UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH & LIFE SCIENCES

School of Psychology

The Influence of Context on Object Recognition

by

Mark Edmund Auckland

Thesis for the degree of Doctor of Philosophy

April 2005

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH & LIFE SCIENCES
SCHOOL OF PSYCHOLOGY

Doctor of Philosophy

THE INFLUENCE OF CONTEXT ON OBJECT RECOGNITION

by Mark Edmund Auckland

The thesis explores how non-target objects influence object recognition. In all five experiments, sets of non-target objects are used to generate 'scene' contexts and these are presented so that they surround individual target objects. The foci of investigation are (1) whether scene context effects with multiple objects exist, (2) if they exist are they perceptual or due to response biases, (3) what role does the distribution of attention play in the generation of scene context effects, and (4) what is the time-course of their generation?

Experiments 1-3 found that target objects were named more accurately when non-target objects were semantically related (context-consistent) than semantically unrelated (context-inconsistent). However the magnitude of the context effect was mediated by visual attention. A significant effect was only achieved when all objects (targets and non-targets) were within an attended region and not when non-targets fell outside of this region.

Experiment 4 used a paradigm conceptually related to the Reicher-Wheeler paradigm to provide a measure of response bias. A six-alternative forced-choice response design demonstrated a significant influence of scene context even after the data were corrected for response bias; suggesting a perceptual/representational locus to the scene context effect generated by non-target objects on target objects.

Experiments 4 and 5 also manipulated the time-course of the onset of non-target objects relative to target objects. The results showed that at least 52msec was required for the presence of non-target objects to influence recognition of target objects. In other words, the scene context effect for multiple non-target objects requires at least 52msec to accumulate.

In summary, scene context effects for multiple non-target objects on target objects directly influence the representational processes of target recognition. Furthermore their magnitude is dependent on the distribution of attention across the visual field and the temporal relationship of non-targets and targets. How these factors influence the modelling of object recognition is also considered.

CONTENTS

LIST OF FIGURES

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

Chapter 5:

Chapter 6:

LIST OF TABLES

Chapter 3 (Experiment 1):

Chapter 3 (Experiment 2):

Chapter 4 (Experiment 3):

Chapter 5 (Experiment 4):

Chapter 5 (Experiment 5):

Acknowledgements

Professional:

This thesis was supported by an ESRC studentship.

The data from Experiments 1 and 2 were presented at The 43rd Annual Meeting of the Psychonomic Society in Kansas City, November 2002 and at the European Conference of Visual Perception in Paris, September 2003. The data from Experiment 4 were presented at the Experimental Psychology Society Summer Conference in Reading, July 2003 and 44th Annual Meeting of the Psychonomic Society in Vancouver, November 2004.

Personal:

Many people have accompanied me through some or all of the period whilst I have been conducting this research, and the unfortunate ones have also put up with me whilst I have been writing up this thesis. My great thanks to all of these people who have helped to make the time since starting my research the happiest I have known. In no specific order these are:

Claire, Matt, Martin, Ailsa, Cara, Helen, Ina, Luke, Shui, Sarah & my family.

I should especially like to thank Kyle for putting up with my continuous questions while designing and running experiments, and Nick for persevering with my "Aucklandisms" during writing up.

Finally, I would like to thank Helen for her tolerance and support during the seemingly never-ending process of writing up.

Thanks!

Chapter 1

The Influence of Context on Object Recognition

Outside the laboratory letters typically make words, and words exist in sentences on a page; eyes, noses and mouths generally exist within faces; and keyboards, monitors and computers appear together on office desks. Put another way, object parts (and letters) appear in whole objects (and words), and objects that are our current focus of attention are normally surrounded by semantically consistent objects rather than in isolation. Our recognition systems benefit from this coherence, producing an advantage known as a context or superiority effect (e.g. Bar, 2004; Biederman, 1981; Davenport & Potter, 2004). Experimentally superiority effects and have been found with words (Cattell, 1886; Reicher, 1969; Wheeler, 1970), faces (Homa, Haver, & Schwartz, 1976), objects (Weisstein & Harris, 1974) and scenes (Biederman, 1972).

It is a manifestation of the scene context effect that forms the focus of this thesis. In particular, this research is concerned with how sets of non-target objects, appearing around target objects, influence the perception and recognition of these targets. Displays containing collections of objects are scenes in the sense that naturalistic scenes usually contain multiple semantically related objects, in addition to their spatial relationships and global background. It is the possibility that these semantically related objects can influence the perception of targets that forms the starting point for this thesis. Despite currently available evidence of the scene context effect current theories of object recognition do not consider how non-target objects influence the perception of a target. That this is the case, and why it is so, will be discussed later in Chapter 1.

Chapter 1 begins with a discussion of superiority effects. In general superiority effects reveal the need to consider contextual influences on the processing of visual targets. However, currently available evidence on the scene superiority effect is inconclusive. Studies, using naturalistic scenes, cannot clearly show that the contextual non-target influence facilitates the perceptual or representational processes of target recognition rather than biases the participant response. This unresolved issue of scene superiority/response bias provides one of the key motivations behind the whole thesis and it has important implications to the role of context within the wider framework of recognition. To provide a basis upon which the relationship between context and recognition can be understood current models of object recognition are outlined in the latter sections of this chapter.

Context and Superiority Effects:

The scene context effect is different from all other contextually driven superiority effects in the degree to which the mechanisms that lead to its generation are still open to debate. The problem is further confused by studies which have not attempted to eliminate response bias and must therefore use the term 'context effect' rather than 'superiority effect'. In order to understand why a scene context effect may not be a genuine superiority effect, I start by reviewing the main groups of superiority effects. What will become clear is that word, face, and object superiority effects result from perceptual/representational processes, and while scene context effects may have similar origins, it is equally likely that they are driven by complex response factors that bias naming (e.g. interference, STM capacity) and recognition (e.g. image quality, attentional load).

Word Superiority:

The word superiority effect was first reported in 1886 by Cattell, who found that participants could report more letters from a valid English word displayed for 10msec than from a random letter string. However, these results could have demonstrated improved ability in remembering letters from the English words rather than ability in identifying them. Reicher (1969) and Wheeler (1970) used the two-alternative forced choice (2AFC) paradigm to address this issue methodologically. In their studies they sequentially presented a word or random letter string, a visual mask, and a choice of two letters. One of the two letters (e.g. K) was shown in the word or letter string. This correct letter was then displayed along with a false choice (e.g. D) above the masked location in the original word or letter string where the target appeared. Participants were required to identify which of the two choice letters (e.g. K or D) they had seen previously in the word or letter string. Both alternatives could be substituted to form a word in the word condition (e.g. WORK and WORD) making each choice equally viable. By controlling for such guessing strategies, this design allowed for the effect of response bias to be eliminated so that the context letters could be shown to directly influence the perceptual processing of the target. Both with and without a pre-cue (which displayed the 2AFC choices prior to the word/letter string) participants were better able to identify letters when they were part of a valid word rather than part of a random letter string. This pattern of results is referred to as the 'word-nonword effect'. Reicher also included a condition whereby a single letter was displayed in place of a word or letter string. It was expected that this would yield the best performance due to the absence of distractors, yet error rates were still higher than in the 'word' condition. This finding is referred to as the 'word-letter' effect.

McClelland and Johnston (1977) extended these findings to show that letters in pronounceable nonwords (e.g. TRAG) are identified more accurately than in unpronounceable nonwords (e.g. ATGR) even if formed from the same elements. However, valid words are still identified more accurately than pronounceable nonwords indicating that psycholinguistic coherence facilitates the perceptual processing of the letters within the word or letter-string. Word superiority effects have been accounted for in connectionist models of recognition (e.g. McClelland & Rumelhart, 1981; Mozer, 1991). The presence of top-down processing allows activation from the word to spread back to the level of the letter processing.

Object and Configural Superiority:

Weisstein and Harris (1974) found an effect similar to word superiority when they asked participants to discriminate between four alternative diagonal line segments (Figure 1: a-d) under varying 'contextual' conditions (Figure 1: e-g). All conditions contained the same number of horizontal and vertical lines. When the target line formed part of a three dimensional image with the context (e.g. Figure 1: e) it was reported more accurately than when it was part of a two dimensional configuration (e.g. Figure 1: f-g). Weisstein and Harris proposed that this is inconsistent with a model using only bottom-up processing in which elementary features are detected before the overall structure. They suggested a holistic analysis, a feedback loop within a connectionist structure, or detectors for both simple and complex features as potential ways of resolving this inconsistency. All of these are ideas that have since been utilised by one or more object recognition theories (see below).

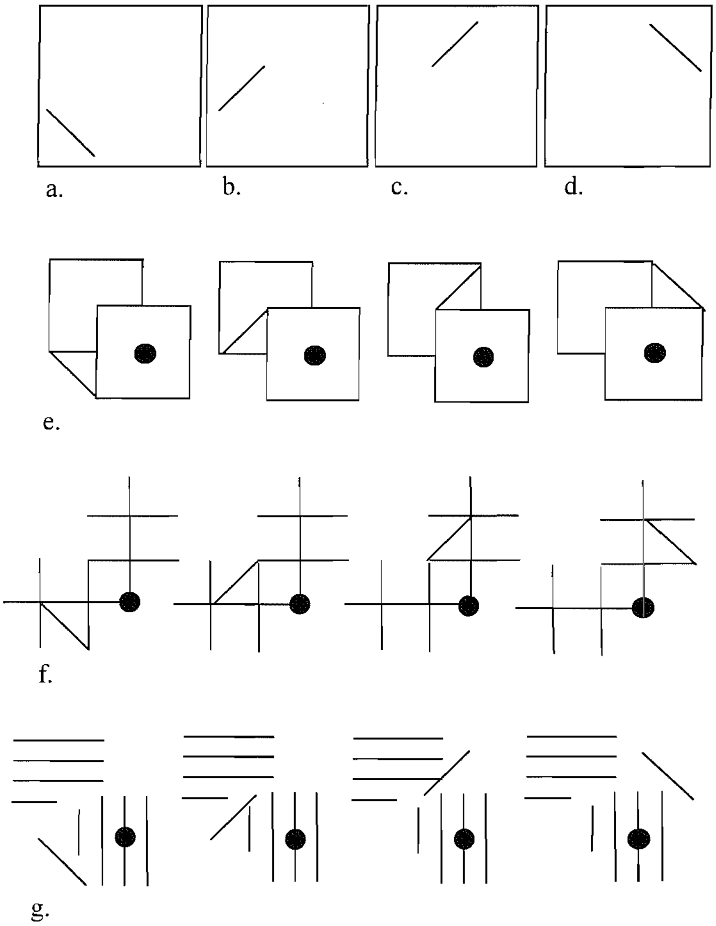


Figure 1: Examples of target and context stimuli used by Weisstein and Harris (1974)

Unlike Reicher's (1969) word-letter effect, Weisstein and Harris (1974) found that any increase in non-target 'irrelevant' stimulation reduced accuracy further. However, an object-line effect has been found under specific conditions (Enns & Gilani, 1988; Williams & Weisstein, 1978) in which target lines are identified more accurately within a context of other lines than when presented in the visual field alone. McClelland and Miller (1979) proposed that it was not the three-dimensionality of the objects per se that generated the object superiority effect, but the structural relevance of the line segment within the object. This view is more consistent with the results of Enns and Gilani (1988) who found that three-dimensionality could contribute to the object-line effect but was not necessary to its generation. More important was the level of discriminability created by the context layouts (line patterns) with and without the target-line.

The configural superiority effect (Pomerantz, Sager & Stoever, 1977) also demonstrates improved performance when discriminating simple configurations (e.g. arrows or triangles) rather than the configural parts (e.g. a diagonal line). It has been proposed (Palmer, 1999) that the configural parts combine to create emergent features and that the activation from these is greater than that of the parts. Whilst the results support the more rapid discrimination of these emergent features, neither the reason they might be easier to detect, nor the precise mechanisms involved, is clear.

Face Superiority:

Facial features are more accurately identified when they are processed following a face than a scrambled face (Homa, Haver & Schwartz, 1976; van Santon & Jonides, 1978). Such an effect appears similar to the word-nonword effect of word superiority but if a facial feature (e.g. a specific mouth or nose) is presented first it is more quickly discriminated when followed by a scrambled face stimulus than a face (Mermelstein, Banks, & Prinzmetal, 1979). They proposed that facial arrangements encode more features with less featural detail than scrambled face arrangements, because they generate a 'perceptual gestalt' or completeness of image. This gestalt biases selective attention away from specific facial features to focus upon the whole image, however when a single target has been encoded (e.g. a mouth) it is advantageous to direct attention to a single feature. The gestalt, or holistic processing hypothesis, has gone on to dominate the study of face recognition (e.g. see Tanaka & Farah, 2003 for review).

The relationships between features provide a source of information to facial arrangements referred to as configural feature information (Sergent, 1984; also referred to as: 'relational' - Diamond & Carey, 1986; 'second order' - Rhodes, 1988) that is not available with scrambled faces. The results of these relationships were demonstrated by Young, Hellawell and Hay (1987) using stimuli in which the top half of one well known face was fused with the bottom half of another well known face. The composite photographs produced novel configural cues and caused difficulty in recognising the original celebrities. Tanaka and Sengco (1997) have provided evidence that these cues are also integrated holistically along with the feature information.

The advantage gained from the facial configuration in recognising facial features generalises to the perception of other objects in the same spatial arrangement as features in an upright face. Davidoff (1986) found that as long as the typical facial feature arrangement was maintained, the facilitation was present when facial features were replaced by cars, telephones or leaves. From this it would seem that a facial gestalt unit can emerge without the presence of genuine facial features. This result suggests a low-level activation that is based in part upon location/arrangement. The facial arrangement of features allows a 'perceptual gestalt' to be generated at an early stage. This gestalt biases the distribution of attention across more features and establishes the relationships between them. The whole influences the perception of the parts through the facial superiority effect but the absence of a single feature advantage, similar to the word-letter or object-line effects, highlights that not all superiority effects operate according to identical principles.

In conclusion there is strong supporting evidence for each of the above superiority effects. Effects have been found using empirical methods based upon concepts devised by Reicher (1969) and Wheeler (1970) to eliminate response bias. It is this ability to isolate the influence of context alone, and its effect on guessing, that makes clear the interactions of context with the perception of individual components.

Context Effects and Scene Superiority:

In contrast to the word, face and object superiority effects, which have an undeniably perceptual/representational locus, scene context effects produce a more complex pattern of findings. Initial studies were interpreted within a perceptual/ representational framework but more recent studies have cast doubt on this conclusion. In this section I consider these studies.

Research demonstrated that objects could be more rapidly detected when displayed within a meaningful scene (Biederman, 1972; Biederman, Glass & Stacey, 1973) though Biederman's (1981) distinction of five classes of relations later refined the definition of a well-formed scene. The general physical constraints of *support* and *interposition* reflect that objects do not float in the air, and that opaque objects will occlude the area behind them. These constraints can be used regardless of object knowledge. In addition to these physical constraints, there are three others: *probability* refers to the likelihood of an object being in a scene; *position* refers to objects that occupy a specific position within a scene; and *size* limits objects to a familiar size. According to Biederman these latter three relations require access to semantic knowledge or referential meaning of both the object and the scene in order to be specified successfully. He claimed that when one or more of the five principle relations is violated the scene structure is weakened and the facilitating scene schema (e.g. kitchen, street) is consequently less likely to be invoked. These schemas were positioned above objects within the recognition hierarchy and could be activated either through the identification of several objects or through emergent properties within the scenes. Their activation then facilitated the selection of objects semantically related to that schema. Object recognition performance supported this hypothesis. Error rates and response times increased in an object detection task when a scene was briefly displayed (150msec) dependent upon the number of relations violated (up to a maximum of three), implying a processing advantage for objects within scenes that maintained relational consistency (Experiment 1: Biederman, 1981). Well-formed, coherent scenes are processed more quickly and accurately.

Other scene influences have also been reported. Palmer (1975) examined the impact of an extended scene presentation (2 sec) prior to the brief display of a target. It was found that when the visual scene was of a target consistent context (e.g. a kitchen and a loaf of bread), fewer naming errors were made than if no scene was presented, and the error rate was highest when the visual scene was target inconsistent (e.g. a kitchen and a mailbox). This consistency advantage relates directly to Biederman's (1981) class of probability, and is the constraint that is most commonly tested in investigations of the context effect. Although the observed effect appears similar to that exhibited by both the word (Reicher, 1969) and object effects (Enns & Gilani, 1988) there is an important methodological difference. Palmer's (1975) design allows the context to be viewed for a sufficient length of time to allow some context objects within the scene to be processed as targets themselves. A similar confound is also found in work on object recollection by Intraub et al. (1996) and Gottesman and Intraub (1999). In these studies the authors identify the phenomenon of boundary extension, a process of memory distortion by which individuals extend a visual scene they have viewed to include consistent objects they have not seen. For example a book is assumed to have been in a college professor's room because it is consistent with their typical environment. In this instance the scene influence has a negative, or misleading, effect. Context may have a pre-recognition influence as a result of time to process context items (e.g. objects within a scene) as targets themselves but that is not what is under investigation in this thesis. To identify if context directly interacts with the recognition process a brief or simultaneous display of context and target will be required.

Eye movement would also have been present in Palmer's (1975) study, and has been examined specifically by Friedman (1979) and Loftus and Mackworth (1978). These studies produce data relevant to the scene consistency advantage as they find that participants fixate earlier and dwell for longer periods on objects inconsistent with their scene, and that participants also return to refixate those objects more frequently. These results could be interpreted as being inconsistent with the scene consistency advantage if they are seen as a facilitation of the processing of objects that violated the probability relation. However, it seems more likely that objects that are consistent with the scene can be processed more efficiently, with less fixation time. The context inconsistent object would be identified as an odd-object-out, and would be given more processing time because it is the area of most interest.

More recent research has begun to address the question of whether scene context effects play an important part in the early stages of normal target-scene viewing. Boyce and Pollatsek (1992) found that when the background was a coherent scene rather than a 2-D pattern it facilitated target naming (Experiment 1). This gain occurred both when the scene was displayed 75msec before the target was flagged to the participant and when the background was altered simultaneously with the flagging of the target (Experiment 3) suggesting that the scene schema, or context, can be extracted in advance if the opportunity presents itself, but can also be extracted during scene viewing to influence object identification. Davenport and Potter (2004) have demonstrated that target-scene consistency generates a significant object naming advantage, even when target and scene are displayed simultaneously and for brief exposures (80msec).

There have been a few studies which replicate the complex effects of scenes in the relatively simple paradigm of object arrays. In addition to his scene studies, Biederman (1981) also conducted an Experiment (4) using this methodology in which participants were provided with a target name, then presented with a brief display of 3-6 objects in a clock-face layout and required to detect whether a target was present or absent. This design focuses on the probability class of relation by manipulating whether the target is semantically related or unrelated to the context items. It was found that participants were more accurate with low probability targets than with high probability targets (using d' to account for response bias), but can be explained by the extended response times (>600msec). These long responses may have resulted in eye movements, and thus findings similar to those in the eye movement studies of Friedman (1979) and Loftus and Mackworth (1978). Consequently participant attention will have dwelt on the more interesting, less-consistent, low probability items.

Superiority effects are defined as those that influence perceptual/representational processes. Demonstration of these superiority effects requires that an effect of context remains after the elimination of response bias. Very few contextual studies have analysed response bias and therefore most of these studies have demonstrated a scene context effect without determining whether it is a scene superiority effect.

Rather than using the methods of Reicher (1969) and Wheeler (1970) Davenport and Potter (2004) used a naming experiment in which they counted erroneous responses semantically related to the context as false alarms, which were then deducted from the hits. However, their design was not a forced choice experiment so very few false alarms were found amidst a large number of non-responses. Based on the low rate of semantic errors, they claimed response bias had a negligible effect. Biederman (1981) also eliminated response bias as an explanation for the context effect as he indicated that both false alarms and misses had increased. However, Hollingworth and Henderson (1998) highlighted that this context effect in Biederman's (1981) Experiment 1 could be due to a failure to control for participants' increased likelihood of answering "yes" (present) in semantically consistent catch trials. The average false alarm rates across base and violation conditions therefore resulted in an over-estimation of sensitivity for the detection of consistent objects. Hollingworth and Henderson (1998) replicated Biederman's (1981) Experiment 1 and corrected for this over-estimate by calculating detection sensitivity separately for consistent and inconsistent conditions. They found that the context effect was removed, and concluded that the effect could be entirely explained by response bias. This conclusion formed the basis of their Functional Isolation hypothesis. Further experiments (Hollingworth & Henderson, 1998, 1999) challenged whether the processing of perceptual information was facilitated by a consistent context if response bias was controlled. No advantage was found when the target was context consistent, and the reliable trend was for improved performance when the target was context-inconsistent.

There are design issues with Hollingworth and Henderson's (1998, 1999: see Chapter 5) experiments, and therefore the debate currently remains unresolved.

Summary of Contextual Influence:

Context effects occur reliably across many different types of perceptual stimuli. In the case of the perception of single words, faces or simple objects, the evidence points to a perceptual locus for context effects. Furthermore, in each case it is possible to generate plausible models that account for context effects. However, in the case of the scene context effect it is less clear whether context effects have a perceptual locus or result from post perceptual processes that are affected by response biases. It is even possible that there may be discrete perceptual and response bias effects driven by contextual processing in scenes.

Current models of multiple object recognition do not account for the results from context research. There is an acceptance of the scene context effect on recognition tasks, but not of whether this is a genuine scene superiority effect. However, even if contextual stimuli do not influence the perception of targets, the process by which non-target items are identified pose questions about how recognition is accomplished.

Models of Object Recognition for Single Objects:

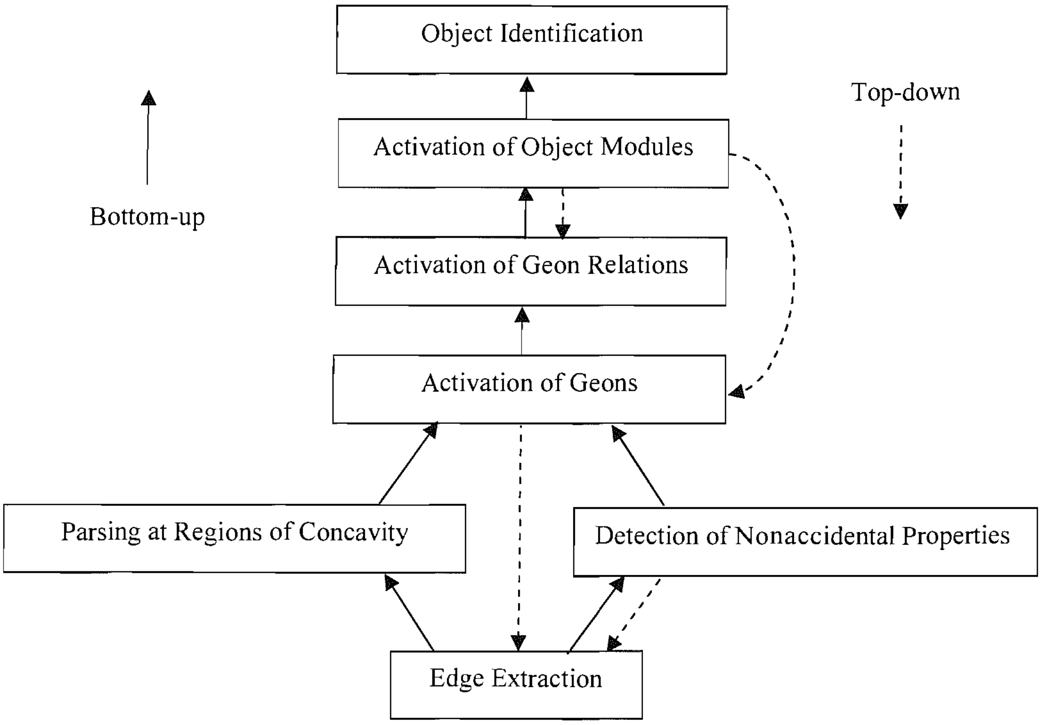
The generation of the scene context effect, even if not from a naturalistic scene, is linked with recognition on two levels. Firstly, the multiple contextual objects within a scene must be processed and it is not known whether these non-targets are processed via the same system as the target, or via a secondary mechanism. Secondly, contextual information from these multiple non-target objects may influence the representational processes involved during target recognition. Therefore, this section examines the three main types of object recognition model.

Structural Models of Object Recognition:

Structural, part-based or piecemeal models of object recognition constitute one of the major paradigms in the field. Processing occurs not through attempting to match the whole object to a stored representation, but through the decomposition of a target into its component parts (or features) and their interrelations. The stored representations themselves are assumed to be analytic, being made up of a list of parts and of their relations specified explicitly and independently of the parts themselves (e.g. “above”). A principal benefit from such a system would be its invariance to size, left-right reflection, and rotation in depth once the identities of the component parts can be established. In addition, this type of system needs only a small number of potential 'parts' to form a large number of different objects. It is faster and less difficult to match a simple geometric shape from a limited selection than to attempt to compare an entire object against many complex representations stored in memory.

An established example of the structural description paradigm is Biederman's (1987) Recognition-by-Components theory (RBC). It is based upon the idea that objects can be constructed from various arrangements of a limited number of 'primitive volumetric components', which he named 'geons'. Biederman states that 36 geons are sufficient to express the estimated 30,000 objects found in basic visual categorisations (approximately 3000 categories multiplied by an average of 10 exemplars per category). Initially the perceptual process extracts edges from the optical image based upon surface characteristics (e.g. luminance, texture) to generate an internal line drawing. Non-accidental properties (collinearity, curvature, symmetry and co-termination) of the drawing are then detected, and regions of concavity act as the principal guide to parsing the image. The resulting geometric components activate stored geon representations, which can be matched against identity lists in memory. Relationships between the geons are also established and matched against the analytic representations held in memory. Reliable recognition is achieved when there is a good match for both the geons activated and the relative relationships between geons (see Figure 2).

Within RBC geon relationships may not be essential for a successful match if the collection of geons making up the target object is sufficiently distinct. Only when an object utilises geons found in several other objects would there be risk of a geon 'illusory conjunction' (e.g. identifying a bucket instead of a mug - see Figure 3 : adapted from Treisman & Gelade, 1980). In these conditions, the different geon relationships would determine the identity.



Object Identification
Top-down
Activation of Object Modules
Bottom-up
Activation of Geon Relations
Activation of Geons
Parsing at Regions of Concavity
Detection of Nonaccidental Properties
Edge Extraction

Figure 2: Presumed processing stages in RBC theory (adapted from Biederman, 1987)

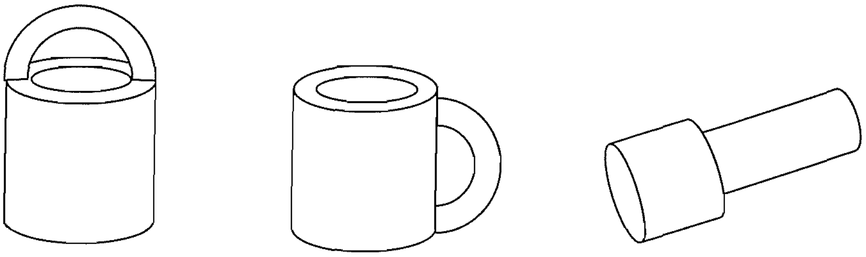


Figure 3: Similar geons can be combined in different ways to produce a bucket instead of a mug, but they will never be mistaken for a torch, which consists of two different components.

Support for this explanation is provided by evidence that individuals can identify pictures of objects formed from just two or three of their basic components (Biederman, Ju & Clapper, 1985 cited Biederman, 1987), and can recognise line drawings of objects as quickly as full colour detailed slides (Biederman & Ju, 1986, cited in Biederman, 1987). Also, the degradation of line drawings does not have a significant effect on recognition as long as it does not delete areas of concavity or co-termination vital for the construction of individual geons (Biederman & Blickle, 1985, cited in Biederman, 1987).

Structural systems such as the RBC encounter performance difficulties when a viewpoint renders a target's geons or geon relationships difficult to establish. For example, when viewed from above, the volumetric definition of many components of a car could not be ascertained and important geons would be obscured (e.g. the wheels). This is not a problem with the paradigm as Palmer, Rosch and Chase (1981) have demonstrated that objects can be more readily identified in some orientations (e.g. canonical) than others. Biederman (1987) suggests that the orientations that maximise performance activate geon combinations shared by fewer objects. Surface characteristics (e.g. colour, luminance and texture) can also provide contour or shape information to allow geon derivation. In addition, diagnostic surface information (e.g. bananas are yellow) may be used to further reduce the number of potential matches. Structural systems are primarily driven by bottom-up processes; however, there is scope for the inclusion of top-down processes within such a model. At the lower levels, top-down activation of a geon or of certain non-accidental properties may assist in edge extraction. At later processing stages, the top-down activation of an object module (see Figure 2) via an alternative source (e.g. context) offers potential benefits to the activation of geons or their relations.

Questions have been raised as to whether structural systems are models of identification or categorisation (e.g. Palmer, 1999). Although the 36 geons of the RBC can sort even novel stimuli into basic level categories (e.g. trees or birds), they do not provide the fine detail required to identify non-distinct exemplars within a category (e.g. oak or yew, blackbird or sparrow). This problem might be resolved using a finer gradation of quantitative parameters in geon descriptions in order to decompose objects into a larger number of more precisely sculpted parts. However, the impact on model efficiency and feasibility of the additional processing must be considered.

Biederman (1987) states the capacity for activating and matching geons is high as they are processed in parallel. This implies that the matching is done via pre-attentive processing as focused visual attention (a limited resource) would restrict capacity, and there is empirical support that some aspects of shape (Donnelly, Humphreys & Riddoch, 1991) and 3-dimensionality (Enns & Rensink, 1990) can be processed at an early stage. With only geons activated, an RBC object is conceptually similar to a pre-attentive object file (Wolfe & Bennett, 1997) containing a loose bundle of unbound features. These assumptions suggest that an increase in the number of geons alone would not disable the model's effectiveness. However, empirical research also links focused visual attention to the binding of features (e.g. Treisman, 1996; Wolfe & Bennett, 1997; Wolfe & Cave, 1999) and this binding is reflected in the relationships between geons. An increased number of geons will inherently result in an increased number of geon relationships. RBC (Biederman, 1987) does not explicitly refer to the attentional requirements of geons and geon relationships, although the positioning of the latter in the model (Figure 2) does suggest that geon relationships are processed at a higher level than simple geon activation. Binding was an issue directly addressed in RBC's neural network successor JIM (Hummel & Biederman, 1992). Binding was achieved via synchrony, but as geon numbers were increased so that stimulus complexity approached levels of realistic proportions the model encountered instances of 'accidental synchrony' (similar to illusory conjunctions - Treisman & Gelade, 1980). It was proposed that even using synchrony as a binding mechanism, selective visual attention would have to be integrated into the model to limit the likelihood of accidental conjunctions.

Hummel and Biederman (JIM - 1991) indicate that the activation of geon relationships requires selective visual attention; thus geons may be processed in parallel without capacity restriction but their interactions cannot.

View-based Models of Object Recognition:

View-based models of object recognition utilise holistic representations. Unlike the analytic representations used by structural models, view-based representations define feature relations relative to a single reference point within a spatial view (Hummel 2000). Objects are identified by feature locations, and the relationships between these features cannot be manipulated if the target object is not correctly aligned with a stored representation. There is evidence for holistic processing that indicates an ability to match objects in which size has been altered, or which have undergone a partial two dimensional rotation. However, an image that undergoes a three-dimensional rotation or left-right reflection cannot typically be matched by a holistic representation. Such systems are viewpoint-dependent and consequently most view-based models require multiple stored representations for each object (for an exception see Lowe, 1987).

The principle behind multiple-view theories (e.g. Poggio & Edelman, 1990; Tarr & Pinker, 1989, 1990) is that the 2-dimensional image is matched directly against a stored representation. The stored representation need not be a detailed or coherent picture as the holistic referential information could be held in the form of a feature map (Palmer, 1999). Thus, salient points, edges and vertices of an image could be used to structure a mapping system of partial templates that could be processed pre-attentively. Hayward and Tarr (1997) claim this model offers a higher degree of specificity than that available to the structural models. More detailed templates or maps may be utilised, although additional perceptual data would then need to be acquired for any matching benefit to be gained.

The number of two dimensional views stored for objects varies between theories. Poggio and Edelman (1990) applied the premise that sufficient 2-D views of an object were equivalent to specifying the 3-D structure. Their learning network utilised relatively few viewpoints (for a multiple-view theory) to approximate an object-specific function that mapped any perspective view into a 'standard' view. When applied to views of different objects, such a function would result in a 'wrong' standard view, equivalent to a failed match against a stored representation. An alternative approach (e.g. Tarr & Pinker, 1989) is that the number of stored views for an object depends upon past experience. Those perspectives maintained within memory are those that are likely to have been seen most frequently, but there is potential to create additional views should novel conditions occur. This approach highlights a potential problem with view-based models in that each object in memory could accumulate many viewpoint representations: a risk increased for non-rigid items or those capable of motion. There would be an inherent load upon memory capacity, and the pool of potential matches to be searched during recognition would be large. Some form of 'housekeeping' mechanism would be required to limit the total number of stored views if such a system were to be feasible.

With either approach, novel orientations and objects will be encountered. The spatial-referencing of features in these models allows these unfamiliar views to undergo a procedure of 'best-fit' matching. The various methods of approximation used within the different theories - for example the method of normalisation used by Poggio and Edelman (1990) - all require time and their influence on performance can be measured.

Empirical evidence for this paradigm is provided by Tarr and Pinker (1989) in a study that demonstrated recognition for two dimensional figures was viewpoint-dependent. During a training phase, participants were shown unfamiliar letter-like figures several times in a limited number of orientations. They were then found to be quicker at recognising the same figures in the orientations for which they had been trained. Recognition of familiar figures in new orientations was achieved, but response times increased as a function of the difference between viewed orientation and the closest familiar orientation. As familiarity with the novel orientations increased through repetition during the task, the effect of orientation on response times decreased. These results suggest that individuals stored the trained views of the stimuli in memory, and further views were stored later to recognise the novel orientations. Tarr (1995) later replicated these findings with 3-D stimuli using depth rotation. More recently he suggested that a view-based model possessed all the requirements for a complete recognition system (including face recognition – Tarr, 2003).

