Classifying Ambiguous Everyday Objects:

The Effects of Stimulus Duration, Visual Field, and Background Context.

Harini Sankar

School of Psychology, University of Nottingham Malaysia, Semenyih

Master's in research in Social and Behavioural Sciences

Supervised by Dr Miflah Hussain

Acknowledgement

Firstly, I would like to express my sincerest gratitude to my supervisor and mentor Dr Miflah for his invaluable guidance and support and patience during my undergraduate and master's study. He is a wonderful and irreplaceable first mentor who kindled within me a passion for research and played a huge role in helping me develop an invaluable set of skills in research and in life. I am extremely grateful for his encouragement and belief in me (especially when I found it difficult to believe in myself). Furthermore, I am grateful for his patience and support through the particularly challenging year due to the covid-19. I look forward to working with him as a colleague in the future. I would also like to thank the University of Nottingham Malaysia and the incredible School of Psychology that has been my home for the past four years and for playing an immense role in my professional and personal development.

I am forever grateful to my best friend Swetha who is the reason I found my love for Psychology and for being there since the beginning as an unwavering pillar of support, encouragement, and inspiration to all my endeavours- academic, emotionally, and otherwise. I couldn't have dreamed of being where I am without her love and support. I am thankful for my friends Nathali Lennon, Kristine, Geet, Wyn and many others at UNM who have encouraged me and kept me sane through my undergraduate and masters. I am particularly grateful for Nathali's friendship and invaluable emotional support. Lastly, I am extremely thankful for parents (Sankar and Latha) for their encouragement and financial support for all my academic and extracurricular pursuits.

Abstract

Visual perception in humans is optimised to function in its habitat. When sensory information is ambiguous, our prior knowledge and expectations about the habitat biases perception. By measuring perceptual biases for ambiguous stimuli, we are able to infer what expectations we have about our environment. Recently, Hussain Ismail et al. (2019) found a perceptual bias in people living in cities, where ambiguous objects (made of one man-made and one natural component images) were more likely to be classified as a man-made object (i.e., manufactured objects such as a vehicle or a house) rather than a naturally occurring object (e.g., flower or animal). They speculated that this bias may be a result of our expectations to see man-made objects more often in our living environment. The aim of this thesis is to examine whether the "man-made bias" is susceptible to factors that are known to alter perceptual biases in vision. The first experiment examined whether the man-made bias is influenced by presentation duration when ambiguous objects ("hybrids") are directly fixated at. We found that a shorter presentation duration (50 ms) that increased the uncertainty of the stimulus increased the magnitude of the man-made bias compared to a longer duration (150 ms), in line with predictions of Bayesian perception. Experiment 1 also demonstrated that the man-made bias is replicable when object classification is measured with an online experiment, using a new collection of object stimuli on a new set of participants. Experiment 2 measured the same perceptual bias by presenting hybrids in participants' peripheral vision. Although we found a significant, *albeit* small, perceptual bias, we did not find any effect of stimulus duration on the man-made bias in the periphery. In Experiment 3, participants viewed hybrids at fixation, and we measured the effect of background scenes on which hybrids were superimposed on. The superordinate semantic category of the background scene (man-made or natural) did not affect the magnitude or the direction of the

man-made bias, indicating that background scenes did not alter the detectability of semantically congruent component images in the hybrids. Overall, we show that the man-made bias is resistant to changes in presentation duration, visual field in which objects are viewed and the semantic congruency of the spatial context.

Table of Contents

1. Chapter 1 – Background

1.1. Selectivity to low-level and mid-level features

Our visual system can effortlessly and rapidly process perceptual information rather accurately from our visual field, despite the transient and inherently noisy input that it receives. Humans are adept at navigating and orienting themselves within any environment, be it a busy street in a large city or a thick forest path leading to a waterfall. Therefore, it is of ecological importance that we perceive the spatial structure of the environment and the typical objects that occur in the environment to navigate and interact within such environments successfully.

