image presentations and if the target object was not clearly asymmetrical along its vertical axis.

For trials in which an orientation change occurred, a reliable consistency effect was found, with better performance for inconsistent object trials compared to consistent object trials, 54.7% and 49.5% respectively ($p<.005$). This result was also reflected in response times, with changes detected more rapidly to inconsistent objects (1321ms) than to consistent objects (1468ms) ($p<.005$). The data indicated a clear 147ms advantage for the processing and retention of information relating to inconsistent objects over consistent objects.

This evidence confirmed the conclusions of the previous Experiment 1 that semantic consistency influenced performance on change detection, facilitating inconsistent objects. However, accuracy rates indicated a substantial proportion of incorrect responses. Almost 1 in 4 trials were inaccurately reported to have contained a change, which suggested that a proportion of correctly detected changes could be attributed to guessing. In this way, accuracy levels for correctly detected changes might have been an overestimation of genuine change detection.

Also, the hypothesis that the results were influenced by participants selectively attending to inconsistent objects compared to consistent objects was compatible with the results obtained. An inconsistent object was likely to change with a probability of 50%, compared to the lower probability of change in any one consistent object, including distractors in the scene. When viewing a scene, participants could have directed their efforts to inconsistent objects according to their increased likelihood of change.

Hollingworth and Henderson attempted to control for this in Experiment 3 by modifying the scene components. A consistent distractor which would sometimes be the changing object was included in each scene, so that every scene contained at least one consistent object which could change. This would make it more difficult for participants to predict which object would be likely to change, even from previous presentations of the same scene background. A second modification involved the inclusion of an inconsistent distractor in consistent scenes, which previously contained only consistent objects, although this inconsistent distractor never changed orientation. Each scene contained an inconsistent object so that the probability that an inconsistent object viewed in a scene would be the changing object was reduced.

Therefore, in consistent trials, there were at least three objects in each scene, only two of which could change, the consistent target object and the additional consistent distractor, with a non-changing inconsistent distractor. In inconsistent trials, there were again two objects which could possibly change, the target inconsistent object and the consistent distractor. If participants were selectively attending to inconsistent objects in scenes, performance for consistent objects would decrease, as participants would fail to detect changes to consistent targets, with an inconsistent distractor included in every scene. Even if participants responded perfectly to all changes to inconsistent objects, these were only one quarter of all change trials (one inconsistent target change condition, one consistent target change conditions and two consistent distractor change conditions). In the remaining three quarters of change trials, an inconsistent object appeared but was not the changing target. In addition, there were an equal number of 'no change' trials so the probability of a viewed inconsistent object changing across views dropped from 50% in the previous experiment to 12.5% (1 in 8).

Accuracy for the 'no change' condition was approximately 84.3% correct, compared to between 36% and 45% correct for the orientation change condition. The data indicated high accuracy for 'no change' trials and lower accuracy for 'change' trials, which would possibly support the proposal that the higher accuracy in change conditions in Experiment 2 was offset by the lower accuracy for 'no change' trials. A reliable 8.7% difference in accuracy was observed between consistent 'change' trials and inconsistent 'change' trials ($p<.05$). As a response bias was found, with participants more likely to respond that a change had occurred when the trial was consistent, a non-parametric measure of sensitivity, *A'*, was calculated. The significant inconsistent object advantage in accuracy increased for the *A'* measures for consistent object trials (0.68) over inconsistent objects (0.76) (p<.005). The response times evidenced a non-significant difference, with faster mean times for inconsistent trials than consistent trials, 1422ms and 1535ms respectively. These results provided convincing evidence of a disparity between the processing and retention of information relating to consistent and inconsistent objects at durations as short as 250ms.

Further conclusions could be reached by analysing these data. The inclusion of additional objects in the scenes appeared to reduce accuracy in both consistent and inconsistent trials. Although the experimental procedure was identical in Experiments 2 and 3, the addition of both consistent and inconsistent distractors modulated a decrease in accuracy from 49.5% to 36.3% for consistent trials and from 54.7% to 45.0% for inconsistent trials. The results for Experiment 3 would have been even less accurate if three of the original participants with accuracy rates below 60% had not been replaced. The inclusion of additional distractors, even though the background remained identical, affected performance on this task, suggesting that the number of objects present in a scene, a measure of scene complexity, affected accuracy. This result has repercussions for the comparison of data from experiments using different stimuli, indicating that the complexity of scenes would require careful analysis before comparing results across experiments.

The authors proposed possible explanations for an inconsistent object advantage, which have been discussed previously, including the memory schema hypothesis, the attentional attraction hypothesis and the attentional disengagement hypothesis. The memory schema hypothesis proposes that semantically consistent objects are represented as normalised items in a scene representation, while inconsistent objects are retained in a more detailed veridical representation. This assumption predicts better memory for details of inconsistent objects rather than consistent objects. Hollingworth and Henderson (2000) argued against this hypothesis, as a subsequent experiment

(Hollingworth and Henderson, submitted) indicated that the inconsistent object advantage was not influenced by inter-stimulus interval, as would be predicted.

The attentional attraction hypothesis proposes that covert attention can be drawn to regions of conceptual difficulty in a scene. As there is no reliable evidence of inconsistent objects being fixated earlier than consistent objects in scenes, covert attention would need to be dissociated from overt attention for covert attention to be preferentially attracted by regions of semantic inconsistency. Also, the region of conceptual difficulty would need to be identified within the 250ms presentation of the first scene image.

Finally, the attentional disengagment hypothesis proposes that covert attention is deployed to regions of interest based on visual features and, although not automatically drawn to regions of semantic or conceptual inconsistency, attention may be captured once fixated, resulting in a more detailed representation of an inconsistent object than a consistent object. This hypothesis is compatible with research indicating that inconsistent objects are not initially saccaded to sooner than an adequately controlled consistent object, but are fixated more often and for longer than control objects. However, for this proposal to explain the results of Hollingworth and Henderson's (2000) experiments, the inconsistent targets would need to have been fixated during the 250ms presentation of the first scene image.

These experiments provided rare evidence of an inconsistent object advantage in scene processing, during brief presentation durations through the application of change detection paradigms. It was considered important to investigate whether the findings were robust and replicable, perhaps with more realistic stimuli rather than simplistic line drawings. Additionally, the absence of information relating to target object eccentricity or participant fixation position made it impossible to determine whether facilitation for inconsistent objects could have been generated by foveal or extrafoveal processing of semantic information. The following experiments modified the procedures, to be more appropriate for use with photographic stimuli and to include fixation control, in an attempt to replicate evidence of an inconsistent object advantage in the detection of changes in complex, naturalistic scenes.

4.2 Experiment 5: Introduction

The purpose of this experiment was to replicate the findings of Hollingworth and Henderson's (2000) Experiment 2. They evidenced an inconsistent object advantage in a two-exposure change detection task, where participants viewed each version of a scene image once only and then made a two-alternative forced-choice decision as to whether a change had occurred or not. Even under these limited viewing conditions, a significant advantage was found for change detection to inconsistent objects compared to consistent objects.

In the current experiment, the stimuli set from Experiment 3 was used, consisting of grey scale photographs of natural scenes. A comparable two-exposure change detection paradigm was implemented, with the intention to provide evidence of an inconsistent object advantage using complex photographic stimuli. In addition, a further variable was manipulated. Hollingworth and Henderson's paradigm did not consider whether the inconsistency between the scene and the object could be discriminated from foveal or extrafoveal processing. Although there was a significant difference in performance for consistent and inconsistent objects, without controlling for fixation position it was not possible to conclude whether semantic inconsistency was detected from extrafoveal vision. The current experiment aimed to conclusively determine whether semantic processing in extrafoveal vision was possible, by manipulating the participant's fixation position relative to the target object.

4.3 Method

Participants

A total of 100 participants were recruited for this experiment, of which 27 were male and 73 were female. They were all undergraduate students at the University of Durham, Queen's Campus and were naïve to the purposes of the experiment. Participants were recruited with the opportunity to win £10, awarded to 5 participants at the end of the experiment.

Apparatus

This experiment was run on open-access university PCs, similar to those used in Experiments 1 to 4, with a processor speed of 266MHz and image resolution of 800 x 600 pixels. Participants were provided with an information sheet and a consent form at the start of the experiment and a debriefing sheet upon completion.

Materials

The stimuli were the photographs originally constructed for Experiment 3. For each scene image containing a target object, there was an alternative image without the target, as photographs were obtained of the image backgrounds with the target object removed. Therefore, for each photograph with a consistent or inconsistent object embedded in it, there was a corresponding photograph which was identical except for the removal of the target object. In this way, it was possible to use the existing photographs in this experiment, by displaying the scene containing the target object, followed by the scene without the target, to mimic the disappearance of the target object as the 'change'.

Trials were organised into four possible sets, to allow a maximum number of images to be displayed without any participant viewing the same scene background more than once. The specific trials assigned to each set were checked carefully to also ensure that a participant did not view the same target object twice, even in a different scene. This was controlled for because an object seen to disappear in one trial, subsequently viewed again in another scene, could be subject to different processing compared with a novel object.

In addition to the 16 experimental trials in each of the four trial sets, a further 16 photographs were used as 'no-change' catch trials, making a total of 32 trials in each set. Some of the 'no change' images had been used as practice trials in Experiment 3. These images were supplemented with additional photographs to construct a set of 16 images suitable for use as 'no change' trials. These 'no change' photographs were designed to be indistinguishable from the 'change' trials, as they also depicted household scenes and some of the images had previously been rejected for use as experimental trials. For example a background scene, without the selected target object, which had been rejected for use as an experimental scene on the grounds that the object was difficult to identify, was used as a 'no change' trial.

Within this set of 16 'no change' catch trials, an object which could be considered inconsistent with the scene context had been placed in half of the scenes. If all 'no change' trials contained only consistent objects, the presence of an inconsistent object could have been associated with an increased probability of a change having occurred, as all inconsistent objects in scenes would have undergone a change. By adding an inconsistent object to half of the 'no change' trials, the presence of an inconsistent object failed to be diagnostic of a 'change' trial, although there was still an increased probability that an inconsistent object viewed in a scene was the changing target, compared to consistent targets and distractors.

Every scene contained a pre-selected target object which could be consistent or inconsistent, and fixation positions were selected in relation to this target. As the target was not changing in the 'no change' catch trials, fixation position was not manipulated in relation to the target object. One fixation position was selected per scene and all participants viewed the same catch trials with the same fixation position. The fixation positions were selected to correspond to an object in the same proportion of trials as experimental trials, to prevent them being distinguished in any way.

Finally, a total of eight additional photographs, which had not been used before, served as practice trials. These photographs were designed to resemble the experimental trials; four trials contained a change, with two consistent targets and two inconsistent targets changing, and four trials were 'no-change' trials, two containing an inconsistent object. These photographs were not used in the experimental trials.

Design

One of the independent variables in this experiment remained the distance between the participant's fixation position when the scene was presented for the first time and the location of the target object, when a change was to occur. As the scenes were presented for longer in this experiment than in the previous experiments (1 - 4), fixation position could only be confirmed for the first image presentation, which was the time of most interest. The same locations for fixation positions within the scene image were used as in Experiment 3. Similarly, the consistency between the scene and target object was again manipulated. Additionally, the change condition itself, whether a change occurred or not, was a third independent variable. The dependent variables of interest were the participant's accuracy in detecting changes and their response latencies.

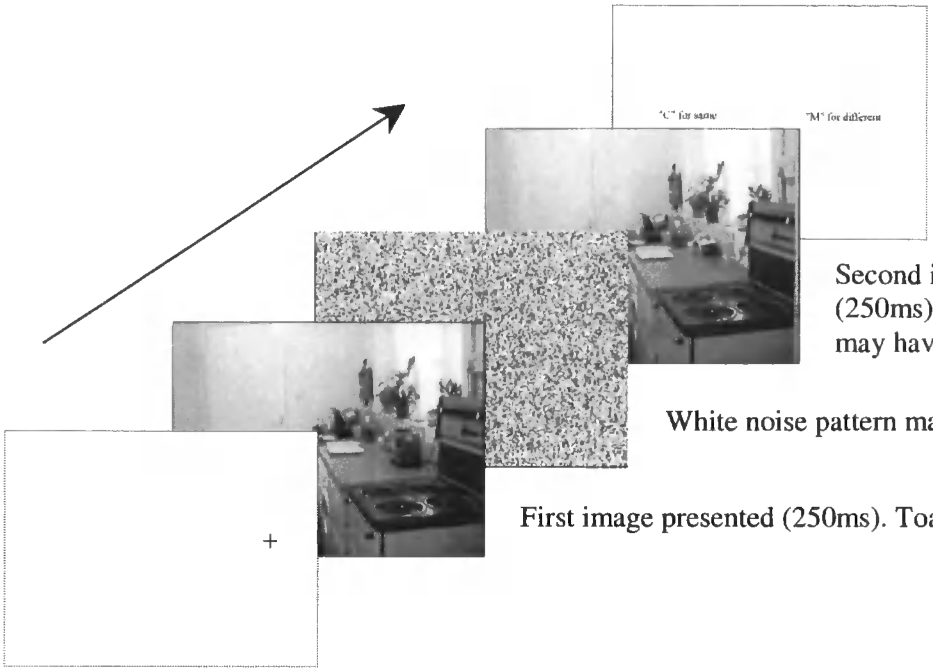
The presentation of the experimental trials was randomised independently for each participant. As participants reported whether a change had occurred or not, rather than selecting between two different objects with possibly slightly different probabilities of being present in the scene, their responses were not subject to the same possible response biases as in Experiments 1 to 4. In this experiment, participants were instructed to make a specific response if they had detected a change, making the issue of interest the extent to which they were sure they had detected a change before responding positively.

Procedure

Participants were provided with an information sheet and a consent form and were able to complete the experiment themselves without requiring further assistance, although the experimenter was available throughout to respond to any queries. The instructions emphasised that they should seat themselves at an arm's length from the screen (approx 60cm). As the viewing distance was only approximate, the size of the target objects and scenes in degrees of visual angle could only be estimated. The participants were instructed to prepare to detect a change, which was described as a major change in the composition of the scene, consisting of a single object appearing or disappearing, although actually the target object always disappeared rather than appeared in the scene. In this way, participants were discouraged from responding positively to any minor change they perceived between the two presented scene images.

Each trial, both in the practice trials and experimental trials, followed the procedure summarised in Figure 4.1. A fixation cross was presented for 1000ms and participants were instructed to fixate it. They were informed that if a change did occur, this fixation cue would sometimes indicate where the target object would appear. Then a scene image was presented for 250ms. If the trial was a 'change' trial, the scene would include either a consistent or an inconsistent target object. If the trial was a 'no-change' trial, then the scene would contain either a consistent or an inconsistent object which did not change and the first and second scene presentations would be identical.

A white noise mask was presented for 80ms after the first image, followed by the second image presentation for 250ms. In a 'change' trial, the second image would be the corresponding background scene with no target object, resulting in the disappearance of the target object between the first and second image presentations. Immediately after the second presentation, participants were reminded to press the 'M' key on the keyboard if they had detected a change and 'C' if they had not. There was an inter-trial interval of 1000ms before the next trial began. When the practice trials were completed, participants were instructed to begin the experimental trials which followed the same procedure.



Participant selects whether a change did occur (different) or did not occur (same).

Second image presented (250ms). Target object may have been deleted.

White noise pattern mask (80ms).

First image presented (250ms). Toaster is target object.

Fixation cross presented to direct fixation (1000ms).

Figure 4.1: Procedure for practice and experimental trials.

4.4 Results

Performance for each individual participant was investigated by calculating the accuracy for correctly detecting changes (hit rate) and for falsely detecting a change when none occurred (false alarm rate). In order to determine that participants were performing better than chance, the hit rate of each participant was plotted against their false alarm rate in Figure 4.2. If a participant were pressing response buttons randomly, then the hit rate would be approximately equal to the false alarm ra [illegible] Figure 4.2. The presentation of the data in this way allowed the investigation of whether higher hit rates were associated with a tendency to respond positively to the change detection task in general, regardless of whether a change had been detected, resulting in a higher false alarm rate.

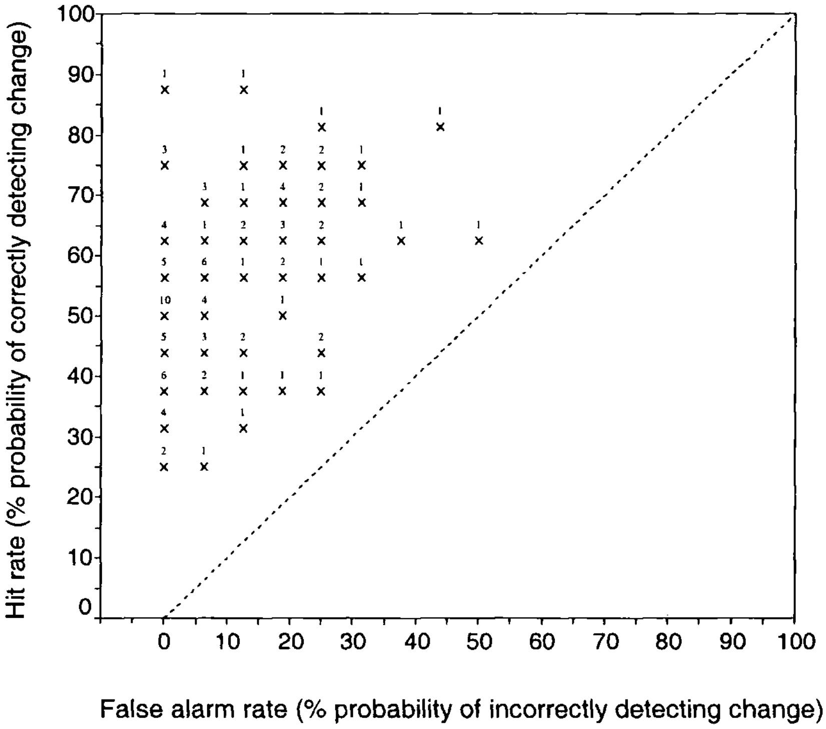


Figure 4.2: Scatterplot of hit rate by false alarm rate.
The number above each data point indicates the number of participants represented by that point. The reference line indicates chance performance where hit rate equals false alarm rate.

As all points in the graph were above the reference line indicating chance level, the hit rate was always higher than the false alarm rate and often by a large margin, indicating that participants were correctly detecting changes more often that incorrectly reporting them. In fact, 40% of participants reported no false alarms at all, always correctly detecting that no change had occurred. These data indicated that when participants reported having detected a change, this response was relatively accurate and unlikely to be a random guess.

Therefore, chance performance on this same/different task would not be 50%, as participants were using reasonably strict criteria from which to determine whether a change had occurred. The mean hit rate across participants was 54.6%, indicating that they were correctly detecting a change when one occurred in more than half of the trials across all fixation positions. By comparison, the mean false alarm rate was approximately 9.9%, showing that participants incorrectly reported a change in fewer than 10% of the 1600 'no change' trials.

Distance of the target object from fixation

The following analysis investigated how the manipulation of fixation position influenced the accuracy of participants in detecting changes. Figure 4.3 illustrates how accuracy decreased when fixation position was manipulated. When the directly fixated target object changed, accuracy was highest at over 85%, falling to 30% when the target object changed furthest from the participants' fixation position. This 30% accuracy was still impressive, considering that the change occurred far in extrafoveal vision and that the false alarm rate was 10%. This level of performance indicated that, even when participants were fixated far away from the changing target during the first image presentation, they could still detect a change reliably better than chance. The decrease in accuracy according to fixation position was found to be statistically significant by a binary logistic regression analysis ($\chi^2(1)=278.6$, $p<.001$).

Consistency of the target object

No significant main effect was found for consistency ($\chi^2(1)<1, p=.48$). Changes to consistent targets were detected correctly on 55.5% of the trials, compared to 53.9% for changes to inconsistent targets.

Interaction of fixation position and consistency

Figure 4.4 illustrates accuracy according to the consistency of the target object and the participants' fixation position. Although accuracy for consistent changing targets was higher than for inconsistent changing targets when presented directly at fixation, there was no reliable effect of consistency at further fixation positions. No reliable interaction between consistency and fixation position was found ($\chi^2(1)<1, p=.86$).

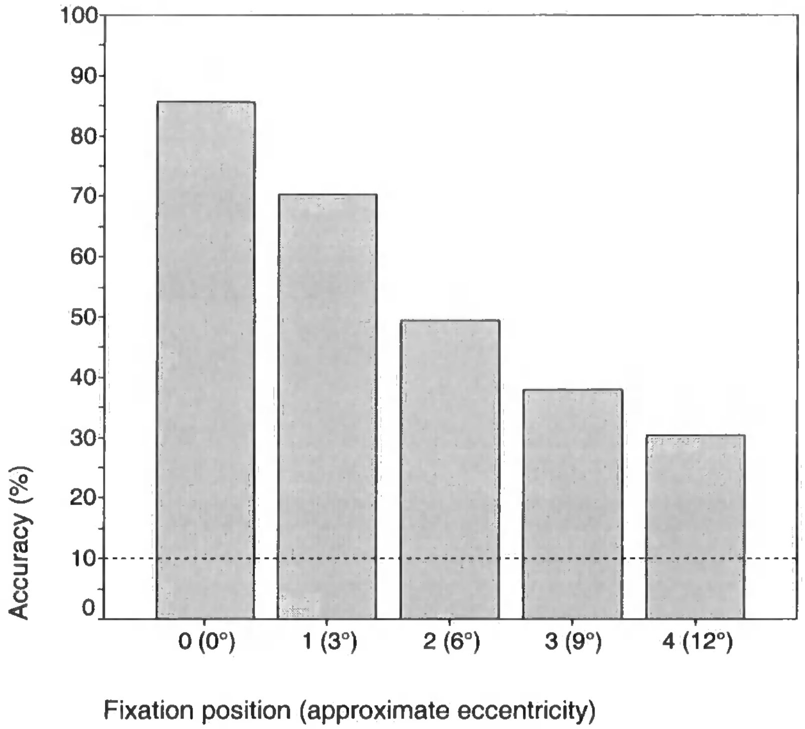


Figure 4.3: Graph showing the change in accuracy as distance between the target object and participants' fixation increases to 12°. False alarm rate indicated.

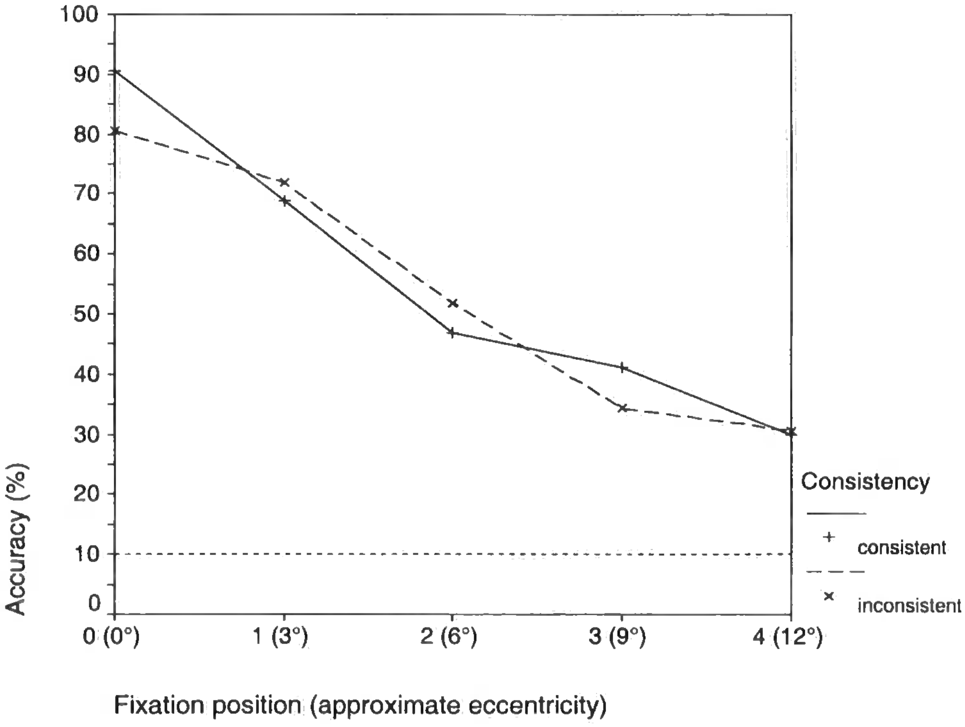


Figure 4.4: Graph showing the change in accuracy by fixation position and target object consistency. False alarm rate indicated.

When the target object changed at the same location as the participants' fixation position, accuracy was 10% higher for consistent objects than for inconsistent objects. This difference was analysed using a binary logistic regression analysis and found to be statistically significant ($\chi^2(1)=6.62$, $p=.010$), indicating that changes to consistent targets were significantly more likely to be detected than changes to inconsistent targets when the participant was directed to the target object. No significant differences were found between consistent and inconsistent trials at any other fixation position. Even at the furthest fixation position, accuracy at detecting changes remained over 20% above the false alarm rate.