Hybrid Part-View Models of Object Recognition:

Structural and view-based models demonstrate complementary strengths and limitations. Whilst neither type of model is generally accepted as fully explaining recognition, each offers supporting empirical evidence under particular conditions: a.) when a current view is matched to a view stored in memory, categorisation is fast and uses little or no attention (holistic); b.) if there is no ready match, then categorisation/identification requires a process of decomposing the objects into parts, which is slower and utilises attention (analytic); and c.) for type-specific detail of familiar objects, fine templates may be used, which benefit from extended data processing (holistic). Attempts to combine these paradigms have led to research into hybrid models of object recognition (Hummel, 2001; Hummel & Stankiewicz, 1996; Tarr & Bültoff, 1998).

Hummel (2001) maintains the distinction of two processing pathways. His model adopts a fast, holistic, view-based mechanism that can be utilised if the target is observed from a familiar viewpoint. Empirical evidence (Hummel, 2001; Hummel & Stankiewicz, 1996) suggests the resultant image will be viewpoint-dependent. For the processing of non-familiar targets (or targets from non-familiar views) a part-based mechanism similar to Biederman's (1987) RBC is employed to provide flexibility regarding viewpoint and allow for the encoding of object identities into long-term memory. Both pathways take the perceptual data from the image simultaneously, but the view-based mechanism is not limited by an attentional bottleneck because the holistic nature of the representation allows recognition to occur even when the object is not attended (Hummel, 2001). The structural mechanism is restricted by visual attention, and the target must be selected prior to processing. Hummel's model suggests that the familiarity of the object is the factor that mediates which pathway is utilised. However, more recent research by Thoma, Hummel and Davidoff (2004) proposed that analytic processing is used whenever the attentional resource is available, and that holistic processing is used only when there is no alternative. Another plausible account is that both pathways are used and the information is combined. Other researchers have also suggested that environmental factors, for example the decomposability of the object or its meaningfulness, may play a role in deciding which processing route is utilised (e.g. Smith, Dror, & Schmitz-Williams, in press).

Tarr and Bülthoff (1998) take a different approach to combining the strengths of part and view based models of object recognition. Rather than maintaining the distinction, their model attempts to fully integrate the two paradigms. They maintain the principle of interpolation across views to compare the similarity between images in order to make a match with those stored in memory, and extend it to include exemplars. When major changes occur in the three-dimensional views or exemplars (e.g. features come in or out of view), this interpolation may fail, so qualitatively different images of the same object are encoded. Linking together multiple images of the same object provides implicit structural information between the features, which potentially has more detail and greater flexibility than RBC. Tarr and Bülthoff also propose utilising a medial axis derived from the object silhouette to create a skeletal description of the object. This would be a coarse topological guide that would remain stable and constrain search space but could only provide limited explicit structural information.

Models of Object Recognition for Multiple Objects:

Structural, view-based and hybrid recognition models perform poorly or not at all when presented with multiple stimuli simultaneously. They can process an object in isolation, and this can include combining basic parts into wholes. However, these models are not designed to process more than one object at a time. For scene-based contextual influence to occur object recognition processes need to be considered both in the extraction of the non-target information, and in how that information is used to aid the recognition of the target. The three models already examined may provide some general ideas for information extraction, but other models are required to address how multiple objects can be processed to produce context effects. In this section two alternative areas (psycholinguistics and neuropsychology) are explored for theories that might be transferred to form a multiple object basis for a recognition model.

Psycholinguistic Models of Recognition:

Psycholinguistic models of recognition aim to identify letters and words rather than objects. Not all of these models are limited to targets in isolation. McClelland and Rumelhart (1981) do not provide a multiple stimuli model, but are included here because they provide a detailed explanation for word superiority and the groundwork for the MORSEL model (Mozer, 1991) that follows. McClelland and Rumelhart (1981) proposed the interactive activation (IA) model, a connectionist network which focussed on the identification of four letter words and the letters from which they were composed. It consisted of three levels, the first of which was a feature layer. Visual input directly activated any of the twelve line segments (nodes) in the four positions in which a letter could occur. Bottom-up processing transferred this information via excitatory and inhibitory connections onto the second level, that of the letter layer. The individual letters within the 104 letter nodes (26 letters x 4 positions) would be activated once the segments of a given letter were detected, and the activation pattern would be stabilised and sharpened according to a 'winner-takes-all' system of mutual inhibition and competition. In addition to the bottom-up activation, the letter layer also benefited from feedback loops allowing excitatory connections from the uppermost word layer of the model. The 1000 word nodes in the IA model received excitatory connections from one letter node for each of the four spatial positions within the word (e.g. 'A' in the first position excites 'ABLE' and 'ACTS'). For each of the four positions, inhibitory connections were received from the other 25 letters in an already represented position. As with the letter layer a 'winner-takes-all' network of mutual inhibition selected the word that emerges with the majority of the activation.

Feedback loops provided excitatory connections back from activated words to their constituent letters (e.g. the word 'ABLE' excites letters 'A', 'B', 'L' and 'E') in the relevant positions. These links allowed the IA model to replicate the specific psycholinguistic visual context effects found by Reicher (1969 - word-nonword effect) and McClelland and Johnston (1977).

Mozer (1991) extended psycholinguistic and connectionist theory with his model for multiple object recognition and selection (MORSEL - see Figure 4). In the shape detection module information is taken via a 36 x 6 retinotopic map that has five elementary features. Information then progresses upwards through levels of maps with decreasing dimensions and increasing feature types. The six levels that form this recognition network, called BLIRNET (because it builds location invariant representations of multiple letter strings), replace both the letter and word layers of

McClelland and Rumelhart's IA (1981) model. Rather than letters or words the output layer of BLIRNET has been trained to respond to letter-clusters (arrangements of three letters) that may form all or part of a word. For example 'EST' would be activated by 'BEST' or 'ESTIMATE', but not 'STEP' or 'SET', which uses the same letters in a different order. These clusters account for three letters in four consecutive slots, which may include adjacent combinations (e.g. ABC) or three letters separated by a single letter (e.g. AB_D). Asterisks were used to indicate spaces or the end of words. Thus a single word could activate a multitude of letter-clusters.

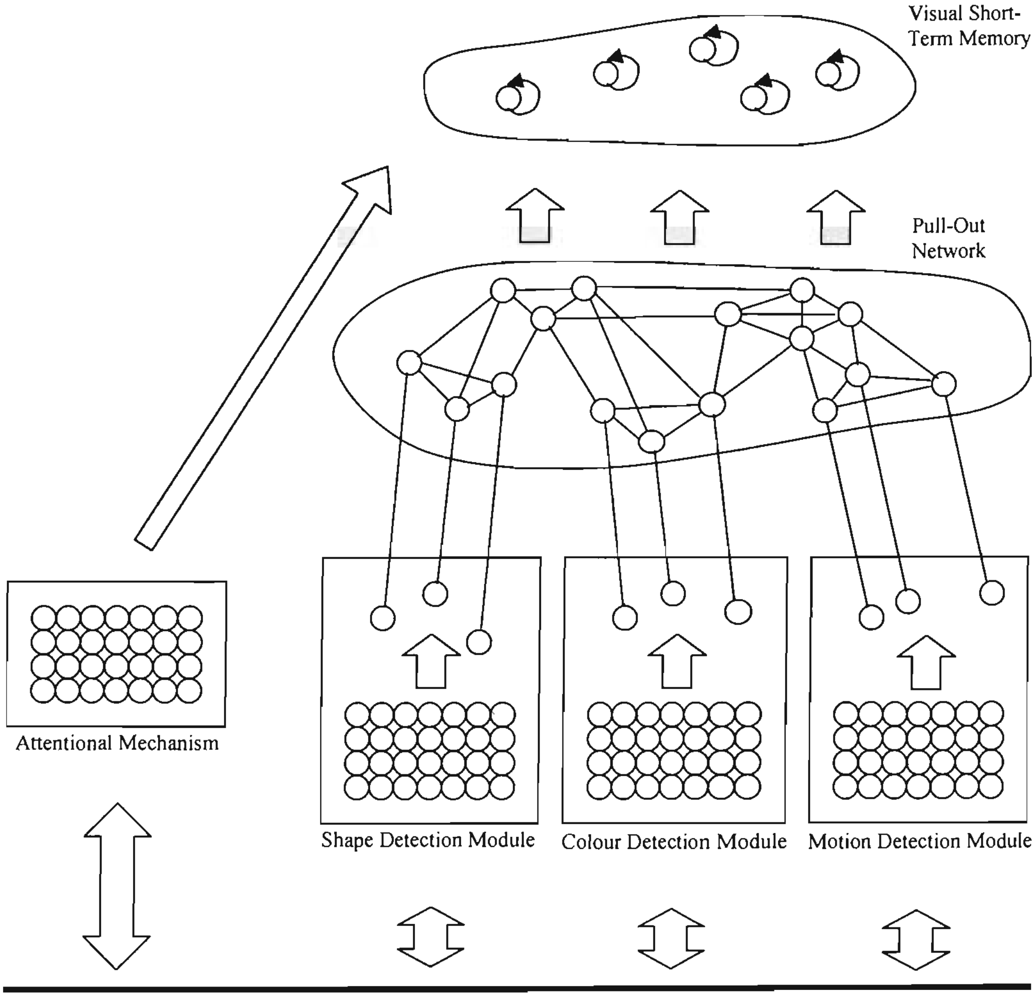


Figure 4: Outline of MORSEL (adapted from Mozer, 1991)

The large amount of information from the letter-clusters enables the simultaneous representation of multiple words. Providing the words are not too similar, or too numerous, there should be sufficient data within two letter-cluster units (collections of letter-clusters activated by a single word) to successfully reconstruct the relevant identities of the words. This is done in the Pull-Out Network (PO net), into which the letter-clusters are output from the shape description module.

One of the primary purposes of the PO net is to remove activation noise. To achieve this, letter-clusters are orthographically matched into the most consistently achievable overlapping strings (e.g. ABC is matched between *AB and BC*). Excitatory connections are formed between compatible neighbours, and inhibitory connections between incompatible neighbours, with special cases for clusters that form word endings. In addition to the orthographic matching MORSEL also attaches semantic feedback to the letter-clusters from the PO net word level via 'semlex' units (lexical representations of semantic features not shared by different words with similar meanings). The resulting winner-takes-all network is distributed, with a top-down element, and is capable of replicating word superiority effects (McClelland & Johnston, 1977; Reicher, 1969) and partial processing of multiple stimuli.

The attentional mechanism is credited with four principal tasks within the MORSEL model. First, it controls the order in which words (targets) are selected by fixing upon their location. Second, it assists in reducing crosstalk between simultaneously displayed words by focusing attention on one word at a time. Mozer notes that whilst the PO net allows processing of multiple words, this generates interference between words that prevents efficient matching. BLIRNET allocates visual attention to remove interference in a manner that serves to bias, rather than inhibit processing towards certain letter-cluster units. The attentional mechanism does not prevent unselected units from receiving processing as the focusing of attention can take between 50msec (Treisman & Gelade, 1980) to 200msec (Colegate, Hoffman & Eriksen, 1973) to occur. Attention is also considered necessary for the recovery of location information and for the co-ordination of information from the other independent detection modules (e.g. colour and motion).

In this sense the role of attention in MORSEL is similar to that within Treisman and Gelade's (1980) Feature Integration Theory.

Although MORSEL is presented primarily as a word recognition model, Mozer (1981) proposes that the principles outlined would perform equally as well for simple two-dimensional objects. However, this proposal is not expanded upon in detail and it is not immediately clear how the orthographic and semlex elements might translate into object recognition. Also, this is a model that aims to isolate the target from distractors. The basic framework (PO net) allows crosstalk between words, but the attentional mechanism works only to restrict interaction between stimuli. The structure of this model may offer some possible explanations for contextual effects that have not been fully explored.

The connectionist models often used within psycholinguistic theory have a biological validity due to the nature of their neuronal network structure, and they can be trained to replicate experimental results with some degree of success. They have also been transferred to other paradigms of recognition (e.g. JIM – Hummel & Biederman, 1991). However, their inherent complexity has often led to limitations being placed on the type of stimuli they can process (e.g. number of words or simple images) that force questions to be raised about ecological validity.

Contributions from Cognitive Neuroscience Input to Object Recognition:

Recent research has suggested that the recognition process may not begin with fine detail features such as edges and vertices as assumed by bottom-up models that focus on the representation and recognition of single objects. Instead it suggests that coarse perceptual information is available more rapidly than information about fine detail. Bar (2003) suggests that anatomical 'shortcuts' from the visual areas direct to the prefrontal cortex (PFC) conduct low spatial frequency information for early processing of a target image. These feed-forward activations are based upon coarse detail and cannot lead to recognition with a high degree of certainty. Studies also indicate that whilst the PFC is capable of differentiation between categories (Freedman, Riesenhuber, Poggio & Miller, 2001) representations stored in the inferior temporal cortex (IT) are required for within-category discrimination (e.g.

Tanaka, 1993). Thus, unless the target is particularly distinct, the low frequency analysis can only reduce the number of potential candidate objects under consideration. The representations of these 'best guesses' are then activated in the IT to provide integrative and top-down feedback to the bottom-up process. This secondary pathway direct from the visual area to the IT consists of high spatial frequency data (e.g. fine detail) that takes longer to process than low frequencies (Schyns & Oliva, 1994; however see Oliva & Schyns, 1997). Once the selective process is complete the remaining potential matches are inhibited.

A feed-forward/feedback mechanism is not utilised in all cases of recognition. Bar (2003) states that when recognition is easy, the PFC may not have sufficient time to develop the facilitative feedback before the bottom-up process has selected the correct representation. Alternatively, the target may be sufficiently distinct that the high frequency detail is not required. Recent research (Bar, 2004) suggests that a scene can be processed simultaneously with the target, along lower frequency pathways, in the parahippocampal cortex (PHC). This parallel processing allows initial guesses of the target to be filtered according to the activated context frame/schema. Thus, a 'fridge' interpretation of a grey rectangular blob would be preferred to a 'safe' if the scene is known to be a kitchen. Neurological evidence indicates two peaks of activation (130msec/230msec) in the PHC, suggesting an 'initial guess' followed by a post-recognition activation of conceptual knowledge (Bar & Aminoff, 2003). Also, clinical studies have found that recognition is slowed but still present in patients with an impaired top-down function due to lesions in the frontal cortex (e.g. Richer & Boulet, 1999).

Evidence for coarse-to-fine processing in objects and scenes has been reported previously (e.g. Schyns & Oliva, 1994), as has the use of low spatial frequency information in recognition. However, findings by Oliva and Schyns (1997), Schyns and Oliva (1999) and Bonnar, Gosselin and Schyns (2002) demonstrated that a coarse-to-fine progression may be the wrong interpretation of how these low frequency data are used. Oliva and Schyns (1997) used a briefly presented (30msec) hybrid image that combined high and low spatial frequencies to prime different scenes. They found that both these scenes could be primed effectively with the same image indicating that the time course of low level scale processing imposed little or no constraint on the selection of which scale frequency was used for scene recognition. Their second experiment went on to show that selection was governed according to task dependent diagnostic information (coarse or fine) at a spatial scale. Manipulation of participant attention to one of these frequency channels, or diagnostic scales, could therefore alter the information that was perceived. Similar results are demonstrated by Bonnar et al. (2002) through the use of the ambiguous painting by Dali entitled 'Slave Market with the Disappearing Bust of Voltaire'. They used an adaptation experiment in which participants were submitted to 200 white noise fields filtered to contain either a high or low frequency response profile. Following this they were presented an image that could be perceived either as a bust of Voltaire or a pair of nuns. All members of the higher spatial frequencies group reported seeing Voltaire whilst the significant majority of the low frequencies group reported seeing nuns.

Schyns and Oliva (1999) proposed that pre-attentive manipulation of frequency channels may be driven by a mechanism capable of rapid object categorisations. A study by VanRullen and Thorpe (2001) provides empirical support for a feed-forward categorisation model that might perform this function. Their research required participants to categorise complex scenes in order to detect the presence or absence of a category: 'means of transport' or 'animal'. An image was flashed onto a screen for 20msec (no masking) and response was via a touch sensitive pad that was released if the target category was present. Results indicated a median reaction time slightly above 350msec in both categories, and a response time limit of 250msec that yielded a performance rate above chance. By deducting the minimum time required to generate a physical response (80-100msec : Kalaska & Crammond, 1992 as cited in VanRullen & Thorpe, 2001) they established that the necessary visual mechanisms for this kind of categorisation required no more than 150msec. Their timing is supported by event-related potential evidence (Fabre-Thorpe, Delorme, Marlot & Thorpe, 2001 as cited in VanRullen & Thorpe, 2001).

Bar's (2004) neuroscience model of contextual object recognition does take into account the influence of scene on the target perception. It provides a mechanism by which scene superiority could function, and has demonstrated consistency in ERP and fMRI readings for relevant brain areas. However, although there is support from other studies regarding the concepts of frequency channels and rapid categorisation along neural short-cuts, no mechanism is provided for the object or scene recognition. Dependent upon the model adopted to explain the primary recognition route the rapid pathway may simply be a lower threshold version of the same image. Likewise, there is little detail how scenes are processed or how positional relationships within a scene affect the context effect. Also, this model does not demonstrate whether a scene superiority effect exists, or whether scene context effects are simply response biases.

Summary of Object Recognition Models:

The five perspectives of the recognition process summarised in this review represent alternative theories and methodologies. They do not provide a comprehensive critique of the literature, but they do illustrate the principal hypotheses within the field, and demonstrate that there is both conflict and overlap between paradigms. There is growing support for models that integrate contrasting methods of visual processing (Bar, 2003; Hummel, 2001; but see Tarr, 2003) and bottom-up processing is seen alongside top-down connections and bi-directional pathways (Bar, 2003; Mozer, 1981). It is also clear that visual attention plays a major role within the processes of object recognition (Hummel, 2001; Mozer, 1981; Thoma et al., 2004). However, the role of visual attention in the generation of the scene context effect has not been previously investigated. This thesis aims to address whether visual attention is required in the generation the scene context effect as utilisation of this limited resource has potential implications for recognition processes and models. Chapter 2 therefore provides a review of the literature on selective attention.

Summary:

The aims of this thesis are: (1) to determine whether scene context effects with multiple objects exist; (2) to determine whether scene-based contextual influence is part of the perceptual/representational part of target recognition or due to response bias; and (3) to establish whether visual attention is required to generate a scene context effect. They are not to provide a new model of recognition, although their results may impact current recognition models.

The role of visual attention has not been previously examined in the generation of the context effect. As attention is presently target focused in the majority of object recognition models the allocation of a share of this resource to the context may require change. Thus, by utilising visual attention context processing could enter into a relationship with recognition without needing to directly affect the perceptual/representational processes because target and non-target processing would share a limited resource.

None of the recognition theories presented provide an account for scene context effects, though some do suggest routes by which contextual information might be utilised (Bar, 2003, 2004; Mozer, 1981). Scene-based contextual facilitation that was demonstrated to act directly on the object recognition process would highlight a clear relationship between context and recognition processes and raise two main questions for current theories: first, how are multiple, non-target stimuli processed? Second, how is the non-target information integrated?

Should a relationship between context and object/target recognition not be demonstrated then alternative questions are raised. The scene context effect has still been shown in a number of studies (e.g. Biederman, 1981, Davenport & Potter, 2004). How the non-target stimuli are processed remains a valid question relevant to the both the recognition literature and attention literature (Chapter 2).

Chapter 2: Visual Attention and Multiple Object Selection

In Chapter 1, the ambiguity over whether scene context effects resulted from perceptual/representational processes or from a response bias was discussed. Given the available data, both types of account can explain the empirical data. However, before running experiments to distinguish between perceptual/representational and response bias accounts, Chapter 1 has also given reasons why the role of visual attention in the generation of scene context effects needs to be explored. Despite the fact that visual attention must play a role in scene perception, there is a lack of research into how it influences the better reporting of target objects when shown in consistent contexts relative to inconsistent contexts. The aim of Chapter 2 is to examine the feasibility for pre-attentive and attentive contextual effects, and to show that we must consider visual attention when exploring how scene context effects emerge.

Scenes are typically composed of multiple objects, and these objects are set within a visual field that can be defined in terms of relational juxtapositions of objects. When viewing scenes, attention can be distributed broadly across an extent of the visual field, including the whole visual field, or can be focussed on specific objects. When specific objects or object relationships become associated with a scene, then eye movements can be guided or attention implicitly distributed to focus on critical objects and locations (Oliva, Torralba, Castalhano & Henderson, 2003). The fact that attention can manifest itself in so many different ways during scene perception suggests that we should consider how attention comes to select objects. Scene context effects may be moderated by the attentional load being borne by the system, but there is also evidence that at least some aspects of the sensitivity to context itself can lead to reconfiguring of orientation and focusing of attention (Chun & Jiang, 1998, 1999). These results suggest a dynamic interaction between contextual information and attention in the course of object recognition.

The Attentional Filter: Is a single filter placed early or late in processing?

Broadbent (1958) proposed that an attentional filter divides the visual process into an early, multi-stimulus, parallel processing phase, and a later phase in which filtered stimuli are processed individually. Exactly what processing occurs before and after filtering is disputed between theories (see below), though it is generally accepted that dedicating the limited capacity of selective attention to one stimulus at a time (serial processing) allows for higher levels of analysis of the selected stimuli than can be achieved prior to attentional selection. An implicit aspect of the filter-based principle is pre-selective processing, as stimuli must be partially processed for there to be something to select from. Neisser (1967) is generally credited with this concept, which he named pre-attentive processing. This principle has resulted in one of the longest running debates within cognitive psychology: at what stage of visual processing is the filter positioned, and what can and cannot be processed before attentional selection.

In viewing scenes, a target will typically be selected from amongst multiple context items. This chapter begins by looking at how some of the single filter theories divide visual processes between pre-attentive and selective attention capabilities. Much of the relevant evidence comes from experiments in visual search. The visual search paradigm has been used to try and establish when attentional selection occurs along an early-late continuum. The tasks used typically require a participant to locate the presence or absence of a pre-defined target amongst a set of distractors in which all stimuli are simple geometrical shapes or textual characters. If the stimuli are processed pre-attentively and in parallel, the number of distractors will not affect the response time. Conversely, if the search requires the serial application of selective attention, then larger display sets will increase the time taken. Treisman and Gelade (1980) made particular use of this method when formulating and demonstrating their Feature Integration Theory of Attention.

Early Selection Theories: Feature Integration Theory:

Feature Integration Theory (FIT) sought not only to define what processing tasks required selective attention, but also to outline what that attention was used for. Treisman and Gelade (1980) proposed that individuals analyse the visual field based upon functionally separable dimensions (e.g. colour or orientation) that are internally represented as 'feature maps'. Within the visual field, a stimulus will possess features along such dimensions (e.g. red or vertical) that can be represented at the appropriate spatial locations within the relevant maps. Their study showed that when participants had to decide whether a pre-defined target was present or absent in a visual search task, display set size made no difference to participant response times (a flat search slope) if the target differed from the distractors along a single dimension (see Figures 5 and 6). However, when a target was distinguished from distractors through a conjunction of features from more than one dimension (see Figure 7) an increase in set size led to an increase in participant response time (a steep search slope – see Figure 8). This pattern of results indicated that a search within a single feature map could be achieved using parallel, pre-attentive processing, but that combining features reliably from across several feature maps required the allocation of selective attention.

Treisman (1982) concluded that the joining of features could occur pre-attentively, but in the absence of selective attention, the features were combined more or less randomly, producing illusory conjunctions (e.g. 'seeing' a red cross when only a blue cross and red circle were presented). She claimed (Treisman, 1986) that whilst feature maps may preserve spatial relations of the visual field, such information is not directly available for identifying complex search targets. Consequently, any attempts to combine features across maps would not be guided by location and would be a matter of chance. In FIT focused attention is applied via a master map that specifies feature locations and is linked to the relevant feature map. This master map allows the simultaneous selection and binding of features present in the same location, and avoids the creation of illusory conjunctions. The bound features are entered into a temporary object representation (or file) that can be matched against those stored in memory. Therefore, within FIT, selective attention is required to ensure the reliable binding of feature dimensions using location (see Wolfe & Cave, 1999 for a review; also see Wolfe & Bennett, 1997).

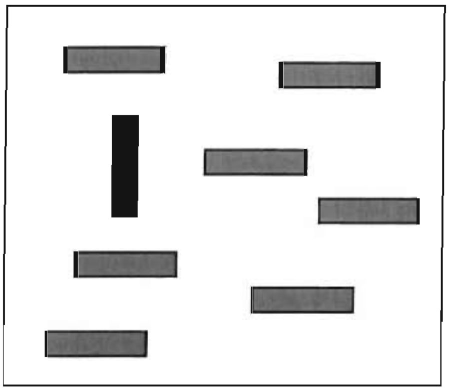


Figure 5: A feature, or 'pop-out' search in which the target is a single red vertical amongst green, horizontal distractors.

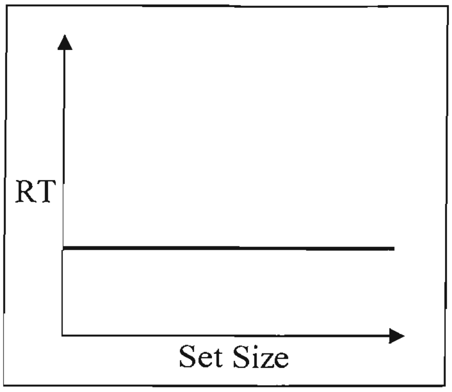


Figure 6: A flat search slope indicating response time does not increase with display set size.

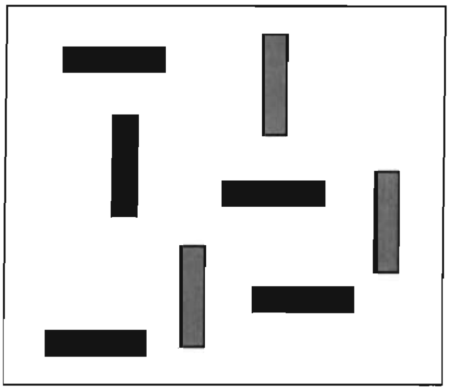


Figure 7: A conjunction search in which the target is a single red vertical amongst a mixture of green vertical and red horizontal distractors.

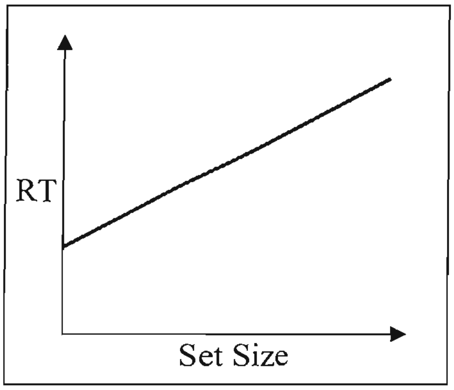


Figure 8: A steep search slope indicating a linear increase in response time as display set size is increased.

It is generally accepted that some basic visual properties can be identified without selective attention (see Wolfe & Bennett, 1997). Such basic properties consist of limited features (e.g. orientation, size, etc. - Treisman & Gelade, 1980), line terminations and intersections (Julesz, 1984), closure (Donnelly, Humphreys & Riddoch, 1991), and basic 3-dimensional shapes (Enns & Rensink, 1990). However, these simplified properties are considerably less complex than the stimuli encountered in object or scene recognition. Although FIT does allow for featural information to be extracted from multiple stimuli simultaneously the absence of feature relationships in pre-attentive processing will prevent higher levels of processing. As features (even basic form features) alone are unlikely to access semantic level information directly it is unclear how FIT can be developed to allow explanation of scene context effects.

The initial FIT was a bottom-up model in which perceptual information was passed upwards from simple to more complex levels of processing. It has now been adapted (Treisman, 1993) to integrate feedback from the master map to guide the search process. However, the FIT has always acknowledged top-down involvement both in the form of prior expectations and knowledge of target context. In extreme instances (e.g. capacity overload), when there is insufficient attention to perform the binding function, FIT may use this additional information about which features are typically conjoined together (Treisman & Gelade, 1980; Treisman, 1982). In such circumstances top-down input assists in selecting the most probable conjunctions from amongst disjunctive features across assorted feature maps (e.g. a forest context would suggest that a vertical is more likely to be brown than blue). These data are obtained via a back-up mechanism and do not assist the perceptual processes of the search, but instead act to maximise guessing performance when reliable conjunctions cannot be made.

Early Selection Theories: Guided Search:

In visual search models such as FIT, stimuli are matched against a pre-defined target, and with conjunction searches this can only reliably be done through selective processing. Within FIT the master map can specify the number of locations that meet the feature conjunction criteria, but it does not generate an order by which to search. This results in the FIT utilising a random search pattern to control the allocation of visual attention.

The Guided Search model (Wolfe, Cave & Franzel, 1989) provides an alternative solution and explains data for conjunction searches that were inconsistent with the FIT model. Wolfe et al. found that with conjunctions of colour and form (target of a red O amongst green Os and red Xs), participants were recording search slopes that were considerably shallower than those predicted by Treisman and Gelade (1980). They also showed that response times for triple conjunction searches, even when the target shared two features with each distractor (e.g. a large red O amongst large red Xs, large green Os and small red Os), were lower than for standard conjunction searches and produced very shallow search slopes. Guided search accounts for this pattern of results by positing that during parallel search, each feature map (e.g. size, colour, form) activates the spatial location that matches the search target (e.g. size = large, colour = red, form = O). The activation from each feature map is then combined within an 'activation map' that is also organised spatially. The most activated areas in the activation map indicate the most probable target locations, and the serial search can then be guided by this information, rather than selecting randomly (see Figure 9).

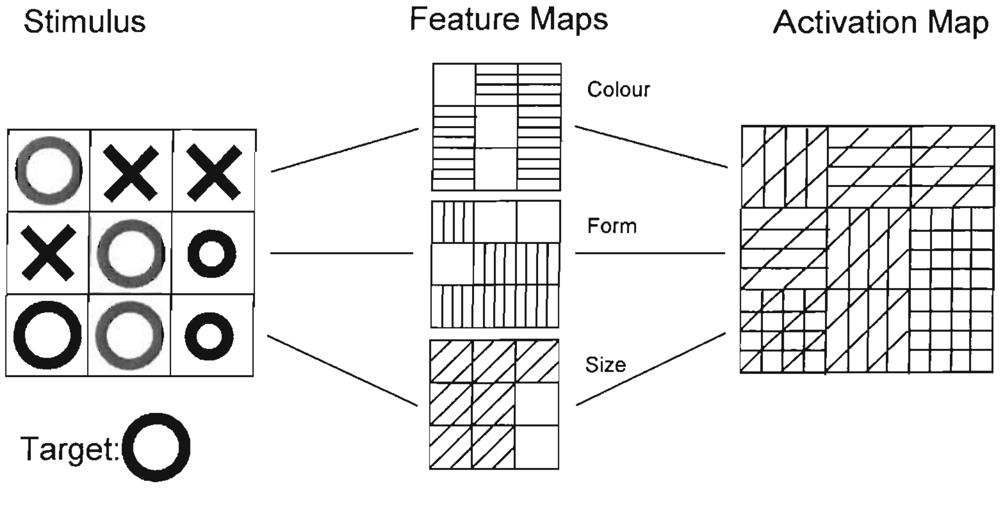


Figure 9: Guided search model for triple conjunction searches (adapted from Wolfe et al., 1989)

Guided Search (Wolfe et al., 1989) is effective in recognising (or detecting) the target when the features have been specified in advance (e.g. large, red circle). Unlike FIT, in which matching can occur after binding, this model utilises top-down information at a feature map level to match against the retinal data. It is this pre-attentive process during the early visual stages that provides the data for the activation map. However, the combined activations are not equivalent to binding – that still requires selective attention.

Guided Search 2.0 (Wolfe, 1994) expands the model by demonstrating the categorical nature of both bottom-up channels and top-down activation. The influence of categories can be best exemplified by considering orientation. Bottom-up processing is considered to be filtered by categorical attributes (e.g. steep or shallow, right or left tilt) rather than specific degrees of orientation (detail possessed at the retinal level). Likewise top-down information regarding target criteria would be based upon 'steepness' and 'rightness'. As a result of such categorisation, when the target to be found is specified as "steep", the activation of a diagonal line on an orientation feature map will be greater for lines with orientations close to the individual's stored exemplar of steep (regardless of slope direction). Support for such categorisation at a pre-attentive stage has been found using visual search techniques to demonstrate that a 0 degrees target (i.e. vertical) amongst distractors tilted 20 degrees left or right is difficult to detect because all could be considered 'steep' (Wolfe, Stewart, Friedman-Hill & O'Connell, 1992). The categorisation of features indicated that more than just basic properties could be processed pre-attentively. However, Wolfe and Bennett (1997) demonstrated that 'shape' could only be formed from the conjunction of other attributes with the allocation of selective attention. The inability to process shape would make a pre-attentive recognition process (target or context) difficult.

Guided Search 2.0 (Wolfe, 1994) also utilises a weighted sum of activations in the activation map, and empirical evidence has demonstrated that certain low-level perceptual information is able to capture attention in preference to others (e.g. abrupt onset – Jonides, 1981; Jonides & Yantis, 1988; Yantis & Jonides, 1984). Feature maps of onsets, contrasts or motion could bias selective attention towards highly activated spatial locations without a pre-specified search item. The use of the activation map to guide visual attention may also explain why attentional capture effects (e.g. sudden onset) can be disabled when the participant is engaged in secondary tasks involving focused attention (Yantis & Jonides, 1990; Warner, Juola & Koshino, 1990).

Wolfe (2003) states that the existence of higher-level functions at early processing levels might be explained by the bi-directional nature of the visual pathway. He proposes that pre-attentive processing may utilise a rapid abstraction of the visual scene to allocate attention, and that later selective processing may have access to fine detail not used during pre-attentive stages. This proposal is similar to the abstractive, pre-attentive feed-forward mechanism combined with 're-entrant pathways' suggested by Di Lollo, Enns and Rensink (2000).

Late-Selection Theories:

Late-selection models (e.g. Deutsch & Deutsch, 1963; Duncan & Humphreys, 1989) introduce the attentional filter later along the visual processing pathway than the theories described above. Duncan and Humphreys (1989) propose that the perceptual description of an object allows the construction of a hierarchical assembly of structural units. At the top of this hierarchy resides a structural unit that represents the entire scene, whilst beneath this are smaller, divisional units. An example would be a human body, subdivided into a head, torso, and limbs, with a hand further subdivided into a palm and fingers (Marr & Nishihara, 1978, as cited in Duncan & Humphreys, 1989). These structural units are described with both physical (e.g. location, motion, colour etc.) and semantic (e.g. categorisations based on meaning) properties, and the process of forming this description is parallel and resource-free.