The process of visual perception begins when light that is often reflected off from objects in our environment falls onto the retina, where photoreceptors convert the light energy into neural signals. This signal is transmitted via the lateral geniculate nucleus (LGN) to the cortical areas in the brain that process visual information - starting from simple feature processing in the occipital regions of the brain (e.g., striate and extrastriate cortices), to higher-level processing in areas such as the temporal and parietal cortices (Deyoe et al., 1996; Ferster & Miller 2000; Kaas & Krubitzer, 1991). Electrophysiological studies on cats and primates have shown that each of these cortical areas is known to process specific aspects of visual information (Hubel & Wiesel, 1959, 1968). The Lateral geniculate Nucleus (LGN), striate cortex (V1) and extra-striate cortices (e.g., V2) are well-known to selectively encode simple visual features such as contrast, spatial frequency and orientation of edges (Hubel & Weisel 1962; Sachs et al., 1971). Apart from having neurons highly selective to specific features, in early occipital regions (e.g., V1) we can also find mechanisms that integrate feature values (such as the orientations of several edges) to produce a coherent percept such as the average orientation of a set of edges (Hubel & Wiesel,

1965). The integration of feature values occurs through connections of neurons within various levels of the occipital regions as well as inter-cortical connections between these regions (Hess et al., 2003).

The external visual field is mapped onto the retina in a one-to-one fashion. Information from specific parts of the visual field always trigger neural signals in specific parts of the retina. This organisation, known as 'retinotopic mapping', is preserved throughout several stages of the early visual processing pipeline, such as in the LGN and V1, so that adjacent parts of the visual field are represented in adjacent neurons (or neural populations) within the same region (Engel et al., 1997; Schneider et al., 2004). Beyond V1, towards the higher-level regions such as the V4 and the ventral occipital cortex (VO), this mapping is not as fine-tuned as it is in the striate and extrastriate cortices (Brewer et al., 2005; Felleman & Van Essen, 1991). In regions where retinotopic mapping is observed, neurons or small neural populations respond to visual signals coming from unique spatial locations in the visual field, known as "receptive fields". So, not only are these neurons selective to features, but they also respond selectively to specific spatial locations. The receptive field of neurons that receive input from the fovea are the smallest, but as the eccentricity (angular distance from the fovea) and the neuron location in the cortical hierarchy increases, the receptive field size also increases (neurons in higher levels of the hierarchy have a larger receptive field) (Engel et al., 1997; Hubel & Wiesel, 1965). Findings regarding the changes in receptive field sizes along the cortical hierarchy largely stem from single-cell electrophysiology research done on cats and non-human primates (Hubel & Wiesel, 1961,1968). However, Smith et al. (2001) further supported this organisation with receptive field sizes estimated from functional Magnetic Resonance Imaging (fMRI) data obtained from the human brain.

Classically, visual information is thought to be processed by the different cortical areas in a feedforward manner, such that early cortical areas like V1 process rudimentary, low-level features like the orientation of edges and subsequently, cortical areas higher up in the hierarchy process increasingly complex features such as the global shape and semantic categories (Felleman & van Essen,1991; Hochstein & Ahissar, 2002). However, feature processing is not strictly a feedforward mechanism as input from the higher levels have been found to modulate lower-level processing through feedback connections (Hochstein & Ahissar, 2002; Rao & Ballard, 1999). Some evidence for this feedback mechanism comes from studies of perceptual learning that show that easier learning occurs at higher cortical levels which subsequently facilitate learning at lower levels (Hochstein & Ahissar, 2002).

1.2. Selectivity to high-level features

In order to examine low-level feature selectivity, studies investigating feature selectivity often use simplistic and artificial stimuli such as Gabor patches and dot diagrams. However, the stimuli we experience in real life are much more complex and composed of many contours that vary in contrast, orientation, and spatial frequency. We perceive structures beyond low-level features from these stimuli, such as the shape of an object (i.e., whether it's a square or a circle). We also assign semantic descriptors to these stimuli (e.g., the gender or race of a face, and whether an object is animate or inanimate) (Oliva & Torralba, 2001). The cortical areas involved in encoding these complex features in humans as well as other primates have been extensively studied using functional magnetic resonance imaging (fMRI) over the last few decades. Many studies have found multiple cortical regions predominantly implicated in the processing of complex features of real-life images. The Fusiform Face Area is one such specialised cortical area implicated in selectively processing faces (Kanwisher et al., 1997). Similarly, other

specialised brain areas have been discovered, such as the Parahippocampal Place Area that selectively encodes scenes, or more specifically layouts of scenes (Epstein & Kanwisher, 1998), the lateral occipital cortex that selectively encodes objects such as tools and animals (Grill-Spector et al., 1998), and even an area in the ventral occipital cortex that responds to selectively to buildings (Aguirre et al., 1998). Further, the Lateral Occipital Cortex and the posterior fusiform gyrus is implicated in shape recognition and has been found to respond to intact images of objects compared to scrambled images (Malach et al., 1995). A recent study found that the LOC is also sensitive to the information about interactions between multiple objects (Kim & Biederman, 2011).