Counterbalanced objects

The design of the experiment allowed the comparison of both performance for the same object according to the consistency of the scene it had been presented in and performance for the matched consistent and inconsistent targets in the same scene background. As the data met the assumptions for a parametric test, with approximately normal distributions, no outliers and equal variances, a matched samples t test was conducted to investigate whether consistency influenced accuracy. The analysis found no significant difference between performance for the same objects in consistent and inconsistent scenes ($t(31)<1$, $p=.76$). Of the 32 target objects, 19 displayed better performance in the consistent scene, 12 showed better performance in the inconsistent scene and one target showed no difference at all.

The data were also analysed to determine whether there was any difference in performance for the matched consistent and inconsistent objects appearing in the same scene backgrounds. As the data were distributed approximately normally, with no outliers and equal variances, a matched samples t test was conducted. Again, no significant difference was found ($t(31)<1$, $p=.61$). Of the 32 object pairs, 17 showed an advantage for the consistent object in the scene, nine displayed an inconsistent object advantage and six showed no difference at all.

Response times

No effects of consistency were expected in the participants' reaction times. Fixation position influenced the speed of response, with faster responses when changes occurred closer to fixation. Also change condition (change or no change) was expected to affect response times, producing longer response times for 'no change' trials than for 'change' trials. The data were tabulated to investigate these predictions (Table 4.1).

Table 4.1: Summary table of mean response times (in ms) by fixation position and target object consistency for correct trials only, including mean response times for 'no change' trials.

	0 (0°)	1 (3°)	2 (6°)	3 (9°)	4 (12°)	Mean	No change mean
Consistent	726	786	891	851	973	814	911
Inconsistent	750	747	883	926	941	819	916
Mean	738	766	887	885	957	816	914

Response times appeared to increase at further fixation positions but no reliable effect of consistency was found, either overall or in interaction with fixation position. Consistent targets were responded to faster than inconsistent targets at fixation positions 0 and 3, with inconsistent targets producing faster responses at fixation positions 1, 2 and 4. As expected, 'no change' trials produced longer response times than the means for change trials. A univariate ANOVA by participants of the data confirmed these conclusions, finding a significant effect of fixation position ($F(4,639)=5.90, p<.001$) with trials in which the target object was presented closer to fixation producing faster response times than trials in which the target object appeared further from fixation. However, no significant main effect was found for consistency ($F(1,639)<1, p=.55$) and no significant interaction was found between fixation position and consistency ($F(4,639)<1, p=.67$).

Object size

The target objects in this experiment were retained in the same categories as identified in Experiment 3. The scene images used in both experiments were identical and they were displayed at the same resolution on equivalent monitors. However, as the viewing distance in this experiment was only approximate, the size of the target objects in visual angle could not be calculated accurately. Objects classified as small had pixel areas less than 4,000. Similarly, medium sized objects ranged between 4,000 and 8,000 square pixels in size, with large objects having areas greater than 8,000 pixels.

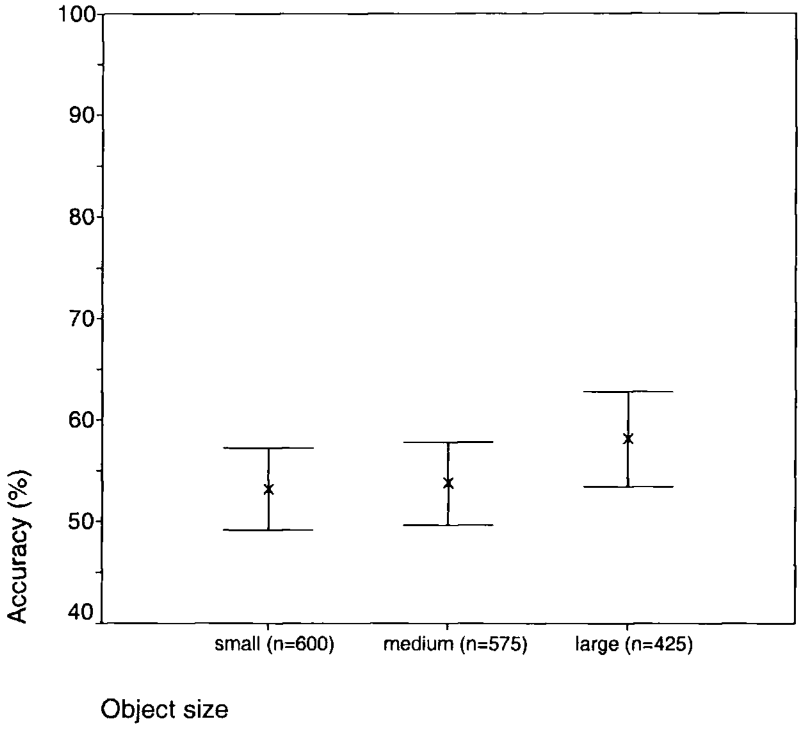


Figure 4.5: Error plot indicating mean accuracy and error bars (95% confidence intervals) for different object sizes.

Figure 4.5 indicates that the manipulation of object size had less effect on accuracy in this experiment than in previous experiments. While an overall increase in mean accuracy was found for large objects compared to small and medium sized objects, this was not large enough to be statistically significant. Overall accuracy was less than in the brief presentations experiments, but was still over 50% correct for changes to small objects, even though this measure was averaged across all five fixation positions. Using a binary logistic regression analysis, a significant main effect of fixation position was found ($\chi^2(1)=278.6$, $p<.001$) but object size was not statistically significant ($\chi^2(2)=3.02$, $p=.22$). No significant interactions were found between the variables.

The data were tabulated in Table 4.2. A slight 2.74% advantage was found for small consistent targets compared to small inconsistent targets, although this difference was not statistically significant ($\chi^2(1)<1, p=.50$). The same pattern was seen to a greater extent for medium sized objects, with a 6.82% difference between accuracy for consistent and inconsistent targets, but this difference was also non-significant ($\chi^2(1)=2.68, p=.10$). The pattern was reversed for large objects which showed a 7.58% advantage for inconsistent objects over consistent objects, a difference which again failed to reach statistical significance ($\chi^2(1)=2.50, p=.11$). To summarise, consistent targets which were small or medium in size produced slightly higher accuracy than comparably sized inconsistent targets, while large targets displayed the opposite trend of slightly higher accuracy for inconsistent targets than consistent targets, although none of these trends reached statistical significance.

Table 4.2: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	54.74	57.00	54.42	55.50
Inconsistent	52.00	50.18	62.00	53.87
Mean	53.26	53.74	57.98	54.69

4.5 Discussion

The results of this experiment indicated that participants performed the task accurately and rarely claimed to detect a change when none had occurred. The accuracy of several participants was especially impressive, with no false alarms being reported and accurate detection of genuine changes. As expected, accuracy decreased when participants were directed to fixate regions of the scene further away from the changing object but even at the most extreme fixation position, approximately 12° from fixation, accuracy was still above 30% on average. As false alarm rates were below 10%, the hit rate detection values were an accurate reflection of the proportion of changes genuinely detected by participants.

There was no evidence to suggest a difference in change detection for consistent and inconsistent targets. An analysis of accuracy according to fixation position and consistency found a significant advantage for consistent targets compared to inconsistent targets when presented at fixation (*p*=.01), as had been found in previous experiments, but no reliable effect at any other fixation position. Investigating performance for matched targets also failed to provide evidence of a consistency effect. No difference was found between the same object in a consistent and an inconsistent scene, or matched consistent and inconsistent objects in the same scene.

Response times did not show any effect of consistency but increased as the distance between the participant's fixation position and the target object increased. When the participant was directed to fixate the target object, response times were fastest, with delayed responses for targets presented in extrafoveal vision. No interaction between the effects of consistency and fixation position was found.

Object size was also investigated to compare accuracy for small, medium and large objects. No significant differences in performance were found across objects sizes in this experiment, unlike the previous brief presentation experiments. This result could suggest that the disappearance of an object was a salient enough visual event for object size to fail to significantly influence its detection. Also, no effect of consistency was found on accuracy for targets of any size.

Comparing results to Hollingworth and Henderson's (2000) Experiment 2

The results of this experiment can be directly compared to those of Hollingworth and Henderson's (2000) Experiment 2. To begin with, accuracy in detecting that no change had occurred was approximately 78% in Hollingworth and Henderson's experiment, compared to accuracy rates of approximately 90% for the current experiment. This discrepancy suggested that the change detection response data in their Experiment 2 could have included a significant proportion of guesses.

Hollingworth and Henderson found a statistically significant difference in accuracy between detecting changes in orientation occurring to consistent and inconsistent objects in scenes. Change detection was more accurate when the target object was inconsistent, with an accuracy of 54.7% compared to 49.5% for changes to consistent objects ($p<.05$). A significant effect was also found for response times, with faster responses to inconsistent trials (1321ms) than to consistent trials (1468ms) ($p<.005$). Response times in the current experiment, with a mean of 816ms for 'change' trials, were much shorter than those for Hollingworth and Henderson's experiment and failed to exhibit an advantage for inconsistent objects.

In the current experiment, detection rates were over 50% for both consistent and inconsistent trials but no significant difference was found. A number of differences were identified between the two experiments and these will be considered in further detail to investigate possible explanations for the differences in results. Several differences were related to the experimental procedure.

To begin with, the current experiment introduced a fixation cue which was manipulated to determine whether performance varied according to fixation position. This manipulation allowed the investigation of both foveal and extrafoveal processing of semantic information. As Hollingworth and Henderson did not manipulate fixation position, it was not possible to determine whether the advantage for inconsistent objects was the result of foveal or extrafoveal processing of the object's semantics.

If the eccentricity of the target's presentation relative to the participants' fixation position was restricted, an advantage for consistent or inconsistent targets could be explained by facilitation in foveal or extrafoveal vision. For example, a reliable consistent object advantage was found in several experiments when the target was presented at fixation. A similar effect facilitating inconsistent targets may have existed within limited extrafoveal eccentricities when using simplified line drawings. In the absence of information relating to the eccentricity of presentation of the target objects in Hollingworth and Henderson's experiment, this possibility cannot be excluded.

The change used in the current experiment was the disappearance of the target object, compared to the left-right orientation change used in Experiment 2 by Hollingworth and Henderson. In their previous experiment (Experiment 1), Hollingworth and Henderson found that differences in change detection for consistent and inconsistent objects were more pronounced for a change in object orientation than in the alternative deletion-addition change condition. However, a reliable effect of consistency was still found for the deletion-addition condition, so the application of a different change condition in the current experiment cannot explain why an inconsistent object advantage was not found. This deletion-addition change condition was selected as the experimental stimuli were suitable to create this manipulation.

The presentation of a white noise mask in the current experiment also differed from the pattern mask used by Hollingworth and Henderson. A pattern mask was most appropriate for the line drawing stimuli they employed but the use of photographic stimuli in the current experiment required a more complex mask, such as a white noise mask. The mask was presented for 80ms to match the mask duration in Hollingworth and Henderson's Experiment 1. The longer mask duration of 80ms was employed in the current experiment, rather than the shorter duration of 30ms used in Experiment 2, as there was concern that 30ms could be too brief to prevent apparent motion effects in the scenes viewed.

A final difference between the experimental procedures was that, in the current experiment, the second scene image was presented for 250ms only, like the first scene image, after which participants viewed a display reminding them of the response buttons on the computer keyboard ('C' for 'no change' and 'M' for 'change'). However, in Hollingworth and Henderson's Experiment 2, the second scene image remained visible until the participant responded. This was considered unsuitable for the photographic stimuli, as scene viewing would be more extensive than for line drawings if the participants performed a search of the second image, looking for a possible change. This increase in viewing time could interfere with the response of interest, which was the perception of whether a change had occurred, without encouraging extensive consideration. Participants were instructed to respond according to an immediate reaction or a 'gut response' of whether a change had occurred, without needing to report the nature of the change. It was hoped that accurate data could be obtained on whether participants had been aware, either consciously or otherwise, of a change which could have occurred far from their fixation position.

This difference in procedures may have modulated the decrease in mean response time in the current experiment compared to Hollingworth and Henderson's experiment. Participants did not view the second scene image indefinitely and would have had nothing to gain from delaying a response, so more rapid responses were obtained. As the second image was not present for inspection, participants in the current Experiment 5 responded much sooner than those in Hollingworth and Henderson's experiment.

It is also worth noting that accuracy in the current experiment was at least equal to that in Hollingworth and Henderson's experiment, with equal detection rates and fewer false alarms, indicating that the shorter presentation of the second image did not decrease accuracy significantly. The use of complex photographic stimuli, containing a greater number of possible targets, could also have been expected to detrimentally affect accuracy but clearly did not. This comparison also suggested that the faster responses found in the current experiment were not associated with a relative decrease in accuracy. Admittedly a more salient change condition was used in the current experiment but this was balanced by a shorter presentation time of the second image and the use of complex photographic stimuli.

The experimental differences described so far were unlikely to have modulated the absence of an inconsistent object advantage in the current experiment. Hollingworth and Henderson interpreted their results as evidence that the information relating to inconsistent objects processed during the first 250ms presentation could be preferentially retained across a mask. For participants to be better able to detect a change to an inconsistent object's orientation upon the presentation of a second altered image, information relating to the inconsistent target would need to be processed during the first 250ms image presentation. An inconsistent object facilitation would be required in either the processing of information in the first presentation of the scene or in the retention of this information over brief periods of time (less than 1000ms), to generate an inconsistent object advantage in change detection. Therefore, the experimental differences described previously such as the nature of the change, the type of mask used and the presentation time of the second image could not have affected the detection of changes, occurring as they did subsequent to the critical encoding of inconsistent object information during the first presentation.

This consideration leaves the nature of the stimuli themselves as an explanation for the failure to find an inconsistent object advantage. As argued previously, the photographic stimuli were more complex than line drawings but performance in both this current experiment and Experiments 3 and 4 indicated that performance did not suffer because of this, possibly because the objects in the photographs were easier to recognise. It would seem unlikely that the nature of the stimuli as photographs, rather than line drawings, would have caused the differences in results between the current experiment and that of Hollingworth and Henderson.

Another possible explanation could be that the change detection task encouraged recall of the objects and the scene as a whole, which could be especially true if the participants were allowed to observe the second image for an unlimited period of time. In contrast, Experiments 1 to 4 could be performed simply by recalling reasonably vague details relating to a single object. If participants were consciously attempting to attend to all scene objects and encode a veridical representation of the image, which would be particularly necessary to detect an orientation change and less so to detect the deletion of an object, the increased visual information in the photographs may have interfered with this process.

Using simple line drawings could facilitate change detection, as the visual scene as a whole would be simplified and recalling the information upon the extended inspection of the second scene image would also have been easier. This assumption could indicate that the results obtained by Hollingworth and Henderson were limited to simplistic scenes containing few component objects. Although no evidence was found to support the enhanced detectability of line drawings of objects in scenes compared to photographs of objects in scenes in Experiments 1 to 4, the comparisons made only considered a paradigm involving object identification from brief presentations rather than change detection.

The following experiment was designed to replicate Hollingworth and Henderson's Experiment 1, in which an object deletion-addition change condition was implemented. As Experiment 1 involved the repeated alternation between the two scene images until the change was detected, the dependent variable of interest was the response time, which could be more robust to semantic influence according to the scene stimuli. The processing of information from extrafoveal vision was not directly investigated but the eye movements of participants were recorded while they carried out the task, in order to conduct an investigation of not only the duration of time required to detect the change but the pattern of saccades which occurred during the trial.

4.6 Experiment 6: Introduction

This experiment attempted to replicate the deletion-addition condition of Hollingworth and Henderson's (2000) Experiment 1, which found an inconsistent object advantage in speed of change detection. In Hollingworth and Henderson's experiment, the standard flicker paradigm was applied, alternating two scene presentations until the participant terminated the trial. There were two change conditions, a deletion-addition condition and a left-right orientation change condition. In both conditions, changes to inconsistent targets were detected more rapidly than changes to consistent targets.

In the current Experiment 6, the photographic stimuli used in the previous two-exposure change detection paradigm were applied to the cycling scene paradigm. These images recreated a deletion-addition change condition only, with the target object appearing and disappearing across subsequent views of the two test images. Two images of the same background scene were displayed for each trial, only one of which contained the target object. The semantic relationship between the target object and the scene background was manipulated.

The current experiment measured response time to detect the deletion-addition change from cycling experimental images, as did Hollingworth and Henderson's Experiment 1, but additionally monitored participants' eye movements while they performed this task. The response time indicated when participants responded to the change they detected, if present. By monitoring eye movements, it was also possible to determine whether eye movements were affected by the 'flickering' of the target object and to calculate the exact time at which participants fixated the target object. This information would indicate an accurate 'search time' for the detection of the target, without the additional time to respond inherent in manual responses. In this way, it was possible to investigate whether consistency in scenes influenced eye movement behaviour and if saccade patterns obtained for trials in which an inconsistent object was present were affected by whether the object was changing or not.

4.7 Method

Participants

A total of 18 naïve participants were recruited from the undergraduate population of the University of Durham to participate in this experiment. Two of these were male and 16 were female.

Apparatus

Eye movements were recorded using a Fourward Technologies Dual Purkinje Generation 5.5 eye tracker. The resolution of the eye tracker was 10min of arc and the sampling rate was every millisecond. The monitor and the eye tracker were both interfaced with a Philips Pentium III PC that controlled the experiment. The movements of the right eye were monitored but viewing was binocular. Head movements were restrained with a chin rest and two forehead rests.

Each session began with a calibration procedure in which the participant was required to fixate sequentially nine small points, which were arranged in a centrally presented rectangle, marginally larger than the experimental image size. The scenes were presented on a monitor with a visible screen size measuring 38cm by 28.5cm. This subtended approximately 26° by 20° at the selected viewing distance of 85cm. The scenes themselves subtended approximately 16° by 12°.

The accuracy of the eye position measure was checked after every four trials and, if necessary, a new calibration phase was conducted. Manual responses were obtained using a button box unit, similar to those used in Experiments 1 to 5. The eye movement data were analysed off line by a semi-automated procedure. A computer algorithm detected the saccades using a velocity criterion and each record was inspected individually.

Materials

All scene images measured 640 by 480 pixels and were presented centrally on a white background measuring 1024 by 768 pixels. The scene area subtended approximately 16° by 12° in visual angle. The practice trial images used in this experiment were those used in Experiment 5, as were the scene images. The experimental images were derived from the 64 photographic scene-object images designed for Experiment 3. For each image containing a target object, another scene image depicted the identical background without the target object present in the scene. This allowed the presentation of both images in rapid succession to mimic the addition and deletion of the target object from the scene image.

16 different photographs were used as 'no change' trials. These were the same images used in Experiment 5 and were not used as experimental trials in any study. Half of the scenes contained an object which could be considered

inconsistent with the context. Therefore the presence of an inconsistent object would not be predictive of a change in the scene. Unlike Hollingworth and Henderson's 'no change' trials, which were the same experimental images with no change occurring, these 'no change' trials used novel images. As the experimental images were not displayed repeatedly to each participant, the same images could not be used as both 'change' and 'no change' trials.

Design

In contrast to the previous experiments in this thesis, initial fixation position was not varied in this study. Each trial began with the fixation of a central cross before the presentation of the first image. The consistency of the target was manipulated, with half the experimental scenes containing an object considered to be inconsistent with the scene background. Each experimental image contained either a consistent or an inconsistent target placed in the same location in the scene. The size of both consistent and inconsistent objects was matched. The final manipulation was the presence or absence of a change, depending on whether a 'change' or 'no change' trial had been presented.

The data of interest included accuracy rates and response latency. Latencies were recorded in ms, from the presentation of the first scene image to the time at which a response was made. In addition, data were collected relating to eye movements during the trial. Of primary interest was the time taken to fixate the target object, regardless of whether it was changing or not. This 'arrival time' was measured in terms of elapsed time in ms since the first presentation. Other measures of eye movement behaviour were also recorded, such as fixation time on the target object and the amplitude of the saccade directed to the target object.

All possible experimental scenes were organised into sets containing each different scene background once only, to prevent participants from fixating the same scene backgrounds repeatedly. Four experimental sets of 16 'change' trials were created, each of which contained only one consistent or inconsistent object from each pair which could be found in the same location in the same scene background. In this way, four sets of 16 experimental 'change' trials were created, to which 16 control 'no change' trials were added. All participants viewed the same control trials with only the 'change' experimental trials varying across participants. The order in which the experimental trials were presented was randomised and, to ensure accuracy of the eyetracking equipment, the apparatus was recalibrated after every four trials.

Procedure

Participants were recruited from the undergraduate population and individual testing sessions were arranged. Upon arrival, participants were provided with an information sheet which explained the procedure of eyetracking and relevant details of the current experiment. If participants agreed to take part, a consent form was completed. Any questions were answered and the experimenter emphasised that the changes occurring in the experiment were of the nature of a whole object appearing and disappearing.

Participants were seated at a viewing distance of 85cm. A calibration matrix was presented, consisting of nine dots in a rectangular grid which was slightly larger than the scene images. After calibration, the eight practice trials were presented, of which four included a changing target, two consistent and two inconsistent objects. At the end of the practice block, participants were allowed a break and confirmed that they had detected the appropriate changes.

It was emphasised that accuracy was more important than speed and that the deletion-addition change would be obvious once detected, to prevent responses to any minor perceived changes between the scenes. The appropriate set of experimental trials, including both 'change' and 'no change' trials, was then begun and after every four trials, the experiment was halted while a calibration was completed. At the conclusion of the experiment, participants were provided with a debriefing sheet explaining the purpose of the experiment. The entire procedure lasted approximately 30 minutes.

4.8 Results

Preliminary analyses showed that the eye movement data relating to one control scene in the 'no change' condition were not suitable for inclusion in the results. The target selected in the scene was small and poorly defined and many participants did not directly fixate it. For this reason, the data obtained from this trial were not included in the analyses. All participants correctly responded that no change was occurring in this trial.

The remaining data were analysed in terms of accuracy, response time and eye movement behaviour. Less than 1% of trials had to be excluded due to extensive tracker loss prior to the fixation of the target object. For these trials, accuracy and response time measures were still included in the analyses but eye movement data were not.

Accuracy

Table 4.3 displays the percentage of trials responded to correctly according to the change condition and the consistency of the target object. Accuracy for detecting changes was over 88% and performance for correctly detecting 'no change' trials was over 99%. The data indicated that very few false alarms were reported and change detection accuracy was also high. Failure to detect some changes suggested that participants occasionally terminated 'change' trials before detecting the change. However, the overall level of accuracy was still high and no obvious difference was found between performance for consistent and inconsistent scenes.

A binary logistic regression analysis was conducted to investigate any significant main effects of change condition and target consistency on accuracy. Change condition was a significant predictor of accuracy, with higher accuracy in the 'no change' condition than the 'change' condition ($\chi^2(1)=33.5, p<.001$). However, consistency was not a significant predictor of accuracy, with the slightly higher accuracy for consistent targets failing to reach statistical significance ($\chi^2(1)<1, p=.36$). As little difference was found in 'no change' trials according to consistency, an analysis was conducted on 'change' trials only but the 2.15% difference in accuracy between detecting consistent and inconsistent changes was not statistically significant ($\chi^2(1)<1, p=.58$).

Table 4.3: Table showing accuracy (in %) by change condition and target object consistency.

	Deletion-addition change condition	No change condition	Mean
Consistent	89.29	100.00	94.65
Inconsistent	87.14	98.40	92.77
Mean	88.22	99.20	93.71

Response time

Response times for correct trials were used to investigate the effects of the experimental manipulations. The data were presented in Table 4.4 and analysed using a univariate ANOVA (by participants), which indicated that there was a significant 4162ms difference in response time according to change condition, with 'no change' trials producing longer response times ($F(1,68)=84.0, p<.001$). As expected, trials were terminated earlier when participants detected a change, as longer observation was required to determine that no change was occurring.

Table 4.4: Summary table of mean response times (in ms) by change condition and target object consistency.

	Deletion-addition change condition	No change condition	Mean
Consistent	2354	6376	4365
Inconsistent	2197	6493	4345
Mean	2276	6438	4355

The effects of the consistency manipulation were less clear. There was no significant main effect, with similar mean response times across change conditions ($F(1,68)<1, p=.80$). The interaction between the two variables also failed to reach significance ($F(1,68)<1, p=.49$). There was a slight 157ms advantage for changing inconsistent objects, compared to changing consistent objects. When no change was occurring, there was a 117ms difference between the response times for consistent and inconsistent scenes, with inconsistent scenes producing slightly longer response times.

The results of the experimental change condition were analysed separately to the control condition to investigate whether the 157ms mean advantage found for inconsistent objects was statistically significant. Although several outliers were found in the data, as would be expected with an infinite maximum response time, the distributions were approximately normal and variances were equal, so an independent samples t test was conducted to compare the mean response times when the target object was consistent or inconsistent. The difference between response times for consistent and inconsistent change trials did not reach statistical significance ($t(34)<1$, $p=.48$) so there was no evidence that inconsistent changing targets were detected any faster than consistent changing targets.

The same analysis was conducted on the 'no change' data, as distributions were approximately normal and variances were sufficiently equal. The 117ms advantage for consistent objects when there was no change also failed to reach significance, according to an independent samples t test ($t(34)<1$, $p=.62$). These data indicated that there was no difference in response times between consistent and inconsistent control trials. Therefore, the analyses so far provided no evidence that there was any difference in performance on this task according to semantic consistency.