These perceptual descriptions, whilst highly processed, remain outside of awareness until the allocation of selective attention. One might consider these as being activated only in long-term memory representations. In Duncan and Humphreys' (1989) model, it is the access to visual short-term memory (VSTM) that is strictly limited, but it is VSTM access that allows structural units to attain awareness and control immediate behaviour. The selection process is highly competitive, with each unit possessing a 'weight'. Weights are dependent upon the degree of match that a unit shares with a pre-defined template (i.e. target criteria) and VSTM resources are then allocated based upon their relative, rather than their absolute, weights.

The research reviewed earlier argues against a parallel and resource-free process capable of describing semantic properties to a high level. However, the majority of these studies have required participant awareness of the stimulus in order to respond. Duncan and Humphreys' (1989) argue that it is the awareness that inherently incurs attention. That a stored representation is activated outside of awareness does not mean it cannot influence items within awareness. Within a neural network there may be a pattern of spreading activation capable of affecting linked object representations (e.g. Kosslyn, 1994). This would provide another potential route for contextual information to be integrated into target processing.

Summary: Single filters, scene context and object recognition

The earlier the attentional filter is situated in the visual process (e.g. Broadbent, 1958) the more of recognition would occur after selection. This would provide credence for the majority of current recognition models and their single-object-at-a-time approach. However, an early filter would make the pre-attentive processing of multiple non-target stimuli unlikely, at least beyond the detection of very low-level perceptual qualities. Therefore early selection models would not readily allow the generation of a scene context effect, as such an effect requires the extraction of high level semantic information from non-target stimuli. Conversely late selection models (e.g. Deutsch & Deutsch, 1963; Duncan & Humphreys, 1989) argue that stimuli can be identified and semantically described in parallel, only needing to pass through an attentional filter to gain awareness. Thus recognition models with a late filter would be expected to process multiple context items simultaneously.

Beyond The Attentional Filter:

The adequacy of theories that place a single filter at a specific processing locus to explain all of the available data on stimulus selection has come under attack. More recent theories suggest either that a multiplicity of filters at different points in the processing hierarchy, or that the passage of information through a filters is influenced by more global processing factors such as perceptual load.

FeatureGate: A multiple filter model.

FIT and Guided Search are single filter attentional models, both of which favour the early-selection argument. The FeatureGate model of visual selection (Cave, 1999) shares some traits with Guided Search (Wolfe, Cave, & Franzel, 1989) but utilises multiple attentional filters. Each filter consists of an attentional gate that can be opened or closed to allow stimulus information to progress through the levels of processing (roughly corresponding to the organisation of the visual areas within the visual cortex). The lowest level of the visual field is split into spatial regions or "neighbourhoods" within which stimuli compete for selection using bottom-up and top-down systems. The bottom-up system is driven by stimulus properties and ensures that locations with feature singletons (e.g. a single red amongst many green) are favoured. The top-down system is activated when target features are known in advance, and it inhibits gateways for locations with non-target features. The most activated stimulus from each neighbourhood is represented at an intermediate level and new competitive neighbourhoods are formed at a higher level. Within this higher-level neighbourhood, a single stimulus selection is selected by feature matching (top level - see Figure 10).

The use of multiple filters and levels of attentional selection blurs the concept of a dichotomous pre-attentive and attentive boundary. Instead of moving from processing multiple stimuli to a single stimulus, FeatureGate steps from many, to a few, to fewer, to one. While processing is widely distributed across all items at the base level, a large proportion of these stimuli are excluded from the next level. This allows processing to be distributed amongst a smaller number of items, and thus in a sense allows more attention to be allocated per item. Therefore discarded stimuli, particularly those filtered out by the upper levels of the hierarchy, may have been processed to a relatively high level (reflected in the level of the visual cortex). The final output of the FeatureGate model represents focused attention upon a single item. At the levels between the first and the last, attention is distributed amongst a number of items, many of which will eventually be rejected. Thus this processing cannot be considered pre-attentive at these levels. It is selective, yet not focused at a single location.

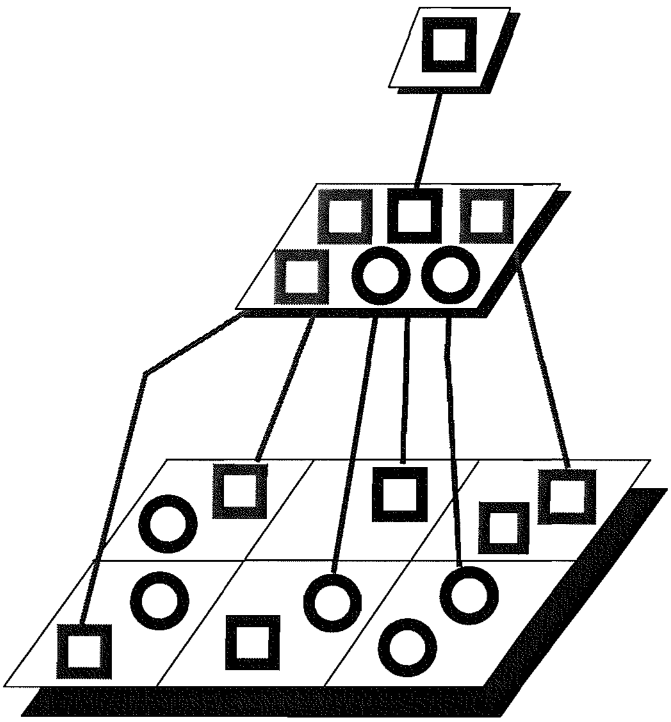


Figure 10: A simplified illustration of FeatureGate's hierarchical structure (adapted from Cave, 1999).

FeatureGate (Cave, 1999; Cave et al. 1999) is not the only theory to consider a non-dichotomous approach to selective attention. Joseph, Chun and Nakayama (1997) conducted research in which an array of Gabor patches was oriented at either +45 or -45 degrees from vertical and was displayed for 150msec. The participant had to detect whether one was oriented differently to the others. Because it was a single feature search task detection was expected in parallel without requiring any attentional resource. An attentionally demanding simultaneous secondary task was used to test this prediction in which participants were required to monitor a stream of letters for a white character. If the single feature task required no attentional resource, then there should be no interference from the secondary task on the primary task. However, it was found that performance in the primary task deteriorated the nearer in time the white character was to the display of the Gabor patches. Joseph et al. concluded that even single feature detection tasks are impaired if a secondary task is sufficiently demanding and that attention is critical even for the detection of 'pre-attentive' features.

Joseph et al.'s (1997) findings question whether any stimulus can be processed without requiring attention of some form. It may be that at the earliest levels of visual processing (e.g. the perception of a sudden onset), the amount of attention required is very small and thus many low-level tasks can be conducted in parallel without nearing attentional capacity. Yantis and Jonides (1990) have shown that even the automatic detection of sudden onsets can be disabled if attention is tightly focused at another location; a result that suggests that the detection of sudden onsets requires attention. Higher levels of processing would require more attention and therefore the number of stimuli attended to would be reduced. The attentional filter may therefore be a series of multiple gates, as in FeatureGate, or it may possess a degree of flexibility dependent upon the stimuli attended. It is this latter principle that drives the concept of perceptual load.

Load Theory:

Lavie (1995) proposed that the degree of perceptual load during visual tasks could explain discrepancies between research supporting early and late views of selective attention. She found that when perception was overloaded the interference from distractors was found to be minimal, as would be expected under early selection. However, if the perceptual load was low, the participant suffered distractor interference comparable to that found in late selection models. Because the most difficult task suffered the least interference, Lavie suggested that the allocation of attention can only be prioritised across different stimuli and tasks, and not set to a specific level for a specific task.

The perceptual selection mechanism outlined above is a passive mechanism as it functions by allowing the relevant stimuli to exhaust the available capacity. Lavie, Hirst, de Fockert and Viding (2004) have proposed a second, more active mechanism to exist alongside the first in order to reject irrelevant stimuli that are processed during conditions of low-load. This second mechanism requires higher cognitive functions (e.g. working memory) capable of correctly prioritising behaviour and minimising distractor interference.

In support of this dual-mechanism approach, Lavie et al. (2004) conducted a series of experiments in which either the perceptual or cognitive functions were manipulated whilst participants attempted to perform a visual task. Results for the perceptual function, replicated those of Lavie (1995), demonstrating that an increase in perceptual load decreased the interference of the distractors. However, for the cognitive function interference increased as cognitive load was increased as the active controlling mechanism was unable to filter distractors. The type of load would therefore seem to have a direct impact on how distractors or non-targets influence performance.

This is a hybrid theory of early and late selection models, with the experimental results dependent upon load. High or low perceptual/cognitive load scores might replicate either early or late selection studies, but this theory also has the capacity to simulate the moderate data in between the extremes.

What determines stimulus selection and how does this relate to scene perception?

The emergent picture is that the pre-attentive and attentive stages described by a single filter explanation do not fully explain the findings of all the previous studies. Instead the amount of pre-focused selective attention allocated to certain non-target objects may vary depending upon how many gates they pass through (FeatureGate – Cave, 1999) or upon the perceptual load (Lavie et al, 2004). The first of these points suggests that standing out in a 'neighbourhood' and matching to a predefined target (if one exists) will aid stimulus selection. Thus, one might assume that in the early stages of visual processing certain distinct objects or areas of interest are selected over others, and given the nature of FeatureGate's processes, these are likely to be far from one another. These areas/objects will receive more processing in the early stages than other non-target stimuli that are eliminated immediately. Only one of these locations, however, will pass through the final stage of selection. There can be wide variation across objects in a scene as to how far their processing proceeds.

If perceptual load plays a role in determining stimulus selection, then factors that reduced perceptual load (e.g. familiarity) would need to be considered. A scene with which a participant was familiar would be of a lower perceptual load than a similar unfamiliar scene. A good example of this is the work by Chun and Jiang (1998, 1999) on contextual cueing. In the initial experiment (Chun & Jiang, 1998) required participants to perform difficult search tasks (e.g. a rotated T amongst rotated L's) during which the spatial layout of some background arrays were repeated in a manner by which they became predictive of the target location. It was found that search times for targets within the predictive contexts were significantly lower than targets in the non-predictive displays. Contextual cueing was not limited to the spatial distribution of distractors. Further experiments (Chun & Jiang, 1999) have shown that contextual facilitation occurs when the predictive element is shape identity or dynamic change information. Chun and Jiang (1998) also suggested that the principles of contextual cueing might be applied towards predictive contexts based upon semantic information.

Chun and Jiang (1998) outline the formation of background arrays into 'context maps' during the repetition phase and describe these as instance-based memory representations acquired through implicit learning mechanisms. These context maps are thought to consist of only coarse visual information, with distractor detail not stored, and only task-relevant information encoded. Interaction between instance-based memory and attention mechanisms then allows these context maps to be used to prioritise attentional allocation based upon previous experience, a process that occurs rapidly and in parallel across the visual field. For this interaction to contribute to search, an abstraction of scene or array must be formed very rapidly so that early-level matching with memory can facilitate attentional selection. Whilst a definitive schema may not be formed in this instance some form of emergent, attention facilitating representation serving a similar purpose may be generated. However, the role of perceptual load in defining stimulus selection may also be influencing these results. By familiarising participants with the background arrays, Chun and Jiang reduced the perceptual load of the distractors in layouts that had been seen before. If the spare resource was used to process these stimuli faster rather than to a higher level then a present/absent decision would be made more quickly with the predictive contexts.

Although the explanation is not certain, Chun and Jiang's (1998, 1999) findings do demonstrate a link between visual attention and contextual influence.

Orienting and Focusing:

The research on visual search suggests that attention is more complex than a simple on/off state (Cave, 1999; Joseph et al., 1997; Lavie, 1995). Evidence supporting this view is also found in the attentional behaviours of orienting and focusing. The spotlight metaphor (Posner, Snyder & Davidson, 1980 – see Cave & Bichot, 1999 for a review) is an attempt to explain the orientation of attention or the way in which visual processing is adjusted in order to attend to a specific area of the visual field. If this orienting is achieved without making eye movements it is known as covert visual orienting (Posner, 1980). One can imagine a circular beam of attention that moves through the visual field in order to select a stimulus within that area. A potential source of guidance for this spotlight might be the activation map provided by the Guided Search models (Wolfe et al., 1989; Wolfe, 1994). Studies (e.g. Jonides, 1981) have indicated that this process of orientation can be controlled both exogenously (reflexively) or endogenously (voluntarily) but that these methods of control are not totally independent of one another (Müller & Rabbitt, 1989). The spotlight metaphor, and many theories that have grown out of it, are based on the assumption that this orienting is responsible for all selection in the visual system, and that it offers no benefit to pre-selective processing. However, selection via orientation does not necessarily mean that all attention is focused upon the target.

The spotlight metaphor was modified by Eriksen & St James (1986), who substituted the spotlight with a zoom lens. They found that the size of the attentional field itself could be altered as well as its position. As the diameter of the attentional field varied, then there was a corresponding variance in the attentional density across the field, which resulted in improved performance when the field was small and more tightly fitted to the outline of the target stimulus, and lower performance when it was spread over a wide area. The pre-attentive abstractions of scenes, suggested by Wolfe (2003), may then use widely focused selective attention prior to a specific orientation to a more narrowly defined region. Even after a narrow target region has been identified, there will be a brief period whilst focusing readjusts during which attention will be distributed to stimuli other than just the target.

Whether these metaphors will remain unchanged in the light of Lavie, Hirst, de Fockert and Viding's (2004) research on load theory remains to be seen. There are some similarities between the two mechanisms they outlined and the concepts of orientation and focusing, but there are also important differences.

Summary:

There are many parallels between the attention and recognition literature. The early selection attentional models (e.g. FIT) correspond most favourably with the part-based recognition models. In such a model (e.g. RBC – Biederman, 1987), identification of an object's parts, the orientations of those parts, and the relationships of the parts to each other would require the binding associated with focused attention. These models propose pre-attentive processing for only basic properties represented within single feature maps. This attentional architecture is compatible with the single-object-at-a-time approach to recognition, except that it would be a process of composition, not decomposition. If pre-attentive processing is not able to identify the more complex aspects of shape (Wolfe & Bennett, 1997), the extraction of semantic knowledge must be delayed until after such early stages. However, visual context studies that utilised brief displays (<100ms) of scenes (e.g. Biederman, 1981; Davenport & Potter, 2004) suggest that the time required for serial selective processing of multiple non-target stimuli make it an unlikely explanation for the context effect.

The concept of a feature map is one that is also used by the view-based recognition models. FIT utilises selective attention to bind information between feature maps, but an effective view-based model may use only feature maps encoding shape/form properties. Evidence that line drawings are recognised as easily as colour photographs (Biederman & Ju, 1985 cited Biederman, 1987), suggest that not all feature maps are essential to object identification.

The late selection models of attention are more compatible with the view-based models of recognition. It is not the processing that loads attentional capacity in Duncan and Humphreys' (1989) model but the transferral of the object's processed representation into awareness or working memory. The transferral or selection criteria may be object or location based but it would have little effect on the target processing itself. Under such a perspective, non-target stimuli are fully processed, though they do not necessarily have a route into awareness.

The multiple filter models (Cave, 1999; Joseph et al. 1997) and the hybrid Load Theory (Lavie et al. 2004) indicate that recent research is moving away from the early/late dichotomy. This reflects a similar shift towards hybrid models in recognition research. Cave, Kim, Bichot and Sobel (1999) suggest that FeatureGate might develop into a model of object recognition through allowing more complex feature combinations at higher levels in the hierarchy. Whilst there does not seem to be a system specific to FeatureGate to match visual targets to the vast array of stored representations held in memory, other models might be attached to complete the recognition process. FeatureGate does process non-targets up to some level, but the extracted information appears to play no role in the model once the irrelevant stimuli have been discarded. Visual context research indicates that non-targets can influence both search and recognition processes (e.g. Chun & Jiang, 1998; Duncan & Humphreys, 1989; Palmer, 1975). A mechanism that used non-target information collected as a by-product of selection might provide a potential explanation for visual context effects. The main difficulty with this possibility is whether the level of information extracted during these early visual stages, being restricted to basic properties, would be of sufficient complexity (see Wolfe & Bennett, 1997). By removing the single locus of division between pre-attentive and attentive processing, these non-dichotomous models do make it difficult to establish how much attention non-target stimuli receive during the selection process. Multiple non-targets may compete for attentional resource with some degree of success.

There is no empirical evidence to demonstrate whether visual attention needs to be allocated to multiple non-target items to generate a scene context effect. Neither the attentional nor the recognition literature can resolve this. Therefore, it is that issue the first two experimental studies in this thesis will seek to address.

Chapter 3: Does Visual Attention Mediate Contextual Influence on Recognition?

Chapter 1 illustrated that a scene context effect is often generated during object recognition tasks and questioned whether this is a scene superiority effect. In addition, semantic consistency within a scene has been demonstrated to influence recognition tasks (Biederman, 1981; Davenport & Potter, 2004; Palmer, 1975), but Chapter 2 has highlighted that visual attention and context may have a dynamic relationship that influences recognition. This chapter begins to address these issues by defining the terms used here to describe scene context effects and presenting the experimental paradigm that will form the basis of the empirical work in this thesis. It will also examine the question of whether visual attention mediates scene-based contextual influence on recognition.

Defining Scene Context and the Use of Arrays:

The majority of previous studies (Biederman, 1981; Boyce & Pollatsek; 1992; Davenport & Potter, 2004; Hollingworth & Henderson, 1998, 1999) have used a coherent and/or naturalistic scene as the target context. Naturalistic scenes are usually complex and contain multiple objects, in addition to a background, that potentially form their own inter-spatial relationships. Biederman (1981) claimed that scene schemas could be activated via information from the identification of specific objects or through scene-emergent factors arising via object (or partial object) relationships. A scene context effect may therefore be formed of object-semantic factors, drawn from the semantic information of related objects, and scene-configuration factors resulting from spatial relationships between objects and global processing. However, previous research has not isolated the objects within the scene or scene-configuration factors to determine if both contribute to the scene context effect. In addition, the complexity and variability of naturalistic scenes makes experimental control difficult. To resolve these issues, this thesis used a specific manifestation of the scene context. Displays of object arrays were presented in which groups of four context items (non-targets objects) were placed around a centrally positioned target object. These arrays removed the scene-configuration factors, allowing the influence of the object-semantic driven scene context effects to be studied in isolation. There has been some previous research with arrays of objects (Biederman, 1981; Henderson, Pollatsek, Rayner, 1987) but not with display times brief enough to prevent eye movement. Therefore, there is little previous indication as to whether object arrays can produce scene context effects without scene-configuration factors.

Previous experiments using naturalistic scenes (e.g. Biederman, 1981; Hollingworth & Henderson, 1998, 1999) also involved an inherent use of a visual search, making it difficult to determine whether context was aiding object recognition or was instead guiding search to locations likely to have the target object. The use of an object array with a central target removed the search element, thus allowing for a better analysis of the scene context effect on object recognition.

The use of naturalistic scenes in generating consistent and inconsistent contextual influences may seem to have a superior claim to ecological validity than the use of multiple object arrays. However, two points must be considered. First, any scene context effects exhibited in objects arrays are likely to be the result of principles learned through the observation of naturalistic scenes. Thus, whilst naturalistic scenes may trigger some additional processes, any result obtained with object arrays will be generalisable to naturalistic scenes. Second, naturalistic scenes are highly complex, and greater experimental control can be attained using limited object arrays. Results are unlikely to be identical between naturalistic scenes and arrays, as arrays will not include the scene-configuration context factors. However, experiments with object arrays will demonstrate whether certain scene context effects can be generated from the object-semantic factors alone, and allow the rather broad concept of scene context effects to be more tightly defined and dissected.

A configurational context defines an additional contextual type found primarily in the work of Chun and Jiang (1998, 1999). This form of context is based solely upon relative location within configuration (e.g. pattern of stimuli), and whilst it may be related to scene context via scene-configuration context this has yet to be demonstrated empirically.

The Manipulation of Visual Attention:

Although not explored in previous context research, visual attention has been manipulated in other recognition studies. The hybrid theory of object recognition (Hummel, 2001) proposes that unattended objects are represented holistically, but attended objects are represented analytically. Empirical support for the representation of unattended images was provided by their ability to act as primes, but being only able to prime identical and scaled images indicated their holistic nature (Stankiewicz, Hummel & Cooper, 1998; Thoma, Hummel & Davidoff, 2004). Attended images, on the other hand, were also able to prime left-right reflections (Stankiewicz et al. 1998) and split versions of the object image (Thoma et al. 2004). These findings suggest that visual attention may activate the featural relationships within object shapes (Thoma et al. 2004). The novel paradigm used in the experiments described below to examine whether visual attention mediates contextual influence on recognition adopts some of the concepts from these recognition studies.

In the experiments presented in this chapter, attention is either focused tightly upon the target object or spread evenly across the entire target/context array. The context items are then varied to either be all semantically consistent with the target object, or all semantically inconsistent, in order to measure the interaction between context and attention. It has already been stated that the potential confound of visual search is removed from this paradigm by centralising the target object. This is important in this study as Chun and Jiang (1998, 1999) found that the configurational context (i.e. the pattern of locations of items in the display) could bias the distribution of attention in visual search. To minimise the potential effect of perceptual load, the location of target and context items remains constant throughout the experiment, familiar objects are used for both targets and context items, and each object is only viewed once as a target (to limit repetition priming).

Based upon results by Palmer (1975) and Biederman (1981) it is predicted that any scene context effect would favour context consistent targets. If a scene context effect is greater when the non-target context objects are attended, then it suggests that visual attention is a mediator in the contextual influence on recognition. However, if the manipulation of visual attention to the non-target context objects does not affect their contextual influence, then it would appear that no attentional resource is required for the processing of context items or for the integration of contextual data.

Experiment 1:

In experiment 1 the primary task was to name the target object, with the two principal manipulations being the semantic relatedness of the non-target objects to the target (context) and the initial allocation of participant attention. Naming was selected as the method of participant response as this is the most common technique used in scene context research (e.g. Boyce & Pollatsek, 1992; Davenport & Potter, 2004; Palmer, 1975). It was considered particularly important whilst establishing a new paradigm for object/context display to maintain a reliable and tested method of response. A secondary colour dominance task was utilised to assist in the attentional manipulation.

Method:

Participants:

Forty-eight undergraduates from the University of Southampton participated in one 25 minute session for course credits. None knew the purpose of the experiment beforehand, and all reported normal or correct-to-normal vision. There were 9 males and 39 females between the ages of 19 and 34 years (mean: 20.57).

Apparatus and Stimuli:

The experiment used a Macintosh Power PC G4 400MHz computer with a 19" ProNitron monitor with a 13msec screen refresh. Participants sat approximately 60cm from the screen in a dimly lit room and responded verbally, via an Electret condenser tie-clip microphone, and via an Apple Pro Mouse, both of which were connected directly to the computer. The order of presentation was randomised individually for each participant.

For each trial, a unique pattern of blue and red coloured regions was generated by computer. Two black concentric circles of radius of 3cm (visual angle 2.68°) and 10.82cm (visual angle 10.26°) were initially displayed at the centre of the monitor to provide a focus cue. These circles were on a white background and remained unfilled for 1300msec. After this focus period the circles were filled with a combination of red and blue patches to match one of four criteria to construct the colour dominance stimuli: i.) both inner circle and total circle showing 75% red, 25% blue; ii.) both inner and total circle showing 25% red, 75% blue; iii.) inner circle showing 75% red, 25% blue but total circle showing 25% red, 75% blue; or iv.) inner circle showing 25% red, 75% blue but total circle showing 75% red, 25% blue. The background in every instance remained white. The filling procedure used a bespoke C program that selected an appropriate number of circle segments and shuffled them about a central column two segments wide, itself randomly positioned around the circle. This basic principle was used in filling both small and large circles. An example of a colour stimulus can be seen in Figure 11.

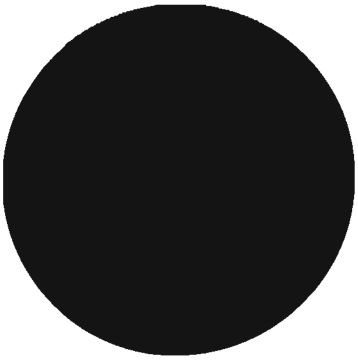


Figure 11: An example of a colour stimulus showing an inner circle of 75% red, 25% blue and a total circle of 25% red, 75% blue.

The naming task stimuli were created from digital colour photographs that had been edited to leave just an object image upon a white background. Whilst previous studies have typically used black and white line drawings, the additional detail offered by such colour images was considered to be more ecologically valid, and were expected to aid object identification in the short exposure times necessary to prevent eye movements. The majority of these images could be fitted within a circle of 3cm radius. One hundred and sixty objects (see Appendix A - CD) were arranged into 32 context groups of 5 semantically related items, and these context groups were divided into 16 pairs. Three objects from each of these context groups were selected to generate 96 target items. Each target could be displayed with four non-target items drawn from their own context group to generate a context-consistent trial, or with four items from the paired context group to generate a context-inconsistent trial. Context group pairings used to create context-inconsistent trials remained paired throughout the experiment (i.e. target from context group 1 with non-targets from context group 2; target from context group 2 with non-targets from context group 1 etc.). The target could also be displayed on some trials without any non-target items. Within the object arrays the target was always placed at the centre of a 23cm x 23cm display region, to be displayed within the area previously delineated by the inner concentric circle, and the non-targets were placed towards the four corners of the display region but within the area delineated by the outer concentric circle. There was no stimulus overlap and the background remained white (see Figure 12).

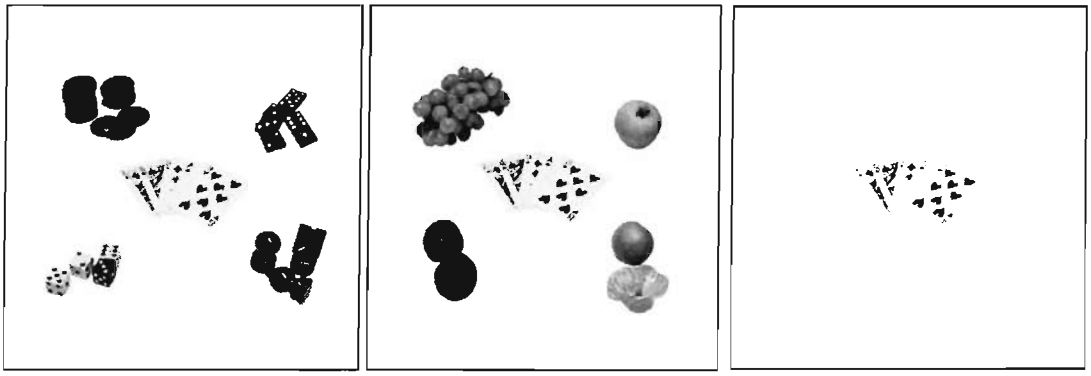


Figure 12: An example of the same target with a consistent context, inconsistent context and no context

Procedure:

In every trial the concentric circles cue provided participants with the visual area on which they were to distribute their attention. This cue was displayed for 1300msec, but the participant's attentional allocation was made endogenously as a result of instructions received prior to the start of the experiment. In half of the trials (blocked first or second half) participants were instructed to pay attention to the area within the inner circle (the narrow focus condition - NF), whilst for the other half they were instructed to pay attention to the whole circle (wide focus - WF). The order with which NF and WF conditions were presented was counterbalanced between participants.

The focus cue was replaced by the colour stimulus, which was displayed for 117msec. The colour dominance task required the participant to identify the dominant colour (red or blue) within the area they had been instructed to attend. However, response was delayed until the end of the trial. The purpose of this task was to reinforce the manipulation of attention by providing the participant with a reason to allocate their attention as instructed.

The offset of the colour stimulus was immediately followed by the onset of the naming stimulus (target and context objects), which was displayed for 78msec. The combined task presentation time was 195msec in order to prevent the programming and execution of an eye movement during stimulus display.

The participant had previously been informed that the target would always be the object at the centre of the screen so that searching would not be necessary. During the trial they were required to name the target object as quickly but as accurately as possible. With the onset of the naming stimulus, a timer was started to record the response time (RT) of the vocal naming response, which was detected via a clip microphone. Naming errors were recorded on a response sheet by the experimenter who was present throughout the session.

The naming stimulus was replaced by a blank (white) screen that remained for a maximum of 3000msec, or until a participant response was detected. The blank display was then replaced with two option boxes marked 'Red' and 'Blue' (left/right randomised within subjects) by which the participant responded to the colour dominance task using a mouse. There was no time-limit for this task, but selection of a box also allowed progression to the next trial. A sequence of displays in a single trial can be seen in Figure 13.

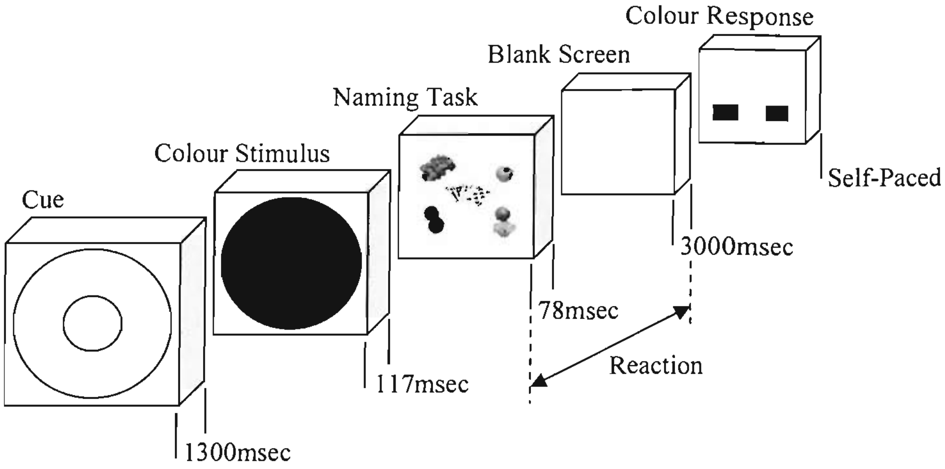


Figure 13: Sequence of displays in a single trial

Experimental trials were divided into six blocks of 16 trials. Three of these six blocks used the three targets from context groups 1 to 16, with each of the three targets in a single context group mapped to a different context type (i.e. target A was context consistent, target B was context inconsistent and target C was no context). Context groups 17 to 32 were used in the same way to generate stimulus sets for the remaining three blocks. Stimulus sets were presented in a random order within the blocks, and to limit repetition priming effects, no participant viewed the same object as a target more than once. Counter-balancing between the participants ensured that every target had an equal chance of being viewed under any condition.

Participants were given a written summary of information on both tasks prior to the experiment and were read a scripted set of instructions at the experiment outset. During this period participants were told to focus on the area contained either within the outer circle or just within the small, inner circle, and were provided an opportunity to practice the colour task, the naming task and the two tasks combined. They were also offered the opportunity to ask questions of the experimenter. The initial experimental trials were each preceded by an additional practice block of 16 trials in order to ensure that the attentional task was well practiced before data were collected, though participants were not informed that these trials would be treated any differently from the experimental trials. After the completion of three experimental blocks the experimenter instructed the participant to change their attentional focus. The participant was then provided an opportunity to practice both the new variation of the colour task, and the colour task and naming task together. A second practice block of 16 trials again preceded the remaining three blocks of experimental trials.

Design:

Three object images from each of the 32 context groups were used as targets to generate 96 trials of recorded data for each participant. Within participants, attention was manipulated so that half of these trials were focused on a small visual area concentrated around the target (NF) and half were focused on a larger area that could include non-target contextual objects (WF). The trials within these halves were further divided equally into those in which the non-target objects were semantically related to the target (consistent) or were semantically unrelated (inconsistent), or in which no context was present. The order in which the small or large visual areas were attended was counter-balanced across subjects, as was the context-type of a specific stimulus set (set). This resulted in basic mixed design with two within-participants conditions (2 x 3) and two between-participants conditions (2 x 3).

An additional independent within-subjects variable was the filler distribution used in the colour dominance task, which could take four possible arrangements of red or blue dominance in small and large visual areas (detailed above), though only the colour dominance in the area to be attended (either red or blue) was used as a factor in the analysis. The colour distribution was randomised across trials.

The three dependent variables measured were the error rate in the naming task, the RT in the naming task, and the error rate in the colour task.

Results:

Performance in the primary naming task was found to be good, with an average error rate across all participants under all conditions of 8.62%, and a minimum error rate for each condition of zero. Participants were less accurate at the secondary colour dominance task, with a mean error rate of 14.52% (min: 2%; max: 34%). If no response was given in the naming task, it was scored as an error. (A response was always required in the colour dominance task.) Reaction time data from the naming task was less reliable than error rates due to problems with background noise and microphone sensitivity. This resulted in some correct naming response trials not recording a time (approx. 10%). However, mean RTs were still achieved for every participant in every condition.

Naming Task Error Rates:

Mean error rates for the naming task by condition can be found in Table 1. Logarithmic transformations have been used previously on error data but are not considered standard practice within context research. In addition, a significant linear relationship was not found between the standard deviations and the means (r = 0.71, ns), and the data are not overtly positively skewed. Therefore, data transformations have not been used.

Table 1: Mean Error Rates (%) in the Naming Task by Attentional Focus and Context Type

Attentional Focus	Context Type		
	Consistent	Inconsistent	No Context
Narrow Focus	7.16 (1.0)	8.46 (1.0)	8.85 (1.0)
Wide Focus	6.77 (1.0)	9.90 (1.2)	10.41 (1.1)

These error rates were subjected to a repeated measures analysis of variance (ANOVA) with within-participant factors of attentional focus (narrow focus and wide focus) and context type (consistent, inconsistent and no context), and between-participant controlling factors of order and stimulus set. There was no significant main effect for attentional focus [$F(1, 42) < 1.0$], but evidence of significant contextual influence on recognition [$F(2, 84) = 5.544, p < 0.01$]. Post-hoc analysis[1] revealed that context-consistent non-target arrays produced significantly better performance than either context-inconsistent [$F(1, 42) = 7.356, p = 0.01$] or no

[1] Three ANOVAs, each of which compared two types of context (e.g. consistent x inconsistent)

context arrays [$F(1, 42) = 11.874, p < 0.01$]. There was no significant difference between the performance of context-inconsistent and no context conditions [$F(1, 42) = 2.778, p > 0.1$]. No significant interaction was found between attentional focus and context type [$F(2, 84) < 1.0$].