1.3. Objects

At this stage, it is imperative to distinguish between what a "scene" is and what an "object" is. Objects can be defined as individual entities that have strict bounding contours (Spelke, 1990). They are recognisable when seen in isolation and we often directly interact with or act upon them, as we navigate in our environment. Scenes contain numerous objects arranged in a spatially coherent manner, and often provide a semantically coherent percept of the environment around us, within which interaction occurs (Henderson & Hollingsworth, 1999; Oliva & Torralba, 2001). For instance, Figure 1.1A shows a collection of objects commonly found in a street scene. Figure 1.1B shows an image of an actual street scene, comprising of objects such as cars, busses, pedestrians, and buildings. As evident from the two images, a mere collection of objects does not make a scene (Biederman, 1972). The particular arrangement of the objects in the scene enables us to accurately identify that it is a street scene.

FIGURE 1.1

Example of objects and a scene

A)

B)

Note. A) A collection of objects that constitute a street scene. B) An image of an actual street scene. All images used are obtained under the creative commons license (See Reference for image citations).

1.3.1. Models of object perception

Most studies investigating perceptual mechanisms have used 2-dimensional (2D) images of objects we encounter, but objects in the real world are almost always 3-dimensional (3D). Still, the participants in those studies are able to accurately recognise them. In fact, even in real life, the image of 3D objects that fall on our retina is 2D in nature. One of the very puzzling questions in the literature of vision science is "How are we able to assign the same semantic label to objects regardless of changes in representations (i.e., 3D to 2D), view, size, lighting conditions and configuration in space?". This problem of recognition resistant to variance has been fundamental in developing the different models that aim to explain object perception (Peissig & Tarr, 2007). Bearing that in mind, the two major types of models of object recognition that will be discussed here are the *structural-description models* and *view-based models*.

The structural-description model of object recognition, which is based on Biederman's (1987) recognition by components model, indicates that complex objects can be broken down into component parts made up of basic three-dimensional units such as cones and cylinders, known as "Geons", and information regarding their spatial relations (Marr & Nishihara, 1978, Beiderman, 1987). For example, a mug can be thought of as consisting of a one-side open cylinder and a curved cylindrical handle attached to the side of it. The relationship between the units is important because if that is altered, it could alter the recognition of the object – if the curved handle was atop the cylinder, it would be recognised as a bucket rather than a mug. Thus, recognition based on this structural description of parts and spatial relations between them remains unaffected by most scenarios where the object's appearance varies (e.g., due to viewpoint, lighting conditions).

However, some changes, such as turning the mug 90 degrees (causing it to resemble a bucket), can affect recognition and result in slower and more inaccurate recognition (Hummel, 2000; Peissig & Tarr, 2007). It is not always plausible to generate structural descriptions for objects especially since the structural descriptions can be altered by different viewpoints, giving rise to different descriptions at different instances (Tarr & Pinker, 1989). To resolve this issue, an image-based model was developed that posited that objects are recognised based on multiple stored 'views' in memory (Tarr & Pinker, 1989). This model theorizes that object recognition is based on matching objects to image templates of numerous possible views stored in our memory. This could explain why recognition of relatively novel objects at different angles may be slow, due to the unavailability of a reference template for that novel view.