Eye movement data

The data relating to participants' eye movements during the search for a change were analysed to investigate whether saccade patterns were influenced by the presence of consistent and inconsistent objects in scenes. Several measures were obtained from the eye movement records, which were categorised as either measures relating to behaviour before target fixation or measures during target fixation. To begin with, measures obtained before target fixation were analysed to investigate whether saccade behaviour was influenced by the consistency of the target prior to its direct fixation.

Before target fixation measures

The time at which the participant first fixated the target, referred to as the 'arrival time' on the object, was measured from the start of the trial. Table 4.5 displays mean arrival times according to consistency and change conditions. A univariate analysis of variance (by participants) indicated a significant main effect of change condition ($F(1,68)=16.7, p<.001$), a marginally significant main effect of consistency ($F(1,68)=3.90, p=.052$) and a marginally significant interaction between the two ($F(1,68)=3.57, p=.063$). 'Change' trials resulted in faster fixation of the target object than 'no change' trials, by a mean value of 854ms.

Table 4.5: Summary table of mean arrival times (in ms) by change condition and target object consistency.

	Deletion-addition change condition	No change condition	Mean
Consistent	1362	2688	2001
Inconsistent	1363	1678	1499
Mean	1363	2217	1753

Two possible interpretations were that a changing object could be perceived in extrafoveal vision and selected as a saccade target more readily than a non-changing object or that the targets in the control scenes were not as salient as those in the experimental scenes. As different images were used in each condition, it was possible that discrepancies between the scenes existed in terms of the salience of the target objects. These differences could have resulted in longer inspection times prior to target fixation for 'no change' trials.

The marginally significant main effect of consistency, with longer arrival times for consistent than inconsistent objects, was solely found in 'no change' trials, resulting in a marginally significant interaction between the two variables. For the experimental scene images, no difference was found between consistent and inconsistent targets but a significant different was found for 'no change' trials ($t(34)=2.09, p=.044$), when inconsistent objects were fixated 1010ms faster than consistent objects. This pattern may not have been caused by the detection of semantic inconsistency but by the fact that the location and salience of consistent and inconsistent targets were not as carefully controlled in 'no change' trials as in 'change' trials.

The scenes used in 'no change' trials were not as strictly controlled as experimental trials and were therefore not suitable for individual analysis. For example, consistent and inconsistent objects were not carefully matched for size to be presented at the same location in the same scene background. The scenes were designed to function as a group, as a comparison to 'change' trials, rather than as individual controls, as would have been the case if each experimental trial had been used in a 'no change' trial also.

It was possible that discrepancies in eye movement behaviour arose from an inadvertent advantage in scene composition favouring the earlier fixation of inconsistent targets. In 'change' trials, with consistent and inconsistent targets placed in the same location in scene backgrounds, there was no difference between target object eccentricity for consistent and inconsistent targets. However, inconsistent targets in 'no change' trials could have been located closer to the initial point of fixation than consistent targets, which would make them likely to be fixated earlier.

Further investigation proved this to be the case, with the mean distance from the initial fixation position being 3.97° for consistent objects and 2.20° for inconsistent objects in 'no change' scenes. This 1.77° difference in target object eccentricity was found to be statistically significant, according to an independent samples t test ($t(34)=9.16, p<.001$). The difference in eccentricity in 'no change' trials suggested that participants fixated inconsistent objects sooner than consistent objects because they were presented closer to the initial fixation position, rather than because of their semantic consistency.

If semantic consistency did modulate the difference in arrival time between consistent and inconsistent targets, it would suggest that semantic information could be processed from extrafoveal vision and used to direct saccades. This suggestion implies that inconsistent objects in extrafoveal vision were more salient targets than consistent objects. However, this assumption would only be true in the control, 'no change' condition.

This hypothesis could be tested by investigating the amplitude of the first saccade directed to the target object. This measure would indicate whether processing prior to target fixation, when selecting the saccade target, was foveal or extrafoveal for consistent and inconsistent targets. Table 4.6 displays the mean saccade amplitudes directed towards the target according to change condition and target object consistency.

Table 4.6: Summary table of mean saccade amplitudes (in °) by change condition and target object consistency.

	Deletion-addition change condition	No change condition	Mean
Consistent	2.83	3.78	3.29
Inconsistent	3.07	3.32	3.18
Mean	2.95	3.57	3.23

Saccades towards the target were larger in the 'no change' condition than in the 'change' condition, contradicting expectations that the addition and deletion of an object could have rendered the target object more salient than a non-changing object. A univariate ANOVA (by participants) indicated that this 0.62° difference was significant ($F(1,68)=5.56, p=.021$). The effects of consistency were less clear, as slightly larger saccades were directed towards inconsistent objects than consistent objects when the object was changing but the opposite pattern was seen for 'no change' trials. No main effect of consistency was found ($F(1,68)<1, p=.67$) and also no interaction between change condition and consistency ($F(1,68)=1.33, p=.25$). The slight differences evidenced between consistent and inconsistent trials were not statistically reliable.

The 0.24° difference between consistent and inconsistent 'change' trials was analysed using an independent samples t test, as variances were equal and distributions were approximately normal. There was no difference in the size of the saccade directed to consistent and inconsistent targets when they were changing (*t*(34)<1, *p*=.53). This result indicated that there was no difference in the salience of consistent and inconsistent changing objects when processed in extrafoveal vision.

Similarly, no significant difference was found for the consistent and inconsistent saccade amplitudes in ‘no change’ trials. The 0.46° difference between consistent and inconsistent trials was not found to be significant according to an independent samples t test (*t*(34)<1, *p*=.35). The slight difference in saccade amplitude indicating shorter saccades directed to inconsistent targets than to consistent targets could have been generated by the inconsistent targets’ closer presentation, relative to initial fixation position, in ‘no change’ scenes. The data suggested that there was no reliable effect of consistency on saccadic amplitude, implying that the shorter time to fixate an inconsistent target in ‘no change’ trials was not associated with an advantage in extrafoveal processing prior to fixation.

Target fixation measures

As participants were instructed to respond when a change was detected, no difference between fixation times on consistent and inconsistent objects in ‘change’ trials would necessarily be expected. Fixations would need to be long enough to observe both versions of the scene image, separated by the mask. Consistency effects on fixation times in ‘no change’ trials could be significant, as longer fixation times than seen in ‘change’ trials could indicate an expectation that the object would change, delaying a saccade to another scene region. Longer fixation times on inconsistent objects in either change condition could also be explained by increased processing difficulty in reconciling the object’s semantic inconsistency with the scene background, resulting in increased first fixation durations compared to those for consistent objects. For these reasons, the investigation of interest would be whether inconsistent objects in ‘no change’ scenes were fixated for longer than inconsistent changing objects.

The data are presented in Figures 4.6 and 4.7 for ‘change’ and ‘no change’ trials respectively. A multivariate ANOVA (by participants) was conducted to investigate the effects of the two independent variables, change condition and consistency, on three measures of fixation time, consisting of first fixation duration, first pass fixation duration and total fixation duration. First fixation durations were the first fixation times prior to the next saccade, regardless of the subsequent fixation position, while first pass fixation durations included all fixations on the target prior to the next saccade leaving the object region. Although in most 'change' trials, participants terminated the trial after the first fixation on the target, having detected the change, the additional measures were included for comparison to 'no change' trials, in which participants could fixate the target object repeatedly.

A significant main effect of change condition was found on first fixation durations ($F(1,68)=9.43, p=.003$) and first pass fixation durations ($F(1,68)=13.5, p<.001$) but not on total fixation durations ($F(1,68)<1, p=.50$). Total fixation durations on 'change' and 'no change' trials were approximately equal but shorter first fixation durations and first pass fixation durations were found for targets which were not changing. This evidence suggested that saccades away from targets in the 'no change' condition were initiated earlier than the time at which 'change' trials were terminated with a manual response. Participants may have fixated the target over more than one scene change before responding that a change had occurred.

Consistency did not have a main effect on any of the fixation measures. There was no effect on first fixation durations ($F(1,68)<1, p=.54$), first pass fixation durations ($F(1,68)<1, p=.40$) or total fixation durations ($F(1,68)=2.17, p=.15$). However, the interaction between change condition and consistency was statistically significant for first fixation durations ($F(1,68)=6.12, p=.015$) and total fixation durations ($F(1,68)=5.52, p=.022$) but not for first pass fixation durations ($F(1,68)=2.36, p=.13$). The data suggested that fixation durations on inconsistent objects did not vary greatly according to change condition but fixation durations on consistent objects were considerably shorter for 'no change' trials than for 'change' trials

To determine whether this effect was caused by a significant difference in fixation times between 'change' and 'no change' consistent trials, 'change' and 'no change' inconsistent trials or both, the data were organised according to consistency and analysed for a significant effect of change condition. This analysis revealed that fixation times significantly decreased for consistent trials when the target object was not changing, for first fixation durations ($F(1,34)=15.15, p<.001$), first pass fixation durations ($F(1,34)=16.3, p<.001$) and total fixation durations ($F(1,34)=5.37, p=.027$). However, the same analysis for inconsistent objects indicated that there were no significant differences in fixation times for first fixation durations ($F(1,34)<1, p=.68$), first pass fixation durations ($F(1,34)=1.96, p=.17$) or total fixation durations ($F(1,34)=1.22$, $p=.28$) between changing and non-changing inconsistent objects. These results suggested that the significant interaction found between consistency and change condition was generated by the significant decrease in fixation times for consistent objects which did not change compared to those which did.

To investigate the effects of consistency further, the data for 'change' and 'no change' conditions were analysed separately. The data relating to 'change' trials in Figure 4.6 showed no evidence of longer fixations on inconsistent targets than consistent targets. Instead, marginally longer fixation times on consistent objects were found, suggesting that the detection of changes was not affected by consistency. The difference in fixation times was greatest for first fixation durations, with consistent objects being fixated for 70ms longer than inconsistent objects. A multivariate ANOVA (by participants) indicated that there was no significant effect of consistency on first fixation durations ($F(1,34)=1.09, p=.30$), first pass fixation durations ($F(1,34)<1, p=.71$) or total fixation durations ($F(1,34)<1, p=.61$). The detection of change was not influenced by the consistency of the target object.

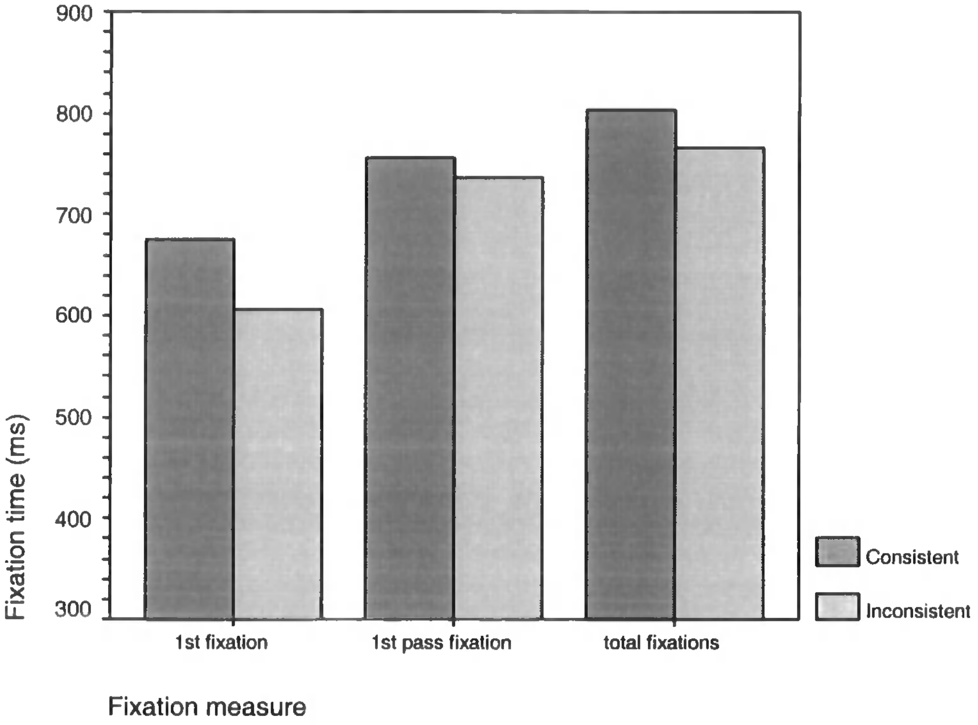


Figure 4.6: Graph showing mean fixation times by target object consistency for 'change' trials.

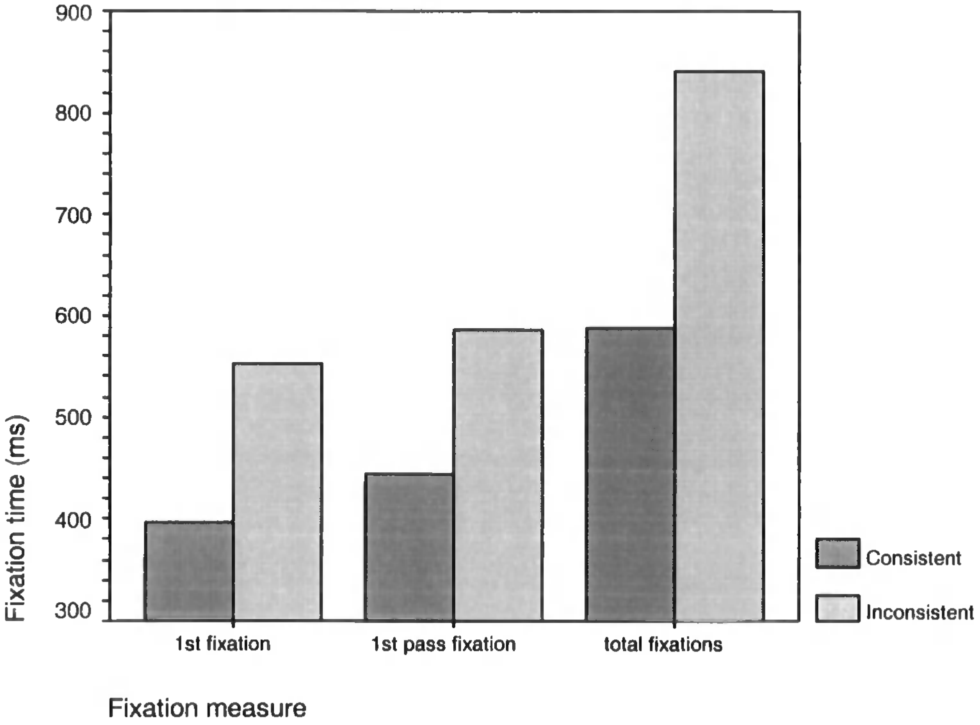


Figure 4.7: Graph showing mean fixation times by target object consistency for 'no change' trials.

More pronounced differences were seen for 'no change' trials in Figure 4.7, with inconsistent objects being fixated for much longer than consistent objects. This difference was 156ms for the first fixation durations, supporting the hypothesis that inconsistent objects are fixated for longer than consistent objects. A 144ms difference was found for first pass fixation durations and the largest difference of 253ms was found for total fixation durations. The analysis of 'no change' trials confirmed a significant effect of consistency on first fixation durations ($F(1,34)=12.3, p=.001$), first pass fixation durations ($F(1,34)=7.43, p=.010$) and total fixation durations ($F(1,34)=13.2, p=.001$). The data suggested that non-changing inconsistent objects were fixated for longer over the course of a trial than non-changing consistent objects. This could indicate an expectation that inconsistent objects were more likely to change, resulting in longer fixations and more refixations before determining that no change was occurring.

Object size

The objects were categorised by size in the same way as in Experiments 3 and 5. However, as the images were not displayed on the same equipment and at the same viewing distance, the object sizes in degrees of visual angle were not the same. In this experiment, small objects subtended less than 1.7° square, medium objects between 1.7° and 3.5° square and large objects more than 3.5° square.

Table 4.7 displays accuracy by object size and consistency. A binary logistic regression analysis indicated a main effect of object size ($\chi^2(1)=4.97, p=.026$), with higher accuracy for larger objects. No main effect of consistency was found ($\chi^2(1)<1, p=.67$) and there was no significant interaction between the two variables ($\chi^2(1)<1, p=.94$).

Table 4.7: Table showing accuracy (in %) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	85.42	92.45	89.74	89.29
Inconsistent	82.46	83.33	100.00	87.14
Mean	83.81	88.12	94.59	88.21

The difference in accuracy between consistent and inconsistent small objects was only 2.96%. The data displayed a 9.12% advantage for medium sized consistent objects compared to medium sized inconsistent objects and a 10.26% advantage for large inconsistent objects compared to large consistent objects. The difference in accuracy between consistent and inconsistent medium sized objects was not statistically significant, according to a binary logistic regression analysis ($\chi^2(1)=2.02$, $p=.16$). However, the difference between consistent and inconsistent large objects was statistically significant ($\chi^2(1)=5.33$, $p=.021$), indicating that changes to large inconsistent objects were detected more reliably than changes to large consistent objects.

This analysis was extended to include response time and arrival time. The data in Table 4.8, presenting response times, indicated no difference according to consistency. Response times decreased slightly as object size increased, indicating a trend towards faster detection of changes to larger objects.

Table 4.8: Table showing response times (in ms) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	2966	2650	2655	2760
Inconsistent	3025	2651	2404	2741
Mean	2998	2650	2536	2751

Similarly, Table 4.9 presents mean arrival times according to object size and consistency. Again, no effect of consistency was found and there was also no obvious effect of object size. The largest difference in time between consistent and inconsistent targets occurred for small objects. Small consistent objects were fixated 329ms before small inconsistent objects but this difference was not statistically significant. According to an independent samples t test, conducted as the variances were equal and distributions were approximately normal, there was no significant difference between mean arrival time for consistent and inconsistent small objects ($t(34)=-1.11$, $p=.27$). The data suggested that there was no significant effect of consistency on arrival time at any object size.

Table 4.9: Table showing arrival times (in ms) by object size and object consistency.

	Small	Medium	Large	Mean
Consistent	1196	1482	1442	1362
Inconsistent	1525	1195	1306	1363
Mean	1372	1347	1368	1363

4.9 Discussion

The results of this experiment indicated that although performance on this task was accurate, no evidence of differential performance between consistent and inconsistent targets was found. Accuracy in detecting the changes was not significantly affected by the consistency of the target object and the same was true of the response times. Although slightly faster mean responses were obtained for inconsistent targets, this difference failed to reach statistical significance.

Additionally, the eye movement data supported these conclusions. The time taken to fixate the target object was not affected by its semantic consistency. The only significant effect of consistency condition was found for ‘no change’ trials, which subsequent analyses found to be modulated by inconsistent targets located closer to the central fixation position than consistent targets. The analysis of saccade amplitude for the saccade directed to the target also failed to indicate that inconsistent objects were more salient saccade targets than consistent objects.

The fixation times on targets confirmed that inconsistent targets were fixated longer than consistent targets, in ‘no change’ trials only. The lack of any difference between fixation times on consistent and inconsistent targets in ‘change’ trials could be attributed to the instructions to terminate the trial once a change was detected. The results as a whole provided no reliable evidence that inconsistent objects were facilitated in this task, contrary to the results found by Hollingworth and Henderson (2000).

Comparing results to Hollingworth and Henderson's (2000) Experiment 1

As expected, accuracy was high in the current experiment, with participants detecting the appropriate change in approximately 88% of trials, compared to 94% for the deletion-addition condition in Hollingworth and Henderson's experiment. This difference could have been caused by participants in the current experiment prematurely terminating trials in an effort to respond as quickly as possible. Also, the photographic scenes were more complex and contained more component objects than the line drawings used by Hollingworth and Henderson. Participants could have been less likely to thoroughly search the more complex scenes before determining whether a change had occurred, compared to simply needing to observe the main items in a simple line drawing scene to determine whether an object was changing. The results of Experiments 3, 4 and 5 indicated that the use of photographic stimuli, compared to line drawings, did not adversely affect performance on both brief presentations and a change detection task. This conclusion suggested that the poorer accuracy found in the current experiment compared to the results of Hollingworth and Henderson's Experiment 1 could not be explained simply in terms of the nature of the stimuli.

A comparison of the response times produced by the current experiment and Hollingworth and Henderson's experiment indicated that changes were detected much earlier in Hollingworth and Henderson's study, with a mean time of 1226ms, compared to 2276ms in the current experiment. This difference could be explained by the composition of the stimuli, with simple line drawings taking less time to be searched fully than complex photographic scenes. Mean search time for 'no change' trials was 6438ms in the current experiment and only 2257ms in Hollingworth and Henderson's experiment, suggesting that the increased complexity of the photographic scenes required longer search times before a negative response could be made. As the addition and deletion of a whole object was still visually salient, the increase in search time implied that the photographs contained more potential target objects to be examined than the line drawings of scenes.

Hollingworth and Henderson's data indicated that changes to inconsistent objects were detected reliably sooner than changes to consistent objects, after 1190ms and 1261ms respectively. The results of the current experiment however failed to find any evidence of faster responses to changing inconsistent targets, with a non-significant 157ms advantage for inconsistent targets. Additionally, eye movement behaviour showed no effects of semantic consistency. This difference in results between the two experiments could only be explained by differences in the stimuli used, as the experimental procedures were the same, except for the use of a white noise mask more suitable for photographic stimuli than a pattern mask.

Differences between the photographic scenes and the line drawings used by Hollingworth and Henderson included the larger number of potential target objects in the photographic stimuli. The increase in scene items could make a target object more difficult to detect within a visual scene. This assumption was supported by the results of Hollingworth and Henderson's Experiment 3, which adapted the line drawing scenes by adding a consistent distractor and an inconsistent distractor to certain scenes. Change detection accuracy decreased from 52.1% in Experiment 2 to 40.7% in Experiment 3, suggesting that the additional objects added to the scenes affected the processing of the target object during brief 250ms presentations. This result implied that the number of potential distractors in a scene could influence accuracy on a task requiring object detection from brief presentations, which could explain the failure to find an inconsistent object advantage, both in the current change detection experiments and also in Experiments 3 and 4.

An additional explanation for the lack of any consistency effect could involve the design of the experimental trials. In Hollingworth and Henderson's experiment, participants viewed each scene eight times, four times with each of one consistent and one inconsistent target object, viewing two 'no change' trials, one 'deletion-addition change' trial and one 'left-right orientation change' trial. As explained previously, displaying the same scene background to each participant could influence performance on the task subsequent to the first presentation, as the image would no longer be novel and a target object seen to change in one trial could be subject to greater focus than a novel target or distractor in a subsequent trial.

Prior experience would be particularly important when participants viewed the scene images of one trial for long durations and had sufficient time to search the scene and foveate many objects within it. As memory for inconsistent objects has been proved to be better than memory for consistent objects, the repeated viewing of the same scene-object images could have resulted in the faster detection of changes in scenes in which a changing inconsistent target had been fixated previously. Whether the memory advantage for inconsistent objects could explain facilitation at briefer presentations, for example in Hollingworth and Henderson's Experiments 2 and 3, remains unclear but it could certainly influence performance on a cycling presentation change detection task, through the faster localisation of likely scene targets.

The eye movement data from the current experiment were analysed to investigate whether the time at which participants fixated the target object was more susceptible to semantic influence than the overall response time. Hypothetically, any tendency to fixate inconsistent objects faster than consistent objects could be overshadowed by the longer fixation of inconsistent objects, which could artificially increase response times. Contrary to this consideration, the eye movement data showed no evidence of a consistency effect on arrival time for the carefully matched consistent and inconsistent objects in 'change' trials. However, a significant advantage was found for inconsistent objects in 'no change' trials, which could be attributed to the lack of careful control over the salience of the target objects, as the targets were not matched for size or eccentricity from central fixation.

In addition, contrary to the assumption that changing objects would be more salient saccade targets, saccade amplitudes directed towards changing target objects were significantly smaller than those directed to non-changing targets (p=.021). There was no significant difference in saccade amplitude towards consistent and inconsistent changing objects (p=.53), which did not support the hypothesis that inconsistent changing objects were more salient saccade targets than consistent changing objects. Subsequent analysis of the 'no change' data indicated that inconsistent targets were located significantly closer to the initial central fixation position, which could explain why they were fixated sooner and after shorter saccades than consistent targets.

When no change was occurring, significantly longer mean first fixation durations and first pass fixation durations were found for inconsistent targets. This result indicated either that inconsistent objects required more detailed processing to reconcile their semantic inconsistency, or the data could reflect an expectation by the participant that inconsistent objects were more likely to change. No significant difference was found between fixation times on consistent and inconsistent changing objects, which was attributed to the termination of the trial by the participant's response.

Finally, the analysis of target object size for 'change' trials produced the following results. The 9.1% advantage in change detection accuracy for medium sized consistent objects, compared to medium sized inconsistent objects failed to reach statistical significance (p=.16). However, the 10.3% advantage for large inconsistent objects over large consistent objects was statistically significant (p=.021). This pattern of a consistent object advantage for medium sized objects and an inconsistent object advantage for large objects was compatible with previous experimental results. Additional analyses on response times and arrival times indicated no significant effects of target object size.