There were no main effects for the between-subjects variables [order; $F(1, 42) < 1.0$: set; $F(1, 42) < 1.0$]. A significant interaction was found between attentional focus and order [$F(1, 42) = 22.043, p < 0.01$] indicating better performance in the first block, whether it be narrow or wide focus. However, moving from a large visual area to a small had less impact on recognition performance than the reverse. A second significant interaction was found between context type and stimulus set [$F(2, 84) = 5.336, p < 0.01$], implying that the majority of object images in stimulus sets A and C produced the expected context-consistent effects, whilst some of stimulus set B did not. Considering that these objects were allocated to their sets arbitrarily, such an effect is likely to be chance, but may indicate that not all objects are equally affected by contextual influence. A final four-way interaction between all independent variables [$F(4, 84) = 6.541, p < 0.01$] is unlikely to influence the main findings but will be examined in the discussion (see Appendix B for all interaction results on Experiment 1).

Naming Task Response Times:

The mean RTs from the naming task can be found in Table 2.

Table 2: Mean RTs (msec) for the Naming Task by Attentional Focus and Context Type

Attentional Focus	Context Type		
	Consistent	Inconsistent	No Context
Narrow Focus	1208.38 (24.11)	1225.08 (24.05)	1213.83 (21.37)
Wide Focus	1204.12 (23.65)	1231.51 (23.35)	1218.14 (22.82)

An analysis was carried out using a mixed design ANOVA with factors identical to those for error rates. No main effect was found for attentional focus [$F(1, 42) < 1.0$] or context type [$F(2, 84) = 1.383, p > 0.1$], nor was there a significant interaction between these within-subjects variables [$F(2, 84) < 1.0$]. There were also no main effects found for the between-subjects counter-balancing variables of order [$F(1, 42) = 1.329, p > 0.1$] and stimulus set [$F(2, 42) < 1.0$]. Significant interactions were found between context type and stimulus set [$F(2, 84) = 3.453, p < 0.05$] and a four-way interaction between all independent variables [$F(4, 84) = 4.498, p < 0.01$].

Colour Dominance Task Error Rates:

Overall performance in the secondary colour dominance task was sufficient to suggest that participants were allocating their attention as directed with a mean error rate of 14.52% across all conditions. This was divided into means of 17.32% (s.e. 1.6) for narrow focus and 11.72% (s.e. 1.2) for wide focus.

These data were subjected to a mixed design ANOVA with attentional focus and order used as factors. A main effect of attentional focus was found [$F(1, 46) = 28.162$, p < 0.01] indicating that significantly more errors were made under the narrow focus condition. There was no main effect of order [$F(1, 46) < 1.0$]. Significant interaction effects were detected between focus and order [$F(1, 46) = 21.609$, p < 0.01], which suggested that moving from the large visual area to the small had a greater detrimental impact on performance than moving from the small area to the large. This result is the inverse to that found for recognition performance in error rates. There was no significant correlation between overall recognition and colour task error rates (r = 0.11, ns), arguing against a trade off between the two tasks.

Discussion:

The improved accuracy of recognition for targets displayed in an array of semantically related non-target objects compared to semantically unrelated non-target objects replicates previous findings of the context consistency effect (Biederman, 1981; Davenport & Potter, 2004; Palmer, 1975). This finding indicates that object-semantic factors alone are sufficient to generate a scene context effect. The absence of a significant difference in error rates between the context inconsistent and no context (baseline) conditions suggests that the presence of unrelated objects did not generate processing interference, and that a consistency-led facilitation occurred within the context effect.

This experiment did not find a main effect of attentional focus, nor a significant interaction between attention and context type. The implication of this is that the contextual items within the arrays can be processed automatically to generate a scene context effect. Holistic processing would explain how inconsistent non-target objects could be processed without generating interference. However, before this explanation is accepted two alternatives must be considered.

There is evidence dating back to Sperling (1960) that iconic memory can hold a considerable amount of visual data for a short time after display offset. Such persistence might allow for covert attention to be redistributed to the wider array even in the narrow focus condition if initial target processing proved unsuccessful. This might then allow contextual influence to occur normally. A second effect of visible persistence would be to provide additional processing time of the target, which would reduce total error rates that might arise through purely bottom-up processes. This is reflected by the fact that there was a definite ceiling effect shown in the recognition naming task (min. 0% errors for all conditions) and a low overall error rate for this task. Iconic assistance to the bottom-up processes would affect all context conditions evenly, but such facilitation would reduce the number of trials that could benefit from context. Although a scene context effect was still found, a smaller effect would make a significant attention by context type interaction less likely.

Both of these alternatives involve the storage of the target and context stimuli within the iconic memory. These possibilities can be tested for by introducing a visual mask into the experimental design immediately after the target and context offset.

The interactions between context and order, and between all four independent variables, suggest that the target objects may not be equally affected by context or by attention, or that they may not all be equally recognisable. Although all items have been chosen to be relatively common objects, some may be more distinct than others. Thorough counter-balancing will ensure that the main effects are no less robust;

however, achieving the main effect may be more difficult if other factors increase the variance. Experiment 2 will therefore begin with a ratings study of the stimuli used, both for familiarity and for semantic relatedness. Targets with low hit rates will also be replaced.

Experiment 2:

Experiment 2 modified the paradigms used in Experiment 1 and introduced controlling factors for object familiarity and semantic relatedness. These latter additions to the methodology were considered necessary following the interactions between context type and stimulus set in the first experiment. Separate ratings studies were conducted on the stimuli to assess the occurrence familiarity and viewpoint familiarity of each stimulus, and the semantic relatedness of each stimulus pairing within a context consistent group. Context inconsistent relatedness was not rated as interference effects of this condition had been found to be minimal, and the required increase in session time would have strained participant concentration. Correlations between these factors and both the context effect and task performance were also examined.

The colour dominance task and the naming task remained unchanged except for the introduction of a pattern mask after the target offset. Such backward masking (Sperling, 1960; Turvey, 1973 as cited in Pashler, 1999) limited the visual processing of the array, in order to increase difficulty and error rates in the naming task. Target stimuli that recorded a mean error rate higher than 20% across all conditions in Experiment 1 (12.5%) were also replaced. These included three targets from the same group (group 26) and therefore this group was replaced in its entirety (see Appendix C for new list).

Part 1: Ratings Studies - Occurrence and Viewpoint Familiarity Rating

Method

Participants:

Ten female post-graduate students from the University of Southampton volunteered to rate the stimuli during a 30 minute session. They were aged between 21 years and 43 years (mean: 27.0).

Apparatus and Stimuli:

The ratings program was written in SuperLab and run on a PC with a 13" monitor. Participant responses were made using a Cedrus 6 button button-box, to which a ratings scale was attached.

The stimuli consisted of the 160 digital photographs of objects used to construct stimulus sets in the naming task, presented on white backgrounds. In this experiment each stimulus was saved as an individual bitmap so it could be displayed at the centre of the screen.

Procedure and Design:

In the first half of the session, participants were required to rate the 160 stimuli for how frequently they saw exemplars of the object outside the laboratory (occurrence familiarity). The rating system was ordinal (1 = Very unfamiliar: 6 = Extremely familiar) and was designed to provide some objectivity to the term 'familiarity' (see Table 3). The familiarity ratings were based on a range of just one month, which is short but was considered valid because an attempt was made to initially select objects that were highly familiar. A long familiarity-span would consequently have resulted in a clustered distribution of ratings towards the high familiarity scores. Objects were presented individually, in a random order, and remained on the screen until they were rated.

Table 3: Ratings Scale Used in Occurrence Familiarity Ratings Task

Rating:	Familiarity:	Occurrence:
1	Very Unfamiliar	Less than once a month
2	Unfamiliar	Approximately once a month
3	Quite Unfamiliar	At least once a week
4	Familiar	Several times a week
5	Very Familiar	At least once a day
6	Extremely Familiar	Several times a day

In the second half of the session, participants were required to rate the same 160 stimuli for typicality of the viewpoint for the object (viewpoint familiarity). The same rating scale was used as for occurrence familiarity. As an example, they were told that a car, if seen from an isometric or side view, would be 'extremely familiar' or 'very familiar' as that is how they would most often see it; if viewed from directly in front or from behind it would be 'familiar' or 'quite familiar'; if viewed from above looking down would be 'unfamiliar'; and from the underneath it would probably be 'very unfamiliar'. Importantly, participants were asked to ignore the familiarity of the object itself in the rating. Objects were again presented individually and in a random order and remained on the screen until they were rated.

On-screen instructions were provided before both halves of the session, and the experimenter was present to answer any questions throughout. Three practice objects were displayed prior to the beginning of both the occurrence familiarity and the viewpoint familiarity ratings trials. Participants rated these objects verbally and with the button box to demonstrate to the experimenter that they understood the rating systems.

Results and Discussion:

The distributions were plotted for the median scores for each object for both types of familiarity and can be seen in Figure 14.

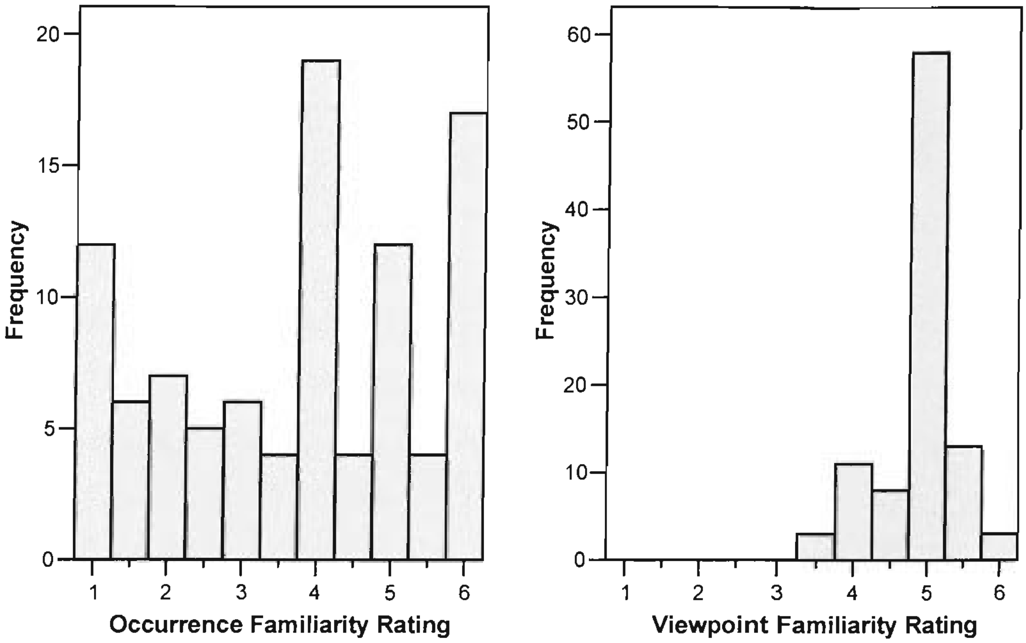


Figure 14: Distributions of median occurrence and viewpoint familiarity ratings by object

There was a wide range of occurrence familiarity found across the objects rated, with the mean positioned at the approximate centre of the ratings scale (mean: 3.72, s.d.: 1.70). Some objects attained low ratings due to the temporal nature of the scale. For example a hammer is distinct and generally familiar, yet its rating was low (median = 1) because it is not frequently used. However, frequency is more relevant to the definition of familiarity than distinctiveness and it provides an objective measure. An ANOVA was conducted on the mean ratings for the stimulus sets used as targets in this experiment (set A: 3.81, set B: 3.58, set C: 3.78). No significant difference between sets was found [$F(2, 93) < 1.0$].

The distribution for the viewpoint familiarity was more tightly clustered at the high end of the ratings scale (mean: 4.90, s.d.: 0.51). This high degree of viewpoint familiarity is not surprising, because most of the objects were intentionally photographed in canonical orientation. Also, for some objects, multiple viewpoints were equally valid (e.g. tennis ball). An ANOVA was conducted on the mean ratings for the stimulus sets A, B and C (set A: 4.97, set B: 4.88, set C: 4.84), and no significant difference was found [$F(2, 93) < 1.0$]. A Pearson correlation analysis between occurrence familiarity and viewpoint familiarity was not significant (r = 0.06).

Summary:

Occurrence familiarity provides a typically lower and more variable rating for stimulus objects than viewpoint familiarity. This finding was expected as occurrence familiarity is more dependent upon individual differences than viewpoint familiarity. The high ratings and relative variability in viewpoint familiarity can be attributed to the use of canonical views for the majority of stimuli. The lack of a significant difference in either familiarity ratings between stimulus sets confirms that no set has a familiarity advantage in the recognition task.

Part 1: Ratings Studies - Paired Semantic Relatedness Rating

Method

Participants:

Nine female students and one male student from the University of Southampton volunteered to rate the stimuli during a 30 minute session. They were aged between 19 years and 43 years (mean: 26.8).

Apparatus and Stimuli:

The ratings program was written in SuperLab and run on a PC with a 13" monitor. Participant responses were made using a Cedrus 6 button button-box, to which a ratings scale was attached.

The stimuli were constructed by forming the 160 digital photographs into the 32 semantically consistent context groups of five objects used for the naming task. Each object image within a context group was then paired with every other image in the group to create 10 bitmaps of 23cm x 13.4cm each displaying two objects horizontally next to one another. There was no occlusion, and across all pairings each object had an equal chance of being shown on either the left or right side of the display. This resulted in 320 stimulus pairs to be rated.

Procedure and Design:

The participants were required to rate the level of semantic relatedness between the two displayed objects. As in the familiarity ratings studies, a six point ordinal ratings system was used with '6' indicating 'very highly related' and '1' indicating 'no relationship'. Semantic relatedness was defined as the likelihood of seeing both objects if you had already seen one of them. For example, the likelihood of seeing a car having seen a petrol pump was very high, whereas the likelihood of seeing a car having seen a duck was not increased as there was no relationship. As both images were presented simultaneously, participants were asked to make the best average for the links between the pair. The 320 stimuli were presented in a random order and displayed until rated by the participant.

On-screen instructions were provided before the session, and the experimenter was present to answer any questions throughout. Three practice pairings were displayed prior to the beginning of either the ratings trials. Participants were required to rate these both verbally and with the button box to demonstrate to the experimenter that they understood the rating system.

Results and Discussion:

The distributions were plotted for the individual pairings and the context groups. Stimulus pairs were plotted using their median ratings. Context groups were plotted using the mean of the median ratings for ten stimulus pairs made from objects within that group. These distributions can be seen in Figure 15.

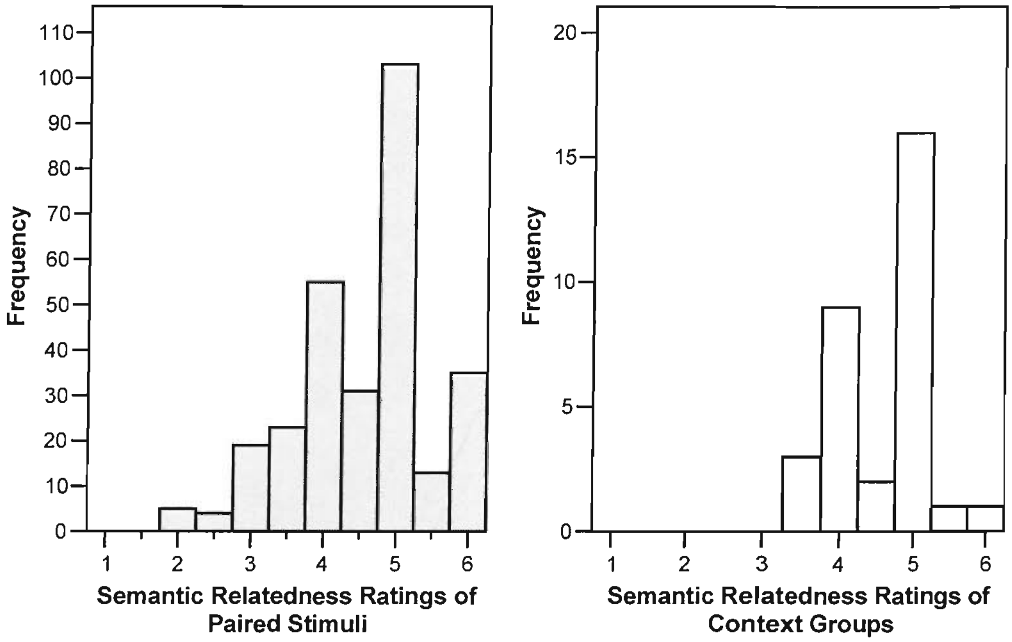


Figure 15: Distributions of median semantic relatedness ratings for paired objects and context groups

Both distributions for semantic relatedness demonstrated a positive skew that was consistent with having drawn the paired stimuli from context-consistent context groups. All the context groups achieved a mean score above the central point on the ratings scale (min: 3.5: mean: 4.59, s.d.: 0.62). The mean ratings for the individual pairs of objects were spread more widely (min: 2.0: mean: 4.56, s.d.: 0.92), however the majority were also located above the central point.

Summary:

The high semantic relatedness rating in all context groups supports their selection as contextually consistent stimuli. Although there is more variability within stimulus pairs, suggesting not all objects are equally related, there remains a high degree of semantic consistency between items.

Part 2: Naming Study

Part 2 of Experiment 2 replicated Experiment 1 with the introduction of a visual mask after the offset of the target/context stimuli in order to limit the use of iconic memory and increase task difficulty. The reviewed stimuli based on the ratings studies in Part 1 were used.

Method:

Participants:

Forty-eight undergraduates from the University of Southampton participated in one 30 minute session for course credits. None knew the purpose of the experiment beforehand and none had participated in Experiment 1 or the ratings tasks. All participants reported normal or correct-to-normal vision. There were 12 males and 36 females between the ages of 18 and 25 years (mean: 19.63).

Apparatus and Stimuli:

The equipment used was identical to that in Experiment 1, as were the focus cue and colour dominance task stimuli. As mentioned above, the identities of some stimuli (12.5%) were changed but the method of construction otherwise remained the same as Experiment 1. Analysis in the first experiment suggested that some of the stimulus sets, particularly set B for context groups 17 to 32, may have produced different results from the others. In an attempt to eliminate these differences, half of the stimulus sets from context groups 17 to 32 were integrated into the first three experimental trial blocks, and half of the stimulus sets from context groups 1 to 16 were integrated in the latter three experimental blocks, with the goal of achieving a better redistribution of targets across the overall experiment.

A 23cm x 23cm pattern mask was constructed that consisted of 1.53cm x 1.53cm squares arranged in a 15 x 15 grid. These squares were sections from a selection of colour photographs that had not been used as stimuli. Each square by itself was considered difficult to identify. Many of these sections had been rotated to further decrease the likelihood of recognition interference from the mask (see Figure 16).

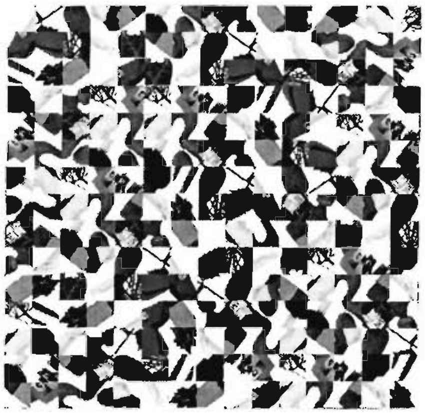


Figure 16: The visual pattern mask

Procedure and Design:

The procedure and design was the same as that used in Experiment 1 with the exception of the pattern mask between the naming stimulus offset and the onset of the colour dominance task response screen. The pattern mask was displayed for 3000ms or until an auditory response was received from the participant via the clip microphone.

As before, there were six experimental blocks of 16 trials, with the trials of each context type (consistent, inconsistent and no-context) distributed evenly between the blocks but presented randomly within them. For half the participants, attentional focus was endogenously manipulated to be wide for the first three experimental blocks, and then narrow for the latter three blocks. For the other half, the order was reversed. Counter-balancing variables of order and stimulus set were again controlled between-subjects. Error rates and response times were recorded for the naming task, and error rates were recorded for the colour dominance task. In addition, any erroneous names were recorded by the experimenter so that the type of error could be explored.

Participants received similar instruction and were allowed the same amount of practice as in Experiment 1.

Results:

Performance in the primary task was found to be much reduced compared to Experiment 1, with an average error rate for all participants across all conditions of 44.57%. Participants were also less accurate at the secondary colour dominance task with a mean error rate of 20.96%. Therefore the addition of the pattern mask decreased performance for both tasks despite its purpose being to remove visible persistence solely in the primary task. This suggests a potential trade-off in performance between the two tasks, but a Pearson analysis found no significant correlation (r = -0.24, ns).

Reaction time data from the naming task was again less reliable than error rates due to background noise and problems with the microphone. As a result, no response time was recorded for some correct trials (approx. 28.53%), and for 12 participants this resulted in zero RTs for at least one condition. These participants were discarded from the RT data analysis (see page 85).

Naming Task Error Rates:

Mean error rates from the naming task can be found in Table 4.

Table 4: Mean Error Rates (%) in the Naming Task by Attentional Focus and Context Type

Attentional Focus	Context Type			
	Consistent	Inconsistent	No Context	TOTAL
Narrow Focus	44.79 (3.3)	47.26 (2.8)	35.29 (2.6)	42.45 (2.7)
Wide Focus	39.84 (3.0)	51.69 (2.9)	48.44 (3.4)	46.66 (2.9)
TOTAL	42.32 (3.0)	49.48 (2.6)	41.86 (2.8)	

As in Experiment 1, error rates were subject to a repeated measures ANOVA with factors of attentional focus (narrow focus and wide focus) and context type (consistent, inconsistent and no context), and between-subjects controlling factors of order and stimulus set. A main effect of attentional focus was found in this experiment [$F(1, 42) = 11.643$, $p < 0.01$] indicating better general performance in the narrow focus condition. A significant main effect of context type was present [$F(2, 84) = 23.218, p < 0.01$], replicating the scene context effect found in Experiment 1. However, a significant interaction between attentional focus and context type was also found [$F(2, 84) = 16.141, p < 0.01$], indicating that attention was influencing the contextual effect.

Post hoc analysis[2] between consistent and inconsistent context conditions found no main effect of attentional focus [$F(1, 42) < 1.0$], but did find a main effect of context type [$F(1, 42) = 31.094, p < 0.01$] and a significant interaction between focus and context [$F(1, 42) = 8.100, p < 0.01$]. An increased influence on performance due to the context consistency effect in the wide focus condition can be seen in Figure 17.

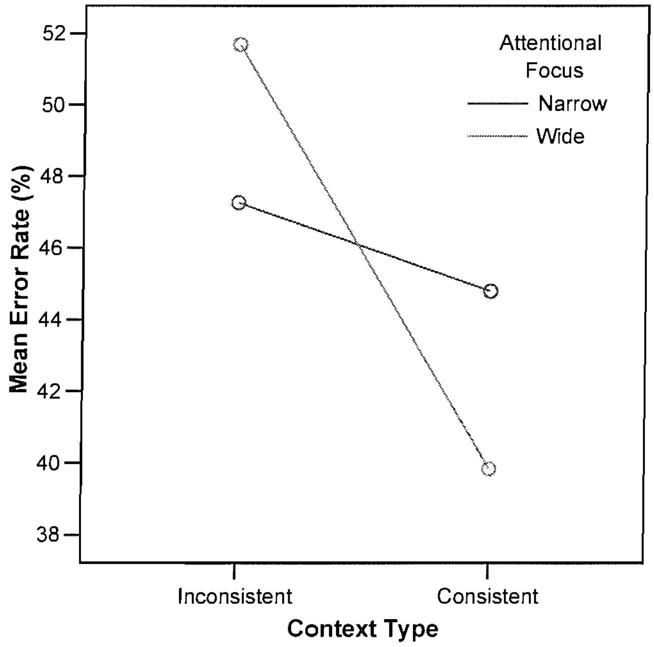


Figure 17: Post-hoc analysis of consistent and inconsistent conditions vs. attentional focus

[2] An ANOVA which removed the no context condition (2 x 2 x 2 x 3)

Significant effects of focus were found in post hoc analyses[3] between the no context and consistent context conditions [$F(1, 42) = 8.458, p < 0.01$], and between the no context and inconsistent context conditions [$F(1, 42) = 27.000, p < 0.01$], highlighting that the initial effect of focus was due solely to the no context condition. A main effect of context between inconsistent and no context was also found [$F(1, 42) = 35.372, p < 0.01$]. There were also significant interactions between focus and context for no context and consistent conditions [$F(1, 42) = 35.922, p < 0.01$], and no context and inconsistent conditions [$F(1, 42) = 7.259, p < 0.01$].

There were no main effects for the between-subjects variables of order [$F(1, 42) < 1.0$] or stimulus set [$F(2, 42) < 1.0$], and no significant interaction between them [$F(2, 42) < 1.0$]. As in Experiment 1, there was a significant interaction between attentional focus and order [$F(1, 42) = 39.648, p < 0.01$]. However, in this study, improvement was greater when moving from a wide focus to a narrow focus (WF = 51.52: NF = 39.84) than from narrow focus to wide focus (NF = 45.05: WF = 41.49). This reversed the effect found previously, suggesting that the visual mask may have removed the ceiling effect and thus allowed the development of a practice effect. A final four-way interaction between all independent variables [$F(4, 84) = 6.965, p < 0.01$] is also present (see Table 5).

Table 5: Mean Naming Error Rates (%) and Standard Errors in Experiment 2 Across All Variables.

		FOCUS					
		Narrow			Wide		
ORDER	SET	Consistent	Inconsistent	None	Consistent	Inconsistent	None
1	1	9.5 (1.3)	7.6 (1.1)	7.3 (1.0)	6.6 (1.2)	9.1 (1.2)	7.4 (1.3)
	2	8.3 (1.3)	8.8 (1.1)	4.6 (1.0)	5.6 (1.2)	5.4 (1.2)	5.9 (1.3)
	3	5.8 (1.3)	8.0 (1.1)	5.1 (1.0)	4.8 (1.2)	7.4 (1.2)	7.5 (1.3)
2	1	5.5 (1.3)	7.4 (1.1)	5.9 (1.0)	7.6 (1.2)	7.5 (1.2)	9.3 (1.3)
	2	6.6 (1.3)	6.8 (1.1)	6.3 (1.0)	6.1 (1.2)	10.6 (1.2)	9.6 (1.3)
	3	7.4 (1.3)	6.9 (1.1)	4.8 (1.0)	7.4 (1.2)	9.6 (1.2)	6.9 (1.3)

[3] Two ANOVAs comparing no context x inconsistent and no context x consistent (2 x 2 x 2 x 3)

Using the experimenter response sheets, naming errors were classified into six categories. Two experimenters classified each erroneous response according to whether it was: i.) perceptually similar to the target; ii.) related to the context; iii.) perceptually similar to the target and related to the context; iv.) in the contextual array; v.) perceptually similar to an object in the contextual array; vi.) no response or unclassified. Inter-rater reliability for these classifications met normally accepted levels (kappa = 0.714). These error types provided a framework into which the number of errors could be placed, and an error breakdown by context and focus could be established (see Table 6).

Table 6: Error Breakdown by Type (%) Across Attentional Focus and Context Type

Error Type	Narrow Focus		Wide Focus	
	Consistent	Inconsistent	Consistent	Inconsistent
i.) Perceptual – target	18.0 (3.1)	16.3 (3.0)	9.9 (1.9)	13.5 (2.6)
ii.) Semantic – context	1.7 (1.1)	1.3 (0.7)	1.5 (0.6)	0.4 (0.3)
iii.) Combined i. & ii.	3.0 (1.6)	0.4 (0.3)	2.4 (1.1)	0.2 (0.2)
iv.) Contextual object	10.1 (3.1)	4.4 (1.5)	8.0 (1.0)	5.4 (1.5)
v.) Perceptual – context	1.6 (0.7)	1.8 (0.7)	3.5 (1.0)	2.7 (0.9)
vi.) No response	65.7 (4.6)	75.8 (3.6)	74.8 (3.6)	77.3 (3.3)

These data on the different error types were subjected to a repeated measures ANOVA with between participants factors of attentional focus, context type and error type. Trials with correct responses were not included in this analysis. There was a main effect for error type [F(5, 235) = 300.020, p< 0.01] but not for focus [F(1, 47) = 1.005, $p > 0.1$] or context [F(1, 47) < 1.0]. A significant interaction was identified between error type and context [F(5, 235) = 3.564, p< 0.05], but not focus and context [F(1, 47) < 1.0], focus and error [F(5, 235) = 2.418, $p > 0.05$ $^{\text{G.Geisser}}$], or between all three variables [F(5, 235) = 1.441, $p > 0.1$]. Individual ANOVAs were used for each error type to perform post hoc analyses on the error by context interaction. These found significantly more of the combined semantic-perceptual errors [F(1, 47) = 4.738, $p < 0.05$] and naming-of-a-context-item errors [F(1, 47) =

5.105, $p < 0.05$] in the consistent context condition, but a greater tendency to not respond [$F(1, 47) = 5.376$, $p < 0.05$] in the inconsistent condition (see Table 7). However, it should be noted that the number of combined semantic-perceptual errors was very low.

Table 7: Means and Standard Deviations of Errors (%) for Error Types Mediated by Context

Error Type	Consistent	Inconsistent
iii.) Combined i. & ii.	2.67 (1.1)	0.28 (0.2)
iv.) Contextual object	9.05 (1.9)	4.88 (1.0)
vi.) No response	70.24 (3.3)	76.56 (3.0)

A difference value for each target object was calculated by deducting the number of hits in the inconsistent context condition from the number of hits in the consistent context condition. This measure of scene context effect demonstrated that the benefit in naming performance due to contextual influence was widely distributed between objects (see Figure 18), with some objects producing negative effects.

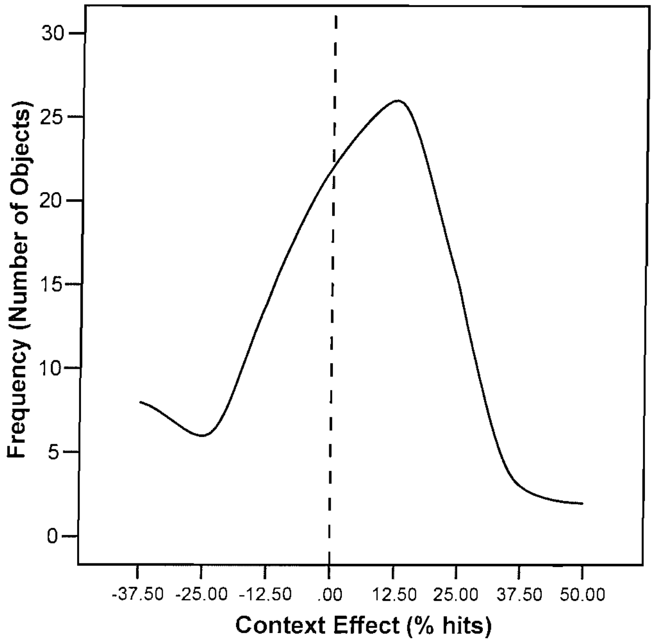


Figure 18: Distribution of target objects by scene context effect

Pearson correlations were conducted with these scene context effect measures against object occurrence and viewpoint familiarity, array semantic relatedness (consistent context) and correct naming (hits) of the object (see Table 8).

Table 8: Pearson Correlations of Scene Context Effect and Potential Mediators

	Hits	Occ. Familiarity	View Familiarity	Sem. Relate
		(target/context)	(target/context)	
Context Effect	-0.103	-0.069 / 0.076	-0.071 / -0.012	-0.064
Hits	-	-0.069 / -0.158	0.119 / 0.081	-0.217
Occ. Familiarity	-	-	0.060 / 0.212	0.012
View Familiarity	-	-	-	0.006

n.b. underlined figures significant to 0.05 (two-tailed)

Neither form of familiarity nor the degree of semantic relatedness within the array (assuming consistency) had a significant influence on the context effect in this experiment. The only significant correlations were between context familiarities (mean item familiarities of array without target), and a negative relationship between the semantic relatedness of the array and the general performance in correctly naming an object.

Naming Task Response Times

The results from Experiment 1 indicated that response times were unaffected by contextual influence, but also that there was no speed accuracy trade-off to explain the improved naming performance. Therefore no effect was expected from the response time data in this study, but it was included both for purposes of replication and completeness. Mean RTs from the naming task can be found in Table 9.

A repeated measures ANOVA with factors of attentional focus and context type, and between-subjects factors of order and stimulus set found no main effects of focus [$F(1, 30) < 1.0$] or context [$F(2, 60) < 1.0$], and no significant interaction between focus and context [$F(2, 60) < 1.0$]. There were also no main effects of order [$F(1, 30) < 1.0$] or stimulus set [$F(2, 30) = 1.097, p > 0.1$]. The only significant interaction was between attentional focus and order [$F(1, 30) = 16.296, p < 0.01$] indicating that whichever focus condition was done first took longer. This suggests a practice effect inherent within the task that influenced speed of response, similar to the effect that was found in the accuracy data. These data replicate the absence of a scene-based contextual influence in response time found in Experiment 1.

Table 9: Mean RTs (msec) in the Naming Task by Attentional Focus and Context Type

Attentional Focus	Context Type		
	Consistent	Inconsistent	No Context
Narrow Focus	1123.96 (43.1)	1106.98 (37.4)	1116.48 (32.1)
Wide Focus	1155.67 (45.3)	1130.43 (44.0)	1157.47 (59.3)

Colour Dominance Task Error Rates

Overall performance in the secondary colour dominance task was sufficient to confirm that participants were allocating their attention as directed with a mean error rate of 20.97% across all conditions (see Table 10).

Table 10: Mean Error Rates (%) in the Colour Dominance Task

Attentional Focus	
Narrow Focus	22.44 (1.6)
Wide Focus	19.49 (2.0)

An analysis was carried out based on a repeated measures ANOVA with attentional focus as a within-subjects factor and order included as a between-subjects controlling factor. No main effect of focus [$F(1, 46) = 3.277, p > 0.05$] or order [$F(1, 42) = 1.060, p > 0.1$], and no significant interaction [$F(1, 46) = 1.134, p > 0.1$] was found.

Discussion:

As predicted, the introduction of the visual mask caused an increase in overall naming error rates. It also resulted in a significant context by attentional focus interaction which demonstrates that visual attention plays a mediating role in generating a scene context effect. This finding suggests that iconic memory was being used to store the entire stimulus array in Experiment 1. By maintaining the scene context after target offset participants may have eliminated any benefit from allocating attention to context items before they appeared.

The context-consistent facilitation in error rates is consistent with previous research using naturalistic scene stimuli (Biederman, 1981; Davenport & Potter, 2004; Palmer, 1975). However, these new findings now show that attention must also be allocated to a consistent context to generate a significant scene context effect. This effect of attention is not just due to the transfer of attentional resource away from target to non-targets, as this would result in a main effect of focus and similar error rates between consistent and inconsistent conditions. In addition, as in Experiment 1, no significant difference in performance was found between the inconsistent and no context conditions in the wide focus condition. Thus, the additional resource must be utilised in a beneficial manner by the semantically related non-targets in order to achieve facilitation.