1.4. Scenes (semantic categories and global properties)

It is known that a scene comprises of numerous objects. However, the arrangement of the objects within a scene provides information regarding the scene and assist in recognition of the scene itself and objects within it. Oliva and Torralba (2006) posited that a scene's coarse spatial properties also known as its global image features can estimate the structural and functional properties of the scene and enable the recognition of the scene's gist. Local features of an object include the colour or orientation of its edges, whereas global features consist of an averaged, or summarised representation of many such local features and/or feature values (Oliva & Torralba, 2006). For instance, in a forest scene image, the local feature would include edges in a range of different orientations, the global feature would comprise of a summary of the distribution of all the orientation values, resulting in a predominantly cardinal global orientation configuration (Coppola et al., 1998; Oliva & Torralba, 2006). Furthermore, Oliva and Torralba (2001) also postulated a few perceptual dimensions by which the spatial properties of scenes can be described and categorised semantically. These dimensions include the degree of naturalness, openness, roughness, expansion, and ruggedness. One dimension used to describe a scene, and is most pertinent to this paper, is the degree of naturalness. The degree of naturalness creates a distinction between man-made (i.e., carpentered) and natural scenes (i.e., with very minimal human-built structures in it, and mostly naturally occurring objects) based on the prevalence of cardinal (vertical and horizontal) orientations and the distribution of spatial frequencies. For instance, although both types of scenes generally have a prevalence of cardinal orientations, this dominance is more pronounced in man-made as opposed to natural scenes (Coppola et al., 1998).

Oliva and Torralba (2001) also categorise scenes based on the level of description of the environment. Subordinate level of descriptions of scenes involves an analysis and recognition of the local structures and objects within a scene (e.g., trees in a forest). Basic level categorisation is

comparatively more abstract than the subordinate level and involves assigning categorical labels to scenes whose objects often have similar shapes and functions (e.g a forest or street). Lastly, the superordinate level of categorisation is the most abstract in terms of description, and scenes within a superordinate category has high variability in terms of the objects that occur in them (e.g. cities and natural landscapes or indoor and outdoor environments) (Oliva & Torralba, 2001). In this thesis, we will be distinguishing the scenes used at the superordinate level, i.e. as natural and man-made environments.

1.4.1. Models of scene perception

One of the earliest theories posited to explain the process of scene perception is the scene schema model (Biederman, 1981; Friedman, 1979; Henderson, 1992). Biederman (1981) likened scene perception to auditory comprehension of sentences. By this comparison, words are analogous to objects, and the sentence represents a scene. In order to understand the meaning of a sentence, the words and their relationship with other words in the sentence matters. Similarly, Biederman (1981) posited that objects within a given scene and its spatial relationship with other objects activate schemas (a memory representation of a prototypical scene) that facilitate the recognition of the overall scene being perceived. For example, in order to identify a park scene, few objects that occur in a prototypical scene of a park such as benches and trees must be present to access the schema of a park. The activation of the park schema is said to further facilitate recognition of other objects within the scene. Subsequently, violations of the semantic and spatial relations between objects can alter the activation of schemas. Biederman (1981) postulated a list of violations that govern the detection of objects within scenes. To illustrate one such violation, if a fire hydrant is atop a mailbox, this violates the positional relationship between the two objects, although they are both objects likely to occur in a street scene.

Biederman et al. (1982) provided empirical evidence that these violations can impair performance of object recognition within scenes. They presented scenes that contained target objects that violated semantic and spatial relationships (such as the position and probability of an object to occur in a given scene) and also other cued objects that did not have any violations. They found that errors in detecting other target objects that violated expectations of semantic and spatial relationships was hindered, as evident from increases error rates and response times. However, detection of other cued objects was not hindered. The mean error in detecting the object and reaction time of detection increased along with the number of violations. Biederman et al. (1982) theorised that the consistency of objects with scenes and their spatial relationships aid in activating schemas of scenes, and this subsequently facilitates the detection of a scene's constituent objects.

Although the scene schema theory has helped question and define the parameters of what constitutes a coherent scene instead of a random collection of objects and provides a plausible explanation of object detection within a scene, Henderson (1992) argued that since there are many ways in which objects in a scene can exist, this would cause difficulty gauging the information of the 'normally' occurring probability and position of objects to activate a schema. Therefore, research gravitates towards a model that suggests that the gist of the scenes are encoded even before the objects in the scene are consciously perceived (Rousselet et al., 2003; Schyns & Oliva, 1994).

A scene's gist provides information of the scene's superordinate category (e.g., man-made or natural). The gist can be perceived by encoding the coarse spatial layout of a scene (carried by low-spatial frequency information) (Schyns & Oliva, 1994) or by encoding a summary of low-level features (Rousselet et al., 2003; VanRullen & Thorpe, 2001), without

necessarily having to identify individual objects in the scene. Here, the spatial characteristics of a scene enables the activation of the scene schema, rather than the objects within it (Schyns & Oliva, 1994). However, this does not imply that low spatial frequency data is always processed first, and that high spatial frequency data is always processed later (Ahissar & Hochstein, 1997; Hochstein & Ahissar, 2002).