The results of this experiment failed to replicate Hollingworth and Henderson's results from Experiment 1, indicating faster detection of changing inconsistent targets than changing consistent objects. The only difference between these experiments was the stimuli applied to the change detection paradigm. For this reason, it was necessary to consider possible explanations for the discrepancy between results and how the two attempted replications of Hollingworth and Henderson's (2000) experiments could be reconciled in terms of the processing of semantic information from scenes.

4.10 Summary of change detection experiments 5 and 6

The two experiments described in this chapter attempted to replicate the findings of Hollingworth and Henderson's (2000) experiments using more naturalistic visual stimuli. However, neither Experiment 5 nor Experiment 6 succeeded in replicating their results. The results obtained did not support Hollingworth and Henderson's conclusions that the detection of changes to inconsistent objects was facilitated in both a two-exposure change detection task or a cyclical change detection task.

Experiment 5 found no evidence of inconsistent object facilitation in the detection of changes during a two-exposure change detection trial. Hollingworth and Henderson interpreted their own significant effect as evidence that semantically inconsistent regions of the scene were preferentially retained across views. They argued that information relating to inconsistent objects was processed during the first image presentation and retained across the mask more accurately than information relating to consistent objects. This facilitation would result in the increased accuracy in change detection for inconsistent objects compared to consistent objects. This effect was observed in both Experiments 2 and 3 by Hollingworth and Henderson but could not be replicated in the current Experiment 5. There was no evidence of an advantage for information relating to inconsistent objects over consistent objects being retained across the mask.

To investigate whether the use of photographic stimuli rather than line drawing stimuli could influence the processing of information from brief scene presentations and its retention across a short interval, the results of Experiments 3 and 4 were considered. Accuracy on this object detection task was significantly improved for photographic stimuli, compared to the same images converted into line drawings, indicating that the ability to obtain object information from brief presentations was not compromised by the use of photographic stimuli. If the simplification of complex scenes to line drawings was necessary for semantic inconsistency to be detected from brief scene presentations, Experiment 4 would have indicated some effect of consistency. Therefore, the fact that Hollingworth and Henderson used line drawing stimuli could not explain why similar results were not obtained using photographs.

The cycling change detection task used in Experiment 6 also failed to support Hollingworth and Henderson's conclusions, indicating the earlier detection of changes to inconsistent changing targets. They argued that information relating to inconsistent targets was retained more efficiently across the mask than information relating to consistent targets. The attempted replication in Experiment 6 did not reach the same conclusions. No evidence of the faster detection of changes to inconsistent objects was found, either in terms of manual responses or eye movement behaviour.

Although the consistency effect found in Hollingworth and Henderson's Experiment 1 could be explained by the selective retention of information across a mask, the detection of a change would also depend on how rapidly the target in question could be fixated. Experiment 6, which recorded eye movements, found no evidence of the earlier fixation of inconsistent targets than consistent targets in photographs, but this possibility could not be rejected for the line drawings used in Hollingworth and Henderson's experiments. For this reason, before concluding that Hollingworth and Henderson's consistency effect was generated by an advantage in the retention of inconsistent object information, it would be useful to determine whether eye movement behaviour could have been influenced by semantic inconsistency in line drawing stimuli.

Although Experiments 5 and 6 addressed slightly different issues of semantic processing, both failed to replicate Hollingworth and Henderson's results. Experiment 5 could not support their conclusions that information relating to inconsistent objects was detected during the first scene presentation, preferentially retained across the mask and compared with the second scene presentation in order to detect the change. Experiment 6 investigated whether inconsistent objects were detected earlier than consistent objects during a longer trial duration. This proposal introduced the consideration that semantic inconsistency could be detected in extrafoveal vision and used to direct saccade scanning.

The inability to replicate Hollingworth and Henderson's findings required an explanation as to why such similar experiments failed to obtain compatible results. A main methodological difference across experiments was the number of trials which each participant viewed. While Hollingworth and Henderson used larger blocks of trials, producing more data, they also allowed each participant to view each scene several times with both consistent and inconsistent targets located in them. Although this might not have selectively facilitated inconsistent trials, it could have improved performance overall.

In the current experiments, the vividness of the images rendered them unsuitable for repeated viewing, as the realistic scenes could have been more memorable than simple line drawings. For this reason, it was considered inappropriate to display each scene more than once in an experimental block of trials and efforts were made to avoid presenting images of the same room from the same viewpoint to any participant. This manipulation of scene images resulted in a large number of participants viewing a relatively small number of experimental trials, which resulted in larger variability across participants. The smaller number of participants in Experiment 6 could have aggravated the problem, with only 4 or 5 participants viewing each scene variant.

However, the most likely explanation for the failure to replicate a consistency effect would involve the nature of the scene stimuli. The photographic scenes used in Experiments 5 and 6 were more naturalistic than the simplistic scenes depicted in the line drawings. The greater number of scene objects could have made a target object more difficult to detect in the photographic scenes, suggesting that information relating to semantic inconsistency could only be processed from brief presentations when simple visual stimuli were displayed.

Alternatively, it could be argued that the manipulation of semantic consistency in the photographic stimuli did not sufficiently illustrate extremes of probability, such as essential and impossible scene targets. Instead, the photographs contained movable household objects which were likely or unlikely to naturally occur in a scene. However, if this explanation could account for the differences in results, it would suggest that the consistency effect found by Hollingworth and Henderson (2000) was not applicable to real-world scene viewing and would only be evidenced within highly constrained and unnatural consistency manipulations.

To summarise the investigation so far, although robust evidence was found by Hollingworth and Henderson (1998, 1999, 2000, 2001) that inconsistent objects were facilitated in tasks involving brief presentations of line drawing scene stimuli, the original experiments in this thesis have failed to replicate this effect using complex photographic scene images. Only when images originating from the Leuven line drawings were used was an advantage found for inconsistent objects. The absence of a consistency effect when viewing photographs could be attributed to the complexity of the visual stimuli, in terms of the increased number of potential targets and/or the manipulation of the consistency relationship.

Applying a two-exposure change detection paradigm to this investigation (Experiment 5) supported these conclusions, as it failed to replicate the consistency effect found with line drawings, when displaying photographs of scenes. Similarly, cycling change detection experiments found that changes to inconsistent targets were detected faster than changes to consistent targets only in line drawings and not in photographs (Experiment 6). The investigation of eye movement behaviour indicated that inconsistent targets in photographs were not fixated any earlier than consistent targets, but this could not be confirmed with the line drawing stimuli. This consideration was investigated in the following experiment.

Chapter 5

Natural Scene Viewing

5.1 Experiment 7: Introduction

As discussed previously, the role of eye movements in the detection of semantic inconsistency in scenes has been subject to extensive investigation. In 1978, Loftus and Mackworth claimed that inconsistent objects were fixated sooner than consistent objects in scenes, but subsequent experimenters have had little success in replicating this effect (e.g. Friedman, 1979, De Graef, Christiaens and d'Ydewalle, 1990, Henderson, Weeks and Hollingworth, 1999). Many possible suggestions have been proposed to explain this inability to replicate the effect reported by Loftus and Mackworth, including differences in scene images, sizes and experimental tasks. For example, many of these experiments investigated whether semantic inconsistency could be detected extrafoveally during scene viewing in anticipation of a memory test. These instructions might have influenced saccade behaviour when viewing the scene images and could also have influenced recall.

In the following experiment, the aim was to investigate whether the presence of inconsistent objects in scenes affected saccade behaviour during scene viewing, under the most naturalistic conditions possible. The scene images included both the Leuven line drawings, used in Experiments 1 and 2, and the photographs of natural scenes and household objects, used in Experiments 3, 5 and 6. It was hypothesised that, by recording eye movement behaviour, it would be possible to establish whether saccade patterns were affected by the presence of inconsistent objects in scenes. The saccade patterns resulting from viewing both stimuli sets were investigated in the search for differences in eye movement behaviour attributable either to semantic inconsistency or to the different types of scenes.

Naïve participants were instructed to view the images normally with no explicit task, in an attempt to replicate natural viewing conditions. These instructions should have ensured that participants were neither actively searching the scenes nor using any specific strategy for viewing them. The intention was to prevent the participants' preconceptions of the experiment affecting their overt behaviour by explicitly emphasising that no additional or subsequent tasks would be involved and that there were no hidden procedures. In this way, participants would have no explicit prior expectations of a further task which could affect how they viewed the scenes.

The aim of the experiment was not to determine whether people *can* process semantic information from extrafoveal vision and use it to selectively saccade to inconsistent objects in scenes, but to determine whether people actually do so when presented with a novel scene image. The intention was to determine whether the eyes were drawn towards inconsistent objects in scenes, compared to equivalent consistent objects, particularly when the objects were appropriately matched, as in the photographic stimuli. Additional to whether semantic consistency could be used to direct saccades under artificial and simplified experimental conditions, a further issue of interest in this experiment was whether similar consistency effects could be found both using complex, realistic visual stimuli and under naturalistic conditions.

The data relating to line drawing scene viewing would provide a further test of the hypothesis raised by Experiments 1 and 2, that the line drawings of inconsistent targets could have been more salient than the drawings of consistent targets, due to inadvertent pictorial differences between the two. If a tendency to fixate inconsistent targets sooner than consistent targets were found with the line drawing stimuli but not with the photographs, this discrepancy would indicate that the consistency effect evidenced in Experiment 1 was unlikely to be caused by semantic consistency. This result would be compatible with the conclusions of Experiments 1 to 4, which showed that a consistency effect could only be evidenced using the Leuven line drawing stimuli.

5.2 Method

Participants

A total of 24 naïve undergraduate students, five males and 19 females, from the University of Durham volunteered to participate in this experiment. They all had normal, uncorrected vision.

Apparatus

The apparatus used for displaying images and recording eye movements was the same as in Experiment 6.

Materials

The scene images for the line drawing stimuli were the ones used in Experiment 1. The photographic scenes used were those designed for Experiment 3 and identical to the ones used in Experiment 6. All scene images measured 640 by 480 pixels and were located centrally on a white background measuring 1024 by 768 pixels. The experimental images subtended approximately 16° by 12°, within the 26° by 20° monitor size, at a viewing distance of 85cm.

In total, the displays consisted of 44 line drawings of scenes and 64 photographs of scenes. Each participant viewed one block of 11 line drawings of scenes and another block of 16 photographic images. Within each of the four possible sets of line drawing stimuli, the 11 images were selected to ensure that each scene background was only presented to each participant once.

For the photographs, each of the four possible sets of trial stimuli contained 16 images, so that each participant did not view the same background or object twice. These precautions were employed to prevent any difference in saccade behaviour resulting from the repeated presentation of the same scene background or object. It was possible that fixation times or probability of fixating a target would be affected by a previous presentation of the same scene background or target object. Unless these precautions were taken, the conclusions reached by the analyses would be compromised.

To summarise, in both line drawing and photographic stimuli, each participant viewed trials which included only one target in each scene background and the same target object was not displayed twice in the same trial set. Both the line drawings and photographs were arranged into four trial sets with each scene presented only once per four participants. Therefore, with 24 participants, each individual scene version was presented to six participants.

Design

The experimental trials were designed in sets so that each participant saw only one version of each scene background. The order in which scenes were presented within each set was randomised. The order of the sets was counterbalanced so that half of the participants viewed the line drawings first and the other half viewed the photographs first. Within each set, a calibration matrix was presented every four trials to ensure accuracy of eye movement monitoring.

The independent variables were the type of scene image presented (line drawings or photographs) and the consistency of the image (consistent or inconsistent target object). The dependent variables of interest were divided into measures of behaviour before the fixation of the target object and after the initial direct fixation of the target. The measures relating to behaviour prior to target fixation consisted of the probability of having fixated the target at least once within the 7000ms scene presentation, the number of saccades executed before fixating the target, the absolute time taken to fixate the target (in ms) and the amplitude of the saccade directed to the target. The measures obtained during or after target object fixation consisted of the first fixation duration on the target, the first pass fixation duration and the total fixation duration. This range of measures was investigated to determine whether the consistency of the target object could be detected before the fixation of the target object and also whether consistency affected subsequent viewing behaviour.

Procedure

Participants attended an individual testing session. They were provided with an information sheet and a consent form to complete if they agreed to participate.

Verbal confirmation was provided that no explicit task was required and participants were encouraged to view the images passively as no extra tasks were included in the experiment.

Participants were seated at a viewing distance of 85cm and, after a calibration phase identical to that used in Experiment 6, the experiment commenced with either a set of line drawing stimuli or a set of photographic stimuli. Each image was presented for 7000ms to allow sufficient time for full examination. At the end of the first block of trials, participants were given an opportunity to ask questions if necessary, while the second block of trials was arranged. When they were ready to continue, the experiment was resumed and participants were given a debriefing sheet at the end of the second block.

5.3 Results

For each trial, saccades were identified individually and the appropriate measures were recorded. The data for the Leuven line drawings and for the photographs were investigated independently and the results have been presented in this way. After consideration of the individual stimuli sets, they were compared to determine whether eye movement behaviour varied according to the type of image displayed.

Leuven line drawings

Data on target fixation could not be obtained in only 20 out of the total number of 264 trials (7.5%), either because the target was not fixated or because tracker loss occurred. The possible relationship between the consistency of the target object and the probability of fixating the target within the 7000ms scene presentation was investigated. A logistic regression analysis was conducted to investigate whether the probability of fixating the target object was affected by target object consistency, but no significant difference was found, with consistent objects being fixated in 91.7% of trials compared to 93.2% for inconsistent objects ($\chi^2(1)<1$, p=.64). This result indicated that within the

7000ms presentation time, both consistent and inconsistent objects were equally likely to be fixated.

Before target fixation

The mean data relating to measures obtained prior to target fixation are presented in Table 5.1. The measures included the number of saccades executed before the target was fixated, the arrival time (within the 7000ms trial duration) at which the target was fixated and the amplitude of the saccade directed at the target object.

Table 5.1: Summary table of mean data relating to measures of saccade behaviour prior to target object fixation.

	Consistent	Inconsistent
Number of saccades	4.5	5.2
Arrival time (ms)	1309	1613
Saccade amplitude (°)	3.7	3.8

The number of saccades taken to fixate a consistent target object was marginally less than the value obtained for inconsistent targets, but this difference was not statistically significant ($t(46)=1.27, p=.21$). The 304ms difference between arrival time on consistent and inconsistent objects just failed to reach statistical significance ($t(46)=1.81, p=.078$) indicating that, contrary to the hypothesis that inconsistent objects could be more salient saccade targets in the Leuven line drawings, they were in fact fixated marginally later than the consistent targets in scenes. There was also no significant difference between the amplitude of a saccade directed at a consistent and an inconsistent target ($t(46)<1, p=.68$). There was therefore no evidence that inconsistent targets were selected as saccade targets from closer or further extrafoveal vision than consistent targets.

The analysis of eye movement data relating to behaviour before the fixation of the target object provided no evidence of differential processing of consistent and inconsistent objects in extrafoveal vision, prior to target fixation. The data did not support the hypothesis that semantic consistency could be detected prior to the direct fixation of the target. The Leuven line drawings used in

Experiment 1 produced an inconsistent object advantage in a brief presentations object recognition task, suggesting that the inconsistent targets were subject to different processing than consistent targets during the 120ms presentation. However, the investigation of eye movement behaviour when viewing these scenes for 7000ms showed no facilitation for the inconsistent targets, indicating that these targets were not more salient and therefore not more likely to attract saccades during passive scene viewing than consistent targets.

Target fixation

The data relating to eye movement behaviour during and after the initial fixation of the target object were analysed to investigate whether inconsistent objects in scenes affected fixation patterns, compared to consistent objects in scenes. The recorded measures were first fixation durations on the object, first pass fixation durations and the total fixation durations, which are reported in Table 5.2.

Table 5.2: Summary table of mean data relating to measures of saccade behaviour during target object fixation.

	Consistent	Inconsistent
1st fixation duration (ms)	383	550**
1st pass fixation duration (ms)	573	718
Total fixation duration (ms)	1020	1244*

* *p*<.05
** *p*<.01

The 167ms difference in first fixation durations indicated that fixations on inconsistent targets were significantly longer than on consistent targets (*t*(34.1)=-2.76, *p*=.008). This difference supported the hypothesis that inconsistent objects were distinguished from consistent objects by participants in terms of processing difficulty. However, the 145ms difference in first pass fixation durations just failed to reach statistical significance (*t*(46)=-2.00, *p*=.052) but indicated the same pattern. Total fixation durations displayed a 224ms difference which was statistically significant (*t*(46)=-2.29, *p*=.027). These data indicated that, as anticipated, inconsistent objects in scenes were fixated for longer than consistent objects.

Object size

In addition to the above analyses, the effects of object size were also investigated. The targets were categorised as in Experiment 1, but the image size and viewing distance were different, affecting the objects' size in pixels and visual angle. Small objects in this experiment had a pixel area less than 4,500, subtending less than 3° square. Medium sized objects had pixel areas ranging between 4,500 and 12,000, subtending between 3° square and 7.5° square, and large objects had pixel areas greater than 12,000, subtending over 7.5° square.

The data were analysed by items and subjected to a multivariate ANOVA, including the same measures obtained before and after fixation as dependent variables and investigating the effects of consistency and size manipulations. The results are summarised in Table 5.3. Considering the measures taken before target fixation first, target object size was found to have no significant effect on the number of saccades executed before target fixation ($F(2,38)=1.90, p=.16$). The effect of object size on the arrival time on the target approached statistical significance ($F(2,38)=2.60, p=.087$), suggesting that larger objects were fixated sooner than small and medium sized objects. No significant effect of object size on saccade amplitude was found ($F(2,38)=1.81, p=.18$), with similar sized saccades directed at small, medium and large targets.

Table 5.3: Summary table of data according to target object size.

	Small	Medium	Large
Number of saccades	5.04	5.45	3.43
Arrival time (ms)	1537	1678	921
Saccade amplitude (°)	3.43	4.21	3.51
1st fixation duration (ms)	570	441	302*
1st pass fixation duration (ms)	604	587	768
Total fixation duration (ms)	1149	907	1419**

* $p<.05$
** $p<.01$

Target object size had a significant effect on first fixation durations ($F(2,38)=3.93, p=.028$), with progressively shorter fixations on larger objects. This effect was not found for first pass fixation durations ($F(2,38)=1.38, p=.26$)

but a significant effect in the opposite direction was displayed in the analysis of total fixation durations ($F(2,38)=5.25, p=.01$). Large targets were fixated for longer in total than small or medium sized targets. There were no significant effects of the interaction between consistency and target object size on any of these variables. The data indicated that target object size did not significantly influence measures obtained prior to target fixation and the only significant effects found were on fixation times, with shorter first fixation durations on large targets but longer total fixation durations, as would be expected.

Photographs

Data were lost from 12.5% of trials, either because the target object was not fixated or tracker loss occurred. This slight decrease in fixation rate compared to the line drawing stimuli could be attributed to the larger number of items in the photographic scenes, resulting in the lower probability that any one object in the scene would be fixated. A logistic regression analysis investigated whether there was any effect of consistency on the probability of fixating the target object. However, no significant effect was found, with the target object being fixated in 89.6% of consistent trials, compared to 85.4% for inconsistent trials ($\chi^2(1)<1, p=.61$).

Before target fixation

The data relating to measures obtained before target fixation are reported in Table 5.4. The number of saccades taken to fixate the target object was investigated separately for trials containing a consistent and an inconsistent target object. The mean number of saccades taken to fixate both consistent and inconsistent target objects was 5.2, so there was clearly no significant difference between them ($t(46)<1, p=.99$). Similarly, the 53ms difference in arrival time on consistent and inconsistent targets was not statistically significant ($t(46)<1$, $p=.77$), indicating that the consistency of the target object did not influence when it was fixated. Saccades to inconsistent targets were found to be slightly larger than those directed to consistent targets but again, this difference failed to reach statistical significance ($t(46)=-1.27, p=.21$).

Table 5.4: Summary table of mean data relating to measures of saccade behaviour prior to target object fixation.

	Consistent	Inconsistent
Number of saccades	5.2	5.2
Arrival time (ms)	1856	1803
Saccade amplitude (°)	3.3	3.7

The measures discussed so far relate to eye movement behaviour prior to the fixation of the target object in the photographic stimuli. As shown, there was no evidence of any difference in saccade patterns for scenes containing a consistent or inconsistent target object. As the target objects were carefully matched to ensure that the consistent and inconsistent objects were comparably sized and presented in the same location in the scene, the absence of any sign of an inconsistent object advantage during the free viewing of the scenes indicated that semantic consistency, as manipulated in these stimuli, did not result in different saccade patterns prior to target fixation.

Target fixation

The measures of fixation time on the target object were also analysed and the data are summarised in Table 5.5. The mean first fixation duration on consistent targets was shorter than the fixation duration for inconsistent targets, being 380ms and 433ms respectively, but this 53ms difference was not statistically significant ($t(46)=-1.56, p=.13$). First pass fixation durations did show a significant effect, with significantly longer fixation times on inconsistent targets than consistent targets ($t(46)=-2.58, p=.013$). Similarly, longer total fixation durations were found for inconsistent targets than consistent targets, with a significant 235ms difference between the mean values ($t(46)=-3.40, p=.001$).

These data indicated that no significant effects of consistency were found on measures obtained prior to the fixation of the target object and that the only effects of the consistency manipulation were on fixation times, with longer first pass and total fixation durations on inconsistent targets. The increased fixation times on inconsistent targets, compared to consistent targets, found in this experiment replicated the robust effect displayed in previous research. This effect also supported the categorisation of the inconsistent targets as semantically incompatible with the scene background, increasing processing difficulty.

Table 5.5: Summary table of mean data relating to measures of saccade behaviour during target object fixation.

	Consistent	Inconsistent
1st fixation duration (ms)	380	433
1st pass fixation duration (ms)	431	549*
Total fixation duration (ms)	775	1010**

* $p<.05$
** $p<.01$

Counterbalanced objects

As the targets used in the photographs were matched, it was possible to analyse the results for scene-target pairs by items. Each dependent measure was analysed to investigate whether there was any systematic variation either between the same target in a consistent and inconsistent context or matched targets in the same scene background. When the same objects in different scenes were analysed, there were no significant differences for any measures relating to behaviour prior to target object fixation. Only first pass fixation durations indicated a significant effect of consistency ($t(31)=-2.18$, $p=.037$), with inconsistent targets being fixated for 555ms compared to only 426ms for consistent targets. The same pattern was found with a 226ms difference for total fixation durations, which just failed to reach statistical significance ($t(31)=-2.00$, $p=.055$). These data indicated that the only significant effect of scene context was on first pass fixation durations for the same target in consistent and inconsistent scenes.

An analysis of paired target objects presented in the same scene background was also conducted. Again, no significant effects of consistency were found for measures obtained before target fixation. The only significant differences between paired consistent and inconsistent targets in the same scene were found for fixation measures. As found in the previous analyses, first pass fixation durations were 131ms longer on inconsistent targets than consistent targets, 555ms and 424ms respectively ($t(31)=-2.61$, $p=.014$). Similarly, total fixation durations were also longer on inconsistent targets (962ms) than on consistent targets (717ms) (*t*(31)=-4.05, *p*<.001). The analysis of matched pairs of targets in the photographic images failed to indicate any evidence of differential viewing behaviour on consistent and inconsistent objects in scenes, prior to their direct fixation.

Object size

The variable of object size was also investigated for photographic trials. The objects were assigned to the same groups as in previous experiments and the object sizes were the same as in Experiment 6, as the viewing distance and scene size were the same in both experiments. Therefore, small objects had a squared pixel area less than 2,500, up to 1.7°. Medium sized objects were between 2,500 and 5,000 pixels square, approximately between 1.7° and 3.5°, while large objects had pixel areas over 5,000, over 3.5° square.

A multivariate ANOVA (by items) investigated the effects of target object size and consistency on the measures obtained both before and during target fixation and these data are presented in Table 5.6. No significant effect of object size was found for either the number of saccades executed before fixating the target (*F*(2,56)<1, *p*=.87) or the arrival time on the target (*F*(2,56)<1, *p*=.85). There was no evidence that the size of the target affected when it was fixated during the course of a trial. However, a significant effect was found for saccade amplitude (*F*(2,56)=7.96, *p*=.001), with large objects being saccaded to from further away than small or medium sized objects. This result indicated that larger objects were selected as saccade targets from further extrafoveal vision than smaller targets.

Table 5.6: Summary table of data according to target object size.

	Small	Medium	Large
Number of saccades	5.53	5.18	5.09
Arrival time (ms)	1932	1810	1753
Saccade amplitude (°)	3.02	3.14	4.36**
1st fixation duration (ms)	477	388	316*
1st pass fixation duration (ms)	541	427	488
Total fixation duration (ms)	1038	751	809

* *p*<.05
** *p*<.01

For fixation measures on the target object, the only significant effect of object size was on first fixation durations (*F*(2,56)=3.37, *p*=.042). The data indicated that smaller targets were fixated for longer than larger targets. However, this effect was not found for first pass fixation durations (*F*(2,56)<1, *p*=.44) nor for total fixation durations (*F*(2,56)=1.90, *p*=.16). Therefore, the significant effect could be explained by larger targets being subject to shorter first fixation durations, as saccades were initiated more rapidly, probably to fixate a different region of the target.