Neither the degree of familiarity nor the degree of semantic relatedness of the array was shown to have a significant correlation with the strength of the contextual effect. These correlations would have suggested an influence from perceptual load but the absence of any correlation may be due to the high levels of all three factors in the stimuli used. Decreased naming accuracy was correlated with an increased semantic relatedness within the array. It has been shown previously that semantic relatedness is required for the generation of a scene context effect (e.g. Biederman, 1981), and this effect is considered to improve recognition performance. However, similar negative effects have been found in naming studies (e.g. Riddoch & Humphreys, 1987). Therefore this result may indicate contextual influence on both perceptual/representational and psycholinguistic processes during the task. The strength of the scene context effect has been demonstrated here to be distributed along a positively skewed bell-shaped curve when plotted against object identity rather than evenly across objects. Further experiments using the same targets and context groups will be conducted to establish whether the scene context effect is mediated by object identity.

The majority of naming errors were perceptual, due to the bottom up processing of the target rather than context. However, scene context did significantly influence three error types. Participants were more likely to not respond in the context inconsistent condition. A spreading activation model of scene context (e.g. Kosslyn, 1992) would suggest that consistency would increase activation, thus raising the likelihood of a decision threshold being surpassed. When the context was consistent with the target, it also induced participants were to make perceptual-semantic errors or name a context item. The perceptual-semantic errors alone might be considered too few to be relevant, however they are the result of perceptual and contextual information. The same is true when a participant names a context item, although it is the wrong perceptual data, it is supported by contextual information. That context type interacts in this manner with errors requiring both sources of information (target perceptual and contextual), and the inconsistent context condition favours non-response, suggests a high decision threshold must be surpassed within naming paradigms.

Contextual effects were not found in the response time data. This may be due to the missing data in some trials, and methods of collection. However, much scene context-based research uses error rates alone (e.g. Cheng & Simons, 2001; Davenport & Potter, 2004; Hollingworth & Henderson, 1998, 1999). Such an absence of RT evidence suggests that context aids only accuracy and not speed in naming tasks.

These results strongly support the hypothesis that scene-based contextual facilitation in recognition accuracy requires context consistency and is mediated by the allocation of visual attention to that context.

General Discussion:

The results of these experiments further support the argument that a consistent context facilitates object recognition (Bar, 2004; Biederman, 1981; Davenport &

Potter, 2004; Palmer, 1975). They also demonstrate that a naturalistic scene is not required for the generation of a significant scene context effect, and that object-semantic factors can produce the effect in isolation.

In Experiment 2, contextual facilitation in accuracy was only found when visual attention was widely spread to include the entire contextual array and not when narrowly focused upon just the target object. This interaction demonstrates the role of visual attention in the mediation of scene-based contextual effects, but does not specify precisely where that role lies. The absence of a significant contextual effect without the allocation of attention suggests that the scene context cannot be processed whilst being unattended, and if we accept Hummel's (2001) description of holistic processing, these results imply that context cannot aid object recognition if it is only represented holistically. However, the brief display time would also suggest that the analytical processing of the context objects in serial was unlikely. Attention may also be used in a binding role to ensure that information from multiple objects remains bound to the appropriate representation and does not interfere with other stimuli. Alternatively the attentional resource could be utilised in integrating the extracted scene context information with the target recognition processes.

There are similarities between these results and those of Goldsmith and Yeari (2003), who found that object-based attention effects are more likely with widely spread attention than with narrowly focused attention. They claim that a central cue excludes distractor objects and weakens their representations, and the same may occur in contextual situations. It may be that object-based attention has a specific role in scene-based contextual influence.

There is evidence of a trend towards a context effect in the narrow focus condition that questions whether contextual facilitation can only occur when visual attention is allocated to the context items. This non-significant trend is weak, and if it does reflect a reliable effect, it may be explained by small amounts of visual attention allocated to the context during the narrow focus condition in some trials. One cause for such misallocation may be the sudden onset capture of attention by the array non-targets. However, it cannot be confirmed that the non-targets in the narrow focus condition did not generate a weak contextual trend without attention.

Experiment 2 demonstrated a decrease in performance in the no context condition when attention was directed to the larger visual area. This could not be due to interference caused by the sudden onset of non-targets (Yantis & Jonides, 1990) as no context items were present. For the same reason, there was no stimulus noise to disrupt target processing through the activation of irrelevant features and internal representations (Duncan & Humphreys, 1989). The lowered performance is consistent with the reallocation of attention away from the target as the diameter of the attentional zoom lens was expanded, so reducing attentional density (Eriksen & St James, 1986). Such a decrease was not shown in Experiment 1 as the no context/narrow focus condition did not demonstrate an initial advantage over the other conditions. It is suggested that narrowly focused attention aids the extraction of perceptual information, however the additional processing time gained by utilising the iconic memory in Experiment 1 allowed virtually all the target perceptual information to be extracted during every condition. An advantage could only be gained if an alternative source of information (e.g. consistent context) could be accessed.

Potentially, the finding that visual attention is a mediator in scene-based contextual facilitation has important implications for object recognition. The majority of object recognition models direct visual attention only to the target object, and this might have been justified had scene context effects been generated without attention. However, significant contextual facilitation occurs through the spreading of attention across the visual context area and away from the target. If contextual information is shown to directly influence perceptual/representational processes of target recognition, then models of recognition will need to integrate scene context effects and the joint relationships with visual attention.

In conclusion, these experiments demonstrate that accuracy in an object naming task is significantly greater if the target was in a context consistent array and if visual attention was spread to include the non-target objects. However, the spreading of attention reduced base-level performance. The manipulation of attention therefore had target-driven negative and context-driven positive outcomes upon performance in the wide focus, consistent context condition. Previously, consistency of context has been viewed as the primary mediator of scene-based contextual facilitation. These results provide further evidence that consistency is a mediator, but that it depends strongly, and perhaps completely, on visual attention. Any study seeking to examine how scene context relates to object recognition will therefore need to control for visual attention.

Chapter 4: How Many Objects Constitute a Scene Context?

Chapter 3 highlights that visual attention plays an important role in the generation of scene context effects. In Experiment 2 both the allocation of visual attention to the context objects and semantic consistency between context and targets were required to achieve contextual facilitation. In that experiment, endogenous cues initially distributed visual attention across entire arrays of four non-targets and one target. However, there is no reason to suppose that the scene context effects generated in Experiments 1 and 2 required processing of all items in the arrays. In Chapter 4, the issue of how many objects are required to generate a scene context effect in the object array paradigm is addressed.

What constitutes a scene context?

Biederman (1981) outlined the relations that characterised a well-defined, naturalistic scene with reference to schema activation. These were: interposition, support, size, position, and probability. However, his account does not fully explain what constitutes a scene context in Experiments 1 and 2. These experiments used arrays of objects rather than naturalistic scenes, and thus provided evidence that an object-semantic driven context effect could be generated without all of Biederman's requirements. *Interposition* was not violated in these studies as occlusion was avoided, although it might be argued that *support* was questioned by the arrangement of stimuli used (Biederman, personal communication). The absence of a background results in the objects positioned in the top left and top right hand corners appearing above those in the bottom left and right corners. They therefore do not meet Biederman's support criteria, despite many of the images having been photographed as groups on a table that is later deleted. Regardless of this apparent lack of support, the context effect was robust. It seems likely that this relation has a greater role in naturalistic scenes than in the object arrays shown in the experiments reported in Chapter 3. The relation of relative *size* is maintained between objects, but *position* loses much of its relevance without a coherent or naturalistic scene in which an object can be placed. The results from the Experiments 1 and 2 therefore indicate that *probability*, or semantic relatedness, and the allocation of visual attention are two of the key mediators in generating a context effect with object arrays. It also demonstrated that scene-configuration factors were not required to generate a significant scene context effect. What is not immediately apparent from these experiments is how many non-target objects are required to create an effective object-semantic driven context.

A single non-target object, or word, is capable of providing target-consistent (semantic-based) facilitation during a recognition task under the correct circumstances, as demonstrated in priming studies (see Neely, 1991 for review). However, using information from a single object from a scene as the basis of a context effect will only be useful if the selected non-target is a good representative of that context. Naturalistic scenes or arrays are typically composed of many objects, each potentially contributing different semantic information to the scene. Therefore, a representation based on a single object from the scene is likely to be less representative of the whole context than a representation based on multiple non-targets that are each partially processed. A representative context is more likely to reflect an individual's stored context maps/networks which in turn will usually assist recognition.

The generation of a context effect can be achieved only when attention is distributed across items (see Chapter 3). However, distributing visual attention across too many contextual items might weaken the context effect. First, if there is a fixed amount of the attentional resource, then when it is distributed between the non-target items, the proportion of attention that can be allocated to an individual non-target is reduced as additional contextual items are included. It is a common assumption in many theories of attention is that visual attention is used in structural processing during the formation of relationships between parts (Treisman & Gelade, 1980; Wolfe & Bennett, 1997; Hummel & Biederman, 1992) and the absence of sufficient attention will prevent the completion of these relationships. If these theories are correct, then such a division of attention between objects would also limit the level of structural processing that could be conducted amongst the context items. Fewer active relationships would reduce the likelihood of non-targets achieving unique matches with stored representations. Thus, whilst more non-targets may increase the breadth of extracted context information, too many items may result in only low-level data (part information) being extracted, and/or the object identifications based on this information being unreliable.

The second potential problem created by spreading attention across too many items is that visual attention is may be used to maintain the integrity of the individual object visual representations. Attention has been shown to play an important role in associating features with locations (Cave & Bichot, 1999; Shih & Sperling, 1996), binding simple units into complex representations (Treisman, 1996; Wolfe & Cave, 1999) and in preventing cross-talk (Mozer, 1991). If the amount of visual attention allocated to each item is reduced below a certain threshold then interference between objects may start to occur.

Given these issues associated with distributing attention across context items, the issue of how many objects can usefully contribute to generating scene context effects becomes important. Although four context objects were shown in Experiments 1 and 2, it is possible that all the advantage for objects in consistent over inconsistent contexts occurred via the processing of one, two or three of the contextual objects. By varying the set size associated with context items in Experiment 3, the magnitude of the scene context effect was compared across set sizes with a view to determining at what point the effect reached its asymptote.

The attentional manipulation:

Endogenous concentric circle cues and a secondary colour dominance task were used to manipulate the initial allocation of participant attention in Experiments 1 and 2. Both these methods required the participants' awareness, intention and control, and consequently may have interfered with other cognitive processes (Posner & Snyder, 1975). Such interference is a potential confound in Experiments 1 and 2 because its influence upon the recognition and naming processes cannot be measured, nor its impact upon the scene context effect. For this reason an exogenous cue was designed.

The purpose of the exogenous cue was to spread visual attention over the entire array, as contextual effects had been shown to require the allocation of attention to context items. The cue consisted of a small circle perimeter that expanded at a steady rate from a point at fixation in order to draw visual attention outwards. The fixation cross was removed at the onset of the cue, and the expanding perimeter was the only visual stimulus (black line) on an otherwise white background. The technique relied upon the reflexive nature of attentional orientation (Jonides, 1981; Müller & Rabbitt, 1989) and focusing (Turatto et al., 2000), and that it is difficult over-ride this automatic tendency without a visual anchor (i.e. fixation point – Turatto et al., 2000). This cuing method is based on the assumption that visual attention would involuntarily follow the expanding circle until it was spread sufficiently wide to include all the context items within the array, at which point the cue was replaced by the naming stimuli.

An exogenous attentional cue does not require a secondary task. It also captures participant attention without the requirement of their awareness or their intended control. This automaticity allows the interference with other cognitive systems to be minimised (McCormick, 1997).

Experiment 3:

Experiment 3 manipulated the number of non-target stimuli distributed about a target in a naming task to examine whether the object-based context effect was influenced by set size. The purpose of this investigation was to examine another potential mediator of contextual facilitation in object arrays, and to infer further knowledge regarding the behaviour of visual attention during contextual facilitation.

With the exception of the attentional cue, the removal of the secondary task, and the manipulation of the context set size the basic naming task paradigm remained the same as Experiments 1 and 2.

Method:

Participants:

Forty-eight undergraduates from the University of Southampton participated in one 40 minute session and received course credits as a result. All subjects reported normal or correct-to-normal vision and English as a first language. None knew the purpose of the experiment beforehand. There were 13 males and 35 females between the ages of 18 and 48 years (mean: 21.60).

Apparatus and Stimuli:

The experiment used a Macintosh Power PC G4 400MHz computer with a 19" ProNitron monitor (13msec screen refresh). Participants sat approximately 60cm from the screen in a dimly lit room and responded verbally.

The 32 context groups of five related stimuli were the same as those used in Experiment 2 (see appendix B), and these were paired in the same manner in order to provide semantically consistent and inconsistent context items (i.e. a target from context group 1 with non-targets from context group 2). From these context groups, four objects from each were used to create targets (stimulus sets A, B, C and D). Despite a slightly lower mean familiarity rating from Experiment 2 (mean: 3.3) an ANOVA indicated that stimulus set D was not significantly different from the other three target sets [$F(3, 124) < 1.0$]. These 128 target objects were matched with one, two, three and four context consistent and context inconsistent non-targets to form 1024 stimulus arrays, each 23cm x 23cm, each of which was stored in a separate image file. Positioning of the context items around the target for set sizes other than four was counter-balanced between the four possible locations.

A multi-coloured mask identical to that used in Experiment 2 was used to prevent participant use of the iconic memory.

Procedure:

The procedure within a single trial was similar to that of Experiment 2. A fixation cross was displayed in the centre of the screen for 480msec. It was replaced by a tiny circle at the fixation point (the exogenous cue), which expanded at a steady rate for 780msec to reach its maximum radius of 10.82cm (visual angle 10.26°). On completing its expansion, the attentional cue was removed, to be replaced by one of the naming stimuli (target and context items).

Piloting had revealed that removal of the colour dominance task resulted in improved participant performance in the naming task relative to Experiment 2; therefore the display time for the naming stimuli was reduced from 78msec to 65msec. Participants were required to name the target object at the centre of the screen as quickly and as accurately as possible. Error rates were scored on a response sheet by an experimenter present during the session, as in previous experiments.

Naming stimulus offset was followed by the presentation of a multicoloured mask for 975msec to prevent the use of iconic memory. Participants were allowed a maximum of three seconds from the onset of the naming stimulus to provide a name before progression to the next trial.

Participants viewed all the target objects across two blocks (context groups 1-16 and context groups 17-32), which were counter-balanced for order between participants. Presentation order was randomised within blocks to ensure that context type (consistent or inconsistent), set size (1, 2, 3 or 4) and context item locations relative to target were unpredictable.

Participants were given scripted instructions prior to the starting of the task, and sixteen practice trials preceded the experimental trials to illustrate two examples of the main condition combinations (e.g. consistent + 2 non-objects). Participants were also debriefed following the experiment.

Design:

Four object images from each of the 32 context groups were used as targets to generate 128 trials of recorded data for each participant. Within participants, half of these targets were viewed in a consistent context condition and half in an inconsistent context condition. The trials within these halves were further divided to allow manipulation of contextual set size (1, 2, 3 or 4) around the target. The stimulus arrays were organised into eight 'session sets', each with 128 different stimuli. Any given object served as a target only once within a single session set. Each participant received one of these eight sets of stimuli, and the eight sets were balanced so that every target would be seen in every combination of context and set size with equal probability across subjects. This resulted in a basic mixed design with two within-participants conditions (2 x 4) and two between-participants conditions (2 x 8).

Error rates were measured in the naming task. Following the absence of an effect for RTs in Experiments 1 and 2, RTs were not recorded in Experiment 3.

Results:

An overall mean error rate of 33.11% suggested that it was high enough to demonstrate an effect if one were present. Table 11 displays these mean error rates and standard errors across each of the four set size conditions (1, 2, 3 and 4) and split between the context type conditions (consistent and inconsistent).

Table 11: Mean Error Rates (%) and Standard Errors by Context Type and Set Size

Context Type	Set Size			
	1	2	3	4
Consistent	32.81 (1.7)	30.60 (2.0)	31.77 (1.6)	33.46 (1.9)
Inconsistent	32.55 (1.8)	33.72 (2.0)	34.12 (1.6)	35.81 (1.7)

This is presented graphically in Figure 19 and suggests a scene context effect in set sizes 2, 3 and 4, but not in set size 1.

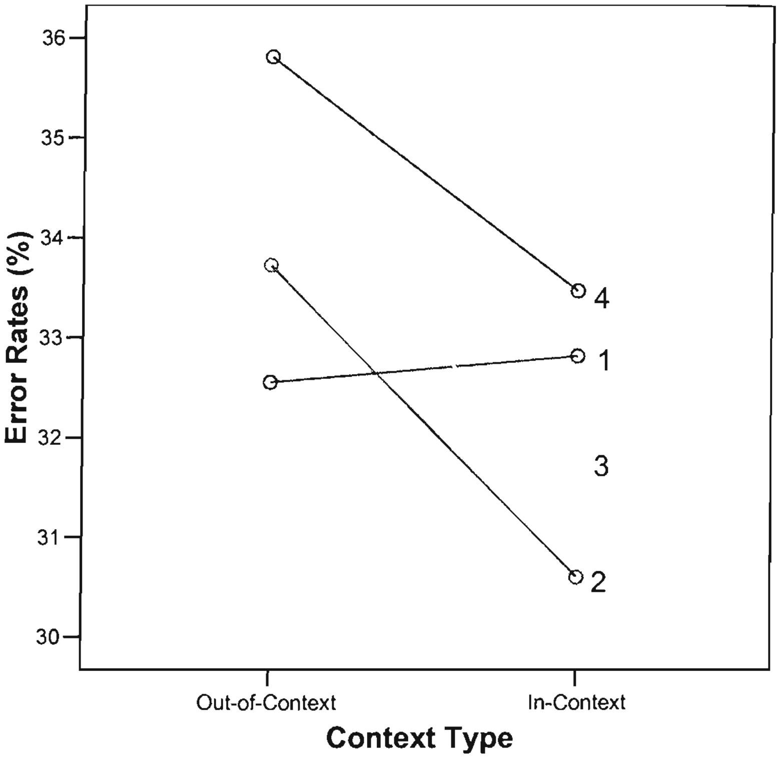


Figure 19: Graph of mean error rate (%) by context type and set size

These error rates were then subjected to a repeated measures ANOVA with factors of context type and set size, and between-subjects controlling factors of order and session set (1 to 8). At the 95% level there was no main effect of context type, but there was a strong trend towards significance [$F(1, 32) = 3.925, p = 0.056$, partial $\eta^2 = 0.11$] supporting a context consistency advantage. There was also no main effect of set size [$F(3, 96) = 1.002, p > 0.1$, partial $\eta^2 = 0.03$], and no significant interaction between context and set size [$F(3, 96) < 1.0$, partial $\eta^2 = 0.02$].

There were no main effects for the between-subjects effects of order [$F(1, 32) < 1.0$, partial $\eta^2 = 0.00$] and session set [$F(7, 32) = 1.972, p > 0.05$, partial $\eta^2 = 0.30$]. A significant interaction between context and session set [$F(7, 32) = 12.700, p < 0.01$, partial $\eta^2 = 0.74$] would appear to be due to the targets for the context consistent condition in groups 1-4 being, on average, more difficult to recognise than the context inconsistent condition. This results in an exaggeration of the context effect in these groups and a negative context effect in groups 5-8 where the context condition is reversed. The counter-balancing ensures that the true effect can still be measured by analysing the difference between the overall effect. Interactions between set size and session set [$F(21, 96) = 36.457, p < 0.01$, partial $\eta^2 = 0.89$] and context, set size and session set [$F(21, 96) = 12.847, p < 0.01$, partial $\eta^2 = 0.74$] were also found as a result of an increased error rate with stimulus set D, relative to stimulus sets A, B and C. Session sets 1 and 5 (same targets) also appeared to be more difficult than the other session sets, though to a lesser degree (see Table 12).

Table 12: Mean Errors Broken Down by Set Size and Session Set

Session Set	Set Size 1	Set Size 2	Set Size 3	Set Size 4
1	67.7	33.3	34.4	34.4
2	19.3	18.2	18.2	54.7
3	23.4	21.9	60.9	29.7
4	25.5	57.8	21.4	26.0
5	60.4	32.8	28.1	27.7
6	19.8	20.3	20.8	54.7
7	27.6	19.7	58.9	29.7
8	19.7	53.1	20.8	20.3

n.b. underlined values show trials where Stimulus Set D was Target

An analysis was performed that replaced stimulus set D with the series mean of stimulus sets A, B and C. The removal of this data reduced the overall mean error rate to 24.64% and decreased the standard error by more than 25% (from 1.277 to 0.950). There was still a main effect of set size [$F(3, 96) = 6.033, p < 0.01$], and interactions between context and session set [$F(7, 32) = 8.124, p < 0.01$] and context, set size and session set [$F(21, 96) = 17.697, p < 0.01$]. However, the interaction between set size and session set was reduced to a trend [$F(21, 96) = 1.625, p = 0.59$] demonstrating its dependence upon stimulus set D. In addition, the main effect of context (a strong trend) was eliminated [$F(1, 32) < 1.0$] but a strong trend towards an interaction between context and set size was found [$F(3, 96) = 2.544, p = 0.061$]. Post hoc analysis based on this strong trend (using a 2 x 2 x 2 x 8 ANOVA) found that the context effect generated by set size 1 was significantly less than that generated by set sizes 2 [$F(1, 32) = 5.812, p < 0.05$] and 4 [$F(1, 32) = 4.136, p = 0.05$], however further post hoc analysis (2 x 2 x 8 ANOVA) found that only set size 2 yielded a main effect of context [$F(1, 32) = 4.418$), $p < 0.05$].

Discussion:

There was sufficient evidence from these error rates to indicate that a scene context effect consistent with that demonstrated in Experiments 1 and 2 was present in Experiment 3. With stimulus set D present, providing a positive context effect in seven of eight session sets, the overall context effect was weaker than in the previous two studies, but both a strong trend and a reasonable effect size (partial η^2) were found. These results of the analysis with stimulus set D (thus reducing variance) suggest that removing set size 1 would have yielded a main effect in context type regardless of other factors.

The interaction between context type and set size was not significant in the initial analysis despite the suggestion that set size 1 performed differently between context conditions compared to the other set sizes in Figure 19. By reducing the variance in the analysis through the substitution of stimulus set D with a series mean of the other stimulus sets a strong trend was found for the interaction between context and set size. Post hoc analysis found that set size 1 generated a significantly weaker context effect than set size 2 or 4, with set size 2 appearing to be the asymptote in this experiment. An object-driven context would ideally be formed of a target plus two semantically related objects. However, it should be remembered that these latter findings are based on both reduced data sets and post hoc analyses of trends rather than a significant result. They should, therefore, be viewed as suggestive.

The context effect may have been confounded with variability in positional certainty across set sizes. In set size 4 all four positions are certain to contain a context object. This probability decreases in set size 3 (75%), set size 2 (50%), and it reaches a minimum in set size 1 (25%). If participants were not distributing visual attention to all four non-target locations on every trial, the probability of visual attention being focussed on empty locations is inversely related to set size. This lack of attended non-target objects would weaken the main effect of context relative to Experiment 2, and explain the absence of an effect in set size 1, despite the fact that the statistical analysis implies an effect is present at set size 1 (Figure 19). The current experiment cannot discriminate between these explanations of why the interaction between set size and context type was not significant, and leaves open questions about whether a single context object can produce a context effect. This issue is returned to in Experiment 5.

It is also possible that the use of an endogenous cue in Experiments 1 and 2 may have strengthened the context effect as they have a longer cueing period relative to an exogenous cue (Müller & Rabbitt, 1989; Yantis, 2000). The secondary task may also have interfered with cognitive systems to increase the influence of non-targets (Lavie et al. 2004). Thus, the removal of these effects by using an exogenous cue may have reduced the contextual facilitation. However, although the scene context effect in this experiment was less robust, its presence indicates that the exogenous cue must have captured the participants' attention, and then held it for sufficient time period to influence recognition. It is therefore considered that the endogenous cue was successful in manipulating participant attention for context/recognition tasks.

Conclusion:

Performance inequalities in session sets and stimulus sets due to targets, their interactions with contexts, or their physical layouts, have made it difficult to draw firm conclusions from this data. However, results suggest that object-semantic contexts require at least the target plus two semantically related items to generate a context effect. Further studies are required to confirm this finding. In addition, the results of this experiment demonstrate that an exogenous cue, as well as an endogenous cue, can be used to manipulate visual attention in order to generate a scene context effect in recognition.

Chapter 5: Using Context Information and Time as a Mediator

The empirical studies presented in previous chapters have explored the generation of contextual facilitation in arrays using a naming task. It is generally accepted that these effects exist and are mediated by contextual consistency within naturalistic scenes (e.g. Bar, 2004; Biederman, 1981; Davenport & Potter, 2004). This thesis has demonstrated similar effects in object arrays, and has also shown that visual attention has a mediating influence on scene context effects driven by object-semantic factors. However, Experiments 1, 2 and 3 do not address the issue of whether the scene context effect is a genuine superiority effect.

If the information extracted from the non-target stimuli is integrated with the perceptual/representational processes that lead to target recognition, the influence can be considered a scene superiority effect. However, observation of a consistent context may not affect perceptual processes operating on the target, yet still bias an individual's guess towards a limited number of options. For example, a participant is more likely to name a frying pan than a football if they glimpse a kitchen scene. Although it is important to establish the nature of scene context in order to understand its relationships with object recognition, the current empirical evidence makes it difficult to distinguish between the two main viewpoints summarised below.

The Interactionist Perspective:

The Interactionist Perspective proposes that the information extracted from the non-target stimuli interacts with the perceptual process during target recognition to generate a superiority effect. There are several accounts that differ from one another in the nature of the interaction.

Hollingworth and Henderson (1999) put forward the Description Enhancement hypothesis based upon Biederman's (1981) experiments. In these experiments a target name was presented to provide the participant with an object to detect, followed by a central fixation point, and then a scene that contained the target. Finally, a mask and a cue were displayed, and the participant was required to identify whether or not the target had been present at the location given by the cue. In each trial the target, or a bystander object, would violate 0, 1, 2 or 3 physical or semantic relations (rules of normal visual behaviour) and participants would be measured upon the accuracy and speed of their response. It was found that violations of semantic relations (i.e. relative size, position, and the probability of something being in a particular scene) hampered performance at least as much as violations of physical relations (support and interposition). Biederman concluded that scene, or schema level semantics were established as quickly as physical relations.

From Biederman's (1981) findings, Hollingworth and Henderson (1998, 1999) suggested that the rapid identification of these scene semantics, and the activation of a schema or memory representation, may actively facilitate the perceptual processing of objects consistent with that schema. Such an influence would provide top-down input that aids the extraction of edges and features (or the combination of features) commonly associated with early-level processing. However, Biederman (1981, personal communication) proposed an alternative perspective in which it is only a few objects in familiar interaction that generate scene-emergent features that provide features not provided by objects in isolation (e.g. a chair partially occluded by a desk). These emergent features arise from groups/scenes and thus facilitate the perception of the setting or schema itself.

These two explanations of the same results have different theoretical implications. The Hollingworth and Henderson (1998, 1999) Description Enhancement hypothesis requires that the schema be activated prior to the target, so that it can facilitate perceptual processing, and thus produce a target representation that will be more detailed or complete as a result of the visual context. The Biederman (1981, personal communication) proposal does not rely on enhancement of early stages of perception such as edge detection. It shares characteristics with a structural recognition model in which the frame of reference has been set above the object level. Thus, the objects themselves do not need to be identified (only their interactions), nor does the schema need to be activated prior to the target. In this proposal, contextual influence can occur at the point of representation matching rather than with perceptual processing, with emergent features providing additional activation (see below). For this reason the two versions will be referred to as the Early Description Enhancement hypothesis (Hollingworth & Henderson, 1998, 1999) and Late Description Enhancement hypothesis (Biederman, 1981, personal communication).

A third account includes a number of similar theories under the collective title of the Criterion Modulation hypothesis (Hollingworth & Henderson, 1999). The general hypothesis stems primarily from the work of Palmer (1975), Friedman (1979), Rumelhart and McClelland (1981) and more recently Kosslyn (1994), and shares with Biederman (1981) the views that the semantic details of a scene can be processed extremely rapidly and often result in the activation of a scene schema. The principle difference between this and Description Enhancement is that Criterion Modulation makes no claim to the emergence of new features either at an object or an interactive level. In this explanation the influence of the scene schema upon object detection or recognition occurs when attempting to match the target object against stored representations held in memory. Regardless of whether a view-based or object-based model of recognition is used, there must be a stage whereby incoming information is compared against potential matches. If a threshold is surpassed, then identification is achieved. Under the Criterion Modulation model, the activation of a specific schema will lower the activation thresholds of those objects semantically linked to that schema. This modulation may occur via a spreading of activation (e.g. Kosslyn, 1994), reducing the amount of information required to activate objects related to the schema, and increasing the likelihood that context consistent objects are selected.

Criterion Modulation can operate with a traditional scene schema (Biederman, 1981), but there is no reason why it needs to be restricted to only functioning through the activation of a schema. Within a connectionist hierarchy, a schema would utilise vertical connections (schema-to-object/object-to-schema) that activated links to objects on a lower level or allowed objects below to activate the schema. However, such a network could also connect objects horizontally (object-to-object) by semantic links. Activation of one object could result in activation being transferred directly to related objects, lowering the threshold requirements for all without the requirement of a schema. In addition, this would increase the likelihood of a scene schema achieving its activation threshold.

The Late Description Enhancement hypothesis shares more elements of this connectionist approach than the Early Description Enhancement hypothesis. Rather than seeking to directly influence perceptual processes, the emergent features can be seen as a hierarchy level between objects and schemas. These features can themselves be activated, and through the network of spreading activation affect representational processes during matching.

The Criterion Modulation hypothesis, especially with horizontal and vertical connections, causes an activated context item to spread its activation to semantically related objects. The viewing of multiple related items would increase this effect, raising the probability that one of these objects, or a related item, was selected. It is the interaction with the perceptual processes that limits the selection of the stored representation to those meeting the physical characteristics of the target. The combination of perceptual and contextual information may be achieved through a parallel simultaneous constraint satisfaction model (Boyce & Pollatsek, 1992). Alternatively, the concept of an activation map similar to that in the Guided Search attention model (Wolfe et al, 1989; Wolfe, 1994) may be applied to the stored representation selection process. The two 'feature' maps providing activation for such a search would be perceptual and contextual, and this would provide an additive integration. There is little empirical evidence to favour either approach relative to the other. Both models would inherently encounter difficulties with high levels of similarity for perceptual or semantic factors.

Isolationist Perspective:

The alternative view-point is based upon the Functional Isolation hypothesis (Hollingworth & Henderson, 1998, 1999). It does not deny the presence of scene context effects, or that there may be a semantic relationship between an object and the scene in which it was displayed. However, it proposes that scene context effects are not due to the facilitation of perceptual processing or matching of descriptions against stored representations. The processing of the context is done in isolation from that of the target throughout the recognition process, and the scene context effect is nothing more than a response bias that does not interact with perceptual processing.

As noted in Chapter 1, Hollingworth and Henderson (1998) claimed that the effect demonstrated by the Biederman study (1981) was a result of response bias rather than sensitivity, and thus it was not a perceptual effect. In their later experiments (1998, 1999) they examined whether a consistent context aided perceptual processing when response bias was controlled. In these studies a central fixation was followed by the brief presentation of a scene that contained several objects (e.g. a garage forecourt). A mask was then displayed before participants had to select which one of two objects from the same category had appeared in the scene (e.g. sports car and saloon car). Their hypothesis was that if a consistent context had assisted the accumulation of perceptual information, then performance would be better under this condition. The contextual validity of both forced choice responses was the same, preventing strategic guessing and eliminating response bias, thus any effect would be due to sensitivity. No performance improvement was found in the context consistent condition, and a reliable trend was detected for improved performance in the context inconsistent condition. Based on these results, Hollingworth and Henderson argued against the interactionist perspective.

Hollingworth and Henderson (1998, 1999) compared their methodology to the Reicher-Wheeler (1969) paradigm, which had been used to examine word superiority effects on letter recognition, and was an accepted technique for eliminating response bias. However, there are important differences between the two paradigms. In the Reicher-Wheeler paradigm, a letter, for example an 'A', is presented within a word (e.g. 'FARM') or a non-word (e.g. 'RADE'). The target letter had no semantic significance of its own to highlight it from the other letters. It was the arrangement of the letters that provided the context rather than the individual semantic properties of each item. Hollingworth and Henderson's studies differed from this pattern. By presenting the target name after the stimulus display the context consistent condition may have eliminated response bias in the same manner as Reicher's (1969) presentation of 'CAFÉ' and 'FARM' as all objects were context-valid. However, the context inconsistent condition was not equivalent to 'RADE' and 'BATE' as the target object would have had an inconsistent semantic property that would highlight it from the other items. Once a mismatch was noticed visual attention could be directed to the 'odd-object-out' and would aid perceptual data extraction. A display time of 250msec allowed plenty of time for this detection to occur. Such a predictive strategy in the inconsistent context condition would have eliminated any context consistent advantage, and explained Hollingworth and Henderson's trend towards a context inconsistent advantage.

In addition, Reicher's participants did not have the added difficulty of distinguishing the target from a perceptually similar exemplar from within a category. Because Hollingworth and Henderson (1998, 1999) used objects from the same category as the two forced choices in a given trial (e.g. a saloon car and an estate car), their participants had to choose between two options that were both in the same context and perceptually similar. Of the different hypotheses described above, only the Early Description Enhancement hypothesis predicts that context should improve accuracy in this choice, because under this hypothesis the additional information from context can directly enhance the perceptual processes of recognition. The lack of a consistent scene context effect suggested that perceptual processing (e.g. extraction of edges and features) was not directly enhanced by consistent context. However, the lack of a context effect in this forced-choice task is still consistent with the Criterion Modulation hypothesis. Semantic-activation could be provided by the context to both forced choice options prior to the matching of representations. Bar's (2003) evidence that the pre-frontal cortex can categorise, but not identify within categories, is therefore consistent with the absence on a context effect when choosing between exemplars. This limitation suggests that contextual information cannot assist when perceptual similarity is high and context validity is the same between options. This argument is also true for the Late Description Enhancement for the Hollingworth and Henderson's experiments due to their lack of combined/occluded target and contextual objects.