Rapid serial visual perception tasks (RSVPs) have widely supported the view that gist perception occurs rapidly. Such tasks involve flashing a series of scenes for extremely short durations and measuring viewers' accuracy in categorising these scenes (Crouzet & Serre, 2011; Joubert et al., 2008; Rousselet et al., 2003). Joubert et al. (2007) showed that participants could successfully categorise scenes that were presented for as short as 26ms into superordinate categories, i.e., as natural and man-made. Greene and Oliva (2009) presented images of natural landscapes, and participants had to categorise the scenes according to their basic categories (e.g., as a desert, lake, mountain, ocean) and their global attributes regarding the spatial and functional properties of the scene (e.g., naturalness, openness, transience, depth, navigability). They found that participants correctly categorised the images when presented for remarkably short durations of 19ms for global properties and 30ms for basic properties.

1.5. Effect of prior knowledge on perceiving low-level / mid-level features

What we ultimately perceive is not simply a product of the encoded stimuli but also our conscious interpretation of the inputs we receive (Helmholtz, 1925). The visual input that we receive from our environment is inherently ambiguous (Kersten et al., 2004). This ambiguity arises when there are multiple possible scenes or objects that can be perceived given the same sensory input, or when different sensory inputs result in the same percept, forcing the visual system to make guesses (intelligently) about the likeliest possible occurrence (Mamassian et al.,

2002). Optical illusions allow us to observe this ambiguity more explicitly. For example, the famous duck-rabbit illusion where we are able to perceive the drawing as a duck when viewed from one angle and a rabbit when viewed from another (Jastrow, 1899). Helmholtz (1925) posited that perception could be thought of as an unconscious inference that requires implicit knowledge of the environment to infer the properties of the objects and resolve perceptual ambiguity. Therefore, perception is widely modelled as a probabilistic inference using the Bayesian framework (Kersten et al., 2004; Mamassian et al., 2002). The Bayesian framework describes the optimal way of combining sensory information with existing knowledge and provides a reliable framework for understanding how we resolve perceptual ambiguity. This framework relies on two main components: 1) a prior distribution that represents the preferences/expectations we hold for some representations of an object or visual features and the contextual information presented along with the sensory input, and 2) a likelihood function that accounts for the sensory responses, i.e. how likely we are to detect and perceive the input (Mamassian et al., 2002). By combining the priors and the likelihood, the Bayes theorem provides a posterior probability distribution enabling the selection and inference of the most probable stimulus. The standard deviation of the prior probability distribution dictates the magnitude of dependency on the sensory information and the prior. When the standard deviation is lower-indicating a strong preference for a particular representation- the reliance on this prior is stronger than the sensory information (the likelihood) and it skews the posterior probability distribution towards the peak of the prior distribution. When the standard deviation is higher-indicating a weaker prior- the reliance on the sensory information increases, shifting the posterior probability distribution towards the peak of the likelihood function (Mamassian et al., 2002). Although it is rather tricky to experimentally manipulate the standard deviation of a prior

directly, we can gauge the dependency of our visual system on priors by altering the amount or the quality of sensory input received by our visual system, in other words, changing the likelihood function (Stocker & Simoncelli, 2006).

Priors that seemingly reflect knowledge of the visual environment's statistical regularities has been found to bias perception for a range of visual features. With regards to low-level features, Girshick et al. (2011) found that people showed a bias towards perceiving edges of near-cardinal orientations as more cardinal (horizontal or vertical) rather than oblique (slant) oriented. This bias reflects the implicit knowledge we have of the orientation statistics of the environment, which show a prevalence in horizontal and vertically oriented edges compared to oblique (slant) edges (Girshick et al., 2011). Similarly, for more complex features such as shape perception, it was found that perception tended to bias towards convexity than concavity depending on the lighting direction (Stone et al. 2009; Sun & Perona, 1998). They found that when the circles presented were illuminated by light coming from the top-left direction, participants perceived them as convex. When the light originated from the bottom right, participants perceived them as concave circles instead. This illustrates the prior assumption of lighting direction that reflects the fact that most light sources, like the sun and artificial lighting, are located above us (Mamassian & Goutcher, 2001).