The interaction between consistency and target object size did not produce any significant effects on any of the measures identified above. Object size only affected the dependent measures, as a main effect, on saccade amplitude and first fixation durations. Therefore the data did not indicate any difference between performance on consistent and inconsistent targets of any size.

Comparing line drawings and photographs

The results of the analyses are provided in Table 5.7, as a comparison between the data resulting from line drawings and photographs. The probability of fixating the target object during the course of the trial was lower for photographs than line drawings. This difference could be explained by the greater number of items in the photographic images and the scenes' greater complexity. With a larger amount of visual information present in these images, the target objects might not have been directly fixated during the trial, but the simpler line drawings contained fewer items for fixation, resulting in a higher probability of target fixation.

The number of saccades executed before target fixation did not differ for line drawings and photographs. This similarity indicated that participants could select fixation targets as easily in photographs as they could in line drawings, with no facilitation generated by the simplistic nature of the image. However, target objects in the photographs were fixated approximately 369ms later than the targets in the line drawing images, suggesting that, since the number of saccades executed were approximately the same for both scene types, fixation durations on distractor items were longer in the photographs. This effect could be attributed to the greater amount of visual information available for processing at each fixation. Saccade amplitudes were also found to be approximately equal in line drawings and photographs.

First fixation durations on consistent and inconsistent targets in line drawings evidenced a significant effect of consistency which was not found for targets in photographs. The increase in mean first fixation duration on inconsistent targets in line drawings suggested an increase in processing difficulty when fixating these targets. This difficulty could be attributed to the objects' semantic inconsistency with the scene context or, more realistically, to the difficulty in identifying the object itself from its ambiguous visual features. No significant effect of semantic consistency was found for likely and unlikely objects in photographs.

First pass fixation durations indicated a significant effect of semantic consistency only for photographic images. Inconsistent targets in photographs were fixated for longer than consistent targets. This pattern was also found for both stimuli sets in the total fixation duration measure. The data exhibited reliable effects supporting the longer fixation of inconsistent targets in both line drawings and photographs of scenes, which helps to confirm the inconsistent objects' suitability as incompatible targets.

Also of interest was the indication that targets in line drawings were fixated for longer than targets in photographs, which was clearest at first pass and total fixation durations and which was contrary to the assumption that the increased visual information in photographs would result in longer fixation times. This difference could indicate greater processing difficulty in identifying both consistent and inconsistent targets in line drawings. The shorter fixations on targets in photographic scenes could be explained by the relative ease in identifying a photograph of a household object compared to a line drawing of a less familiar object.

The increased fixation times on line drawing targets could also reflect the scarcity of items in the scenes, resulting in participants refixating the few discrete objects present in the image, including the targets, regardless of their consistency. This hypothesis was generated by the participants themselves during the experimental debriefing, which consisted of an informal discussion about the purposes of the experiment and their experiences during scene viewing. When asked whether they had experienced any differences between the two trial blocks, approximately one third of participants suggested that the line drawings had been displayed for more time than the photographs, as they had perceived the viewing of the line drawings as lasting longer. While they did not always have sufficient time to thoroughly observe the photographic scenes, participants reported the impression that they had searched the line drawings thoroughly and fixated the component items more than once during the 7000ms trials.

Table 5.7: Summary of results for consistent and inconsistent line drawings and photographs.

Measure	Line drawings		Photographs	
	Consistent	Inconsistent	Consistent	Inconsistent
Probability of target fixation (%)	91.7	93.2	89.6	85.4
Number of saccades	4.5	5.2	5.2	5.2
Arrival time (ms)	1309	1613	1856	1803
Saccade amplitude (°)	3.7	3.8	3.3	3.7
First fixation duration (ms)	383	550**	380	433
First pass fixation duration (ms)	573	718	431	549*
Total fixation duration (ms)	1020	1244*	775	1010**

* $p<.05$
** $p<.01$

5.4 Discussion

The aim of this experiment was to investigate whether patterns of eye movement behaviour showed sensitivity to semantic inconsistency in either line drawings or photographs of scenes. For the line drawing images, there was no significant effect of consistency on measures relating to behaviour prior to target object fixation, suggesting that the semantic inconsistency in trials containing an inconsistent target object was not detected before target fixation. Once the target was directly fixated, differences in fixation times emerged, with inconsistent targets being fixated for longer than consistent targets. In the investigation of object size, no effect was found on measures obtained prior to target fixation but again significant effects were found on fixation times, as first fixation durations on large objects were shorter than on small or medium sized objects and total fixation durations were longer.

These data were related to the findings of Experiment 1, which indicated an inconsistent object advantage when the target was presented extrafoveally. It was postulated that visual disturbances introduced into the inconsistent scene images could have caused this effect but this hypothesis was not supported by the results of Experiment 2 in which the images were inverted. The findings of the current Experiment 7 also failed to support this hypothesis, as no difference was found in eye movement behaviour between consistent and inconsistent scene viewing. Any differences in visual salience which could discriminate between consistent and inconsistent scenes during a 120ms presentation would be expected to persist during longer trial durations and to affect the properties of the saccade directed to the target, if not earlier saccade behaviour.

Similar results were produced from the display of photographic stimuli. No difference in saccade behaviour was found prior to target fixation, with the only consistency effects indicating longer fixation times on inconsistent targets than consistent targets. Again, target object size did not interact with consistency but two main effects of size were found. Larger saccades were directed towards larger target objects and shorter first fixation durations were found on larger targets. This difference in fixation times could be explained by the greater need to fixate a different region of a large object and to continue object processing through the execution of an intra-object saccade. In comparison, an inter-object saccade directed to an entirely different region of the scene, once object processing was completed on a small object, could be initiated later than an intra-object saccade. The absence of a consistency effect in this experiment was compatible with the results obtained from Experiments 3, 4, 5 and 6.

Little difference was found between eye movement behaviour for line drawings and photographs. The target object in each scene was directly fixated in the majority of trials and there was no difference between the likelihood of fixating the target at least once during the trial according to its consistency. Photographic targets were marginally less likely to be fixated than line drawing targets, a difference which was attributed to the composition of the scenes, as the photographs contained many more distractor objects.

The time taken to fixate the target object varied slightly according to the scene type, with targets in photographic scenes taking slightly longer to fixate than targets in line drawings. As the mean numbers of saccades executed prior to target fixation were the same for both stimuli types, this difference in arrival time was attributed to increased processing demands, due to a greater amount of visual information, lengthening fixations on distractor objects in photographs. The amplitude of the saccade directed at the target object was also comparable across scene types, indicating that saccades under both conditions were approximately the same size and were not affected by the level of visual detail included in the images. Fixation times on line drawings and photographic targets were approximately equal but total fixation times were slightly shorter for photographs. This discrepancy could be explained either by the greater processing difficulty in line drawings, with photographs of household objects being easier to identify than line drawings of less familiar stimuli, or by scene composition, with the repeated fixation of targets in line drawings when few other items of interest were presented in the images.

The issue of whether semantic information could be detected from extrafoveal vision was related to whether this information, if detected, could be used to direct eye movements. The brief presentations experiments provided no evidence that semantic information could be processed from extrafoveal vision when viewing natural scenes. This scene viewing experiment confirmed this result, providing no evidence that semantic inconsistency, as manipulated in the available stimuli, could be detected in extrafoveal vision when viewing scene images with no overt task requirements. Therefore, semantic information was not believed to be used in the selection of saccade targets, to direct saccades to regions of semantic inconsistency during natural scene viewing.

The current findings were compatible with the results obtained by previous researchers, such as De Graef et al (1990) and Henderson et al (1999), who also failed to find evidence of the earlier fixation of inconsistent objects compared to consistent objects in scenes, as suggested by Loftus and Mackworth (1978). Loftus and Mackworth's significant consistency effect and the subsequent inability to replicate it suggested that the conditions under which inconsistent objects can be detected from extrafoveal vision and saccaded to preferentially over consistent objects must be extremely limited and not replicable using realistic scene stimuli. The robust inconsistent object advantage found in different tasks by Hollingworth and Henderson and attributed to the preferential retention of semantically inconsistent information also contributed to this debate, as the effects have only been replicable using the researchers' own scene stimuli and not more naturalistic images.

The absence of evidence that semantic inconsistency was used to direct saccades in this experiment did not indicate that semantic information could not theoretically be used in this way, under different experimental conditions. Although the data obtained in Experiment 7 conclusively indicated that this did not occur when viewing images naturally, with no explicit intentions, these results did not negate Hollingworth and Henderson's data illustrating reliable consistency effects in different experimental tasks. The data presented in this thesis failed to replicate their effects using more naturalistic stimuli and viewing conditions, suggesting that semantic information may be accessible only under certain conditions. Participants may be able to selectively direct saccades to inconsistent regions of a scene and to make use of extrafoveal visual processing under specific conditions, such as when viewing simplistic stimuli or anticipating the presence of inconsistent objects. This proposal could explain the discrepancy in results across experiments which have failed to provide reliable and compatible results.

Chapter 6

General Discussion

As defined in the introduction, the purpose of the series of experiments described in this thesis has been to address certain questions which have not been clearly answered by existing research. Although researchers such as Hollingworth and Henderson and De Graef and colleagues had investigated the processing of semantic inconsistency from scenes, some methodological problems were identified which have been addressed here. The current experiments investigated the role of foveal and extrafoveal vision by manipulating fixation position, to determine whether semantic information could be processed extrafoveally. Also, the applicability of consistency effects to real-life scene viewing was considered by creating complex photographic scenes for use as experimental stimuli.

6.1 Summary of Experiments 1 to 4

In Experiment 1, the investigation concerned the ability to obtain semantic information about consistent and inconsistent objects in scenes, from a single fixation at variable distances from the target object. The stimuli used were simple line drawings constructed from the Leuven line drawing library. As expected, performance on the object identification task was best when participants directly fixated the target object. The consistency manipulation was found to affect accuracy significantly at two fixation positions, with higher accuracy for *consistent* objects when directly fixated and higher accuracy for *inconsistent* objects when the target was located 3° from fixation. At other individual fixation positions, no significant effect was found but a significant interaction indicated better performance for inconsistent targets at extrafoveal positions.

However, an investigation into the recognisability of the images indicated that semantic processing was unlikely to have caused the consistency effects, as the scenes and targets presented could not be reliably identified by participants during extended scene viewing. Further analysis of the experimental data, using only a high quality subset of the scene images, failed to replicate the advantage for inconsistent objects presented extrafoveally. As these scenes were reliably identified by the majority of participants, they were the most likely candidates for inducing semantic consistency effects. The extinction of the significant inconsistent extrafoveal advantage found with the entire stimuli set indicated that it was unlikely to be generated by the processing of semantic information or the detection of semantic inconsistency.

Inconsistent scenes were created by embedding targets from other scenes into alternative, inconsistent backgrounds. To investigate whether this process could have introduced visual differences salient enough to give rise to the significant effect found, Experiment 2 repeated this experimental procedure with inverted image presentations. The inversion of the scenes was hypothesised to interfere with the identification of both the scenes and the target objects and consequently to inhibit the processing of scene and object semantics. However, this manipulation would not be expected to influence performance if participants were not using semantic information. If participants were simply matching visual features or detecting inconsistent objects better because of visual differences between them and the scenes in which they were located, the apparent ‘consistency effect’ facilitating extrafoveal inconsistent targets would be expected to remain.

Performance for inconsistent objects was no better than that for consistent objects at further eccentricities from fixation. A comparison of accuracy by consistency and fixation position across both experiments (Figure 2.11, page 109) suggested that performance for consistent trials was comparable across the upright and inverted viewing conditions. However, performance for inconsistent trials decreased at all extrafoveal fixation positions between the upright and inverted image conditions, but not when directly fixating the target.

This significant reduction in accuracy, when the images were inverted, was found only for inconsistent targets. This result indicated that the inversion of the scenes selectively influenced performance for inconsistent objects only, suggesting that the advantage evident in Experiment 1 could not be attributed to the facilitated detection of view-invariant visual features of inconsistent objects. Although the analysis of high quality data from Experiment 1 indicated that the detection of semantic inconsistency was unlikely to have generated the inconsistent object advantage, the results of Experiment 2 did not support the hypothesis that sufficient visual differences existed between consistent and inconsistent scenes to explain the effect.

The results of these two experiments indicated that fixation position relative to the target object did indeed influence performance on this task, as proposed. However, contrary to the suggestion that consistent objects would be facilitated over inconsistent objects in extrafoveal positions, inconsistent objects evidenced an advantage in Experiment 1, particularly at closer extrafoveal positions (3° from fixation). Whether this difference was modulated by the detection of semantic inconsistency remains unclear. The investigation of the recognisability of the images called this possibility into question and an analysis of the most appropriate images failed to replicate the consistency effect. However, the inversion of the scenes did extinguish the inconsistent object advantage, as would be expected if it was influenced at least in part by semantic processing.

Additionally, both experiments indicated that consistent objects were responded to more accurately than inconsistent objects when presented at fixation. The effect was significant in Experiment 1 but not in Experiment 2, although the same pattern was evidenced. This finding suggested that a consistent object advantage in foveal vision could be a reliable effect worthy of investigation in subsequent experiments.

Experiment 3 was designed to investigate whether the same effects could be obtained using the same experimental procedure with more naturalistic scene stimuli. Complex grey scale photographs of genuine household scenes were displayed and all evidence of a consistency effect was extinguished. Performance for consistent and inconsistent objects was almost entirely equal at all fixation positions and was better than that obtained with simple line drawings in Experiment 1. The analysis of a high quality subset of the visual stimuli also failed to evidence a significant main effect of consistency. Accuracy for inconsistent targets was slightly, but not significantly, higher than for consistent targets at fixation positions 0 and 1, corresponding to foveal and near foveal vision.

Unlike the results of Experiments 1 and 2, no advantage was found for consistent objects over inconsistent objects when directly fixated in the analysis of the entire stimuli set. This result suggested that the context of a simple line drawing could assist in the selection of a target object in this task but the context of a richer visual scene did not have a similar effect. It was possible that the relative ease of the task when the images were easily identifiable and naturalistic, reflected in the high levels of accuracy, prevented any further facilitation by contextual priming.

The failure to find any evidence of a consistency effect with complex photographs indicated that such effects could be restricted to the simplistic visual stimuli used in laboratory experiments, rather than being applicable to real world scene viewing. In order to determine whether the nature of the images influenced the expression of consistency effects, Experiment 4 replicated the same procedure using line drawing stimuli created from the photographs used in Experiment 3. Performance was significantly worse with these simplified images than with the photographic scenes, with accuracy decreased by at least 10% at each fixation position. This decrease in accuracy indicated that the task was inherently more difficult when viewing simple line drawings than complex and realistic images.

However, there was still no evidence of a consistency effect with these simplified images, which more closely resembled the scenes used by previous researchers. Even when the high quality stimuli were analysed (i.e. those which were recognisable and rated as most consistent and inconsistent), no reliable effects of consistency were found. Performance for inconsistent targets was never higher than that for consistent targets. The marginally significant interaction between fixation position and consistency appeared to reflect higher accuracy for consistent targets when the target was presented at fixation, at position 2 and position 4.

The results of Experiments 3 and 4 indicated that it was not the nature of the photographs themselves that resulted in the abolition of the consistency effect, found with the line drawing stimuli in Experiment 1. The conversion of the photographic scene images into line drawings did not elicit a significant advantage for inconsistent targets. Therefore, alternative explanations were considered, including the possibility that the composition of the photographic scenes, containing many more non-target objects than the Leuven set, affected the ability to process extrafoveal information and facilitate inconsistent object performance.

Object size

The variable of target object size was also investigated, as a post hoc analysis. As eccentricity was found to influence consistency effects, it was hypothesised that consistency effects might also be modulated by object size, which would influence an object's perceptibility at different retinal locations. In Experiment 1, a significant main effect of object size on performance was found, with highest accuracy for medium sized objects, followed by large objects and then small objects. The absence of an increase in accuracy for large objects over medium objects could be explained by the fact that fewer objects were assigned to the 'large' category, which would have affected the reliability of the mean values obtained.

When the size categories were analysed separately, small objects showed no effect of the consistency manipulation. Medium sized objects produced a consistent object advantage when presented at fixation which was almost significant but, at every other fixation position, accuracy for inconsistent objects was slightly but not significantly higher than that for consistent objects.

Large objects displayed a significant main effect of consistency, with much better performance for inconsistent targets than consistent targets when the targets were not presented at fixation. When the targets were directly fixated, accuracy for consistent objects was slightly higher than that for inconsistent objects. This analysis confirmed that the variable of object size could also affect the expression of consistency effects, as medium and large sized objects indicated an increase in accuracy for inconsistent targets presented extrafoveally while small objects did not.

The inverted line drawings presented in Experiment 2 displayed no similar effects of object size influencing consistency effects. Again, accuracy for medium sized objects was higher than that for large objects, with small objects displaying the lowest accuracy, replicating the effects of object size on accuracy in Experiment 1. However, no difference in performance was found according to consistency for objects of any size. This extinction of any consistency effects modulated by object size supports the hypothesis that the inversion of the images interfered with the perception of semantic inconsistency, so performance for consistent and inconsistent targets was equal for all object sizes.

Object size effects were found in Experiment 3, which indicated that object size also influenced performance when viewing complex, naturalistic scenes. Accuracy levels for small and medium sized objects were approximately the same, with higher accuracy for large objects which appeared to be caused by a significant increase in accuracy for inconsistent large objects. The analysis of individual size categories indicated that, like the line drawing images, small objects did not exhibit any effects of consistency. Medium sized objects indicated a significant advantage for consistent targets and large objects showed a significant advantage for inconsistent targets. Again, this analysis indicated that object size could influence consistency effects.

Although small objects showed no effects of consistency, the consistent object advantage for medium sized objects was found across all fixation positions, so eccentricity did not influence the improved performance for medium sized consistent targets. Similarly, the inconsistent advantage for large objects was also displayed at all fixation positions, indicating that the effect was robust across target location. These results were compatible but slightly different to those obtained from the line drawings in Experiment 1.

Experiment 1 exhibited at least a trend towards an inconsistent object advantage for both medium sized and large objects, but for both object sizes, performance when the target was presented at fixation was better for consistent targets than inconsistent targets. The data from the photographic stimuli used in Experiment 3 exhibited a distinction between a consistent object advantage for medium sized targets and an inconsistent object advantage for large targets. The trend towards an inconsistent object advantage for medium sized line drawing targets, rather than the consistent object advantage evidenced with the photographic scenes, could be explained by differences in object sizes.

The terms 'small', 'medium' and 'large' were defined with respect to the range of the stimuli within each experiment, rather than referring to an absolute size across experiments. While medium sized line drawing targets subtended between 7° and 16° square, medium sized photographic targets subtended between 4° and 8° square only. Therefore, medium sized line drawing targets were equivalent in size to large photographic targets, which also displayed an inconsistent object advantage. Small line drawing targets may not have displayed any clear effects of consistency because they contained objects of a larger size range (up to 7° square) and smaller objects would not be expected to display consistency effects.

It could also be significant that performance for foveally presented targets in line drawings was facilitated for consistent targets compared to inconsistent targets, but this effect was not seen in photographs. As explained previously, the semantic relationship between the scene and the target could have helped participants identify and recognise the target object presented in the two-alternative forced-choice when the images were line drawings. However, with more complex and recognisable photographic images, the facilitation provided by the consistent scene context might not have enhanced performance above the level obtained during the direct foveal fixation of a readily identifiable target, which possibly approached ceiling level.

The previous results could be compared to the analysis of object size in Experiment 4, when the images were derived from the photographs but presented as line drawings. Although the objects assigned to each size category were different in this experiment, the same size groupings were used so the results could be compared to those obtained in Experiment 3. Experiment 4 data indicated that performance was significantly affected by object size. There was a monotonic increase in accuracy across increasing object size categories.

Again, small objects displayed no clear effects of object size and no facilitation for consistent targets when presented directly at fixation. Medium sized objects produced a significant consistency effect, with performance for consistent objects higher than that for inconsistent objects at fixation positions 0, 1 and 2 (up to 6°), after which performance fell to chance levels for all targets. However, contrary to previous findings, no consistency effect was evidenced for large objects at all.

These results suggested that the conversion of the images to line drawings did not influence the effects of consistency on small or medium sized objects but extinguished the significant facilitation for inconsistent large objects over consistent large objects. It was possible that the recognisability of the images was compromised, making line drawings of large targets more difficult to identify than line drawings of medium sized or small targets. The data suggested that some distinction existed between photographs and line drawings of larger targets but, as the targets were not necessarily the same in both stimuli, the cause of this is unknown.

Interpreting consistency effects

The advantage evidenced in Experiment 1 deserves to be treated with caution, as additional investigation into both that stimuli set and the photographs indicated that the Leuven line drawings were not immediately recognisable during such brief presentations. Three quarters of the Leuven images consisted of objects and/or scene backgrounds which could not be fully identified and categorised, compared to less than one third of the photographs. For this reason, it was considered that the advantage evidenced in Experiment 1 for objects classified as inconsistent may not have been generated by the detection of semantic inconsistency between the object and the scene.

Some alternative explanations were proposed, including the possibility that the differences between the consistent and inconsistent targets were visual rather than semantic. This proposal was supported by the analysis of the high quality subset of line drawing stimuli, which failed to replicate a consistency effect when only readily recognisable and appropriately rated consistent and inconsistent scenes were presented. However, the inversion of the images in Experiment 2, hypothesised to interfere with semantic processing, also failed to replicate an advantage for inconsistent objects.

The further absence of a consistency effect when displaying photographs and their line drawings could be attributed to the differences between the images used in these experiments. Although Experiment 4 disproved the suggestion that consistency effects were only evidenced with line drawings, additional differences between the scenes existed. The photographic scenes contained many more non-target objects and this increase in visual detail was included in the line drawings of photographs. Compared to the simpler Leuven line drawing images, this additional visual information in the form of discrete distractor objects may have interfered with the extrafoveal processing of the visual detail distinguishing between consistent and inconsistent objects.

Alternatively, the consistency effect in Experiment 1 could be an artefact resulting from the use of visual stimuli which were, in several cases, unidentifiable by the participants. This proposal may explain the results of experiments by other researchers, who used experimental stimuli from the same source, demonstrating clear facilitation for inconsistent objects. It would be advisable to investigate this further, possibly by comparing performance on the Leuven inconsistent objects with non-objects located in the scene. These conditions may be comparable if the inconsistent objects were considered by participants to be object-like figures with no readily identifiable semantic associations.

Even if the consistency effect in Experiment 1 is robust, the same effect was not evidenced using more naturalistic stimuli. It may be argued that the inconsistent items used in the photographic stimuli were not rated as 'sufficiently inconsistent' as they were located 'improbably' rather than 'impossibly'. However, Boyce, Pollatsek and Rayner (1989) argued that the level of consistency or inconsistency between a scene and an object did not modulate the consistency effect found, with the important detail being the plausibility and implausibility of the items, not their predictability. This conclusion would imply that the fact that participants rated the consistent targets as likely and the inconsistent targets as unlikely would be sufficient to exhibit any consistency effects.

The inconsistent objects used in images by other researchers, which were considered impossible rather than unlikely, were not sufficiently realistic to be suitable for use. For example, many inconsistent objects were large, fixed and static, such as a swing, a lectern or a shower head. These objects simply could not realistically be located in inconsistent scenes in real life, such as finding a barge in a street or a fire hydrant in a living room. Therefore, even if a facilitation effect for inconsistent objects is robust under such specific probability manipulations, it could be argued that the finding contributes to our understanding of real-world visual processing in only very limited circumstances. Any consistency effect of applicable and general interest would need to be demonstrated with realistic and plausible stimuli and manipulations.

6.2 Summary of Experiments 5 and 6

In these experiments, the intention was to determine whether the findings of Hollingworth and Henderson's (2000) change detection experiments, which indicated a reliable inconsistent object advantage, could be replicated with photographic stimuli. Experiment 5 used the photographs designed for Experiment 3 and required participants to make a two-alternative forced-choice of whether a change, identified as the appearance or disappearance of a target object, had occurred across two brief scene presentations. There was no evidence that accuracy on this task was influenced by consistency, with the only significant effect being a consistent object advantage when the changing object was directly fixated. This result was compatible with similar effects found with the line drawing stimuli in previous experiments, indicating that the task was facilitated when the scene context and the target object were semantically consistent and the target was directly fixated.

The results of this experiment did not replicate Hollingworth and Henderson's findings of a significant change detection advantage when the changing target was inconsistent. They concluded that semantic inconsistency facilitated the detection of changes to inconsistent targets, but this effect could not be replicated with naturalistic photographs. The methodological differences between the experiments seemed insufficient to explain the absence of an effect, so the scene images themselves were investigated. As the strength of the consistency manipulation did not appear to modulate consistency effects when investigated in previous experiments, it was considered unlikely to affect performance on this task.

Instead, the composition of the photographic scenes, containing many non-target objects, could have influenced the ability to process information relating to consistent and inconsistent targets from a brief 250ms scene presentation. This proposal was supported by a comparison of results from Hollingworth and Henderson's (2000) Experiments 2 and 3. Both experiments involved a two-exposure change detection task but one or two additional non-target objects were added to each scene image for Experiment 3.