Despite recent interest in the field of context research, there has been little progress in establishing whether scene context effects are due to influences upon the perceptual/representation processes of recognition, or are the results of response bias. Davenport and Potter (2004) referred to response bias, and suggested that it was a minor factor in explaining the significant scene context effects they found in their naming task. However, their analysis does not take into account non-responses in what is a non-forced choice experiment, and thus there is ample opportunity for response bias in their paradigm. Experiment 2 in this thesis has demonstrated that there is a potential contextual bias towards response-types, including significantly more non-responses in the inconsistent context conditions. Consequently there is currently no empirical evidence by which to select one of these two hypotheses.

It is also worth noting that Davenport and Potter (2004) utilise a naming paradigm. This is a standard response method in scene context research, and has been used in the three studies presented in Chapters 3 and 4. Naming minimises experimenter interference with the participant response (i.e. by restricting options). However, it allows non-responses, as noted above, and also engages the psycholinguistic function in addition to the recognition processes. Observed effects generated by a task in which two cognitive systems are activated may be the result of either system. Whilst an interactionist approach would maintain that the performance advantage is due to the contextual influence upon the recognition processes, interference to the language system provides an alternative explanation suitable for the isolationist viewpoint. This issue has not been addressed in scene context research; however naming research has examined the influence of semantic relatedness on performance.

Semantic Effects in Naming:

There is empirical evidence that categorical relationships between stimuli influence the naming function. Riddoch and Humphreys (1987) found that categories possessing structurally similar exemplars (e.g. animals, fruit, vegetables etc.) were named slower than structurally distinct items. Within perceptually similar categories, items low in prototypicality were named fastest (e.g. a giraffe), whereas in distinct categories name frequency exerted a strong influence on naming times. These RT effects were found both in picture naming and in word naming, suggesting their source is within the linguistic processes. Riddoch and Humphreys' participants only named a single object at a time in these experiments, but their cascade model is based on activations and inhibitions within the structural, semantic and phonological systems. These systems act in parallel, potentially active from the initiation of the previous stage. Thus the phonological process can begin before structural and semantic stages have been completed. Introducing more activation via contextual items, particularly in a perceptually similar category, would interfere with naming times further.

Similar categorical effects have been found with picture naming errors (Vitkovitch, Humphreys & Lloyd-Jones, 1993) in which a wider range of errors were made for objects from structurally similar categories. This is consistent with having more closely related potential responses that partially satisfy both perceptual and semantic systems. The evidence suggests that activation of categorical co-members, particularly in structurally similar categories, interferes in the naming process. However, such categorical associations possess similar characteristics to the contextual consistency required to achieve recognition facilitation in Experiments 1, 2 and 3. This has been defined as the "semantic relatedness paradox" (Neumann, 1986).

Further evidence of semantic interference in naming is provided by Vigliocco, Vinson, Damian and Levelt (2002). They demonstrated graded naming latencies for object pictures and action pictures modulated by the semantic similarity of exemplars they presented. For example, items from the same category (e.g. clothes) were named slower than those from a closely related category (e.g. body-parts), but the unrelated category (e.g. vehicles) were named fastest.

Starreveld and La Heij (1996) conducted a study that displayed a picture with a word at five time asynchronies (-200msec to 200msec at 100msec intervals). The word could be either the correct name for the picture (e.g. CAT), be orthographically and semantically related (e.g. CALF), be semantically related (e.g. PIG), be orthographically related (e.g. CAR), be unrelated (e.g. PIN), or be a control of capital X's. Participants were required to name the picture. There were orthographic facilitation effects in naming RTs across almost the entire range of word onset asynchronies (-200msec to 100msec). Semantic interference effects were found but only during the periods of -100msec and 0ms. This illustrates that although semantic relatedness influences the linguistic function via images, it does so only over a very limited time course. Bloem and La Heij (2003) extended these findings to show that in a word translation task, context words produced semantic interference where context pictures generated facilitation. They also demonstrated that whilst context words can produce phonological facilitation, context pictures cannot – they can only produce phonological interference.

The cascade model (Riddoch & Humphreys, 1987), the two-stage connectionist model of Starreveld and La Heij (1996; also Bloem & La Heij, 2003) and neuropsychological evidence and computer modelling (Humphreys, Price & Riddoch, 1999) suggest that naming is influenced by context. However, the majority of evidence indicates that rather than being the source of the contextual facilitation in scene processing, it is likely to reduce any picture-based context effect through interference. It is possible that psycholinguistic effects of contextual interference may be weakening the overall context effect in Experiments 1, 2 and 3. This could also explain the absence of RT effects in Experiments 1 and 2, as RT context effects have been found in object detection studies (Biederman, 1981) and RTs have been found to be affected by naming interference (Riddoch & Humphreys, 1987; Vigliocco et al. 2002). It would therefore provide a more rigorous test of contextual effects to examine them under an alternative paradigm that minimised the influence of the linguistic function.

Removing the Response Bias:

The empirical studies in this chapter utilise a non-naming paradigm that addresses the issue of whether scene-based contextual information directly influences the recognition process. In order to do this, a new paradigm was required that allowed the estimation and removal of response bias from the contextual effect.

The Six Alternative Forced Choice Method:

The alternative method selected was a post presentation six alternative forced choice paradigm (6AFC). An array of stimuli similar to that used in previous naming experiments would be displayed, after which participants would be presented with a list of six object names. Three of the choices in the list would be:

a) the target
b) an object perceptually similar to the target (but not contextually related)
c) an object semantically related to the context (but not perceptually similar to the target)

This array of choices would allow perceptual errors and semantic errors to be separated, and semantic errors could be used as a basis for detecting and eliminating response bias. However, a three alternative forced choice response would be insufficient as it allows participants to deduce the context type of the previously viewed object array and modify their response strategy accordingly. For example, in a context consistent trial the target would be semantically related to the context, and therefore related to the semantic error choice. This would not be so in a context inconsistent trial as all three choices would be semantically unrelated. Consequently, selecting one of a semantically related pair (if present) would be an effective guessing strategy that would generate a context consistent advantage. Three additional choices would be provided to address this problem:

d) an object perceptually similar to choice c.
e) an object with no relationship to any previous choice (context consistent trials) or semantically related to choice b (context inconsistent trials).
f) an object perceptually similar to choice e.

These additional options ensured there would be three pairs that were perceptually similar, and one pair that was semantically related, in every trial regardless of context type. They also provide base-line responses without perceptual or semantic bias (i.e. chance errors). Choices would be controlled between lists so that probabilities of appearing in both context consistent and inconsistent trials were equal. The presentation order of the list would be randomised for every trial.

Participants might still use the strategy of selecting one of the two semantically related pairs. This would yield a correct target response on half of the context consistent trials, but would always produce an error in context inconsistent trials. A post-test guessing correction would be used to correct for this possibility. The only reason a participant would select choice 'e' in preference to choice 'd' or 'f' in the context inconsistent trials is if they were favouring the semantic pairs. Therefore, the difference between the number of choice 'e' errors and a mean of choice 'd' and 'f' errors approximates the effect of this guessing strategy for semantic pairs (see Figure 20):

Response	Consistent	Inconsistent
a) Target	Cards - G	Cards
b) Perceptual Error	Paper Fan	Paper Fan - G
c) Semantic Error	Roulette Wheel - G	Banana
d) Error type 3	Tyre	Horn
e) Error type 4	Mobile Phone	Chopsticks - G
f) Error type 5	Calculator	Straws

$$G = \text{Response}_e\,[\text{Out}] - (\text{Response}_d[\text{Out}] + \text{Response}_f[\text{Out}])/2$$

Figure 20: The guessing correction formula, and how it would be applied to a single context consistent and inconsistent example. In this example, the target object to be identified is a set of playing cards. In the context consistent condition, it appears among other objects associated with games, and in the inconsistent condition it appears among different types of fruit.

The appropriate correction would be made by subtracting G from the number of choice 'a' and 'c' errors in the context consistent condition, and from the number of choice 'b' and 'e' errors in the context inconsistent condition.

This paradigm still leaves open an opportunity for response bias, if a participant is unable to perceptually identify a target object but is able to gather some information about the context objects, and then chooses an object name from the list purely because it matches the context. A similar technique to the guessing correction would be used to calculate and remove the effects of response bias. The difference between semantic errors (choice 'c') and baseline errors (choices 'd' and 'f') indicates of the level of response bias due to context. This would be calculated using the context inconsistent trials as these would provide semantic errors un-confounded by semantically related targets. The difference between the number of choice 'c' errors and the mean of choice 'd' and 'f' errors would provide an approximation for the effect of response bias (see Figure 21):

Response		Consistent	Inconsistent
a)	Target	Cards - B/2	Cards
b)	Perceptual Error	Paper Fan	Paper Fan
c)	Semantic Error	Roulette Wheel - B/2	Banana - B
d)	Error type 3	Tyre	Horn
e)	Error type 4	Mobile Phone	Chopsticks
f)	Error type 5	Calculator	Straws

$$B = \text{Response}_c[\text{Out}] - (\text{Response}_d[\text{Out}] + \text{Response}_f[\text{Out}])/2$$

Figure 21: The response bias correction formula, and how it would be applied to a single context consistent and inconsistent example.

The eliminations of bias would then made by subtracting B from choice 'c' on the context inconsistent trials, and B/2 from choices 'a' and 'c' on the context consistent trials (as the response bias would be divided between the two alternatives).

A lower number of total errors when identifying an object in the context consistent condition than in the context inconsistent condition, after correction for response bias, would be evidence against the Functional Isolation hypothesis. However, lack of a significant context effect under these conditions would suggest that contextual facilitation can be explained by response bias alone.

Time Course of Semantic Facilitation:

Starreveld and La Heij (1996) found semantic interference when a word was presented simultaneously or 100msec before a semantically related image, but that orthographic facilitation occurred when a phonographically similar word was presented before or after the image (e.g. picture: CAT – word: CAR). Such research has not been conducted into the time course of semantic facilitation, specifically not into the time course of contextual effects. An additional aim of this experiment was to examine whether the presentation of the visual context prior to the target influences the magnitude of the scene context effect. The display time of the context would remain constant, but there would be no visual mask between the non-targets' offset and the target onset. This early exposure of the context allows the iconic memory to be used to generate an increased temporal window to process contextual information before attention is drawn by the target. It was predicted that when the non-target items were displayed prior to the target, the additional processing time would result in a larger context effect.

Experiment 4:

In this study we used the same picture stimuli that have been used in Experiments 2 and 3. These had already been shown to generate contextual facilitation with a naming paradigm. Previous experiments did not determine whether the scene context effect was due to response bias or was a result of an interaction of perceptual and contextual information. The six-alternative forced choice (6AFC) paradigm adopted for this experiment allowed the control and elimination of response bias, and minimised the utilisation of the naming processes. A temporal manipulation of non-target presentation in relation to the target also explored how time mediates the context effect.

Method:

Participants:

Sixty undergraduates participated in one 30-min session either for course credits or for £3.00 in compensation. All participants reported having normal or corrected-to-normal visual acuity and English as a first language. None had previously seen the pictures.

Apparatus and Stimuli:

A Macintosh Power PC G4 400MHz computer with a 19" ProNitron monitor (13msec screen refresh) controlled stimulus presentation and response acquisition in a dimly lit room. The same context groups as Experiments 2 and 3 were used to generate similar object arrays, with a target fitted within a circle of 3 cm radius (visual angle 2.68°), and four non-targets within a circle of 10.82cm radius (visual angle 10.26°). Context pairings (i.e. context group 1 with context group 2) used previously to create consistent/inconsistent non-targets were maintained. Context groups 33 and 34 were added (see appendix C) and paired with each other, and four stimulus sets were created (A, B, C and D) as in Experiment 3. For each target there were five 23cm x 23cm image files: one containing the target with consistent non-targets, one with the target and inconsistent non-targets, one with only the target, and one each of the consistent and inconsistent non-targets without the target. A visual mask identical to that used in earlier naming experiments was used to terminate visual processing.

Procedure.

Piloting had revealed that the participants performed better on the 6AFC task than on the previous naming task. In order to maintain a sufficient difficulty, display time was therefore reduced from 65msec to 52msec.

Having demonstrated that an exogenous cue could be used in contextual research in Experiment 3, an identical expanding circle technique was used in this study for attentional manipulation. This cue type was chosen in preference to the endogenous cue because it minimised activation of cognitive functions not directly involved with contextual processing. If only the recognition processes were utilised, then the effect can be more reliably attributed to that system. It also ensured that the scene context effect was not exaggerated by interference from the secondary task (see Experiment 2).

All trials began with a fixation cross at the centre of the screen. On removal of fixation the outline of a circle expanded outwards from the centre, at a steady rate, until it reached a radius of visual angle 10.26° after 780msec. On trials in which the non-target stimuli onset asynchrony (SOA) was –104msec or –52msec, the offset of the circle was immediately followed by just the context objects. In the –104msec display condition the offset of the context picture stimulus was followed by a blank screen display of 52msec before the display of the target object stimulus. In the -52 msec display condition, the offset of the context picture stimulus was immediately followed by the onset of the target object stimulus. On simultaneous trials (in which the SOA was 0msec) the offset of the expanding cue was immediately followed by the onset of a combined target object and contextual stimuli, which was displayed for 52msec. In every condition, target and non-targets were each displayed for 52msec, and the offset of the target was followed by the onset of a multicoloured mask displayed for 1000msec. Participants viewed all the target objects across two blocks, which were controlled for order between participants.

The 6AFC was presented at the offset of the mask. Every list contained: a.) the target object; b.) an object perceptually similar to the target; c.) an object semantically related to the context; d.) an object perceptually similar to (c.); e.) an unconnected object (context consistent trials) or an object semantically related to (b.) (context inconsistent trials); and f.) an object perceptually related to (e.). These six items were listed vertically and their order was randomised. Participants responded using a mouse to click on their choice, with auditory feedback being given to incorrect responses.

Participants received scripted instructions prior to the task, and completed 16 practice trials on non-experimental stimuli in order to simulate each condition. They were debriefed following the experiment.

Design:

Four images from each of 33 of the context groups were used to generate 132 experimental trials. (Context group 34 only provided non-targets for context group 33.) Participants were shown each target as a target only once, although all targets also appeared as non-targets in three other trials. Within participants, context type was manipulated so that half these targets were viewed with a consistent context and half with an inconsistent context. The trials within these halves were divided evenly to allow manipulation of context presentation before the target (SOA: -100msec, -52msec, 0msec). Between participants, control variables of order (block presentation) and session set (1 to 6) were used. Session sets were created in pairs so that consistent-context targets in set 1 were inconsistent-context targets in set 2. Stimulus sets (A, B, C and D) were mixed between the three pairs. Participants were shown an equal number of targets from each combination of conditions, and across participants every target was seen in every combination of conditions an equal number of times. This resulted in a mixed design of two within-participant variables (2 x 3) and two between-participant variables (2 x 6).

Results:

Error Rates: (Before response bias correction)

An overall error rate of 33.88% indicates that the forced choice task was difficult enough to avoid a ceiling effect through lack of errors. Table 13 displays the mean error rates across context conditions for the different SOAs before the response bias correction[4], Table 14 displays the error rates across all conditions.

Error rates were subject to a repeated measures analysis of variance (ANOVA) with factors of context type (consistent and inconsistent) and SOA (-104msec, -52 msec, 0msec), and between-subject controlling factors of session set and block order. The contextual facilitation prior to the response rate correction was found to be significant [$F(1, 48) = 14.611, p < 0.01$], and a main effect was also found for SOA [$F(2, 96) = 50.924, p < 0.01$]. Despite having the additional processing time, the –104msec display condition had the worst performance (43.4%), followed by the –52msec display condition (33.4%), with the simultaneous condition having the lowest error rate (28.7%).

[4] The guessing correction has been made on all results.

Table 13. Mean Error Rates (%) in Each Condition Before Response Bias Correction

Context type	Context SOA		
	-104msec	-52msec	0msec
Consistent	38.90 (2.2)	30.65 (2.4)	27.61 (2.7)
Inconsistent	47.80 (2.3)	36.15 (2.4)	29.85 (2.4)

Table 14: Mean Error Rates (%) Broken Down By All Variables

		Context					
			Consistent			Inconsistent	
Session Set	Order	0msec	-52msec	-104msec	0msec	-52msec	-104msec
1	1	37.7	42.7	46.4	53.6	60.0	60.9
	2	24.5	15.0	30.5	28.2	38.2	40.0
2	1	29.1	25.5	50.0	23.6	28.2	50.0
	2	20.0	20.5	32.7	14.5	21.8	40.9
3	1	34.5	50.0	56.8	49.1	56.4	60.9
	2	20.5	25.0	45.9	25.5	33.6	48.2
4	1	39.5	55.0	53.2	39.1	47.3	68.2
	2	26.8	34.5	31.4	22.7	31.8	49.1
5	1	27.7	35.5	26.4	28.2	36.4	48.2
	2	10.0	19.5	19.1	24.5	20.9	33.6
6	1	40.0	30.9	47.3	28.2	47.3	51.8
	2	20.9	13.6	27.3	20.9	11.8	21.8

There was a significant interaction between context type and SOA [$F(2, 96) = 3.367, p < 0.01$] before response bias correction that reflects an increasing contextual facilitation for the context consistent condition with the earlier context displays. However, this is not sufficient to overcome the inherent performance reduction with the -104msec and -52msec SOAs relative to the simultaneous condition. Post hoc analysis was carried out for each SOA condition individually using a further mixed design ANOVA (2 x 2 x 6) to assess whether the contextual facilitation was present throughout. This revealed significant effects for -104msec [$F(1, 48) = 13.638, p <$

0.01] and -52msec [$F(1, 48) = 8.707$, $p < 0.01$], but not for 0msec [$F(1, 48) = 1.373$, $p > 0.1$]. No scene context effect was present when context and target were displayed simultaneously, even without response bias corrections.

There was a main effect of the between-subjects controlling variable for order [$F(1, 48) = 17.535, p < 0.01$] indicating that when items A and B of the contextual sets were viewed in the first half, and C and D were viewed in the second half, performance was worse than when the order was reversed. There was no main effect of session set [$F(5, 48) = 1.898, p > 0.1$] or interaction between order and session set [$F(5, 48) < 1.0$]. A significant interaction was found between SOA and session set [$F(10, 96) = 2.251, p < 0.05$] which suggests that objects were not equally affected by the temporal manipulation, though most did produce increased errors in the -104 msec condition. These differences across objects were also evident in a three-way interaction between context, SOA and set [$F(10, 96) = 2.575, p < 0.01$].

Error Rates: (After response bias correction)

Table 15 displays the mean error rates across context conditions and SOAs after the response bias correction. As no context effect was found in the simultaneous condition prior to a response bias correction, the correction was carried out but it was not deemed appropriate to include this condition in further significance analyses. However, all SOA conditions are used to calculate the contextual difference (inconsistent errors minus consistent errors) per participant in order to produce a measure of scene context effect suitable for graphical representation (see Figure 22).

Table 15. Mean Error Rates (%) in Each Condition After Response Bias Correction

	Context SOA		
Context type	-104msec	-52msec	0msec
Consistent	42.71 (2.4)	31.84 (2.5)	27.[illegible]6 (2.7)
Inconsistent	47.80 (2.3)	36.15 (2.4)	29.[illegible]5 (2.[illegible])

These error rates were subjected to a repeated measures ANOVA with factors of context type and SOA (only -104msec and -52msec), controlled by between-subjects variables of session set and order. Importantly, a main effect of context type was found [$F(1, 48) = 8.407, p < 0.01$] indicating significant contextual facilitation and lower error rates in the context consistent condition after the removal of response bias. There was also a main effect of SOA [$F(1, 48) = 58.339, p < 0.01$] indicating significantly improved performance in the -52msec condition than in the -104msec condition. However, there was no significant interaction between context and SOA [$F(2, 96) < 1.0$].

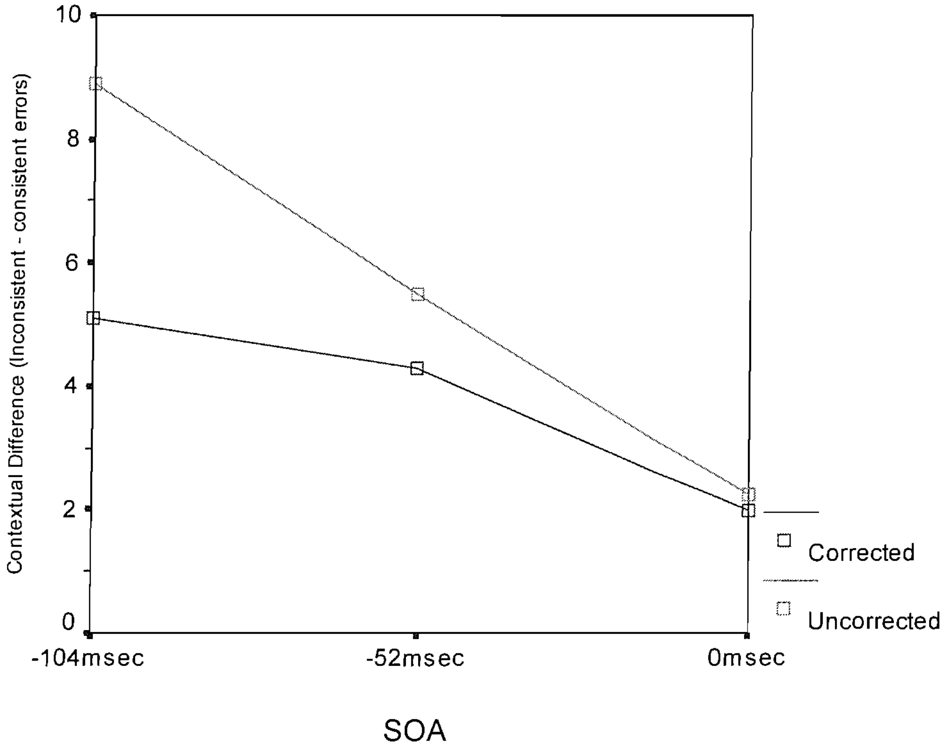


Figure 22: Graph showing contextual difference broken down by SOA before and after response bias correction

A main effect was found of the between-subjects controlling variable of order [$F(1, 48) = 18.377, p < 0.01$] but not session set [$F(5, 48) = 1.826, p > 0.1$]. The order effect might be explained through participants missing out on the practice effect when stimulus sets C and D were received second, if stimulus set D was harder to identify (as suggested in Experiment 3). There was no interaction between order and session set [$F(5, 48) < 1.0$] although there was a significant interaction between

SOA and set [*F*(5, 48) = 2.311, $p < 0.05$], in which session set affected performance only in the -104msec SOA condition. In this condition, session set 3 (mean: 54.1) also scored significantly more errors than set 5 (mean: 31.9) according to a Bonferroni comparison ($p < 0.05$). An interaction was found between context and set [*F*(5, 48) = 3.149, $p < 0.05$], with set almost gaining significance in the inconsistent condition [*F*(5, 48) = 2.330, $p = 0.057$]. Finally, there was a three-way interaction between context, SOA and set [*F*(5, 48) = 5.072, $p < 0.01$]. In the inconsistent condition, all sets showed a standard pattern of increasing errors from 0msec, -52msec to -104msec but this was not repeated in sets 2, 5 or 6 in the consistent condition. There were also fewer sets demonstrating a context effect in the 0msec SOA (sets 1, 4 and 6) than in -52msec (set 4) or -104msec (sets 2 and 6).

As in Experiment 2, the contextual difference for each target object between participants was calculated as a measure of the context effect by deducting the number of errors in the consistent context condition from the number of errors in the inconsistent context condition. The result was a positively skewed distribution for the 6AFC paradigm (see Figure 23) similar to that found using the naming task (see Figure 18 – Chapter 3). This distribution highlights that the context effect is an average of this distribution, and that the effect during any one trial may be quite varied due to the spread. A Pearson's correlation analysis was done between the effects generated by the objects from stimulus sets A, B and C in Experiment 2 and Experiment 5, to examine whether object identities were influencing the scene context effect. Stimulus set D could not be included as it was not used in Experiment 2. Despite similar distributions, no significant correlation was found (r = 0.163, $p > 0.1$).

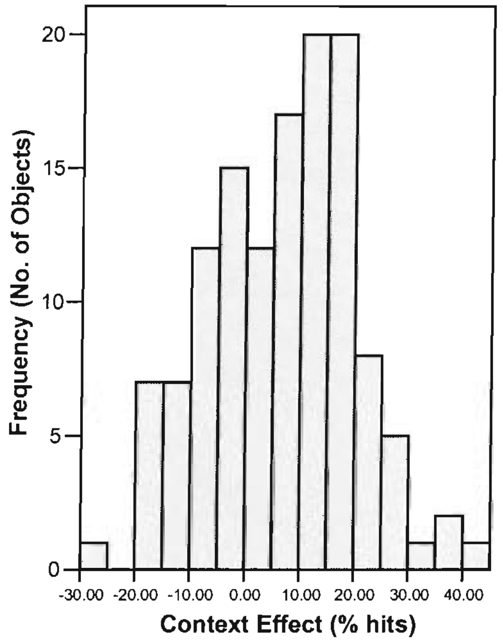


Figure 23: Distribution of target objects by scene context effect

Pearson correlations were conducted with this measure of context effect at an object identity level using the object occurrence and viewpoint familiarity, and array semantic relatedness (consistent context). These were acquired from Experiment 2 ratings studies. Correlations were also conducted with the mean success rates for the target object recognition from this study (see Table 16).

Table 16: Pearson Correlations of Context Effect and Potential Mediators

	Hits	Occ. Familiarity (target/context)	View Familiarity (target/context)	Sem. Relate
Context Effect	-0.088	0.108 / 0.066	-0.089 / 0.031	-0.148
Hits	-	0.099 / -0.123	*0.244* / -0.025	-0.058
Occ. Familiarity	-	-	0.179	-0.007
View Familiarity	-	-	-	0.015

n.b. underlined figures significant to 0.05 : underlined & italicised significant to 0.01 (two-tailed)

Neither familiarity of occurrence or of viewpoint, nor the degree of semantic relatedness for the entire array (assuming consistency) had a significant influence on the context effect in this experiment. The only significant correlations were a relationship between occurrence and viewpoint familiarity of the target (i.e. objects from stimulus sets A, B, C and D), and a strong relationship between viewpoint familiarity and hits that indicates improved performance with greater familiarity of viewpoint.

Error Types:

A separate analysis was conducted on incorrect responses before the response bias correction to examine error type behaviour in conjunction with context type and SOA (see Figure 24). Unlike the naming task in Experiment 2, participants did not have the option to not respond with 6AFC task, and this method automatically classified each error as perceptual, semantic or neither.

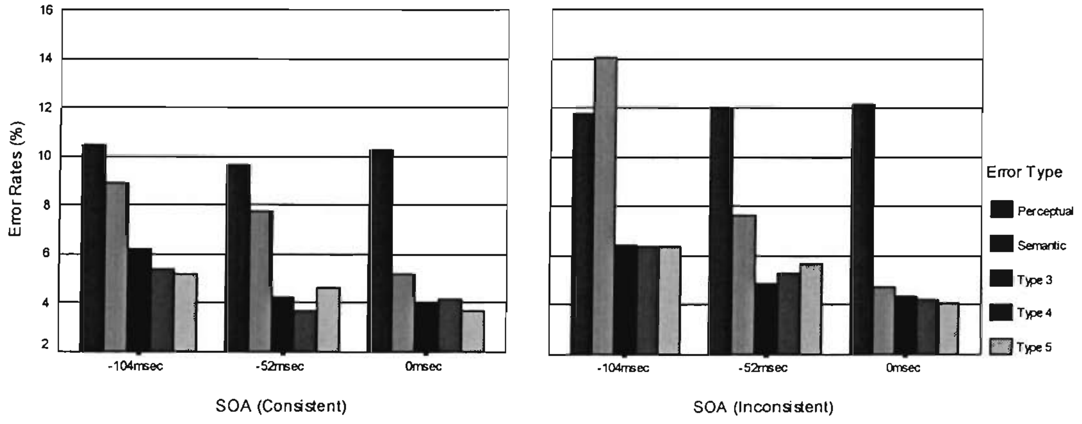


Figure 24: Graphs showing error rates (%) broken down by type, SOA and context

Error type was introduced as a third variable in a repeated measures ANOVA (with context type and SOA), controlled by between-subject variables of order and session set. A main effect of error type was found [$F(4, 192) = 34.304, p < 0.01$], and post-hoc analysis using the Bonferroni method revealed significantly more perceptual errors than any other error type, and more semantic errors than control errors (types 3, 4 and 5 – see example above). Types 3, 4 and 5 did not differ significantly between themselves (see Appendix D).

A significant interaction between error type and SOA [$F(8, 384) = 5.447, p < 0.01$] highlighted the graded increase of semantic errors when non-targets were displayed prior to the target. More of this type of error could be due to increased contextual information, due to more processing time, and/or reduced target perceptual information – an issue addressed in the discussion. There was a trend suggesting an interaction between error type, SOA and context that implied semantic errors were more strongly affected by SOA during the context inconsistent condition [$F(8, 384) = 1.972, p = 0.105$ $^{\text{G-Geisser}}$)]. This interaction was not shown to be significant, but this was partly due to the lack of sphericity in the error condition. There was also a significant interaction between error type and order [$F(4, 192) = 4.174, p < 0.05$]. The lower error scores for order 2 (block 2 then block 1), particularly in semantic errors, may suggest a more focused attentional strategy. However, as only the object identity varied between blocks a strategy shift is unlikely. The spread of errors between participants was wide and a chance distribution of several poorly performing individuals could also have caused this result.

The effect of the response bias corrections on the error type data reduces the semantic error rates, but further exploration of the corrected error type data adds nothing to the overall analysis and is not included.

Discussion:

The 6AFC paradigm used in this experiment has demonstrated that a scene context effect can be generated with an object array using a task that does not require a naming response. Although participants may have utilised sub-vocal naming, the effects of the psycholinguistic system have been reduced relative to standard naming tasks. As in Experiments 1, 2 and 3, fewer errors in identification were made in the context consistent condition than in the context inconsistent condition. The same stimuli and display methods have been used across these experiments, and similar distributions have been attained when the context effect has been split by object identity. However, Experiments 1, 2 and 3 obtained contextual effects when non-target stimuli were displayed simultaneously with the target. In this study the simultaneous display of target and context (SOA = 0msec) failed to produce a significant context effect. This difference may indicate a different type of effect, but it is perhaps more likely that the reduced stimulus exposure time in Experiment 4 limited the time available for generating a scene context effect. The time required to process multiple context objects and generate an effect has not been examined previously and is an issue considered below.

The main effect of contextual facilitation was significant before the elimination of response bias, but importantly it was still significant after the correction had been made. This effect indicates that the scene context effect cannot be completely explained by response bias, and it makes the isolationist perspective difficult to maintain. Contextual objects appeared to influence the perceptual/representational processes directly during this recognition task. This effect supports the claim that context induces an object-semantic driven scene superiority effect. However, response bias does explain a proportion of the scene context effect, and this proportion increases when the context is displayed earlier in this study (see Figure 22). As a result, it would appear that the scene context effect reported in the majority of context studies consists of a combination of two effects. Therefore, contextual influence per se cannot be considered a scene superiority effect, but only if response bias has been measured and controlled for.

It was predicted that presenting the context before the target, without a visual mask, would provide a longer temporal window in which participants could process the non-target items. A larger context effect was expected when the non-targets were displayed earlier. A result matching these predictions was found prior to response bias corrections, but not afterwards, indicating that the longer temporal window aided the bias effect, but not the scene superiority effect. The reason for this pattern may be connected with why semantic error rates were also higher when participants were given more time to process the non-targets, despite the increase in total scene context effect. It is possible that participants were still attending to the context items when the target appeared (Chun & Potter, 1995) or experienced an effect similar to the attentional blink (Shapiro, Raymond & Arnell, 1994). As a result, the perceptual information they would have obtained from the target would have been reduced, potentially forcing them to rely on alternative sources (e.g. the context). This account would not explain why the semantic error rate increased between the –52msec and –104msec conditions, but these differences may be due to an increased difficulty in disengaging relative to the degree of focus or the level of non-target processing. Both of these factors would increase with time.

The absence of a significant scene context effect during the simultaneous presentation of target and context in this experiment could be due to noisy data and therefore should be replicated to be confirmed. That is the aim of Experiment 5. However, it would appear that although object recognition operates effectively at 52msec, it takes longer than this to fully process and integrate the object-semantic factors with a scene context reliably.

Experiment 5:

The primary purpose of Experiment 5 was to explore whether a scene context effect could be generated using a simultaneous presentation of context/target stimuli (0msec SOA), a 52msec display time and a 6AFC response task. The absence of a context effect with such a brief temporal window would replicate the results of Experiment 4 and provide confirmation that time does mediate object-semantic context effects. Experiment 3 also raised issues regarding the positional and layout uncertainty of contextual items in the design, positing that this may have been a cause for the weakening of the context effect. Therefore, a secondary aim was to examine several outstanding questions regarding such uncertainty through the use of blocking and using the 6AFC paradigm rather than a naming task.

Method:

Participants:

Forty-eight post and undergraduates participated in one 40-min session for £5.00 in compensation. All participants reported having either normal or corrected-to-normal visual acuity, and English as a first language. None had participated in any of the previous experiments. There were 18 males and 30 females between the ages of 20 and 46 years (mean 26.33).

Apparatus and Stimuli:

The experiment was conducted on the same equipment as used in previous experiments, with 160 stimuli in 32 context groups of 5 contextually related objects. Context groups 33 and 34 used in Experiment 4 were discarded for balancing purposes. Contextual pairings were maintained between context groups and four objects from each context group were used to form stimulus sets A, B, C and D (as in Experiments 3 and 4). Each of the 128 targets was used in a consistent and inconsistent context stimulus with one, two, three and four context items to generate 1024 23cm x 23cm image files.

The same pattern mask and exogenous attentional cue shown to be effective in Experiments 3 and 4 were used.

Procedure:

The procedure was similar to that in Experiment 4 except that the experiment was broken down into four blocks, each containing 16 practice trials followed by 32 experimental trials. Each block contained trials of a single set size of contextual items (1, 2, 3 or 4) and the order in which participants received the blocks was randomised by the program. The trial structure and 6AFC response were the same as the simultaneous (SOA = 0msec) condition in Experiment 4.