1.6. Effect of spatial and temporal context on perceiving low-level / mid-level features

On the one hand, prior knowledge refers to the pre-existing information which could be implicit and developed due to long term exposure to the environment, such as the preference for cardinal orientations (de Valois et al., 1982; Girshick et al., 2011; Li et al., 2003;), or more explicit, such as the preferential perception of older and male faces (Watson et al., 2016). On the other hand, "contextual information" can be defined as perceptual inputs that appear

simultaneously with ("spatial context") or immediately preceding ("temporal context") another input that is the target of recognition.

To quote Schwartz et al. (2007), "No sensory stimulus is an island unto itself; rather, it can only properly be interpreted in light of the stimuli that surround it in space and time." The spatial context of an objects refers to the information that surrounds it, such as the layout of the scene, other typical objects and their arrangement within a scene (Schwartz et al., 2007). The effect of spatial context has been a widely investigated topic within numerous subfields of cognition. A well-known study of context effects on memory was conducted by Godden and Baddeley (1975), who showed that retrieval of learned words were best when the environmental cues present during retrieval were similar to those that were present during encoding, compared to when they were dissimilar. Contextual effects are known to occur within and between different sensory modalities (e.g., visual, auditory). For example, the McGurk effect is a well-known phenomenon that describes how the presence of a visual context alters the perception of sounds (McGurk & MacDonald, 1976). McGurk & MacDonald (1976) found that visual stimuli of a person's lip movement that resembled pronouncing a specific syllable impacted the auditory perception of another syllable.

Psychophysicists have explored the effects of context on the perception of low-level features. Context effects have been found to affect the perception of orientation, object size and lengths of simple stimuli. This has been demonstrated by various optical illusions (for a review see: Todorović, 2010). For example, the Müller-Lyer illusion shows that the direction of V-shaped flankers on the endpoints of a straight line alters the perceived length of the line such that when the two prongs of the 'V' are facing outward, the line appears larger than when the prongs face inward (like an arrowhead). To quote another example, in the Ebbinghaus illusion, the presence

of large or small circles around a target circle affects the perception of the size of the target circle. If the flanking circles are small, it causes the target circle to appear larger. In the Zollner illusion, parallel lines are perceived to be non-parallel due to the addition of short vertical or horizontal lines on the parallel lines. Luminance has also been found to be affected by spatial context. For example, two squares of equal luminance presented against darker and lighter backgrounds are perceived to have different luminances due to the change in background luminance (Gilchrist, 2006).

As opposed to spatial context, temporal context refers to the information that has been available and observed over time. Like the spatial context, the change in information available over time can affect the perception of the target stimulus. With regards to the effect of temporal context on early visual perception, Gibson and Radner (1937) showed that prolonged exposure to a line slightly tilted from vertical subsequently biases the perception of an objectively vertical line to appear tilted in the opposite direction. Eagleman et al. (2004) showed that a briefly presented flash appears brighter than a nearby patch of light with constant luminance. Furthermore, the increase in temporal asynchrony between the onset of the constant luminance patch and the brief flash was also found to increase the perceived brightness of the flash, indicating that the visual system adapts to the prolonged exposure to a patch of light and alters the perception of a new stimulus presented at a later time.

2. Chapter 2 - Effect of presentation time in the classification of hybrids

2.1. Introduction

How perceptual ambiguity is resolved for higher-level features (e.g., semantic categories) that we see in different contexts has not been thoroughly investigated. Kersten et al. (2004) posit that the resolution of ambiguity within our visual system is Bayesian, where our percept is derived from the prior probability and likelihood estimates. Within this framework, when features of a visual stimulus become noisy (i.e., when ambiguity increases), the dependency on priors increases. Therefore, by manipulating the level of ambiguity of an image, we can gauge the extent to which priors play a role in perceiving and classifying images.