Although consistency effects were found in both experiments, accuracy for detecting a left-right orientation change decreased from approximately 52% for the simpler scenes, to approximately 41% for the scenes containing additional non-target objects. This decrease implied that the number of items in the scenes influenced performance on this task, which encouraged the processing of the entire scene. This finding supports the hypothesis that the processing of information from extrafoveal vision could be influenced by the composition of the scenes and the greatly increased complexity of the photographic scene images could possibly affect the expression of consistency effects.

The results of Experiment 5 could be compared to the results of Experiment 3, both of which investigated the ability to obtain and retain information about a peripherally presented target object. In Experiment 3, performance was accurate over the range of eccentricities investigated from an exposure lasting 120ms, but no difference was found between consistent and inconsistent objects. In Experiment 5, the exposure duration was 250ms for each image, replicating Hollingworth and Henderson's procedure.

Although the fixation position could only be confirmed for the duration of the first image, change detection would require the processing of the target during the first scene presentation. Therefore an estimate could be made about the eccentricity in the first image at which it was possible to process the target sufficiently to detect its disappearance upon presentation of the second image. Both experiments indicated no consistency effect for photographic images containing extrafoveal targets presented for durations up to 250ms. These compatible results suggested that the semantic consistency of objects in scenes was not usually detected from such brief presentations and that objects were not subject to preferential processing on this basis.

Further analyses on target object size also failed to provide evidence of any consistency effects in Experiment 5. Object size did not affect accuracy significantly, which remained approximately 55% for all object sizes. Slight differences were seen when the object categories were analysed separately. Again, small objects showed no evidence of any effects influenced by semantic consistency. However, a non-significant trend towards better performance for consistent medium sized targets than inconsistent medium sized targets was indicated. A similar non-significant trend was found for large targets, suggesting that inconsistent changing targets were detected more often than consistent changing targets. This pattern of a consistent object advantage for medium sized targets and an inconsistent object advantage for inconsistent targets was compatible with the results of Experiment 3.

Experiment 6 investigated the somewhat different issue of whether consistency affected saccade behaviour during a change detection task. In the traditional flicker paradigm employed, the two image versions alternated until a change was detected and the participant terminated the trial. Hollingworth and Henderson presented evidence for the faster detection of changing inconsistent objects than changing consistent objects, measured by response time. However, Experiment 6 again failed to replicate these results, with no significant evidence of the faster detection of inconsistent changing objects, compared to consistent changing objects.

The analysis of saccade behaviour during these trials confirmed the absence of any consistency effects. Measures relating to saccade behaviour prior to the direct fixation of the target object indicated that mean arrival times on consistent and inconsistent objects were equal. Also, the size of the saccade directed at the target was not affected by the consistency of the target object. This result indicated that inconsistent objects were not selected as saccade targets from further extrafoveal vision than consistent targets, or vice versa. Finally, there was no effect of consistency on fixation times on the changing targets. The manual response terminating the trial once a change was detected was believed to have affected the fixation measures by truncating fixations before a saccade was initiated away from the target.

The data were analysed to investigate whether object size affected response time or arrival time on the target. However, no obvious effects of object size or consistency were found on either variable. Response times for small objects were slightly longer than those for medium sized and large objects, but not significantly so. The consistency of the target did not influence response times either. Similarly, the arrival time on the target was clearly not affected by semantic consistency. The data indicated that, although object size modulated performance on brief presentations tasks, this variable had much less influence on scanning behaviour, suggesting that eye movements were not affected by target object size to any great extent.

The attempted replication of two of Hollingworth and Henderson's change detection experiments failed to support their findings that performance for changing inconsistent objects was facilitated over changing consistent objects. There was no evidence of the preferential processing or retention of inconsistent target information from brief presentations in Experiment 5. Similarly, Experiment 6 failed to confirm that changes to inconsistent targets were detected sooner than those to consistent targets, either in terms of response time or eye movement behaviour.

Possible explanations for the discrepancy between the results were considered and again involved the differences between the stimuli used. Although these experiments addressed two different issues within the investigation of semantic consistency, the results still indicated that conclusions reached on the basis of analyses using line drawing stimuli could not be replicated with realistic photographic stimuli. The robustness of these conclusions suggested that any effects of consistency obtained using stimuli which do not reflect natural scene viewing conditions may not be replicable in real life.

6.3 Summary of Experiment 7

This experiment compared eye movement behaviour when viewing simple line drawings of scenes and more complex photographs, in an attempt to determine whether the absence of consistency effects with the photographic stimuli could be attributed to the image type, rather than their composition. Any indication that the line drawing targets were more salient than the targets in photographs would have been evident from observing scan paths on these images, prior to the fixation of the target object. Additionally, differences in saccade behaviour between consistent and inconsistent scenes for either line drawings or photographs could provide evidence of the use of extrafoveally processed semantic information in selecting saccade targets.

No difference was found in the number of saccades executed before target fixation between consistent and inconsistent targets for either line drawings or photographs. This variable was also comparable across the stimuli types, indicating that photographic targets could be selected for fixation as easily as line drawing targets. The time taken to fixate the target also showed no evidence of consistency effects for either scene type, but photographic targets appeared to be fixated slightly later than line drawing targets, by about 350ms. As the number of fixations taken to fixate the target were approximately equal, this increase in time could reflect longer fixation durations on non-target scene regions in photographs. The amplitude of the saccade directed at the target object also failed to evidence significant differences across scene types or consistency conditions. The comparability of saccade size for line drawings and photographs indicated that the increased complexity of the photographic images did not detrimentally affect the processing of extrafoveal detail, by resulting in the execution of shorter saccades whose endpoint could be sufficiently processed prior to fixation.

Overall, few differences were found between saccade behaviour when viewing the two different image types. There was evidence of slight differences in fixation durations but not in the selection of saccade targets. Differences between semantic consistency conditions were not found prior to the fixation of the target object, providing no evidence of inconsistent objects being saccaded to preferentially, compared to consistent objects, in either line drawings or photographs. These results indicated that under these circumstances, any semantic information obtainable from extrafoveal vision was not used in the selection of saccade targets. Significant consistency effects were only found after target fixation, indicating that inconsistent objects were fixated for longer than consistent objects, which was compatible with all previous research. The data obtained in this experiment replicated the robust finding that, once detected, inconsistent objects were fixated for longer than consistent objects, but failed to support the more controversial suggestion that the semantic consistency of non-fixated objects could influence eye movement behaviour during natural scene viewing.

Object size was again investigated in this experiment and only two significant effects of object size were found. Saccade amplitude was influenced by the size of the target, with longer saccades being directed to large targets than to medium sized and small targets. This difference implied that large targets could be selected as saccade targets from further extrafoveal vision and were more salient in visual periphery. This effect was not modulated by consistency in any way. Additionally, target size influenced first fixation durations on targets. Significantly shorter fixations were found on large targets than medium sized and small targets, suggesting that first fixations on large targets were terminated sooner than those on smaller targets, possibly by instigating a further fixation on a different region of the target.

This experiment suggested that consistent and inconsistent objects were processed in an equivalent fashion until they were fixated. This result is compatible with the memory hypothesis and attentional disengagement hypothesis to explain inconsistent object advantages. Consistency effects would be generated through target fixation. Once inconsistent objects are fixated, they are usually subject to longer fixations, possibly in the effort to reconcile their semantic identity with that of the scene in which they are located.

From this study, it was clear that participants did not naturally fixate inconsistent objects any earlier than consistent objects when passively viewing a scene, regardless of its type and composition. This conclusion was supported by the results of Experiment 6, in which participants did not adopt the strategy of searching for inconsistent objects, even though an inconsistent object in the scene would have a 50% chance of changing, compared to a substantial number of consistent distractors, additional to the target, which never changed. The results of these two experiments comparing viewing behaviour over extended presentations confirmed that, subject to the specific scene stimuli investigated, there was no evidence of consistent and inconsistent objects being processed differently prior to direct fixation.

The results described so far indicated that the findings obtained by other researchers demonstrating an inconsistent object advantage were not replicated using different experimental stimuli which were more naturalistic than simple line drawings of sparsely populated scenes. The data exhibiting an inconsistent object facilitation by other researchers appeared robust and reliable but could not be replicated using more complex scenes, although a significant effect was found using the Leuven line drawings.

The Leuven stimuli provided for use in this thesis did not contain all the images used by other researchers, who adapted existing stimuli and supplemented the Leuven set. For this reason, the suitability of other researchers' materials cannot be commented on beyond the analysis of the available images presented here. However, if the findings supporting an inconsistent object advantage using similar stimuli are indeed reliably reported, then this effect can only be replicated under very limited conditions. Therefore, the conclusion must be reached that such an effect is not found with more complex stimuli and is unlikely to occur during real life scene viewing.

6.4 Further research

The results obtained from the experiments contained in this thesis support previous findings that inconsistent objects in scenes are fixated for longer than consistent objects in scenes (e.g. Loftus and Mackworth, 1978; Friedman, 1979), suggesting that semantic inconsistency is detected upon fixation and then requires additional resources to integrate into a memory schema or representation. However, no robust support has been found for the conclusion that objects inconsistent with the scene context can be subject to preferential processing *before* direct fixation. The only evidence in this thesis supporting this proposal was the significant inconsistent object advantage for extrafoveally presented targets in Experiment 1. However, this conclusion was not supported by a number of additional findings.

The semantic identities of both the objects and the scene backgrounds were difficult to recognise, indicating that the effect was unlikely to be caused by the immediate detection of semantic inconsistency. The analysis of the high quality image subset confirmed this conclusion by failing to replicate an inconsistent object advantage when viewing only recognisable scene images. Finally, this inconsistent object advantage was not replicated in any of the subsequent experiments, suggesting that semantic consistency was not detected extrafoveally when viewing natural scenes.

The experimental data partially supported schema hypotheses, which predict that consistent objects will be facilitated when presented in scenes. A significant or slight consistent object advantage was exhibited when the target was presented directly at fixation in most of the experiments which manipulated fixation position (Experiments 1 to 5). Only Experiment 3, displaying complex photographs, failed to evidence this pattern of data. The persistence of this effect, particularly in difficult tasks such as those involving line drawings of scenes or change detection, suggested that the presence of a consistent scene gist did in fact facilitate the identification of the target object and performance on the task. Also, when participants were allowed to saccade around complex visual scenes, shorter fixation times were found on consistent than inconsistent objects, which may reflect an advantage due to the facilitatory context.

The limitations of these experiments gave rise to several additional areas of further research. To begin with, the scenes used to investigate semantic consistency effects need to be appropriate. This means that the scene and object semantics need to be accessible to participants when viewing them for brief periods of time. Additionally, the relationship between objects deemed consistent and inconsistent with the scene context needs to be evaluated in more detail.

Consistent and inconsistent targets need to be reliably rated as such by the participant population and the definition of the 'consistency' manipulation needs to be clarified. A clear distinction needs to be made between objects which could possibly be found in a given context but are unlikely to be and objects which could never be found in a specific context. While the strongest manipulation of defining inconsistent objects as items which would never be found in a given location would appear to be the most desirable, this would be impossible when using natural images. In real life, when objects are found in unlikely places, their presence is by definition possible, even if rarely seen. Therefore, creating scenes in which objects were located in impossible contexts might enhance the consistency manipulation but would not reflect realistic scene viewing conditions.

During the process of acquiring consistency ratings for the scenes used in these experiments, it became clear that this method of obtaining confirmation of object categorisation was subject to a substantial amount of variability. There was not always complete agreement on whether a given target was likely to appear in a scene or not, as the definition of likely and unlikely was determined to some extent by each individual participant. The student body from which participants were recruited contained some individuals who considered few household objects to be extremely unlikely when found in another household scene. As discussed previously, student participants may not have found the items sufficiently surprising or unusual to be considered highly inconsistent, possibly contrary to the views of the general, non-student population. For these reasons, it is important to obtain a more precise and objective measure of association between the objects and scenes.

This relationship could be investigated using a priming technique, similar to that reported by Palmer (1975), in which the time taken to identify an object was affected by the prior presentation of a scene. A related scene would ‘prime’ the identification of the object, facilitating recognition of objects consistent with the scene context. An unrelated context decreased accuracy in object identification, by priming the identification of objects consistent with the scene context. The application of this procedure to the line drawing and photographic stimuli used in these experiments would address important issues in the suitability of both stimuli sets.

By investigating the images in the Leuven set, it would be possible to determine whether the scene backgrounds provided sufficient contextual information to prime the identification of the objects selected as consistent targets. Existing analysis of the line drawing scenes indicated that the scene backgrounds and targets were not always readily identifiable, so this further investigation could confirm whether stimuli were sufficiently recognisable to exhibit priming effects. The appropriateness of the inconsistent objects could also be tested by investigating naming performance on these objects, compared to consistent objects. This investigation could determine whether each scene background generated enough contextual information to distinguish between naming latencies obtained for consistent and inconsistent objects.

This paradigm could also be applied to the photographic stimuli. It could confirm whether the consistent relationship between the consistent objects and scenes was sufficient to elicit a naming latency facilitation, compared to inconsistent objects in scenes. Data exhibiting a distinction between naming latencies obtained for objects preceded by a consistent or an inconsistent scene prime would prove that the categorisation of objects into consistent and inconsistent targets was appropriate. In this way, the assignment of consistent and inconsistent objects, for both line drawings and photographic scenes, could be confirmed by a more objective and rigorous measure of relatedness which would be less subject to individual variability.

Another issue requiring further clarification through future research was the discrepancy between the results of Experiments 1 and 4. Although both made use of line drawings of scenes as experimental stimuli, Experiment 1 found an inconsistent object advantage at extrafoveal locations while Experiment 4 did not. The procedures were identical so the differences can only be attributed to the nature of the visual stimuli. The advantage demonstrated for inconsistent objects when viewing Leuven line drawing stimuli would need to be subject to further investigation.

The inconsistent object advantage in Experiment 1 could be considered genuine, indicating true facilitation for inconsistent objects in scenes. Then, the issue of interest would be to explain why this same advantage was not evidenced for the line drawings of photographs. The target objects used in the photographic stimuli might not have been considered inconsistent enough to stimulate the detection of regions of semantic inconsistency, although fixation times on inconsistent targets were found to be longer than on consistent targets in Experiment 7. This distinction suggested that the semantic relationship between the object and the scene was manipulated appropriately. If the consistency manipulation were not strong enough, consistency effects would not be expected in realistic viewing situations and the advantage in Experiment 1 could be explained as a laboratory phenomenon occurring under only the most specific and limited conditions.

Additional differences between the two sets of line drawing stimuli could explain the lack of an advantage when using images derived from photographs. The complexity of these line drawings was greater than that in the Leuven stimuli set, with a greater number of objects present in the scene and more detailed depiction of the background. This increase in visual detail may have interfered with the extrafoveal processing necessary for detecting objects not presented at fixation and affected the ability to detect regions of semantic inconsistency. In order to investigate this, it would be desirable to create simpler line drawings from those originally created from photographs, possibly by removing textural regions and surplus non-target objects. If these stimuli also failed to demonstrate a similar inconsistent object advantage, it would indicate that the Leuven stimuli were distinguished in a different way, possibly in the creation of the scenes by moving objects from one background to the other.

Alternatively, the inconsistent object advantage evidenced in Experiment 1 may not have resulted from the genuine detection and preferential processing of regions of semantic inconsistency within 120ms. The difference between consistent and inconsistent targets could be unrelated to their semantic identity. Although the results of Experiment 2 suggested that the differences were not simple view-invariant visual details, other options should be considered before concluding that semantic consistency caused the effect.

Consistent and inconsistent targets could have differed in recognisability, so performance could have been affected by the difficulty in identifying inconsistent targets, rather than by their actual identification as inconsistent. Line drawings of inconsistent objects were suggested to be more difficult to reliably identify than objects in consistent scenes (see Appendix A), as a consistent scene appeared to facilitate the identification of some targets. It is possible that the unusual features of the inconsistent targets, which could not be reliably identified even from extended foveal viewing, could be detected from extrafoveal vision, generating the inconsistent object advantage at extrafoveal locations.

Under this hypothesis, the inversion of the line drawing images could have affected performance on inconsistent objects, by removing their 'distinctiveness'. Inverted inconsistent targets would have been more difficult to identify but this would also have been true of consistent targets, which could extinguish any advantage for the relatively less recognisable inconsistent targets. Performance on both consistent and inconsistent targets, rendered equally difficult to identify from inverted images, would be expected to be equal, as found.

Admittedly, this hypothesis relies on objects which were difficult to identify being selectively processed within a scene presentation of 120ms. Although there was no direct evidence that this could occur, the hypothesised increase in salience for inconsistent objects would be in the nature of visual features, rather than semantic features, which could be detected from extrafoveal vision. The study of non-objects in scenes, being discrete object-like items with no semantic associations, could shed light on the matter. Further investigation could include a replication of the experimental design applied to Experiments 1 to 4, providing a comparison of performance for consistent objects, inconsistent objects and non-objects. This investigation could determine whether the advantage found in Experiment 1 was generated by facilitation for inconsistent objects or inhibition for consistent objects, relative to a control condition.

The body of evidence contained in this thesis could be summarised to conclude that although inconsistent objects are fixated for longer than consistent objects, no evidence was found to support the proposed ability to detect them, prior to fixation, from extrafoveal vision. Inconsistent objects were not subject to preferential processing within a single fixation on a scene and no difference was found in saccade behaviour for consistent and inconsistent scenes either when viewing scenes passively or when conducting a change detection task. Participants did not make use of semantic information during scene viewing under these conditions.

These studies could be continued to their logical conclusion to investigate whether participants *can* make use of semantic information voluntarily, when explicitly instructed to do so. Although there was no evidence that participants spontaneously processed semantic information from extrafoveal vision or used this information to direct saccades, it is not clear at present whether this could occur with specific intent. If this were possible, it may explain why only some previous research found evidence for a difference in performance between consistent and inconsistent objects, as the processing of semantic information may have been affected by experimental instructions and individual participants' motivations.

To fully answer this question, further investigations would need to determine whether participants could make use of semantic information available beyond the current fixation position, if explicitly required to do so. To this end, the simplest experimental design would provide explicit instructions for participants to search through a realistic display for items which are semantically inconsistent with the scene context. It would need to be emphasised that there were no additional tasks to be completed at the end of the experiment and the definition of inconsistency would need to be very clear. For example, if participants suspected that inconsistent objects appeared in unlikely locations within the scene, this might trigger a search of unusual spatial locations (such as under a surface or 'hidden' objects behind other items). Therefore, a task which required participants to fixate a semantically unusual item in an ordinary household scene might provide an adequate test of whether semantic information can be processed from extrafoveal vision when viewing complex scenes.

These suggestions for further research should assist in the investigation of the detection of semantic inconsistency in scenes. It seems clear, from the work conducted so far on naturalistic images, that this effect does not appear in plausible and realistic conditions, mimicking natural images within an experimental construct. However, the work contained in this thesis has also contributed to this field by identifying additional variables which need to be considered when investigating semantic consistency effects. These experiments have confirmed that consistency effects can indeed be modulated by target location, relative to fixation, as suggested by De Graef's (1998) investigation. The expression of both a consistent object advantage and an inconsistent object advantage within the same data analysis indicated that the perception of semantic information could be influenced by the eccentricity of the target from fixation.

In addition, the novel discovery of the effects of target object size also needs to be taken into account in the investigation of semantic consistency effects. Although the size of the targets used in these studies was not carefully controlled in the original design, post-hoc analyses have indicated that the size of the target, sometimes together with target eccentricity, can modulate the expression of both consistent and inconsistent object advantages. Further investigation of these effects, using more carefully controlled target sizes, would be desirable and the re-analysis of previous research could indicate whether target size could affect performance using different experimental stimuli. If robust effects exist, this variable could also explain different consistent and inconsistent object advantages evidenced in different experiments.

6.5 Research questions

Although the results obtained from the experiments contained in this thesis have been discussed and issues for further research have been identified, it is also important to consider how the data, as a whole, address the issues identified in the introduction as worthy of further investigation (page 72). The specific issues raised in the introduction will be answered from the available data and the main findings will be summarised.

Do inconsistent objects in scenes evidence a detection advantage upon first fixation on a scene and is this effect influenced by its location relative to fixation position?

Objects considered to be inconsistent with the scene context did demonstrate an advantage under specific conditions, but only when viewing simple line drawings from the Leuven library in Experiment 1. The effect was found to be partially influenced by the object's position relative to the current fixation but not in the manner predicted. Consistent objects showed an advantage over inconsistent objects when they were presented directly at fixation. At other fixation positions, inconsistent objects were detected better and a significant interaction between consistency and fixation position was exhibited.

This effect also appeared to be influenced by object size, as only medium and large sized targets indicated any difference in performance between consistent and inconsistent trials for line drawing stimuli. No effect of consistency was found for small target objects, a small but non-significant advantage for inconsistent targets was found for medium sized objects and a clear advantage for inconsistent targets was found for large objects. These data suggested that targets varying in size over a large range could mask evidence of consistency effects within specific size categories. Analysis of the photographic stimuli confirmed the effects of object size, with small objects again failing to evidence consistency effects, medium sized objects showing an advantage for consistent targets and large objects displaying an advantage for inconsistent targets.

The most reliable effect of consistency was the advantage for consistent targets when presented directly at fixation. This effect was found for the Leuven line drawing stimuli in Experiment 1 and was expressed to a greater or lesser extent in nearly all experiments controlling for initial position (Experiments 1, 2, 4 and 5). Experiment 3, using photographs of scenes, did not exhibit this advantage for consistent targets presented at fixation.

The pervasiveness of this effect across stimuli types suggests that the effect is largely robust, with a consistent scene context facilitating the identification of a consistent object, compared to an inconsistent object, when presented at fixation. The failure to find this effect in Experiment 3 might have been caused by ceiling effects, brought about by the relative ease of identifying a foveally presented target when it was a photograph of a familiar household object. The more difficult tasks of identifying line drawings and detecting changes across scenes could have enabled the effect to become visible.

However, it was unlikely that the inconsistent object advantage found in Experiment 1 was attributable to the semantic relationship between the object and the scene context. The quality of the images was called into question and a high quality image subset failed to replicate the effect. Further research will be required before it can be proved conclusively that the difference between consistent and inconsistent objects, which generated this advantage, was purely one of semantics.

Is such an effect replicable using more realistic stimuli such as photographs?

The investigations into the detection of semantic inconsistency in photographic scenes failed to find any evidence of a significant advantage for inconsistent targets, or even any difference between performance for consistent and inconsistent targets. Simplifying the photographs to create line drawings also failed to produce an inconsistent object advantage, indicating that it was not the photographs themselves which prevented the expression of any consistency effects. The failure to produce consistency differences with complex photographs of scenes indicated that such effects are unlikely to be important in everyday scene viewing.

Experiments 5 and 6, designed to resemble Hollingworth and Henderson's (2000) change detection experiments, also failed to evidence an inconsistent object advantage. The two-exposure design in Experiment 5 produced no reliable advantage for inconsistent targets at any fixation position. The lack of consistency effects was compatible with the results of Experiment 3, which also investigated the extrafoveal processing of consistent and inconsistent targets in briefly presented scenes.

The only consistency effects found were again modulated by object size. In Experiments 3 and 5, which displayed photographic scenes for short periods of time and controlled fixation positions, there was a significant effect or a trend towards a consistent object advantage for medium sized objects, coupled with a significant effect or a trend towards better inconsistent object accuracy for large targets. This pattern was also expressed in Experiment 6, during the cycling change detection task, when accuracy was slightly higher for consistent medium sized objects than for inconsistent medium sized objects, but a clear advantage was found for inconsistent large objects over consistent large objects.

The analysis of the Leuven line drawing stimuli confirmed the suspicion that target size could affect consistency effects, with small objects failing to exhibit any effect. However, it is necessary to remember that the stimuli from which these conclusions on object size effects were based were a limited set, with only approximately 20 target objects within each size category. To add strength to this argument, it would also be desirable to investigate other researchers' results for evidence of size effects, assuming sufficient variability in object sizes, or construct another set of experimental stimuli with more clearly manipulated object sizes for further analysis.

Is there any evidence of preferential early fixation on inconsistent objects compared to consistent objects, in line drawings or in more naturalistic images like photographs?

Experiment 6 found no significant difference in the time taken for participants to terminate consistent and inconsistent trials, unlike the inconsistent object advantage observed by Hollingworth and Henderson (2000). The investigation of eye movement behaviour confirmed this result, providing no evidence of the earlier fixation of either consistent or inconsistent targets. The analysis of object size indicated that changes to medium sized targets were detected slightly more accurately, but no sooner, when the target was consistent. For large changing objects, a significant consistency effect was found, with better accuracy for large inconsistent objects than large consistent objects. Again, there was no significant difference in eye movement behaviour.

Experiment 7 also failed to find evidence that inconsistent objects were fixated earlier than consistent objects, or that saccade patterns were in any way different in scenes containing a consistent or inconsistent target object, prior to its fixation. When instructed to observe the scenes naturally, with no explicit task set, participants did not fixate inconsistent targets sooner than consistent targets. The compatible results of these two experiments confirmed that, under these voluntary conditions, participants did not use extrafoveal processing to detect semantically inconsistent objects for preferential fixation.