Design:

Each participant performed 128 trials, half of which were context consistent and half context inconsistent. For each set of four target objects from the same context group, two were always seen in the context consistent condition. Rather than utilising a mixed presentation of set sizes and randomly distributed non-targets, this experiment blocked set sizes and displayed non-targets in the same positions for each trial within those blocks. Each block contained one target from each context group with a set number of non-targets (1, 2, 3 or 4), and each participant received all four set sizes over four blocks. Context type was randomised between trials but was dependent upon stimulus set, with half of the trials in each block consistent and half inconsistent. The order variable ensured that the assignment of context type to different stimulus sets was counter-balanced between participants with order 1 presenting stimulus sets A and B as consistent and stimulus sets C and D as inconsistent, and order 2 reversing that presentation. The positional layout of the contextual items within a block was constant, but was counter-balanced between-participants in the layout variable.

Every target had an equal chance of being seen in each context type, with each number of context items across all participants, with the positional layout variations also evenly distributed. This resulted in a mixed design with two within-participant variables (2 x 4) and two between-participant variables (2 x 4).

Results:

The overall error rate was 17.3 %, which was lower than Experiments 2, 3 and 4. Table 17 displays the mean error rates and standard errors for each set size across both contextual types. These suggest no effect of a context consistency advantage except in set size 1, with a slight context inconsistency advantage apparent in set sizes 2 and 4.

These error rates were subjected to a repeated measures ANOVA with factors of context type and set size, controlled by between-subjects variables of layout and order. No main effect was found for context type [$F(1, 40) < 1.0$] or for set size

[F(3, 120) = 2.391, $p > 0.05$], and there was no significant interaction between context and set size [F(3, 120) = 1.045, ns]. There were also no main effect for layout [F(3, 40) = 1.120, $p > 0.1$] or for order [F(1, 40) < 1.0]. There was a significant interaction between context and order [F(1, 40) = 85.270, $p < 0.01$], and a significant four-way interaction between all variables [F(9, 120) = 2.441, $p < 0.05$] which indicates fewer errors when stimulus sets A and B were used and more for sets C and D regardless of which context type they were used to represent (see Table 18).

Table 17: Mean Error Rates (%) of Participants for Each Set Size by Context Type

		Set Size		
Context Type	1	2	3	4
Consistent	13.93 (2.4)	18.62 (2.6)	18.36 (2.4)	17.71 (2.1)
Inconsistent	16.67 (2.5)	17.71 (2.6)	18.36 (2.5)	17.06 (2.1)

Table 18: Mean Error Rates (%) Broken Down Between All Conditions

					Context				
			Consistent				Inconsistent		
Layout	Order	1	2	3	4	1	2	3	4
1	1	0.0	14.6	17.7	6.3	15.6	17.7	16.7	21.9
	2	13.5	22.9	17.7	22.9	11.5	13.5	19.8	11.5
2	1	22.9	14.6	19.8	14.6	26.0	20.8	30.2	13.5
	2	22.9	32.3	32.3	34.4	20.8	25.0	21.9	26.0
3	1	12.5	14.6	13.5	15.6	20.8	34.4	17.7	19.8
	2	20.8	21.9	16.7	20.8	12.5	8.3	7.3	16.7
4	1	5.2	9.4	11.5	8.3	19.8	12.5	20.8	16.7
	2	13.5	18.8	17.7	18.8	6.3	9.4	12.5	10.4

Discussion:

No significant main effect of context type was found in this experiment. This null result replicates the findings from Experiment 4 in which the target and context were presented simultaneously (SOA = 0msec) for 52msec. Such replication indicates that the previous finding was not due to chance and that a temporal window of more than 52msec is required to generate contextual facilitation in a recognition task. This window would include time for processing of the context items and for the generation of the effect itself (e.g. integrating information or activating a schema). The absence of a context effect in these studies does not demonstrate that the object-semantic factors of a scene context must be displayed for more than 52msec prior to target presentation, but suggests that they do need to be displayed in order to allow processing and effect generation prior to target matching ('contextualising'). It appears from the generally high level of performance in this task that contextualising non-target information requires more time than target processing, though how much of this is due to the processing of context items cannot be ascertained from these results. This new knowledge will have implications on the role that object-semantic factors and scene-based context can play during recognition.

The total error rates were lower in this study than in Experiment 4 despite using a similar paradigm and identical stimuli. This may have been due to a more motivated participant pool receiving payment rather than course credits. Alternatively it could be due to the blocking of the set size condition, removing some of the uncertainty from the task. Such lower error rates may have weakened a context effect but is unlikely to have eliminated it completely, as Experiment 1 demonstrated an effect can be achieved with error rates below 10%.

General Discussion:

Experiment 4 makes an important distinction regarding the nature of the scene context effect. By eliminating the response bias and maintaining a significant contextual facilitation, the 6AFC paradigm demonstrates that the consistency advantage cannot be explained by the Functional Isolation hypothesis (Hollingworth & Henderson, 1998, 1999). The significant effect after correction suggests an integration of perceptual and contextual information, therefore supporting some versions of the interactionist view. Models of object recognition will have to include mechanisms for using contextual information to resolve perceptual ambiguity, just as participants have done in this task. However, it is important that having demonstrated that a specific manifestation of a scene context can influence perceptual/representational processes, it is not assumed that all scene context effects reflect only perceptual effects.

Although Biederman (1981), Hollingworth and Henderson (1998, 1999) and to a lesser degree Davenport and Potter (2004) acknowledge that response bias plays a role in context, none have offered an integrated perspective that combines a scene superiority effect and a bias effect. The difference in scene context effect magnitude before and after the response bias correction in Experiment 4 indicates that response bias does account for a proportion of the normally observed context effect. Results indicate that this proportion increases when perceptual information regarding the target decreases, and consistent contextual information remains constant or improves (due to longer processing time). Given that the object-driven scene superiority effect will result from some form of interaction between contextual and perceptual data, the bias effect may result from the same mechanism(s) in which the perceptual data is weak or absent. A lack of perceptual data for the target would leave decision making dependent upon contextual information, and as such any item related to the context is likely to be selected regardless of the target. This would appear identical to a decision based upon response bias. Alternatively, the effect may be due to a different mechanism. These issues require further investigation.

Whilst the response bias effect will not play a direct role in the relationship between scene contexts and object recognition models, it must not be ignored. It is likely to play a major role in 'real-world' situations, especially when display time is less brief. Therefore a coherent model of contextual influence will eventually need to combine both the superiority and response bias effects.

The studies reported in this chapter demonstrate that a scene context requires a temporal window longer than 52msec to generate a context effect driven by object-semantic factors. A difference in the time courses between context and target processing may be due to different processing mechanisms (e.g. perhaps analytic and holistic). In addition, it seems likely that contextual information would be formed from multiple context items, whereas target processing focuses on a single item, although Experiment 3 was unable to demonstrate this conclusively. Finally, contextual processing in an array must not only identify the non-target items, it must also integrate the extracted data into the target processing mechanism.

The presence of a significant context effect in Experiment 4 during the conditions in which the context was displayed before the target could be taken as further evidence that the temporal window for context effect generation can be extended by the use of iconic memory. This use of iconic memory was initially indicated by results from Experiment 1. However, this result could also be explained if the integration process or schema activation took time to implement. Contextual item processing may be complete within the 52msec but additional time is required to effectively utilise the information/activation. An increase in context effect when a longer temporal window was provided to the context was also found in Experiment 4 (earlier SOAs). However, that increase may be due to a decrease in perceptual information as a result of an attentional blink effect, the extraction of more contextual information, or a combination of the two. Across all experiments, scene context effects have been found that were robust at 78msec (Experiments 1 and 2), a trend at 65msec (Experiment 3) and non-significant at 52msec for simultaneous displays (Experiments 4 and 5). This supports the view that a longer temporal window may benefit contextual facilitation, but the robust effects in Experiments 1 and 2 relative to Experiment 3 could be a result of the advantages of an endogenous cue and the secondary task used (see Chapter 4). Further investigation is required to explore whether a longer temporal window and increased processing time for the context can strengthen the contextual effect.

The results reported here demonstrate that the scene context effect is made up of two effects. One of these is a scene superiority effect, which results from the direct influence of contextual information upon the perceptual/ representational processes during target recognition. This finding highlights the role of the object-semantic factors within scene context, and how they will need to be considered in models of object recognition that go beyond isolated targets. The other effect is a response bias. These experiments have also shown that time mediates scene context effects, with an effect only generated when the temporal window was greater than 52msec. However, this result does not mean that context items cannot be processed completely within that period. These novel findings, and those from earlier chapters, provide the foundations for future research in context, and for conjecture into the mechanisms at work in contextual processing (see Chapter 6).

Chapter 6: General Discussion

The primary aims of this thesis were to determine whether scene context effects could be generated using only multiple objects; whether contextual information could directly influence the perceptual/representational processes used during target recognition for objects in scenes; and whether visual attention played a role in generating scene context effects. Experiments 1 and 2 examined the role of attention in generating scene context effects; Experiment 3 examined the number of objects required to generate scene context effects; and Experiments 4 and 5 examined whether scene context effects remain significant when a forced choice procedure allows the discrimination of perceptual/representational factors from response biases that might affect naming.

Across all the experiments the stimulus displays used were object arrays rather than naturalistic scenes, as used in most other studies on this issue (e.g. Biederman, 1981; Palmer, 1975; Davenport & Potter, 2004). Naturalistic scenes are immensely complex, being constructed of objects in functional spatial relationships and backgrounds, and previous experiments had found it difficult to determine whether context effects in recognition were generated merely by the presence of the context objects, or by the specific configurations in which they appeared. The position and support relationships found in naturalistic scenes may enhance context effects, but the experiments presented here show that scene-configuration factors are not necessary to generate a scene context effect; all that is necessary is the presence of related visual objects. Arrays and naturalistic scenes can therefore be considered specific cases of a broader 'scene' classification.

Biederman (1981) and Bar (2004) support the view that scene contextual effects within naturalistic scenes can be activated through the recognition of key objects within the scene. Such context effects do appear to be formed of object-semantic factors and scene-configuration factors, and it would be surprising if the object-semantic factors driving the context effects demonstrated here did not also appear when the objects are part of a coherent scene. It is left to future experiments to determine the degree to which this contextual facilitation is shaped and changed by the scene-configuration factors generated by object relationships.

Scene Context Effects: Superiority Effects or Response Bias?

Chapter 1 highlighted that current models of object recognition do not provide a satisfactory account of context consistent facilitation on recognition. The majority of these models have not been constructed to process multiple stimuli (or scenes) or to take advantage of the contextual relationships between objects. The Functional Isolation hypothesis (Hollingworth & Henderson, 1998, 1999) provided object recognition modellers with a theoretical reason to ignore the existence of scene context effects when considering object recognition. It was based on the claim that context was processed independently of the target, and that the scene context effect was entirely due to response bias, thus allowing the perceptual/representational processes within the recognition system to be separated from contextual information. However, as described in Chapter 1, the empirical evidence supporting the Functional Isolation hypothesis leaves room for some types of contextual facilitation, and Experiment 4 showed conclusively that a scene superiority effect in recognition can be generated by contextual information.

Scene context effects were shown to consist of two separate effects. A novel 6AFC response paradigm was developed for use with the object array, to allow response bias to be calculated and controlled. Experiment 4 demonstrated a robust main effect of context before the response bias corrections, and a reduced (although still significant) scene context effect after the correction was applied. These findings indicated that, contrary to the Functional Isolation hypothesis (Hollingworth & Henderson, 1998, 1999), contextual facilitation on recognition cannot be explained by response bias. The effect remaining after bias has been controlled for is a scene superiority effect that results from contextual information directly influencing the perceptual/representational processes of target recognition. Providing evidence of such a superiority effect within object arrays is an important step in context/ recognition research, and meets the aim of this thesis. It also highlights that a complete model of object recognition will need to take account of object-semantic contextual influences within scenes as well as target processing.

Response bias did account for part of the total context effect (before correction) in Experiment 4. Response bias effects do not interact with perceptual processes as they are guesses based upon contextual information. The total scene context effect was formed from both the response bias and superiority effects.

The addition of a superiority effect and a response bias effect to form a total scene context effect, suggests that previous research on contextual effects in scenes has been reporting the influence of two effects rather than one. Unless response bias is controlled for using a method based upon the concepts of the Reicher-Wheeler paradigm, the impact of each effect on the detection and perception of objects in scenes cannot be isolated. Future models of contextual processing will also need to address this dual-effect to provide a coherent perspective. Further investigation will be necessary to explore the specific mechanisms which generate the effects themselves.

Is Visual Attention Required for Contextual Processing Within Scenes?

The allocation of visual attention to context items was shown to be necessary to generate an object-semantic driven context effect within a scene. Previous research has provided empirical evidence that the absence or presence of visual attention can directly influence the mechanism used in object recognition (e.g. Thoma, Hummel, & Davidoff, 2004). It has also been suggested that configurational context can be used to direct visual attention during visual search (Chun & Jiang, 1998, 1999; but see Chapter 2). However, within the domain of research on scene perception, the effect of visual attention on the context effect had been neither previously investigated nor controlled for.

Experiments 1 and 2 used a novel endogenous cueing paradigm to manipulate participant attention so that it was either narrowly focused on the target object, or widely spread across the entire array. Experiment 2 found fewest errors during a naming task when attention was narrowly focused on the target object and there were no context items. When a complete array was presented, accuracy was greatest when attention was focused on the wide visual area, and the context items were semantically related to the target. The superior performance of the narrow focus, no context condition indicates that the reduction of visual attention focused directly upon the target does reduce accuracy. This deterioration occurs regardless of whether the dispersal of attention is voluntary (e.g. wide focus condition) or involuntary (e.g. sudden onset of a non-target). Contextual facilitation partially offsets this decrease in performance when context items are allocated visual attention, and are semantically consistent with the target. However, these studies showed visual attention needs to be allocated prior to stimulus onset via the cue, as little or no effect was found in the narrow focus conditions when non-targets captured attention through onset.

A novel exogenous cue was used to ensure the allocation of visual attention to the entire object array in Experiments 3, 4 and 5. When target and context stimuli were presented simultaneously, the scene context effects using the exogenous cue were weak (Experiment 3) or non-existent (Experiments 4 and 5). Endogenous cues have been shown to produce longer lasting attentional effects than exogenous cues (Müller & Rabbitt, 1989; Yantis, 2000). These differences in results may be due to visual attention having been directed for a longer time at the context items by the endogenous cue than by the exogenous cue. The reported scene context effects in Experiment 4 were in a condition in which the context was presented before the target, so that it could more easily capture attention. Alternatively, the secondary task presented after the endogenous cue may have interfered with the cognitive filtering process, and increased the influence of the non-targets (Lavie et al, 2004). The types of cue do not provide the only explanation to the differences in contextual facilitation between the first two experiments and the later three (see below), but they do need to be considered.

The necessity for considering visual attention in the generation of the scene context effect is confirmed by the results reported here, but the specific role that it plays within the contextual system cannot be ascertained by these experiments. It may be that attention is simply required for complex object processing; it may be that attention provides item binding and prevention of cross-talk between multiple context items; or it may be that attention allows the integration of context information. This issue remains a valid area for further study.

The Influence of Time on Contextual Processing:

In Experiments 4 and 5, no scene context effect (combined superiority and bias effect) was found when context items and target were displayed simultaneously, although overall performance at the 6AFC task was good. In these experiments, the contexts and targets were always shown for 52msecs. In one condition of Experiment 4 and Experiment 5, contexts and targets were shown simultaneously displayed. In the other conditions of Experiment 4, the presentation of contexts preceded that of targets by a blank SOA of 0 or 52msecs (put another way, the onsets between contexts and targets were -104, -52 or 0msecs). A mask was always shown at target offset (see Figure 25).

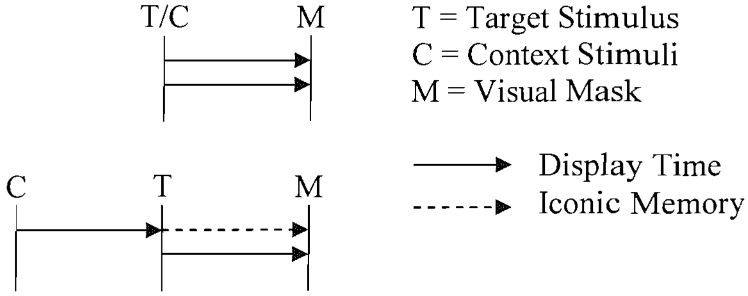


Figure 25: Temporal windows of contextual processing (C to M) in simultaneous and non-simultaneous target/context presentations

The experiments showed no evidence of a significant scene context effect for simultaneous presentations of targets and contexts, an intermediate effect when the context was presented 52msec before the targets, and a larger context effect when contexts were presented 104msec before targets. This display time of 52msec provides an estimate of the minimum time required to generate a scene context effect in the 6AFC paradigm. This does not demonstrate that contextual processing was not completed within this time period. Information from fully processed contextual items may not have been integrated, or contextual schemas may not have been activated, prior to the mask onset within the 52msec window. Alternatively the temporal window for contextual processing may have been extended using iconic memory whilst the target was displayed in the non-simultaneous conditions (see Figure 25).

A comparison across experiments suggests that time may influence the strength of the context effect. Experiment 2 used a display time of 78msec whereas Experiment 3 used 65msec. The scene context effect found in Experiment 2 was strongly significant whereas the context effect was only a strong trend towards significance in Experiment 3. However, the problem in stating that the difference in display time was the factor distinguishing the robustness of effects in Experiments 2 and 3 is that the comparison is compromised by the use of different attentional cues. An endogenous cue was used in Experiment 2 and an exogenous cue was used in Experiment 3.

Does Scene Context Influence Errors?

In Experiment 2, a naming task was used. Errors were classified (using two raters with an inter-rater reliability of kappa = 0.714) into perceptual (similar to the target), semantic (related to the context), perceptual-semantic, background (a context item), background-perceptual (perceptually similar to a context item) and non-response. Non-response was the most common error type recorded under all conditions; however more non-response errors were found with inconsistent contexts.

In examining errors made to consistent and inconsistent contexts, two comparisons are worthy of note. First, more non-responses were made to inconsistent than consistent contexts. Second more perceptual-semantic and background errors were made to consistent than inconsistent contexts. In other words, consistent contexts generated more responses (with fewer errors) but with errors being made to objects related to the targets both perceptually and semantically.

In Experiment 4 (and 5, although these data are not reported), the 6AFC paradigm allowed perceptual and semantic errors to be automatically recorded, and non-responses were not permitted. Context consistency did not influence error type; however the SOA of context presentation did affect semantic errors. Although perceptual errors remained constant across conditions, semantic errors increased significantly the earlier the context was displayed. This increase in semantic error rates mirrored an overall decrease in accuracy.

The results of Experiments 2 and 4 can be considered together. They seem to show that consistent contexts lead to more accurate naming and recognition, but when errors are made, they are made to objects semantically or perceptuo-semantically related to targets.

Other Issues:

Familiarity, Semantic Relatedness and Object Identity:

There was no evidence that familiarity, the magnitude of semantic relatedness within a consistent condition, or object identity influenced the scene context effect. Ratings of familiarity of occurrence measured how often an object was seen in an individual's daily life, and familiarity of viewpoint measured how close the stimulus image was to a typical view of an object. Ratings were attained both for targets and for contexts (mean rating of the remaining four objects). No correlations were found with the contextual difference (consistent hits – inconsistent hits) in either the naming paradigm (Experiment 2) or the 6AFC paradigm (Experiment 4). Familiarity of viewpoint for the target did show a significant positive correlation with overall performance (hits) in Experiment 4. This relationship suggests that the familiarity of viewpoint measure detected a factor capable of influencing perceptual processing that did not affect contextual processing. However, there is insufficient basis to draw conclusions from the data.

Results from Experiment 2 and Experiment 4 indicate that the magnitude of semantic relatedness within the contextual group, providing the group is context consistent, does not influence the magnitude of the scene context effect. One explanation is that the mean relatedness in the context groups is not widely distributed enough, and is sufficiently high, to conceal any such influence. Alternatively, if a scene schema is activated and utilises a spreading activation network (e.g. Kosslyn, 1994; Rumelhart & McClelland, 1981), the activation provided to linked objects might be of similar amounts. Such an activation distribution might be expected from the activation of a schema (Biederman, 1981) or context frame (Bar, 2004). This pattern might not be expected were objects linked horizontally, by pathways of varying semantic strengths, capable of transferring different levels of activation. However, whilst good indicators for possible explanations, further study is required to examine these hypotheses based on correlation data. There was also no correlation found when comparing contextual difference sorted by object identity between Experiments 2 and 4.

Naming and Non-Naming Paradigms:

Context effects were demonstrated with object arrays using both naming (Experiment 1, 2, and 3) and non-naming (Experiment 4) response paradigms. This finding indicates that scene context effects do not require linguistic processing, and the absence of the psycholinguistic processes from visual contextual facilitation would be consistent with previous naming research. Such studies (e.g. Bloem & La Heij, 2003) report only semantic interference from contextual images during naming.

The Stimulus Sets:

The total stimulus set (32-34 contextual groups) was relatively small, and to avoid repetition, participants saw each target only once as a target. However, a relatively large number of participants were used in each experiment. Stimulus sets A, B, and C were equivalently rated for familiarity, but stimulus set D (used in Experiments 3, 4 and 5) was rated of lower familiarity than sets A, B and C (though not significantly). However, this difference did raise the error rate in the experiments in which stimulus set D was used. The different stimulus sets were counterbalanced across the other factors so that the effect of stimulus set could not be misinterpreted as some other effect. Nonetheless, subtle effects which may have been significant (e.g. interaction between set size and context in Experiment 3) may have been lost. Set D was used in Experiments 4 and 5, despite the problems highlighted in Experiment 3, as continued use of the same stimulus sets was considered more important to the studies.

How Do We Account for Scene Context Effects?

Several accounts of how scene context effects are generated have been outlined in Chapter 5. Interactionist explanations propose that contextual information directly influences the perceptual/representational processes during target recognition. The Early Description Enhancement hypothesis (Hollingworth & Henderson, 1999) suggests that this interaction is through the activation of a scene schema that then aids the basic perceptual processes (e.g. extraction of edges and features). The Late Description Enhancement hypothesis (Biederman, 1981) and the Criterion Modulation hypothesis (based on Friedman, 1979; Kosslyn, 1994; Palmer, 1975; Rumelhart & McClelland, 1981) maintain that activation from the schema interacts with target recognition during the matching process. The Functional Isolation hypothesis (Hollingworth & Henderson, 1998, 1999) proposes that contextual and target data are processed separately, and context effects can be explained entirely by response bias.

Results from this thesis demonstrate that a scene superiority effect can be generated by contextual objects, thus providing support for the Interactionist perspective theories. However, response bias did account for part of the total scene context effect and there were indications that target and context processing utilised different mechanisms. These latter findings are more closely associated with an Isolationist viewpoint. No previous study has clarified the nature of the two effects (superiority and response bias) which form the total scene context effect and integrated them to form a coherent model.

The role of visual attention has not been previously investigated regarding scene context effects and therefore has not been included in earlier accounts of contextual facilitation. Neither have accounts for contextual effects been made that consider their processes relative to models of object recognition. Evidence from these experiments of contextual influence upon perceptual/representational processes during target recognition, and the use of visual attention to generate scene context effects, validates the exploration of models that include both contextual and target processing.

The Attentional Cascade Model:

Visual attention has been shown to play a significant role in generating scene context effects (Experiment 2), and in influencing the mechanism used by the recognition system (Thoma, Hummel & Davidoff, 2004). Hybrid models of object recognition (e.g. Hummel, 2001) suggest that a target is processed via two pathways. If visual attention is present, then an analytic representation is used, but if attention is absent, a holistic representation can be utilised. Recent research by Shih and Sperling (2005) on visual attention and memory also indicates that stimulus information is processed by two mechanisms. They examined participant performance on implicit and explicit memory tasks in conditions during which an attentional blink (AB) was produced. It was established that the AB was a consequence of working memory (WM) having not finished encoding the initial stimulus before the next stimulus was presented. Working memory utilises visual attention, and thus if a stimulus captures a large proportion of a participant's attention during its processing there would be less available for subsequent items. This description of WM is similar to the mechanism required by structural models of recognition in order to perform serial processing of targets (see RBC theory – Biederman, 1987). Shih and Sperling (2005) also found that stimuli activated long-term memory (LTM) automatically, without accessing WM. These were unaffected by the AB effect. This second route suggests a rapid, low-resource activation pathway suitable for the contextual processing part of the recognition system.

Shih and Sperling (2005) have used their findings to contribute towards a cascade model of attention and memory. This model links two areas of cognitive psychology that have not been theoretically closely connected. As stated in Chapter 1, attention has not been extensively explored with regards to recognition and the dual processing routes from Shih and Sperling's model can be modified appropriately (see Figure 26).

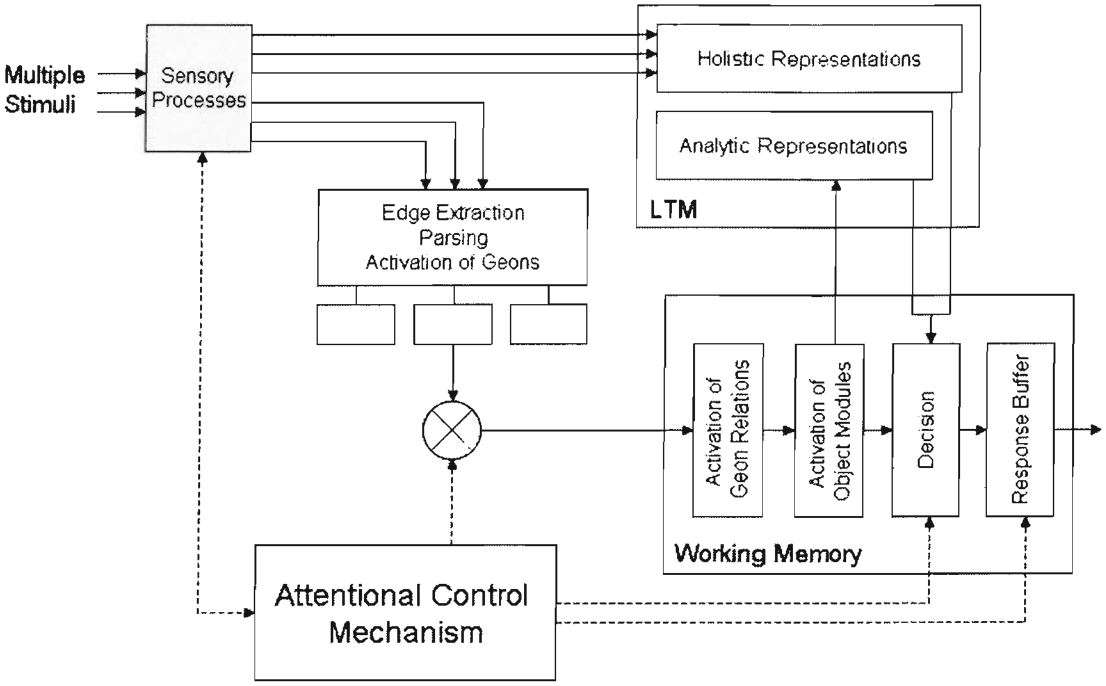


Figure 26: An adaptation of Shih and Sperling's (2005) cascade model of memory and attention to model recognition and attention.

The attentional control mechanism sets the initial spread of attention across the visual field, and controls the selective focusing onto a target. In the experiments within this thesis, the initial spread of attention has been manipulated using attentional cues. Visual attention is not required, or the task requires sufficiently low levels of resources, that the early stages of object processing (edge extraction, parsing, activation of geons) can be achieved prior to target selection. Only those structural processing tasks utilising working memory (e.g. activation of geon relations) require focused visual attention. These processes access the analytic representations in LTM. Experiment 2 found that visual attention was necessary to generate a significant scene context effect. The model reported above indicates that stimuli (target and non-targets) are processed simultaneously via a holistic pathway.

Such a route would not utilise visual attention to process stimuli, but may require it to integrate the extracted information into the decision area within WM. These processing mechanisms would work simultaneously, with contextual data using holistic representations and target data using primarily analytic representations, which could be integrated during the decision process. Such a model could produce both scene superiority and response bias effects from a single system. The absence of sufficient perceptual target data via the WM pathway would place more dependency upon contextual data, as suggested by Experiment 4. The broad activation of all objects linked to that context would result in a response bias effect.

Although the adapted cascade model was developed to account for contextual processing, it also fits data for single object recognition based on the findings of Stankiewicz, Hummel and Cooper (1998) and Thoma, Hummel and Davidoff (2004). The allocation of visual attention to the single stimulus would potentially allow both processing pathways to be utilised. Importantly, the WM pathway would enable the structural processing demonstrated in left-right reflection (Stankiewicz et al. 1998) and split image (Thoma et al, 2004) priming. If the single stimulus was unattended, the WM pathway would not be available and structural processing could not be used. The target would therefore have to rely only on the information provided via the holistic processing pathway.

A Neuropsychological Model:

Moshe Bar (2004) put forward a neuropsychological explanation of scene context effects that also utilised multiple routes of processing (see Figure 27). The target recognition part of Bar's model is similar to an object recognition mechanism he put forward in 2003 (Bar, 2003). A low spatial frequency representation of the target stimulus is projected along magnocellular pathways to the prefrontal cortex (PFC) from the early visual areas (V2 and V4). The transfer is rapid, but provides only a coarse image. This image allows generic object recognition (categorisation) which reduces the number of possible stored representations for the target (initial guesses). The reduced pool of options is then projected along further low spatial frequency paths to the inferior temporal cortex (IT). It is in this area, when all the information has been acquired, that an identification decision will be made. Detailed target perceptual data is projected directly from V2 and V4 to IT along high spatial frequency pathways. This route allows more accurate identification but is slower than the low spatial pathways, and thus processing along this direct pathway may benefit from top-down facilitation from completed target categorisation.

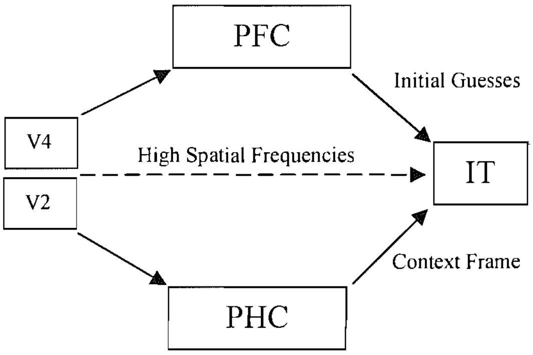


Figure 27: Schematic illustration of Bar's (2004) Model for Contextual Facilitation

A low spatial frequency image of the scene is also projected from the visual areas to the parahippocampal cortex (PHC) and retrosplinal cortex (RSC), though these do not appear to be as direct as the pathway to the PFC. There is evidence that it receives both visuo-spatial (posterior parietal cortex in the dorsal stream) and visual shape input (area TE and perirhinal cortex). The PHC, the primary visual area of scene context, uses the resultant information to identify the most likely 'context frames'. It does this by extracting a global scene from the coarse image or by selecting key objects from within it. These two routes to schema activation reflect the object and scene based pathways identified by Biederman (1981). Context frames are used to generate a sensitised set of possible objects, and to provide information regarding those objects probable spatial arrangements. Low spatial frequency routes are then used to project these sensitised sets into the inferior temporal cortex (IT). The interaction between the sensitised set provided by the context, and the limited pool of options provided by the PFC categorisation of the target, may provide a single recognition decision before the high spatial frequency target information reaches IT.

Bar's (2004) model claims to be an account of naturalistic scene context effects, although the PHC is described as a 'multiplexer of associations' for objects and there is no obvious way that these associated objects are categorised into context frames. Nor does it demonstrate how global images would be identified as specific contexts. It highlights again that principles utilised in a model for naturalistic scenes can be equally applied to object arrays. Context frames/schemas are used to generate the context effect, rather than direct semantic links between objects, but their even activation pattern between context items would reflect no effect of magnitude in semantic relatedness from consistent contexts (Experiments 2 and 4).

Experiments 4 and 5 demonstrated that reliable target processing could occur in a shorter temporal window than reliable context effect generation. This model provides two potential explanations for this finding. Although scene context effect generation uses the rapid, low spatial frequency pathways, the route from the visual areas to the PHC and RSC is less direct than from V2/V4 to the PFC. Information is transferred from more brain areas, suggesting a more complex integration process (including the activation of a context frame). Such complexity may require the longer temporal window. In addition, the feed-forward mechanism utilised to categorise target data provides participants with a best guess prior to either context information, or high spatial frequency target perceptual data. In the experiments reported in this thesis, a categorisation (e.g. shoe, apple, mug) would have been sufficient in the majority of trials. Thus, brief display times may have forced participants to use only this mechanism. The absence of a scene context effect during these display times would be due to no sensitised set of options from the PHC having been projected to the IT when a decision was made.

Perceptual target data, from two pathways, and contextual data are integrated in the IT within this model. As in the attentional cascade model (above), the presence of both forms of information has the potential to generate both scene superiority and response bias effects (see Experiment 4). A presence of contextual data without perceptual target data would generate only response bias effects. The proportions of these effects and the precise conditions required to manipulate them require further investigation.

Bar's (2004) model offers a neuropsychological framework for contextual processing. His assertion that low spatial frequencies are processed first is not universally held. Oliva and Schyns (1997) have studied the effects of frequencies and found that individuals appear to be able to process them simultaneously. This is an area that requires further research.

Are Scene Context Effects a Case of Priming?

Semantic priming studies have typically focused on words not objects, utilised successive rather than simultaneous displays of non-target and target, and employed single non-targets as primes. There are exceptions to each of the previous norms, though none similar to the experiments presented in this thesis. Although four non-target objects were presented in the majority of object arrays, if these are activating a scene schema (Bar, 2004; Biederman, 1981) these cannot be considered to act as four independent primes. For example whilst a hair-clip may be semantically related to a ribbon-bow it would not assist the activation of a 'birthday' scene schema that might also contain a ribbon-bow target. Even if scene schemas are not utilised there is the potential for the spread of activation between semantically related context items (Kosslyn, 1994), which could raise activation levels of the overall scene context, that is not accounted for if context items are considered independently. The target itself may contribute to this in a context consistent condition. Whilst the precise mechanisms of contextual processing within arrays cannot be ascertained from these studies, if the scene context effects found are due to priming, it is a new and specific form.

The existence of an overall coherence within a consistent array makes it both different from independent multiple stimulus sets, and similar to a naturalistic scene. Results from arrays formed of multiple non-targets each semantically related to a target, but unrelated to each other, would not generalise with such validity. As noted earlier, both Biederman (1981) and Bar (2004) state that scene context effects can be generated through the activation of key objects within a scene. The extraction of such objects from within a scene is the process replicated by use of object arrays.