In addition, the results from Experiment 7 indicated that, under normal viewing conditions, semantic information was not used to select potential saccade targets when viewing either the Leuven line drawing stimuli or the photographic stimuli. This finding was particularly interesting in the case of the Leuven stimuli, which demonstrated an inconsistent object advantage in Experiment 1. If this effect were modulated by visual differences between consistent and inconsistent scenes, these visual differences might have been detected extrafoveally and used to direct saccades to regions of visual inconsistency, rather than semantic inconsistency. The failure to fixate inconsistent targets any sooner than consistent targets indicated that any visual differences or even semantic differences present between consistent and inconsistent scenes were not used to direct saccades to these regions of increased interest.

Is there a distinction between whether we can process and use semantic information under specific conditions and whether we actually do so during the course of real-world viewing?

Unfortunately, this question cannot be answered in full. Experiment 7 found no evidence that participants spontaneously fixated inconsistent targets sooner than consistent targets while passively viewing either simple or complex scenes. Experiment 6 investigated whether inconsistent objects which were changing were any more visually salient as saccade targets than consistent changing objects. No difference was found in saccade behaviour prior to the fixation of the changing target object. Even when a useful strategy would be to search for inconsistent objects because, if present, they had a much higher probability of being the changing target, participants did not selectively saccade towards those objects sooner than the consistent objects.

Although it cannot be assumed that participants considered the strategy, it is of interest that more use of extrafoveal vision was not made to complete a task in which unusual activity in a discrete object was to be detected. However, it could be argued that the task required participants to search for an object changing in visual terms, so little incentive was provided to search for or to investigate any regions of semantic inconsistency detected. In this way, it would be of interest to investigate further whether participants' expectations and motivations would influence performance in detecting semantic inconsistency from extrafoveal vision, as suggested previously.

The absence of any evidence to the contrary indicated that semantically inconsistent targets were not detected prior to direct fixation in the experiments conducted. It was clear that under normal viewing conditions, there was no natural tendency to search for regions of semantic inconsistency for immediate or preferential fixation, regardless of the stimuli type. However, it cannot yet be determined whether such actions are possible given the appropriate stimuli and the motivation to do so.

6.6 Final conclusions

To summarise, the original evidence considered in this thesis has mostly failed to replicate the results of other researchers that inconsistent objects in scenes are facilitated in different visual tasks. No evidence was found to support the hypothesis that semantic information could be processed from extrafoveal vision, either in brief scene presentations or during extended scene viewing. The available evidence indicated that extrafoveal processing of potential saccade targets involved visual processing alone and evidence of semantic processing effects was only detected in foveal vision. The sole reliable consistency effect exhibited in these experiments suggested that a consistent scene context could facilitate the identification of a consistent object when it was directly fixated.

Attempts have been made to reconcile the discrepancies between different areas of research and suggestions for further work have been identified to clarify the remaining issues. Also, additional variables which may influence the role of consistency have been identified in the manipulation of target object size and target retinal location. The investigation of these factors in previous and future work could further our understanding of elusive consistency effects in scene viewing.

Appendix A

Investigation into the Suitability of the Experimental Images

A.1 Introduction

This study investigated whether the scene stimuli used in the experiments contained in this thesis were appropriate experimental images. A preliminary pilot study investigated whether the target objects and scenes were appropriate prior to their use in experiments. The extended analysis of the scene images was motivated by the debriefing of participants after Experiment 1 which suggested that participants were unable to fully identify the line drawing scenes. Participants claimed that some objects presented in the two-alternative forced-choice display were difficult to recognise and that they were selecting between items they could not identify.

Additionally, the scenes themselves were not always clearly defined, containing few diagnostic objects to facilitate identification. The difficulty or impossibility of recognising either the scene background or the target object would influence the perception of semantic consistency, as the identity of both the scene context and the target would be required to determine whether they were semantically compatible. For this reason, the clarity and 'recognisability' of the experimental images used in these experiments were investigated further.

Both the Leuven line drawings and the photographic scenes were investigated to ensure both that the scenes and target objects could be recognised and also to confirm the consistent and inconsistent relationships determined by the experimenter. Performance differences in experimental tasks could only be attributed to semantic consistency if these two conditions were met by the stimuli. All the Leuven line drawings and all the photographic scene images used as experimental stimuli were presented to participants, who were required to name both the target object and the scene.

This naming task ensured that the image backgrounds and their component target objects were recognisable, which would be a prerequisite to determining whether the scenes were semantically consistent or inconsistent. Participants were then required to rate the likelihood of finding the specified object in the given location in the scene. These ratings were obtained to ensure that the consistent and inconsistent targets selected by the experimenter were considered equally consistent and inconsistent by the experimental population.

The preliminary data were used to select experimental images which contained appropriate consistent and inconsistent targets. Objects which were not reliably rated as consistent and inconsistent to match the experimental selection were replaced with other objects more compatible with the experimental manipulation. However, given the restrictions of the provided Leuven stimuli set, such substitutions were not always possible and certain targets remained which did not meet the desired levels of recognisability and consistency ratings.

The data presented in the extended analyses described here relate to the final set of experimental images used in all experiments in this thesis and not the rejected scene images. Using these data allowed the identification of the most appropriate scene images for experimental use. These images were defined as those which were reliably identified by the majority, if not all, of the participants and which also received suitably consistent or inconsistent ratings.

A.2 Method

Participants

There were a total of 18 participants who were all undergraduate or postgraduate students at the University of Durham. Seven of them were male and 11 were female .

Apparatus

The experimental scenes were presented on paper, with the relevant consistent or inconsistent target object clearly circled. Response sheets were provided, in which participants indicated their identification of the scene background, the target object and the rating of the consistency between the scene and the target. The Leuven line drawings and the photographic scene images were presented in individual booklets. Each image measured approximately 10cm by 8cm and was identical to the images used in the experiments, except that the target object in each was circled, to identify it.

The 44 line drawing images, consisting of 11 different scene backgrounds, each containing one target out of four possible objects (two consistent and two inconsistent), were displayed in a random order. The 64 photographs, consisting of 16 different scene backgrounds containing one of four target objects were also displayed in a random order in a separate booklet. The experimental images (without circled targets) can be found in Appendix B.

Design

The consistency relationship between the target object and the scene background was manipulated in both the line drawings and photographic scene images. The dependent variables, the scene and target identification and the consistency ratings were evaluated and, when possible, inappropriate objects were replaced. The results presented here are of the final set of images used in the experiments.

The line drawings and photographs were presented separately to the participants in a counterbalanced order, with half viewing the line drawings first and the other half viewing the photographs first. The presentation order of scenes within each block of line drawings and photographs was randomised to prevent images with the same background being presented in immediate succession. This random order was the same for all participants.

Procedure

Participants were provided with the booklets containing the scene images, with the appropriate response sheets, and were instructed to complete one before commencing the second. They were allowed to view the images for as long as necessary before identifying the scene background and circled target object in writing. The self-paced procedure would indicate whether the difficulty in identifying scenes and targets under experimental conditions was affected by the brief presentations of the images or the composition of the images themselves.

Finally, participants were required to make a judgement, on a scale of 1 to 5, on the likelihood of finding the target object in the scene, as depicted in the image. This score was the 'consistency rating' between the scene and the object. A rating of '1' was classed as 'very unlikely' and '5' was rated 'very likely'. It was expected that inconsistent targets would be rated close to 1 and consistent targets would be rated close to 5.

The participants were instructed to work as quickly as possible and to concentrate on their first impressions when identifying the objects and scenes. If they were unable to identify any scene image or object, they were required to either provide their best guess or omit the question if necessary. As the scene images were printed and sized smaller than in the experimental presentations, the resolution of the images was reduced and could have been less clear than when displayed full screen on a monitor. This limitation was particularly true for the photographic scenes and, in some cases, participants claimed to be unable to identify the target object.

Under these circumstances, where possible, participants were allowed to view the experimental image on a computer monitor, as presented in the experiments. If this presentation enabled them to identify the object, the participants were allowed to complete the relevant answers. No further assistance was provided and, if still unable to identify the object, they were instructed to omit the relevant item. This procedure was not always possible for some of the participants, due to time restrictions.

A.3 Results

The results were analysed to determine whether the scenes and objects had been identified correctly and rated appropriately, according to the experimenter's classification. Although the images were presented in a random order to the participants, the data were organised according to alphabetical order to collate the results. The data relating to the line drawings will be presented first.

Line drawings

The scene backgrounds in the Leuven stimuli were provided with corresponding consistent targets so alterations were not made to those images on the basis of these results. For the inconsistent targets, selected by the experimenter, the consistency ratings could be used to replace unsuitable targets whenever more suitable alternatives were available. The data presented relate to the final stimuli selection.

The data obtained from the participants are summarised in Table A.1. The images were grouped in sets of four, sharing the same scene background but containing different target objects. Table A.1 presents the scene and object names selected by the experimenter first, followed by any alternative names provided by the participants. The mean consistency rating for each item, averaged across participants, is also provided.

The scene name and object names selected by the experimenter were not provided to the participants. Participants identified the scene and the object first and then rated the likelihood of finding the specific object in the scene as depicted on a 5-point scale, with a score of 1 being rated as 'very unlikely' and 5 rated as 'very likely'. Scene names were considered to be appropriate if they retained the semantic meaning of the scene and inappropriate if they described a different scene or failed to capture the semantic element of the image presented. For example, the use of the word 'lounge' to describe a living room was an appropriate name but a clearly defined room, such as a kitchen or bathroom, described simply as a 'room' was not.

The appropriate identification of the scene is important for the detection of semantic consistency as an inappropriate scene meaning would alter the relationship between it and the target object. For example, the 'waterfront' images, which included a consistent barge as a target, were often identified inappropriately as a 'street' or an 'industrial estate'. Unless the scene was identified accurately, the barge would be considered highly inconsistent with the scene background. Inappropriate scene names were identified with an asterisk (*) in Table A.1 and the ratings based on these identifications were removed from the calculation of the mean if the ratings were affected by the misidentification.

The identification of the target object was also inspected as any targets not appropriately named could have received inappropriate consistency ratings. All the names suggested by participants are presented in Table A.1 and those which incorrectly identified the target were labelled with an asterisk (*). Several targets could not be recognised at all, such as the objects labelled by the experimenter as 'books' in the bedroom scene but which defied any label for seven participants. Similarly, the lectern in the gymnasium was not named at all by six participants. The semantic relationship between these targets and the scenes in which they were located was compromised by the inability to reliably identify them when viewed for unlimited time.

The ratings on the likelihood of finding the target object in the scene were used to calculate a mean consistency rating. Individual values provided by the participants were excluded from the calculation for one of two reasons. If the participant failed to identify either the scene or the target object, any rating provided by them for that image was not used to calculate a mean value. Ratings were also withheld when participants identified either the scene or the target object inappropriately.

It was not possible to exclude all ratings obtained when the identification of the scene or object was incorrect because, for certain images, very few participants identified both correctly. If the misidentification of the scene or target object did not greatly influence the consistency judgement, the rating provided by the participant was used in the calculation of the mean consistency value. For example, the radiator was a consistent target in the laboratory and workshop scenes, both of which were misidentified by some participants as a 'classroom' or even a 'kitchen'. However, a radiator could be a consistent target in any indoor room, so the misidentification did not affect consistency ratings. In this way, the mean rating averaged the values provided by participants who plausibly identified the scenes and target objects. The number of values which were excluded due to inappropriate identifications were provided with each mean consistency rating.

The results presented here were those obtained from the final, refined set of stimuli, after preliminary results indicated that certain inconsistent targets needed to be replaced. As the consistent targets were selected by the Leuven researchers who created the stimuli, these were not altered by the experimenter according to the ratings obtained in this study. Several inappropriate inconsistent targets were replaced on the basis of this analysis. One of these was an inconsistent loudspeaker in the workshop scene, which was rated as very consistent with the scene. This object was replaced by a sink, which could be embedded into the scene more realistically than other alternatives but also failed to obtain a sufficiently inconsistent rating.

Similarly, an inconsistent climbing net was used as a preliminary target in the gymnasium scene but was rated as consistent with the context. This target was replaced by another object, the stool, which also received a higher consistency rating than would be desirable. However, the number of possible targets which could be located in scenes was limited, so no further target substitutions occurred.

The data raised concerns about consistent targets also, but no action was taken to substitute objects selected as consistent targets prior to this investigation. One of the targets in particular, the gas cylinders in the workshop, was difficult to identify. Even participants who correctly labelled the target as 'oxygen cylinders' or 'gas holders' did not rate the item as highly consistent with the scene. It seemed likely that the lack of exposure to a similar scene in real life resulted in the relatively low item consistency rating of 3.67. A similar difficulty in identifying the waterfront scene background resulted in 11 out of 18 participants rating a consistent target barge as highly inconsistent with the 'street' or 'industrial estate' scene.

Recognising and rating the images

From these results, it was clear that, when the scenes and objects were identified correctly, they were indeed considered consistent and inconsistent as assigned by the experimenter. Only one consistent target object had a mean item consistency rating less than 4 out of 5, with the mean total consistency rating for all consistent targets being 4.65 (*SD*=.35). Similarly, only two inconsistent targets received a mean item consistency rating higher than 2 out of 5, with the mean total consistency rating being 1.41 (*SD*=.66). These data indicated that the scene images presented to the participants were rated as suitably consistent and inconsistent when identified correctly.

However, it was a cause for concern that the scenes and targets were not reliably identified by all participants. Only 11 of the 44 scene images produced mean item consistency ratings which were calculated from every participants' individual ratings. This result indicated that the validity of at least one individual's consistency rating was questionable in three quarters of all the trials.

Of the 11 scenes reliably identified by all participants, 7 contained a consistent object, suggesting that the presence of a consistent scene background could assist in the correct identification of the target or vice versa, during extended scene viewing conditions. Only four of the reliably identified scenes contained an inconsistent target. In fact, in over one third of all the images (15/44), three participants or more failed to identify either the scene or the target appropriately, resulting in their individual consistency ratings being excluded from the calculation of the mean item rating. This investigation indicated that the scene images were not readily identifiable to participants viewing them for extended periods of time, suggesting that the images might not have been recognisable during brief 120ms experimental presentations.

Although some common household scenes like the bedroom and the kitchen were reliably identified by all participants, more unusual or unfamiliar scenes proved more difficult to recognise and were identified inappropriately more often. For example, the waterfront scenes in particular were not identified as such by many participants. Up to 11 ratings were excluded from the calculation of the mean consistency rating because the scene had been identified as a 'street' or an 'industrial estate'. Other indoor scenes, like the laboratory and the workshop, were labelled generically as 'rooms' by some participants, as there was little to distinguish between them and they contained few diagnostic objects to assist in their identification. The absence of diagnostic objects in other backgrounds also caused participants to incorrectly identify the scenes. For example, the bathroom was identified as a kitchen by one participant, possibly because the only distinguishing feature, without the shower head and basin which acted as consistent targets, was the shower cubicle, if recognised as one.

Some target objects were also difficult to identify. A high level of agreement was found for the identification of some targets such as the stool and the radiator. Other objects, like the basketball hoop and vaulting horse, were given different but appropriate names by almost all participants. For both these types of objects, all participants understood the semantic identity of the object, regardless of the name provided, so these names were considered appropriate. The ratings obtained could be used to calculate the mean item consistency rating.

In contrast, the identification of certain objects resulted in little agreement. The following objects were most ambiguous in identity. The item labelled as a lectern was originally a consistent object in a chapel scene (not used in this thesis) and was placed as an inconsistent object in the gymnasium scene. None of the participants identified it accurately but all considered it inconsistent with the scene.

An object resembling a cash register or weighing scales was originally a consistent object in a supermarket scene (not used in this thesis) and was used as an inconsistent target in the bathroom and dining room. Up to six participants were unable to name it in these scenes. A third ambiguous object was a consistent rather than an inconsistent target. The items placed on a shelf in the bedroom scene, which were tentatively identified as books by the experimenter, proved equally ambiguous for the participants. Seven participants were unable to name the items and other suggestions were limited to boxes of some description.

The difficulty in identifying and naming these objects has repercussions on the process of rating the semantic relationship between the objects and the scenes. Although some participants attempted to provide a rating of consistency even when they had failed to identify the scene or the object, the likelihood of finding an object in a specific scene cannot be determined if either cannot be recognised. As only one quarter of the scenes were identified appropriately by all participants, the value of these images in determining the effects of semantic inconsistency is questionable.

The results so far have suggested that only a limited proportion of one quarter of the images provided no difficulty to the participants in identifying the scene background, the target object and determining the semantic relationship between the two. The remaining scenes could be criticised, as at least one participant had been unable to provide an appropriate item consistency rating due to the recognisability of the scene or the target. In the majority of cases, the scenes either contained ambiguous target objects which were difficult to identify or backgrounds which were indistinct, preventing participants from correctly labelling them, even when viewing the images with unlimited time.

The analysis of the entire stimuli set calls into question their use as experimental stimuli to investigate semantic consistency effects.
To address this problem, the data were used to identify a higher quality subset of stimuli, which were reliably identifiable and also sufficiently consistent and inconsistent, to investigate semantic consistency effects. This analysis is included in Experiment 1 (page 92).

Table A.1: Table displaying the recognisability results of the line drawing scene images.

The first scene and object names were selected by the experimenter. Additional names were provided by participants. Items marked with an * were potentially inappropriate responses. The mean rating of consistency is reported for each image, together with the number of ratings excluded from the calculation due to inappropriate responses.

Scene code and consistency	Scene name	Object name	Mean item consistency rating (no. of excluded ratings)
Bar-c1 Consistent	Bar Dining room*, Kitchen*, Lab* Don't know*	Beer taps Taps Sink Don't know* x5	4.42 (5)
Bar-c2 Consistent	Bar Dining room*, Kitchen*, Lab* Don't know*	Light(s) Hanging lights Don't know*	4.87 (2)
Bar-i1 Inconsistent	Bar Dining room* Kitchen* Lab* Don't know*	Shower head Shower Light* Lamp* Don't know*	1.00 (5)
Bar-i2 Inconsistent	Bar Dining room* Kitchen* Lab* Don't know*	Towel rail Towel Tea towel Towel rack	1.92 (4)

Bath-c1	Bathroom	Shower head	4.67
Consistent	Shower	Shower	(2)
	Don't know* x2	Shower piece	
		Light*	
		Don't know*	
Bath-c2	Bathroom	Sink	4.93
Consistent	Shower	Sink & tap	(3)
	Kitchen*	Water tap	
	Don't know* x2	Tap	
		Don't know* x2	
Bath-i1	Bathroom	Loudspeaker	1.00
Inconsistent	Shower	Speaker	(3)
	Don't know* x3	Sound equipment	
		Security camera	
		CCTV	
		Don't know*	
Bath-i2	Bathroom	Cash register	1.17
Inconsistent	Shower	Cash till	(6)
	Kitchen*	Weighing scales	
	Don't know* x4	Scales	
		Kitchen scales*	
		Toy*	
		Iron*	
		Don't know* x6	

Bed-c1	Bedroom	Picture	4.88
Consistent		Painting	(0)
Bed-c2	Bedroom	Books	4.40
Consistent		Boxes	(7)
		Small boxes	
		Don't know* x7	
Bed-i1	Bedroom	Sink	1.27
Inconsistent		Tap	(4)
		Tap & sink	
		Model*	
		Don't know* x3	
Bed-i2	Bedroom	Beer taps	1.00
Inconsistent		Tap(s)	(5)
		Sink	
		Don't know* x5	

Concert-c1	Concert hall	Lights	4.77
Consistent	Theatre	Speaker	(3)
	Cinema	Camera	
		Lighting	
		CCTV	
		Security camera	
		Sound equipment	
		Don't know* x3	
Concert-c2	Concert hall	Loudspeaker	4.81
Consistent	Theatre	Speaker	(1)
	Cinema	Light	
		Camera	
		CCTV	
		Don't know*	
Concert-i1	Concert hall	Basketball hoop	1.06
Inconsistent	Theatre	Basketball board	(0)
	Cinema	Basketball net	
		Basketball rim	
		Netball thing	
		Netball net	
		Goal post	
Concert-i2	Concert hall	Shower head	1.00
Inconsistent	Theatre	Shower piece	(2)
	Cinema	Shower	
		Lamp*	
		Light*	

Dining-c1 Consistent	Dining room Study Lounge Living room Breakfast Room with table Room*	Chair	5.00 (0)
Dining-c2 Consistent	Dining room Study Kitchen Living room Breakfast Room with table Room*	(Half a) picture Don't know*	4.56 (1)
Dining-i1 Inconsistent	Dining room Study Lounge Living room Breakfast Room with table Room*	Towel rail Towel Towel rack	1.88 (1)
Dining-i2 Inconsistent	Dining room Study Lounge Living room Breakfast Room with table Room*	Cash register Cash till Weighing scales Scales Kitchen scales* Iron* Toy* Don't know* x 4	1.42 (5)

Gym-c1	Gymnasium	Basketball hoop	4.94
Consistent	Gym	Basketball net	(0)
	Sports hall	Basketball rim	
		Basketball board	
		Netball target	
		Netball goal	
		Netball thing	
Gym-c2	Gymnasium	Vaulting horse	5.00
Consistent	Gym	Gym equipment	(0)
	Sports hall	Gym horse	
		Gymnastics box	
		Horse	
		Sport equipment	
		Activity box	
		Jump box	
		Thing that gymnasts jump over	
Gym-i1	Gymnasium	Lectern	1.33
Inconsistent	Gym	Pot plant*	(6)
	Sports hall	Plant*	
	Don't know*	Don't know* x6	
Gym-i2	Gymnasium	Stool	2.40
Inconsistent	Gym	Walking frame*	(6)
	Sports hall	Don't know* x6	
	Don't know* x3		

Kitchen-c1 Consistent	Kitchen	Double sink Sink Tap(s)	5.00 (0)
Kitchen-c2 Consistent	Kitchen	Towel rail Towel Towel rack	4.65 (0)
Kitchen-i1 Inconsistent	Kitchen	Loudspeaker Speaker Sound equipment CCTV Security camera Don’t know* x2	1.40 (2)
Kitchen-i2 Inconsistent	Kitchen	Truck Toy truck Lorry Toy Vehicle	1.50 (0)

Lab-c1	Laboratory	Radiator	4.63
Consistent	Lab,		(1)
	Science lab,		
	Kitchen* x4		
	Room*		
	Workshop		
	Don't know*		
Lab-c2	Laboratory	Stool	4.77
Consistent	Lab		(4)
	Science lab		
	Kitchen* x5		
	Room*		
	Gym*		
	Don't know* x2		
Lab-i1	Laboratory	Basketball hoop	1.07
Inconsistent	Lab	Basketball ring	(2)
	Science lab	Basketball rim	
	Kitchen (x5)*	Basketball net	
	Classroom	Netball goal	
	Room*	Netball net	
	Gym*	'Hocky thing'	
	Don't know*	Don't know*	
Lab-i2	Laboratory	Swing	1.14
Inconsistent	Lab	Recreational structure	(3)
	Science lab	Don't know*	
	Classroom		
	Workshop		
	Kitchen* x4		
	Room*		
	Don't know*		

Play-c1	Playground	Swing	4.69
Consistent	Garden	Recreational structure	(1)
	Park	Don't know*	
Play-c2	Playground	Climbing net	4.06
Consistent	Garden	Ladders	(0)
	Park	Net frame	
		Climbing frame	
		Play thing	
		Climbing structure	
		Activity apparatus	
Play-i1	Playground	TV unit	1.00
Inconsistent	Garden	TV with stand	(0)
	Park	TV equipment	
		TV/VCR	
		TV stack	
		Entertainment system	
		Hi-fi system	
		Computer	
Play-i2	Playground	Vaulting horse	1.47
Inconsistent	Garden	Gym equipment	(2)
	Park	Gymnastics box	
		Gym horse	
		Horse	
		Activity box	
		Sport equipment	
		Jump box from gym	

Wfront-c1	Waterfront	Barge	4.83
Consistent	Dock	Ship	(11)
	Street*	Boat	2.47 – all
	Factory*		ratings
	Airport*		
	School*		
	Industrial building*		
	Industrial estate*		
Wfront-c2	Waterfront	Truck	4.69
Consistent	Street*	Lorry	(1)
	Factory*	Vehicle	
	Airport*	Don't know*	
	School*		
	Industrial building*		
	Buildings*		
	Industrial estate*		
Wfront-i1	Waterfront	Windbreak	1.13
Inconsistent	Street*	Tent	(2)
	Factory*	Don't know* x2	
	Airport*		
	Road*		
	Industrial building*		
	Industrial estate*		
Wfront-i2	Waterfront	Vaulting horse	1.00
Inconsistent	Street*	Gym equipment	(1)
	Factory*	Gymnastics box	
	Airport*	Vaulting bench	
	Road*	Gym horse	
	Industrial building*	Gym thing	
	School/road*	Sport equipment	
	Industrial estate*	Horse	
		Jump box from gym	
		Don't know*	

Work-c1	Workshop	Gas cylinders	3.67
Consistent	Laboratory	Oxygen cylinders	(5)
	Lab	Gas	
	Science lab	Oxygen tanks	
	Workbench	Gas holders	
	Room*	Gas pumps	
	Classroom*	Giant fire extinguishers*	
		Don't know* x4	
Work-c2	Workshop	(Part of) radiator	4.00
Consistent	Lab	Heater	(2)
	Science lab	Don't know* x2	
	Classroom		
	Room & workbench		
	Don't know*		
Work-i1	Workshop	Double sink	3.88
Inconsistent	Lab	Sink/tap	(2)
	Science lab	Don't know*	
	Room & workbench		
	Classroom		
	Don't know*		
Work-i2	Workshop	Basketball hoop	1.00
Inconsistent	Science lab	Basketball board	(0)
	Lab	Basketball rim	
	Classroom	Basketball net	
	Workbench	Netball thing	
	Room*		

Photographs

The results obtained from the photographic scenes are summarised in Table A.2. The data were again organised into sets of four images, each set depicting the same room. The backgrounds to the four images were not necessarily identical. Each matched consistent and inconsistent target pair of images (e.g. the sponge and the book in the bathroom) shared the same scene background, in which both images were identical except for the identity of the target. However, the corresponding consistent and inconsistent target pair of images in the same set (e.g. the shampoo bottle and the photo frame in the bathroom) did not necessarily contain the same background, generally depicting the same room but not necessarily the same region or from the same viewpoint.