Summary:

More information is required regarding how the scene context effect itself is generated, and about how the multiple non-target items are processed. The two models presented in this chapter are not mutually exclusive; indeed there is considerable overlap. However, they present a basis for developing more detailed explanations of scene-based contextual processing that can benefit from both cognitive and neuropsychological research.

Future research:

The results reported in this thesis have highlighted the role of visual attention and the temporal processing window in scene context effect generation, and have demonstrated that the scene context effect is formed from both a superiority effect and a response bias effect. These findings are important as they expand the knowledge of contextual processing beyond semantic relatedness, and because they integrate it with object recognition and attention. However, the studies presented have raised several issues that can be formed into four distinct lines of further research. The first of these concerns questions relating directly to contextual processing time, and the role of visual attention in scene context and context set size. The second concerns the generalisation of the results from object arrays to naturalistic scenes. The third relates to how the context items are processed and the scene context effect is generated. The fourth is the question of how contextual frameworks are formed.

Influence of Time, Attention and Set Size:

These results suggest that the strength of the context effect may be affected by the length of processing time allotted to the context stimulus, because the response bias effect was stronger when the context items were displayed earlier before the target in Experiment 4, and the most robust observed context effects in Experiments 1 and 2 when the longest display times (78msec) were used. However, the increased bias effect could have been due to an attentional blink effect that reduced perceptual data, and the strong effect in Experiment 2 may have been a result of the longer cueing period of the endogenous attentional cue relative to the endogenous cue (Müller & Rabbitt, 1989; Yantis, 2000) used in Experiment 3, and the cognitive interference from the secondary task (Lavie et al. 2004). A single study that manipulated the temporal window for context effect generation (equivalent to display time for this study) based upon the experimental design used within this thesis would allow these issues to be addressed. It would utilise an exogenous cue to ensure visual attention is spread across the visual field, an object array (a target and four non-targets), and a 6AFC response paradigm. Context stimuli would be either consistent or inconsistent with the target. Three levels of context display time would be used (65msec, 78msec and 91msec) with context and target onset simultaneous to prevent attentional blink effects. The target would be displayed for 65msec in every condition. To avoid ceiling effects in performance at this display time, the target would be degraded, and a visual mask would replace both the target and context stimuli at their offset. An alternative would be to pilot target only trials to establish the degradation required to achieve similar levels of performance at each display time. However, my findings indicate that the nature of the scene context effect may be dependent upon target perceptual data (Experiment 4), thus it would be preferable to maintain a constant target perceptual input to each condition.

It would be expected that if the scene context effect was influenced by the longer temporal window allotted to the contextual stimulus then the size of the effect would increase as non-target display time was raised. If a main effect of display time is detected, the 6AFC response paradigm will reveal whether processing time has more of an influence upon the scene superiority effect or the response bias effect. The 65msec condition would also provide partial replication for Experiment 3. This study reported a weak effect, but was the only experiment to generate any effect using both an exogenous cue and simultaneous presentation.

Experiment 2 demonstrated that visual attention must be allocated to context items to generate a significant scene context effect. This finding suggests that attention may need to remain allocated to context items for a minimum period of time, perhaps throughout the temporal window of contextual processing. To explore whether visual attention remains widely spread to the context, or is re-allocated to the target, a probe experiment is proposed. The experimental design would be based upon Experiment 2, but would use an exogenous cue to manipulate visual attention (wide and narrow focus) and a 6AFC response paradigm rather than a naming task. The exogenous cue would require modification to cue the narrow visual area, but this would be done with a slower expansion of the circle from the fixation cross. Stimulus display time would be decided based upon the results from the previous experiment (e.g. 65msec). Non-probe trials would therefore examine whether the role of visual attention was unchanged when the cue and response methods were varied. The 6AFC paradigm would also allow a main effect of attention to be examined relative to the scene superiority and response bias effects.

In the probe trials, probes could be presented either at stimulus onset (e.g. 0msec), approximately mid-way between onset and offset (e.g. 26msec), or just before offset (e.g. 52msec) based upon stimulus display time. Probes could also be presented in one of the four context item locations, or at the target location, and would take the form of an arrow pointing up or down. The arrows would be displayed within an object from the array to minimise effects of sudden onset. Participants would be required to report the direction of the arrow as quickly and as accurately as possible on the appearance of a probe, although their primary task would be to identify the target object in the 6AFC task. If the probe performance was unchanged between the SOAs, it would indicate that attention was not varying over time. However, if performance in the context probes decreased over time, and performance in the target probe increased, that would suggest visual attention re-allocated away from the context to focus upon the target object.

A third experiment based directly upon this thesis would be a replication of Experiment 3, to investigate the influence of set size upon the context effect. Although the graph (Figure 19) and re-analysis of data without stimulus set D indicated that three or more array items (including the target) were required to generate a context effect, this conclusion was only suggestive. If a robust scene context effect has been demonstrated with an exogenous cue and a specified context display time (e.g. 65msec) in the first of these three proposed studies, then this experiment can be done with only set size as an unknown. Previously, inequality in stimulus set familiarity generated a high degree of variance, so stimulus objects would be more carefully matched. In addition, more context groups would be generated and more participants would be used. The experimental design would remain unchanged except for the use of the 6AFC paradigm rather than the naming task. Use of this response method provides a consistency across these three studies, and would allow any scene context effect at each set size to be examined at a superiority/response bias level. For the scene context effects found to be more than a basic priming effect, at least two non-targets need to be utilised in generating the facilitation.

Generalisation of Object Arrays to Naturalistic Scene Displays:

Naturalistic scenes do not generalise per se to object arrays due to the lack of inter-object relationships in arrays (Biederman, 1981), but scene context effects are formed from a combination of object-semantic factors and scene-configuration factors. This thesis has isolated the object-semantic factors through the use of arrays, and shown these are sufficient to generate a scene context effect. Demonstrating that these factors are also applicable to naturalistic scenes would further increase the impact of this research. In addition, the use of similar experimental designs may identify where differences occur between array and naturalistic scene recognition, and how the scene-configuration factors affect contextual processing. The research would use five item object arrays that are integrated with a coherent, naturalistic scene instead of a plain white background. One study would seek to replicate the findings of Experiment 2 using the endogenous cue to manipulate visual attention (wide or narrow) and a naming response paradigm. The second study would replicate Experiment 4 using an exogenous cue to spread visual attention across the entire array and a 6AFC response paradigm. However, rather than present the context items before the target, they would be presented simultaneously for three different display times (52msec, 65msec, and 78msec). The target would be presented for 52msec with sufficient degradation to prevent a ceiling effect in performance, and a visual mask immediately replacing it after offset. These display times have direct comparisons with arrays in those experiments proposed in the first section of studies. Results could be examined to confirm whether the scene context principles established through object arrays were maintained in naturalistic scenes. Specific differences could then be explored to indicate whether all naturalistic scenes generate additional facilitating relationships, or if such scene-configuration factors occur only in specific instances. A lack of difference between naturalistic scenes and arrays regarding scene context effects would suggest that most facilitation is due to specific objects rather than their spatial relationships.

Processing of the Contextual Items:

Although the experiments presented in this thesis have illustrated the integrative role of contextual processing with recognition and attention, they do not specifically explore the processing mechanism itself. It is therefore important to establish the level of processing that is conducted upon contextual items within the object arrays. An initial study would aim to determine whether sufficient perceptual information is extracted from the contextual stimuli to allow participants to differentiate which of two images they have viewed previously. In a number of probe trials, participants would be tested upon the non-target objects rather than the target using a 2AFC task. Condition one probes would consist of one of the contextual items from the target/context display presented alongside a perceptually dissimilar, but semantically related image. These alternatives share few perceptual characteristics. Condition two probes would present a contextual item alongside an alternative exemplar of that object that is perceptually similar (e.g. a different pair of scissors). In this condition, both the visual outline and features are closely matched. The semantic relatedness of the alternative choice in both conditions means that it is equally likely to have been present in the display context and thus there will be equal levels of contextual facilitation across the two conditions. Any difference in recall across the groups would therefore be due to perceptual processing. Consequently, these conditions provide distinct levels of perceptual differentiation and would allow the extension of Stankiewicz, Hummel and Cooper's (1998) paradigm with a comparison between a target receiving tightly focused attention and unattended distracters. Stankiewicz et al. (1998) found such a task could be achieved with coarse perceptual representations even if the context items received no attention (condition 1). If differentiation is also demonstrated between perceptually similar alternatives (condition 2), it will indicate that the perceptual representations need not be limited to crude information, but are capable of detailed discrimination.

A second study would seek to determine whether the perceptual information known about a context item was viewpoint-specific by testing differentiation between different viewpoints of previously viewed objects. The experimental design would use test trials formed of four blocked four-alternative forced choice selection response probes (one for each context item). Two of the choices presented would be dependent upon the results from the first study. In choice one, the image would be a perceptually similar object viewed from an orientated viewpoint in which perceptual similarity was reduced (e.g. 90°), and in choice two the viewed image would be aligned to maintain maximum perceptual similarity with the context object. If higher-level differentiation was not achieved in the earlier experiment, then an object that is perceptually dissimilar to a context item would be used. Choices three and four would be images of the previously viewed context item. Choice three would be the object viewed from an orientated viewpoint similar to that used in choice two (e.g. 90°), whereas choice four would be an image identical to that previously seen by the participant. The correct selection would be choice four. As in the previous experiment, contextual matching of the alternatives would ensure that difference in recall is due to perceptual processing. Comparison between choices 1 + 2 and 3 + 4 replicates the concept of the first experiment and would confirm that the result was not due to object identity rather than perceptual information. Poor discrimination between 3 and 4 would suggest that identical and orientated images were being equally differentiated from the non-seen alternatives, and would imply that perceptual information was stored in a solely non-viewpoint specific manner. If high levels of differentiation were achieved (in favour of choice four) then it indicates that information about context items is stored in a viewpoint-specific format. These two studies would determine not only how much could be extracted from a context item, but also whether this information was extracted from each item in a five-object array. Such experiments would provide the empirical foundations for constructing a model of the contextual processing mechanism that generates the scene context effect.

The Formation of Contextual Frameworks:

The fourth line of future research would be into how contextual relationships are formed. We are not born with a mental library of these schemas or semantic links, nor are we explicitly taught the vast majority of them, yet their influence is robust. Individuals come to learn what is semantically inter-connected through experience and this is mentally encoded in contextual networks. Chun and Jiang (1998, 1999) have demonstrated that configurational context can be implicitly learned through the repetition of consistent information in order to aid visual search. They present the concept of context maps, which are instance based and which weight the importance or salience of component objects. As an incoming image matches the current instance held in memory, objects will be prioritised according to these weights and attention allocated accordingly. In this manner, visual search is facilitated. They also state that the context map resolves a difficult problem of how to allocate attention in a complex scene, and their results correspond to "regions of interest" in real-world scenes (Rensink, O'Regan, & Clark, 1997). The context maps they used in their studies were based upon stimulus location, stimulus identity and stimulus movement patterns, and they highlight the need for further research upon semantic based context maps.

A study is planned that examines the formation of new contextual maps by constructing new relationships between previously unrelated objects. This research would create a contextual instance that would exist only within the experimental environment. The learning would be done implicitly, with participants presented a target/context display and told to identify the target object using a 6AFC response whilst being kept unaware of contextual manipulations. However, feedback would be provided after each response. Three learning groups would be used. Group one would have their consistent target stimuli presented with the same context items in the same locations. Group two would have their consistent target stimuli presented with the same context items in randomised locations (one of four positions). Group three would have their consistent target stimuli presented with one of four contextual exemplars (all of the same name) for each item in randomised locations. In all of the conditions, inconsistent target/context stimuli would be randomly generated from amongst the objects, with no more than one used from each 'real' semantic set. Such inconsistent trials must be included to ensure repetition priming is not responsible for improved performance. During training, learning will be continuously assessed through 100 trial blocks within the 1000 trial epochs (based on error rates). At the completion of training, three further testing blocks would be run. The first additional testing block would test all groups with familiar context items in randomised locations. This test would examine whether group one have encoded their context networks by layout, or whether their learning is transferable to the more complex situation. The second testing block would be a repeat of the training condition for each group. The third testing block would test all groups with new exemplars (of the same name) of familiar context items displayed in familiar (group one) or randomised (group two or three) locations. Ability to demonstrate scene context effects with the perceptually new object will indicate that the contextual network is not limited to perceptual matching, but is based upon object identity. The nature of these tests not only explores how contextual relationships are encoded, but also whether the form of that encoding is dependent upon the way in which information is presented.

Conclusion:

The aims of this thesis were to determine whether a scene context effect could be generated using multiple objects alone; whether contextual processing directly influenced the perceptual/representational processes during target recognition; and whether visual attention played a role in generating scene context effects. It has been demonstrated that a scene context effect can be driven by object-semantic factors. It has also been shown that this scene context effect consists of a superiority effect, which does influence perceptual/representational processes, and a response bias effect, which does not. The presence of this scene context effect aids recognition performance in naming and non-naming tasks, but is dependent on the allocation of visual attention to the context items, context consistency with the target, and a temporal processing window greater than 52msec.

Evidence for the scene superiority effect and the response bias effect was provided by Experiment 4, in which a 6AFC response method was used to calculate and control for response bias in error rates during a recognition task. Fewer errors were recorded when the target was displayed with semantically consistent non-targets than with semantically inconsistent non-targets, both before and after the correction for response bias. The presence of an effect after the correction demonstrates that contextual information can directly influence perceptual/ representational processes during target recognition. However, that the total scene context effect was reduced in magnitude by the correction indicates that part of the contextual effect was accounted for by response bias.

The demonstration of a superiority effect establishes a relationship between scene/context processing and target/object processing that has implications for both fields of research. Models of object processing will need to consider how contextual information is integrated, and accounts of contextual effects may benefit from exploring the visual processes involved in extracting information from a scene. In addition, previous scene context research that has not effectively controlled for response bias (e.g. Davenport & Potter, 2004; Palmer, 1975) has reported the total scene context effect. Unless studies utilise a measure similar to the 6AFC, it is impossible to isolate how much of the total effect is scene superiority effect and how much is response bias.

The role of visual attention in contextual facilitation was demonstrated in Experiment 2. Naming errors of a target increased when attention was distributed away from the target and across a wide area by an endogenous cue, relative to being cued to a narrow area surrounding the target. Contextual facilitation was almost sufficient to offset this decrease, but was generated only when context items (placed within the wide visual field) were allocated visual attention early via cue manipulation and were semantically related to the target. Visual attention is a limited resource; therefore, to benefit from contextual processing there must be a decision to distribute attention widely despite an initial cost to target processing. However, it was also found in Experiment 2 that in the narrow focus conditions in which context items were displayed, errors were higher than when no context was presented. If attention cannot be maintained on a single target, perhaps due to sudden onset, a strategy to distribute that attention early is an ecologically practical decision as "real" targets are seldom encountered in isolation.

It is clear that visual attention plays a role in generating the scene context effect. However, the specific role it plays within the contextual system may be one of non-target processing, item binding and prevention of cross-talk between multiple context items, or the integration of context information. Determining the role of attention requires further investigation.

The generation of a scene context effect was also shown to require a temporal window of greater than 52msec (Experiments 4 and 5). The time necessary for generating the context effect was longer than that required to process the target under similar experimental conditions. The time difference is consistent with claims that contextual and target processing utilise different mechanisms. These mechanisms are more complex than simply alternate recognition pathways, as both models presented offer two routes for processing a single target object, and it is likely that the contextual system would be able to utilise information from more than just one recognition pathway. Bar's (2004) research already indicates cross-modal input as well as visual data into the PHC and RSC. The longer temporal window also provides new support for the concept of context frames/scene schema activation.

Appendix A

Contextual sets and their objects:					
SET 01	Batteries	Audio Tape	Headphones	Walkman	CD
SET 02	Can Opener	Wooden Spoon	Knife	Egg Whisk	Spatula
SET 03	Razor	Toothbrush	Soap	Sponge	Toothpaste
SET 04	Wine Glass	Corkscrew	Candle	Cutlery	Wine Bottle
SET 05	Sieve	Cheese Grater	Frying Pan	Dish	Saucepan
SET 06	Floppy Disks	Mouse	Telephone	Desk Light	Keyboard
SET 07	Bread	Sausages	Eggs	Onion	Bacon
SET 08	Stapler	Holepunch	Scissors	Highlighter	Sellotape
SET 09	Teabags	Mug	Biscuits	Teaspoon	Kettle
SET 10	Bad. Shuttle	Tennis Ball	Baseball cap	Trainers	Sports socks
SET 11	Hairbrush	De-odorant	Comb	Nail Clippers	Flannel
SET 12	Pen	Paperclips	Bull-dog clip	Pencils	Ruler
SET 13	Notepad	Calculator	Glasses	Book	Folder
SET 14	Iron	Clothes Pegs	Coat Hanger	Wash. Powder	Jug
SET 15	Mobile Phone	Keys	Money	Credit Cards	Wallet
SET 16	Carrot	Mushrooms	Pepper	Potato	Spring Onions
SET 17	Banana	Orange	Grapes	Plums	Apple
SET 18	Cards	Dominoes	Dice	Chess Piece	Poker Chips
SET 19	Hammer	Spanner	Pliers	Nails	Screwdriver
SET 20	Camera	Sunglasses	Sun-cream	Train Tickets	Postcard
SET 21	Washing up Liquid	Scrubbing Brush	Tea Towel	Scourer	Jay Cloth
SET 22	Umbrella	Shirt	Shoes	Tie	Briefcase
SET 23	Gloves	Boots	Hat	Scarf	Backpack
SET 24	Dustpan & Brush	Dustbin	Air Freshener	Duster	Hoover
SET 25	Cigarettes	Packet of Crisps	Pint Glass	Bottle	Peanuts
SET 26	Flower	Stones	Fir-cone	Leaf	Pine twig
SET 27	Present	Card	Party Hat	Balloons	Party Popper
SET 28	Paint Brush	Roller	Paint Can	Sandpaper	Wallpaper
SET 29	Necklace	Watch	Ring	Broach	Ear Rings
SET 30	Tablets	Antiseptic Cream	Pipette	Plasters	Bottle
SET 31	Lipstick	Nail Varnish	Mirror	Perfume	Make-up
SET 32	Plant pot	Fork	Secateurs	Seeds	Bulbs

Note: Context sets used in Experiment 1.
Images can be found on attached CD.

Appendix B

Mean Naming Error Rates (%) and Standard Errors in Experiment 1 Across All Variables.

		FOCUS					
			Narrow			Wide	
ORDER	SET	Consistent	Inconsistent	None	Consistent	Inconsistent	None
1	1	8.6 (2.3)	10.2 (2.5)	4.7 (2.5)	6.3 (2.4)	13.3 (3.0)	18.0 (2.6)
	2	6.3 (2.3)	5.5 (2.5)	5.5 (2.5)	13.3 (2.4)	9.4 (3.0)	12.5 (2.6)
	3	4.7 (2.3)	7.0 (2.5)	7.8 (2.5)	7.0 (2.4)	14.8 (3.0)	10.2 (2.6)
2	1	3.1 (2.3)	10.9 (2.5)	19.5 (2.5)	5.5 (2.4)	7.8 (3.0)	7.8 (2.6)
	2	12.5 (2.3)	7.8 (2.5)	11.7 (2.5)	5.5 (2.4)	4.7 (3.0)	5.5 (2.6)
	3	7.8 (2.3)	9.4 (2.5)	3.9 (2.5)	3.9 (2.4)	9.4 (3.0)	8.6 (2.6)

Appendix C

Contextual sets and their objects:

SET 01	Batteries	Audio Tape	Headphones	Walkman	CD
SET 02	Can Opener	Wooden Spoon	Knife	Egg Whisk	Spatula
SET 03	Hammer	Spanner	Screwdriver	Nails	Pliers
SET 04	Camera	Sunglasses	Sun-cream	Train Tickets	Postcard
SET 05	Sieve	Cheese Grater	Frying Pan	Dish	Saucepan
SET 06	Floppy Disks	Mouse	Telephone	Desk Light	Keyboard
SET 07	Gloves	Boots	Hat	Scarf	Backpack
SET 08	Dustpan & Brush	Dustbin	Air Freshener	Duster	Hoover
SET 09	Teaspoon	Mug	Biscuits	Teabags	Kettle
SET 10	Bad. Shuttle	Tennis Ball	Baseball cap	Trainers	Sports socks
SET 11	Present	Card	Party Hat	Balloons	Party Popper
SET 12	Paint Brush	Roller	Paint Can	Sandpaper	Wallpaper
SET 13	Notepad	Calculator	Glasses	Book	Folder
SET 14	Iron	Clothes Pegs	Coat Hanger	Wash. Powder	Jug
SET 15	Lipstick	Nail Varnish	Mirror	Perfume	Make-up
SET 16	Plant pot	Trowel	Flower	Seeds	Bulbs
SET 17	Banana	Orange	Grapes	Plums	Apple
SET 18	Cards	Dominoes	Dice	Chess Piece	Poker Chips
SET 19	Razor	Toothbrush	Soap	Sponge	Toothpaste
SET 20	Wine Glass	Corkscrew	Candle	Cutlery	Wine Bottle
SET 21	Washing up Liquid	Scrubbing Brush	Tea Towel	Scourer	Jay Cloth
SET 22	Umbrella	Shirt	Shoes	Tie	Briefcase
SET 23	Bread	Sausages	Eggs	Onion	Bacon
SET 24	Stapler	Highlighter	Scissors	Holepunch	Sellotape
SET 25	Cigarettes	Packet of Crisps	Pint Glass	Bottle	Peanuts
SET 26	Teddy	Dolly	Toy car	Building blocks	Childrens' Book
SET 27	Hairbrush	De-odorant	Comb	Nail Clippers	Flannel
SET 28	Pen	Paperclips	Bull-dog clip	Pencils	Ruler
SET 29	Necklace	Watch	Ring	Broach	Ear Rings
SET 30	Tablets	Antiseptic Cream	Syringe	Plasters	Bottle
SET 31	Mobile Phone	Keys	Money	Credit Cards	Wallet
SET 32	Carrot	Mushrooms	Red Pepper	Potato	Spring Onions

SET 33	Lollypops	Chocolate Bars	Jellies	Fudge	Chocolate Raison
SET 34	Wardrobe	Dining Chair	Desk	Bed	Arm Chair

Note: Contextual sets used in Experiments 2, 3, 4 and 5. Sets 33 and 34 were used only in Experiment 4.
Images can be found on attached CD.

Appendix D

Post-hoc Analysis of Mean Error Rates Across Error Types in Experiment 4 Using Bonferonni Method.

(I) ERROR	(J) ERROR	Mean Difference (I-J)	Std. Error	Sig.
1	2	3.005	1.010	.046
	3	6.041	.585	.000
	4	6.187	.552	.000
	5	6.104	.605	.000
2	1	-3.005	1.010	.046
	3	3.036	.792	.004
	4	3.182	.819	.003
	5	3.099	.879	.009
3	1	-6.041	.585	.000
	2	-3.036	.792	.004
	4	.145	.254	1.000
	5	.063	.393	1.000
4	1	-6.187	.552	.000
	2	-3.182	.819	.003
	3	-.145	.254	1.000
	5	-.083	.227	1.000
5	1	-6.104	.605	.000
	2	-3.099	.879	.009
	3	-.063	.393	1.000
	4	.083	.227	1.000

REFERENCES

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15(4)*, 600-609.

Bar, M. (2004). Visual Objects in Context. *Nature Reviews Neuroscience, 5(8),* 617-629.

Bar, M. & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38(2),* 346-358.

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

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy and J. R. Pomerantz. (Eds.) *Perceptual Organization (pp. 312-253).* Hillsdale, NJ: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94(2),* 115-117.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97(1),* 22-27.

Bloem, I. & La Heij, W. (2003). Semantic facilitation and semantic interference in word translation: Implication for models of lexical access in language production. *Journal of Memory & Language, 48,* 468-488.

Bonnar, L., Gosselin, F., & Schyns, P. G. (2002). Understanding Dali's Slave Market with the Disappearing Bust of Voltaire: A case study in the scale information driving perception. *Perception, 31,* 683-691.

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, & Cognition, 18(3)*, 531-543.

Broadbent (1958). *Perception and Communication.* New York. Pergamon Press.

Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research, 62,* 182-194.

Cave, K. R. & Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review, 6(2),* 204-223.

Cave, K. R., Kim, M-S., Bichot, N. P., & Sobel, K. V. (1999). Visual selection within a hierarchical network: The FeatureGate Model. Unpublished manuscript.

Cattell (1886). The time taken up by cerebral operations. *Mind, 11,* 220-242; 377-392; 524-538.

Cheng, E. K. & Simons, D. J. (2001). *Perceiving the internal consistency of scenes.* Poster session presented at annual meeting of Psychonomic Society, Orlando, FL.

Chun, M .M. & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28-71.

Chun, M .M. & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science, 10(4),* 360-365.

Chun, M. M. & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance, 21(1),* 109-127.

Colegate, R. L., Hoffman, J. E., & Eriksen, C. W. (1973). Selective encoding from multielement visual displays. *Perception & Psychophysics, 14,* 217-224.

Davenport, J. L. & M. C. Potter (2004). Scene consistency in object and background perception. *Psychological Science, 15(8),* 559-564.

Davidoff, J. (1986). The mental representation of faces: Spatial and temporal factors. *Perception & Psychophysics, 40(6),* 391-400.

Deutsch, J. A. & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70(1),* 80-90.

Diamond, R. & Carey, S. (1986). Why faces are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107-117.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of re-entrant visual processes. *Journal of Experimental Psychology: General, 129(4),* 481-507.

Donnelly, N., Humphreys, G. W., & Riddoch, M. J. (1991). Parallel computation of primitive shape descriptions. *Journal of Experimental Psychology: Human Perception & Performance, 17(2),* 561-570.

Duncan, J. & Humpreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96(3),* 433-458.

Enns, J. T. & Gilani, A. B. (1988). 3-Dimensionality and discriminability in the object-superiority effect. *Perception & Psychophysics, 44(3),* 243-256.

Enns, J. T. & Rensink, R. A. (1990). Sensitivity to three-dimensional orientation in visual search. *Psychological Science, 1(5),* 323-326.

Eriksen, C. W. & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40(4),* 225-240.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291,* 312-316.

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.

Goldsmith, M. & Yeari, M. (2003). Modulation of object-based attention by spatial focus under endogenous and exogenous orienting. *Journal of Experimental Psychology: Human Perception & Performance, 29(5),* 897-918.

Gottesman, C. V. & Intraub, H. (1999). Wide-angle memories of close up scenes: A demonstration of boundary extension. *Behaviour Research Methods Instruments & Computers, 31(1),* 86-93

Hayward, W. G. & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception & Performance, 23(5),* 1511-1521.

Henderson, J. M., Pollatsek, A., & Raynor, K. (1987). The effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception & Performance, 13(3),* 449-463.

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(2-3),* 319-343.

Homa, D., Haver, B. & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition, 4(2),* 176-185.

Hummel, J. E. (2000). Where view-based theories break down: The role of structure in shape perception and object recognition. In E. Dietrich and A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans & Machines (pp. 157-185).* Hillsdale, NJ: Erlbaum.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition, 8,* 489-517.

Hummel, J. E. & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99(3),* 480-517.

Hummel, J. E. & Stankiewicz, B. J. (1996). Categorical relations in shape perception. *Spatial Vision, 10(3),* 201-236.

Humphreys, G. W., Price, C. J., & Riddoch, J. (1999). From objects to names: A cognitive neuroscience approach. *Psychological Research, 62,* 118-130.

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory & Language, 35(2),* 118-134.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. In J. B. Long and A. D. Baddeley (Eds.), *Attention and Performance IX (pp. 187-203).* Hillsdale, NJ, Erlbaum.

Jonides, J. & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics, 43(4),* 346-354.

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

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience, 7(2),* 41-45.

Kosslyn, S. M. (1994). *Image and Brain.* Cambridge, Massachusetts & London, England, The MIT Press.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception & Performance, 21(3),* 451-468.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133(3),* 339-354.

Loftus, G. R. & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 565-572.

Lowe, D. G. (1987). 3-Dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31(3),* 355-395.

McClelland, J. L. & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception & Psychophysics, 22,* 249-261.

McClelland, J. L. & Miller, J. (1979). Structural factors in figure perception. *Perception & Psychophysics, 26(3),* 221-229.

McClelland, J.L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88(5),* 375-407.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception & Performance, 23(1),* 168-180.

Mermelstein, R., Banks, W., & Prinzmetal, W. (1979). Figural goodness effects in perception and memory. *Perception & Psychophysics, 26,* 472-480.

Mozer, M. (1991). *The Perception of Multiple Objects: A Connectionist Approach.* Cambridge, Massachusetts & London, England, The MIT Press.

Müller, M., M. & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception & Performance, 15(2),* 315-330.

Neisser, U. (1967). *Cognitive Psychology.* Englewood Cliffs, NJ: Prentice-Hall.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner and G. W. Humphreys

(Eds.), *Basic Processes in Reading: Visual Word Recognition (pp. 264-336).* Hillsdale, NJ: Erlbaum.

Oliva, A. & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72-107.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. *Proceedings of the IEEE International Conference on Image Processing, Sept, Barcelona, Spain,* 14-17

Palmer, S. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3(5),* 519-526.

Palmer, S. E. (1999). *Vision Science.* Cambridge, Massachusetts & London, England, The MIT Press.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long and A. Baddeley (Eds.) *Attention and Performance IX (pp. 135-151),* Hillsdale, NJ, Erlbaum.

Pashler, H. E. (1999). *The Psychology of Attention*, Cambridge, Massachusetts & London, England, The MIT Press.

Poggio, T. & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263-266.

Pomerantz, J. R., Sager, L.C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception & Performance, 3(3),* 422-435.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Posner, M. I. & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium (pp. 55-85).* Hillsdale, NJ, Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109(2),* 160-174.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1995). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8(5),* 368-373.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81(2),* 275-280.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 9,* 44-63.

Richer, F. & Boulet, C. (1999). Frontal lesions and fluctuations in response preparation. *Brain & Cognition, 40(1),* 234-238.

Riddoch, M. J. & Humphreys, G. W. (1987). Picture naming. In G. W. Humpreys and M. J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach (pp. 107-143).* Lawrence Erlbaum Associates.

Schyns, P. G. & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science, 5,* 195-200.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *The British Journal of Psychology, 75,* 221-242.

Shapiro, K. L., Raymond, J. E. & Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance, 20(2),* 357-371.

Shih, S. & Sperling, G. (1996). Is there feature based attentional selection in visual search? *Journal of Experimental Psychology: Human Perception & Performance, 22(3),* 758-779.

Shih, S. & Sperling, G. (2005). Modulating effect of the target saliency. Manuscript in preparation.

Smith, W., Dror, I., & Schmitz-Williams, I. (in press). The effect of decomposability and meaningfulness on the representation and processing of visual information in mental rotation. Journal of Mental Imagery.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74(11),* 29.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence of dual representation of object shape. *Journal of Experimental Psychology: Human Perception & Performance, 24(3),* 732-744.

Starreveld, P. A. & La Heil, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22(4),* 896-918.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science, 262,* 685-688.

Tanaka, J. W. & Farah, M. J. (2003). The holistic representation of faces. In M. A. Peterson and G. Rhodes (Eds.), *Perception of Faces, Objects and Scenes (pp. 53-74),* Oxford University Press.

Tanaka, J. W. & Sengco, J. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25,* 583-592.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychological Bulletin & Review, 2(1),* 55-82.

Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. A. Peterson and G. Rhodes (Eds.), *Perception of Faces, Objects and Scenes (pp. 177-211),* Oxford University Press.

Tarr, M. J. & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67,* 1-20.

Tarr, M. J. & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21(2),* 233-282.

Tarr, M. J. & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1(4),* 207-209.

Thoma, V., Hummel, J. E., & Davidoff, J. (2004). Evidence for holistic representation of ignored images and analytic representation of attended images. *Journal of Experimental Psychology: Human Perception & Performance, 30(2),* 257-267.

Treisman, A. M. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception & Performance, 8(2),* 194-214.

Treisman, A. M. (1986). Features and objects in visual processing. *Scientific American, 255(5),* 114-125.

Treisman, A. M. (1993). Representing visual objects. In D. E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV (pp. 163-175).* Hillsdale, NJ, Erlbaum.

Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology, 6,* 171-178.

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12(1),* 97-136.

Turatto, M., Benso, F., Facoetti, A., Falfano, G., Mascetti, G., G., & Umilta, C. (2000). Automatic and voluntary focusing of attention. *Perception & Psychophysics, 62(5),* 935-952.

VanRullen, R. & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra rapid visual categorisation of natural and artifactual objects. *Perception, 30,* 655-668.

van Santon, J. P. H. & Jonides, J. (1978). A replication of the face-superiority effect. *Bulletin of the Psychonomic Society, 12(5),* 378-380.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition, 85,* 61-69.

Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19(2),* 243-259.

Warner, C. B., Juola, J. F., & Koshino, H. (1990). Voluntary allocation versus automatic capture of visual attention. *Perception & Psychophysics, 48(3),* 243-251.

Weisstein, N. & Harris, C. S. (1974). Visual detection of line segments: An Object-Superiority Effect. *Science, 186,* 752-755

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59-85.

Williams, A. & Weisstein, N. (1978). Line segments are perceived better in coherent contexts than alone: An object-line effect. *Memory & Cognition, 6,* 85-90.

Wolfe, J. M. (1994). Guided search 2.0 – A revised model of visual search. *Psychonomic Bulletin & Review, 1(2),* 202-238.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences, 7(2),* 70-76.

Wolfe, J. M. & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37(1),* 25-43.

Wolfe, J. M. & Cave, K. R. (1999). The psychophysical evidence for a binding problem in human vision. *Neuron, 24,* 11-17.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception & Performance, 15(3),* 419-433.

Wolfe, J. M., Stewart, M. I., Friedman-Hill, S. R., & O'Connell, K. M. (1992). The role of categorization in visual-search for orientation. *Journal of Experimental Psychology: Human Perception & Performance, 18(1),* 34-49.

Yantis, S. (2000). Goal-directed and stimulus-driven determinants of attentional control. In S. Monsell and J. Driver (Eds.), *Attention and Performance XVIII (pp. 73-103).* Hillsdale, NJ, Erlbaum.

Yantis, S. & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance, 10(5),* 601-621.

Yantis, S. & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic selection. *Journal of Experimental Psychology: Human Perception & Performance, 16(1),* 121-134.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception, 10,* 747-759.