Across image sets, the backgrounds were also different. For example, the room depicted in the four images described above (bath-c1, bath-c2, bath-i1 and bath-i2) was entirely different to the bathroom depicted in the following four bathroom images (bath-c3, bath-c4, bath-i3 and bath-i4). Because different rooms were displayed in each set of four images, participants could view at least one scene image from each set of four, without viewing the same scene background twice.

Table A.2 provides data on the identification of both scene backgrounds and target objects in the photographic scenes. The first name provided was selected by the experimenter and is followed by any alternative names suggested by participants. Any inappropriate names, identified according to the same criteria as in the line drawing analysis, were marked with an asterisk (*). Mean item consistency ratings were calculated from all ratings provided by participants who identified both the scene and the target object appropriately. The number of ratings excluded from each calculation, due to misidentification or failure to identify the scene or target object, was also displayed.

Again, the results displayed refer to the final set of images selected for use as experimental scenes. The preliminary results of this investigation identified images which were inappropriate and which were subsequently replaced for the final stimuli set. The relevant target objects which were not rated as sufficiently inconsistent in the selected scenes were quickly identified and removed from the set of experimental images.

Recognising and rating the images

The results presented in Table A.2 indicated that the scene images were highly recognisable. 46 out of 64 images (72%) had an item consistency rating calculated from all participants' responses, due to no inappropriate identification of the scene or object which influenced the consistency judgement. This result compares to only 25% for the line drawing stimuli. It was clear that a much larger proportion of scenes were identified unambiguously in the photographic stimuli than in the line drawings. Of the 18 scenes which were not reliably identified by all participants, nine were consistent scenes and nine were inconsistent, proving that there was no bias in correct identification for consistent and inconsistent scenes.

Only eight images in the entire data set (12.5%) were not appropriately identified by three or more participants, compared to over a third of trials in the line drawing stimuli (34%). When three or more participants had their individual ratings excluded from the calculation of the item mean, the reason for the exclusion was the failure to identify or the misidentification of the target object. The scene was always recognised appropriately.

There were only two exceptions to the appropriate identification of all scene backgrounds. The child's playroom was called a lounge or bedroom and the hall was identified as a lounge, all by only one participant. This high degree of accuracy and compatibility across participants was attributed to the presence of objects other than the target which were diagnostic of the room's purpose and identity. Also the high quality photographic images displayed realistic images of actual scenes, rather than artistic simplifications, which could have facilitated recognition.

Although not all targets were reliably identified by all participants, performance was still better than for the line drawings. The relative ease of identification could be explained by the nature of the targets as familiar household objects. The few targets which could not be identified were often small in size, such as the book and the sponge in the bedroom scene. Small items could have been difficult to resolve from an image printed on paper, rather than the larger display on a monitor used under experimental conditions. Although participants were allowed to view the images on a computer monitor if they felt the resolution or size of the image was insufficient to identify the object, in practice, time constraints meant that some participants did not make use of this facility. In the remaining cases, a reliable misidentification of the target caused the error, such as the participant who failed to recognise the garden gnome and labelled it a 'mini statue', resulting in an inconsistent rating when placed in a garden.

Individual ratings were again obtained on a scale of 1 to 5 and the item ratings for each consistent and inconsistent target were appropriate, as determined by the experimenter. No consistent object was rated as inconsistent (below 3) and no inconsistent object was rated as consistent (above 3). All scenes had a higher mean item consistency rating for a consistent target than the matched inconsistent target in the same scene. The mean difference in consistency ratings was calculated to be 2.88 (*SD*=.74), indicating that a consistent object received a much higher item rating than the matched inconsistent object for that scene.

Five of the 32 consistent targets obtained a mean item rating below 4, but 11 inconsistent targets had a mean item rating above 2. This difference suggested that while the consistent targets were more reliably considered to be highly consistent with the scene context, the same objects placed in different scenes were less likely to be rated as highly inconsistent. This proposal was supported by the calculation of the mean total consistency rating for all consistent targets as 4.58 (*SD*=.46). A less extreme mean total rating and greater variability were found in inconsistent ratings, with a mean value of 1.70 (*SD*=.56).

Although each individual rating was appropriate to the target object, some of the mean item ratings for consistent and inconsistent targets did approach the mid-rating of 3. The definition of a rating of 3 was the 'possible' likelihood of finding the target object in the specified scene. As all the targets were household objects, they could all 'possibly' be found in any of the household scenes depicted, compared to the line drawings in which images of outdoor or very large objects were placed indoors. In this way, the photographs would have contained inconsistent target objects which were unlikely, but possible, to be found in the scene, rather than impossible objects.

A further explanation for the distribution of ratings spread across the range, rather than clustered at the extremes, was supplied by one participant. Upon rating the likelihood of finding a garden gnome on an office desk as '4' (quite likely), the participant added the comment, "I'm a student!" As the majority of participants were students, whose college rooms contained items usually found in different household rooms, such as kitchens and bathrooms, it was possible that objects which were rated as highly inconsistent by the experimenter may not have been judged as highly inconsistent by undergraduate participants. Other participants verbally confirmed that the objects placed in unusual locations were all possible, especially in a student house. The hypothesis that consistency ratings would vary according to the housing standards enjoyed by the participants could not be tested empirically, as there were insufficient mature students to serve as a comparison. However, this possibility would be worthy of further investigation.

To summarise, while consistent objects were quite reliably recognised and rated as being located appropriately, inconsistent objects were judged as less inconsistent, possibly seeming less inappropriate in the living environments of the student participants. These data allowed the selection of stimuli which were both reliably identified and rated as highly consistent and inconsistent. This high quality subset of stimuli was used to further analyse the data obtained in Experiments 3 and 4 for reliable consistency effects when displaying highly appropriate scene images.

Table A.2: Table displaying the recognisability results of the photograph scene images. The first scene and object names were selected by the experimenter. Additional names were provided by participants. Items marked with an * were potentially inappropriate responses. The mean rating of consistency is reported for each image, together with the number of ratings excluded from the calculation due to inappropriate responses.

Scene code and consistency	Scene name	Object name	Mean item consistency rating (no. of excluded ratings)
Bath-c1 Consistent	Bathroom	Sponge Scrubber Don't know* x2	4.5 (2)
Bath-c2 Consistent	Bathroom	Shampoo bottle Bath oil Bubble bath Bottle Don't know* x3	5.00 (3)
Bath-i1 Inconsistent	Bathroom	Book	2.42 (0)
Bath-i1 Inconsistent	Bathroom	Photo frame Picture Picture frame Photographs	1.73 (0)
Bath-c3 Consistent	Bathroom Toilet	Toilet roll Loo roll Loo paper Bog roll	4.94 (0)
Bath-c4 Consistent	Bathroom Toilet	Toilet brush Loo brush Brush Toilet scrubber	5.00 (0)
Bath-i3 Inconsistent	Bathroom Toilet	Teapot Kettle	1.06 (0)
Bath-i4 Inconsistent	Bathroom Toilet	Gnome Garden gnome Mini statue	1.25 (0)

Bed-c1 Consistent	Bedroom	Alarm clock Clock Radio alarm clock Clock radio Electronic alarm clock	5.00 (0)
Bed-c2 Consistent	Bedroom	Bedside lamp Lamp Table lamp	5.00 (0)
Bed-i1 Inconsistent	Bedroom	Shoes Trainers	1.44 (0)
Bed-i2 Inconsistent	Bedroom	Football Ball	1.22 (0)
Bed-c3 Consistent	Bedroom	Shirt	4.88 (0)
Bed-c4 Consistent	Bedroom	Book Don't know* x6	4.83 (6)
Bed-i3 Inconsistent	Bedroom	Calendar	2.65 (0)
Bed-i4 Inconsistent	Bedroom	Sponge Soap Teapot* Bowl*, Don't know* x8	2.00 (10)

Child-c1 Consistent	Playroom Kids' room Toy room Child's play room Lounge*	Football Ball	3.88 (1)
Child-c2 Consistent	Playroom Kids' room Table Bedroom*	Globe Earth toy	4.00 (0)
Child-i1 Inconsistent	Playroom Kids' room Toy room Child's play room Bedroom*	Kettle	1.06 (0)
Child-i2 Inconsistent	Playroom Kids' room Bedroom*,	Rucksack Bag Don't know* x6	2.36 (6)
Child-c3 Consistent	Playroom Kids' room Nursery Toy room Child's bedroom	Teddy bear	5.00 (0)
Child-c4 Consistent	Playroom Child's bedroom Toy room Nursery Kids' room	Winnie the Pooh bear Pooh bear Teddy bear Jack in the box*	4.94 (0)
Child-i3 Inconsistent	Playroom Kids' room Play area Child's room Toy room	Toaster TV* Don't know* x3	1.08 (4)
Child-i4 Inconsistent	Playroom Child's playroom Toy room Kid's room	Milk jug Water jug Jug	1.59 (0)

Dining-c1	Dining room	Fruit bowl	4.62
Consistent	Kitchen	Basket	(0)
	Breakfast table	Fruit holder	
	Dining table	Bowl	
	Table		
Dining-c2	Dining room	Milk jug	4.59
Consistent	Dining table	Water jug	(0)
	Breakfast table	Jug	
	Breakfast room	Jar	
	Kitchen		
	Table		
	Living room		
Dining-i1	Dining room	Watering can	2.29
Inconsistent	Dining table	Water bottle	(0)
		Water can	
Dining-i2	Dining room	Winnie the Pooh bear	2.00
Inconsistent	Breakfast room	Teddy bear	(0)
	Breakfast table		
	Kitchen		
	Table		
	Living room		

Garden-c1 Consistent	Garden	Gnome Garden gnome Garden statue Mini statue*	4.27 (1)
Garden-c2 Consistent	Garden Back garden	Football Ball	4.69 (0)
Garden-i1 Inconsistent	Garden	Desk lamp Lamp Table lamp Study lamp	1.08 (0)
Garden-i2 Inconsistent	Garden	Bedside lamp Lamp	1.00 (0)
Garden-c3 Consistent	Garden Front garden Outside house Garden steps Step by garden	Watering can Water bottle	3.59 (0)
Garden-c4 Consistent	Garden Back garden	Gnome Mini statue* Don't know*	4.00 (2)
Garden-i3 Inconsistent	Garden Front garden Outside house Garden steps	Fruit bowl Fruit basket Bowl of fruit Bowl	1.31 (0)
Garden-i4 Inconsistent	Garden Back garden	Toilet brush Toilet scrubber Loo brush Brush Don't know* x2	1.10 (2)

Hall-c1 Consistent	Hall Hallway Front door Porch Lounge*	Rucksack Bag Satchel School bag	4.08 (0)
Hall-c2 Consistent	Hall Hallway Entrance hall Front door Porch Lounge*	Umbrella	3.83 (0)
Hall-i1 Inconsistent	Hall Hallway Front door Porch Lounge*	Globe Earth toy	1.25 (0)
Hall-i2 Inconsistent	Hall Hallway Porch Front door Lounge*	Broom Brush	2.33 (0)

Hall-c3 Consistent	Hall Hallway Front door Stairway Lounge*	Shoes Trainers	4.25 (0)
Hall-c4 Consistent	Hall Hallway Entrance Front door Lounge*	Wellington boots Wellies Boots Letters*	3.91 (1)
Hall-i3 Inconsistent	Hall Hallway Front door Lounge*	Alarm clock Clock radio Clock	1.17 (0)
Hall-i4 Inconsistent	Hall Hallway Porch Entrance Front door Lounge*	Waste paper bin Waste paper basket Waste bin Basket Bin	2.58 (0)

Kitchen-c1 Consistent	Kitchen	Kettle	5.00 (0)
Kitchen-c2 Consistent	Kitchen	Broom Brush	3.83 (0)
Kitchen-i1 Inconsistent	Kitchen	Football Ball	1.33 (0)
Kitchen-i2 Inconsistent	Kitchen	Umbrella	2.22 (0)
Kitchen-c3 Consistent	Kitchen	Toaster	5.00 (0)
Kitchen-c4 Consistent	Kitchen	Teapot Kettle	4.82 (0)
Kitchen-i3 Inconsistent	Kitchen	Teddy bear	1.65 (0)
Kitchen-i4 Inconsistent	Kitchen	Toilet roll Toilet paper Bog roll Loo roll	1.53 (0)

Living-c1 Consistent	Living room TV room Sitting room Lounge TV area Front room Shelves	Video player Video recorder Video VCR DVD player	4.94 (0)
Living-c2 Consistent	Living room Sitting room Lounge	Photo frame Pictures Photographs Don't know*	4.91 (1)
Living-i1 Inconsistent	Living room TV room Sitting room Lounge Bookshelves TV unit	Paper tray Tray In box tray Desk tray In/out basket Don't know* x6	2.18 (6)
Living-i2 Inconsistent	Living room Front room Sitting room Lounge	Shampoo bottle Bubble bath Bottle Plastic bottle Don't know* x5	2.11 (5)

Office-c1	Office	Paper tray	4.82
Consistent	Study	Work tray	(0)
	Workstation	In tray	
	Working room	In box tray	
		Desk tray	
		Paper basket	
		Post tray	
		Letter tray	
		In/out basket	
Office-c2	Office	Calendar	4.88
Consistent	Study		(0)
Office-i1	Office	Video player	2.13
Inconsistent	Study	Video recorder	(1)
	Computer room	Video	
		VCR	
		Don't know*	
Office-i2	Office	Shirt	2.59
Inconsistent	Study		(0)
Office-c3	Office	Desk lamp	5.00
Consistent	Study	Lamp	(3)
	Desk	Don't know* x3	
	College room		
	Table		
Office-c4	Office	Waste paper bin	4.69
Consistent	Study	Waste paper basket	(0)
	Work room	Waste basket	
		Bin	
		Basket	
Office-i3	Office	Gnome	1.17
Inconsistent	Study	Statuette*	(2)
	Desk	Mini statue*	
	College room		
Office-i4	Office	Wellington boots	1.67
Inconsistent	Study	Boots	(1)
	Desk	Wellies	
		Files*	

A.4 Discussion

The purpose of this study was to determine how suitable the images used in the experiments contained in this thesis were in the investigation of semantic consistency effects. Semantic consistency can only be determined from the identification of both the scene context and the target object and a direct comparison of their semantic compatibility. Therefore, the ease of scene and object recognition were evaluated, together with ratings of the semantic relationship between the scene background and the target object.

The investigation of the Leuven line drawing images confirmed the suspicions raised by the debriefing process after Experiment 1, that the scenes and the target objects were not readily identifiable in all trials. The study indicated that only one quarter of all the experimental images were reliably identified by all participants under unlimited viewing conditions. This result suggested that, under experimental conditions involving brief presentations, the images were unlikely to have been sufficiently clear to determine the semantic consistency between the scene context and a target object. This conclusion calls into question the significant consistency effect found in Experiment 1, suggesting that alternative visual differences must have generated that effect.

These data challenge the assumption that the Leuven line drawings used in Experiment 1 were appropriate experimental images for the investigation of semantic consistency effects. Scene images derived from this source have been used by other researchers for similar experiments, many of which assumed the detection of semantic consistency within brief presentation times. It is well established that scene identity or gist can be determined quickly from a brief presentation. However, the inability to unambiguously identify these scene backgrounds, as investigated in this thesis, from extended and unlimited viewing conditions must question whether scene context could be derived under experimental conditions and whether the differences in performance attributed to semantic consistency could indeed have been generated by its detection. The difficulty in object identification reported by participants further compounds this problem.

For these reasons, the proposal that the line drawings investigated in this study were appropriate for use in the investigation of the processing of semantic information and the relationship between an object and its background from brief scene presentations must be treated with caution. This investigation has indicated that any experimental manipulations based on the categorisation of these objects as consistent and inconsistent targets in scenes may not be fully justified. The assignment of objects into these categories risks being considered a random manipulation, if the identity of the scene and the object could not be reliably detected in all experimental stimuli. As it would be impossible to determine whether an object was likely to appear in a given context unless both the context and the object's identity were clear, the assumption of the detection of semantic consistency would be false.

For the reasons outlined above, the most appropriate experimental stimuli would be those that were most reliably identified by all participants. The removal of the more ambiguous scene trials would reveal more appropriate and accurate data about the detection of semantic consistency from brief scene presentations. The data obtained from Experiment 1 was analysed in this way, by selecting a total of 20 scenes which were most reliably identified and rated as consistent and inconsistent in this study. The use of reliable scene stimuli would indicate an effect possibly explained by semantic consistency, if the effect remained. Alternatively, the absence of a semantic consistency effect when displaying the most recognisable images would indicate that the effect evidenced from the analysis of the entire data set was not modulated by the detection of semantic inconsistency. In this way, the data obtained in this study were used to evaluate the results obtained in Experiment 1.

The set of photographic images was designed by the experimenter to provide a more suitable stimuli set for investigating semantic consistency effects. The issues raised by the line drawing stimuli relating to the recognisability of the scenes and the target objects were addressed by selecting familiar household scenes and objects. The scene backgrounds were photographs of rooms in genuine homes, to ensure that the images represented naturalistic scenes whose context could be discerned easily. Target objects were selected from ordinary household items, which were located in consistent and inconsistent scenes and should also have been easy to identify. In this way, the scenes depicted in the photographic stimuli were designed to be reliably identified by participants, in order to determine whether semantic consistency could be detected from brief scene presentations.

From the analysis of these scene images, it became clear that the scene backgrounds in particular were very reliably recognised by the participants. There was a large degree of agreement on the identity of the scene backgrounds across all participants. Scenes were almost always identified accurately, with only two exceptions where one participant failed to reflect the specific semantic meaning of the scenes and provided generic labels.

The majority of the target objects were equally well recognised, with the exception of some small objects. On occasion, these proved difficult to identify from the images presented to participants, but would have been presented in a larger display when used in experiments and should have been easier to identify under these conditions. Overall, over 70% of the scene images were appropriately named by all participants and only 12.5% of images were not identified accurately by more than three participants.

One cause for concern, however, was that the consistency ratings provided by the participants were not as extreme as those found for the line drawings. Ratings approximating 1 and 5 indicated highly inconsistent and highly consistent scenes respectively and mean total ratings for all consistent trials and all inconsistent trials were closer to these extremes for the line drawings than for the photographic stimuli. Total consistency ratings for consistent scenes were comparable for line drawings (4.65) and for photographs (4.58), indicating that consistent targets in photographic scenes were considered equally likely as the consistent line drawings in the Leuven stimuli. However the mean total rating for inconsistent photographs (1.70) was greater than that for line drawings (1.41), suggesting that inconsistent targets in photographs were less unlikely to appear in their scenes than inconsistent targets in the line drawing stimuli.

Inconsistent targets in photographs were also consistent targets in another household scene. As all the targets selected for the photographic scene images were movable household items, no target objects could be placed in scenes in which they could not possibly be found. This manipulation would have limited the degree of inconsistency possible from the familiar items. However, the analysis of the ratings confirmed that consistent objects were reliably rated as more consistent than the matched inconsistent objects, indicating that the inconsistent objects were still rated as relatively inconsistent in their scenes.

To account for these ratings in the statistical analysis of experimental results, it would be desirable to identify scene-object pairs which were not considered sufficiently consistent or inconsistent, according to the results obtained in this study. The selection of the 20 most consistent and inconsistent scene images, subject to the condition that they were reliably identified by all participants, would assist the interpretation of experimental data, to determine whether semantic consistency could be detected from brief scene presentations. Although the results of Experiment 3 clearly indicated no semantic effects, this conclusion was investigated further, using the data obtained in this analysis, to attempt to evidence a consistency effect using only the most appropriate stimuli, that is, those which were rated most consistent and inconsistent.

This limitation in the manipulation of semantic consistency in photographic stimuli resulted from the use of realistic scene backgrounds and actual physical objects. As the line drawings were not limited in this way, an image of a large outdoor object, such as a truck or swing, which would normally not be movable, could be placed in a wholly inappropriate and impossible position indoors. In this way, the line drawings were often rated more reliably consistent and inconsistent than the photographs. However, although the line drawings were considered marginally more consistent and inconsistent than the photographs, they were not as reliably identified.

Comparably, the fact that the scenes and objects used in the photographs were realistic household images resulted in their increased recognisability and reliability in identification. However, this familiarity also resulted in difficulty in selecting truly impossible scene-object combinations and highly unlikely objects in scenes. Household objects were often considered suitably consistent with one scene but not excessively inconsistent in an alternative scene, as the object could have appeared misplaced. Therefore, although the scene images were highly appropriate in terms of semantic processing, the manipulation of semantic consistency could not be as strong. The data obtained in this study have highlighted the strengths and weaknesses of both stimuli sets and provided an alternative investigation of experimental data, by identifying the most appropriate stimuli for further analysis in each case.

References

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research, 43*, 277-292.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organisation.* Hillsdale, New Jersey: Erlbaum.

Biederman, I., Glass, A. L., & Stacy, E. W., Jr. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*(1), 22-27.

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

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103*(3), 597-600.

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(3), 531-543.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 556-566.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 207-230.

Carlson-Radvansky, L. A., & Irwin, D. E. (1995). Memory for structural information across eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1441-1458.

De Graef, P. (1992). Scene-context effects and models of real-world perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 243-259). New York: Springer Verlag.

De Graef, P. (1998). Prefixational object perception in scenes: Objects popping out of schemas. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 313-336). Amsterdam: Elsevier.

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

De Graef, P., De Troy, A., & d'Ydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology, 46*, 489-508.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108*(3), 316-355.

Germeys, F., & d'Ydewalle, G. (2001). Revisiting scene primes for object locations. *Quarterly Journal of Experimental Psychology, 54A*(3), 683-693.

Henderson, J. M. (1992a). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology, 46*(3), 319-341.

Henderson, J. M. (1992b). Identifying objects across saccades: Effects of extrafoveal preview and flanker object context. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(3), 521-530.

Henderson, J. M. (1997). Transsaccadic memory and integration during real-world object perception. *Psychological Science, 8*(1), 51-55.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception and Psychophysics, 59*(3), 323-346.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 449-463.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception and Psychophysics, 45*(3), 196-208.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210-228.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127*(4), 398-415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica, 102*, 319-343.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in scenes. *Visual Cognition, 7*(1/2/3), 213-235.

Hollingworth, A., & Henderson, J. M. (submitted). Schema normalisation fails to account for semantic informativeness effects on change detection in scenes.

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 and Review, 8*(4), 761-768.

Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*(5), 1211-1222.

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 565-572.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics, 2*(11), 547-552.

McCauley, C., Parmelee, C. M., Sperber, R. D., & Carr, T. H. (1980). Early extraction of meaning from pictures and its relation to conscious identification. *Journal of Experimental Psychology: Human Perception and Performance, 6*(2), 265-276.

McConkie, G. W., & Loschky, L. (1997, 28-29 January 1997). *Human performance with a gaze-linked multi-resolutional display.* Paper presented at the Proceedings of the Army Research Laboratory Advanced Display and Interactive Displays Federated Laboratory First Symposium, ARL Adelphi Laboratory, Adelphi, MD.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 6*(4), 391-399.

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory and Cognition, 3*(5), 519-526.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*(4), 587-595.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*(3), 426-442.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 199-210.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965-966.

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

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*, 618-660.

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

Saida, S., & Ikeda, M. (1979). Useful field size for pattern recognition. *Perception and Psychophysics, 25*, 119-125.

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8*(5), 374-378.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5*(4), 195-200.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception, 26*, 1027-1038.

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception, 18*, 347-361.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

van Diepen, P. M. J., De Graef, P., & d'Ydewalle, G. (1995). Chronometry of foveal information extraction during scene perception. In J. M. Findlay & R. Walker & R. W. Kentridge (Eds.), *Eye movement research: Mechanisms, processes and applications* (pp. 349-362). Amsterdam: Elsevier.

van Diepen, P. M. J., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception, 27*, 1141-1151.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1995). *The use of coarse and fine peripheral information during scene perception.* Paper presented at the Eighth European Conference on Eye Movements, Derby, UK.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337-355). Amsterdam: Elsevier.

Verfaillie, K., & De Graef, P. (2000). Transsaccadic memory for position and orientation of saccade source and target. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1243-1259.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum Press.