It has been previously established within the literature that animate objects hold an advantage over inanimate objects during attentional capture (Delorme et al., 2010; He & Cheung, 2019; New et al., 2007). For instance, it was found that participants were consistently faster in finding animals than man-made objects in visual search tasks (He & Cheung, 2019). Using a change detection paradigm, New et al. (2007) measured the time taken and probability of detecting the changes in objects (either man-made or natural) when a scene and an alternate version of the same scene were presented serially. They found that observers were quicker and more likely in detecting changes to animals and other animate objects than inanimate objects in scenes. New et al. (2007) explained their findings using the animate monitoring hypothesis. This hypothesis states that the category-specific preference to animate over inanimate objects could be attributed to the evolutionary and ancestral priorities since animals and other humans were the most frequently occurring and most relevant features of the environment and were necessary for interactions that facilitate survival.

Contradicting the animate monitoring hypothesis, is the expertise hypothesis, which states that attentional resources are allocated based on the needs, training, context, and goals that develop based on relative ontogenetic importance (New et al., 2007). According to this hypothesis, humans should have developed an attentional preference towards inanimate objects such as vehicles and buildings given that the current (and recent past) environment that surrounds us contains more man-made objects that we interact with. However, based on their findings, New et al. (2007) posit that attentional preference follows the animate monitoring hypothesis than the expertise hypothesis. As further support for the animate monitoring hypothesis, Calvillo and Hawkins (2016) showed that animate objects still biased attention regardless of the threat status of the object. If the animate monitoring hypothesis was based largely on the higher perceived threat posed by animals, threatful inanimate objects (such as guns) should also be detected as likely as threatful animate objects (such as snakes). However, threatful and non-threatful animate objects were found to have a higher detection rate than threatful and non-threatful inanimate objects, indicating that animate objects were still more likely to be detected regardless of whether the object posed a higher risk of threat or not.

More recently however, Hussain Ismail et al. (2019) found evidence for a perceptual preference for inanimate or man-made objects contrary to the animate monitoring hypothesis. They showed participants a series of hybrid images – a complex ambiguous image containing a man-made object component (e.g., vehicles) and a naturally occurring object component (e.g., animals). They found that observers were more likely to categorize these hybrids as man-made images than natural images, eliciting a bias towards man-made objects. Hussain Ismail et al. (2019) speculated that this ‘man-made bias’ observed could be caused by the prolonged exposure and experience of long-term living in man-made environments.

Previous studies have demonstrated that humans have an implicit perceptual preference for lines oriented horizontally or vertically (cardinal) over lines oriented at a slant angle (oblique) (Girshick et al., 2011). This preference corresponds to the statistical regularities of the environment where cardinal edges occur more frequently than oblique edges. Similar biases have been found for relatively more complex features such as motion within the visual system, where observers tend to underestimate the speed of objects due to a preference for low speeds, resulting from prior knowledge that most objects in our environment are not moving or moving at low speeds (Stocker & Simoncelli, 2006). At the semantic level of the visual system, gaze direction and gender expectations are also shown to be biased (Mareschal et al., 2013; Watson et al., 2016). For example, a 'male bias' was observed where participants were biased towards judging androgenous faces as male more often than female and an 'age bias' was also observed which biased the judgement of age from the faces as being older than their own age (Watson et al., 2016). Given that valence and social dominance of an individual is crucial for creating first impressions of unfamiliar faces (Oosterhof & Todorov, 2008), and the fact that older men in societies typically hold socially dominant positions, the 'male and age bias' could have developed due to frequent representations of older and male faces within the visual system (Watson et al., 2016). Hence, the preference to categorise ambiguous images as man-made objects falls in line with the expectation of a highly carpentered, urban environment where objects that occur most frequently are also manufactured.

Since perceptual biases have been found for features at various levels within the hierarchy of the visual system, low-level feature biases could affect high-level features and semantic biases. For instance, adaptation to curved lines caused a bias in perceiving neutral faces

as smiling or frowning (Xu et al., 2008). Hussain Ismail et al. (2019) explored the possibility that the orientation statistics of the environment could have affected the judgement of hybrid images containing a natural and man-made object. It is known that man-made environments more frequently feature cardinal edges, observed in buildings, vehicles, and other man-made structures. In contrast, natural environments feature similar amounts of cardinal and oblique edges, such as in images of trees and mountains. However, they found that the man-made bias persisted even after filtering out the dominant cardinal orientations from the man-made components in the hybrids, eliminating the possibility that biases to low-level features caused the man-made bias.

In this first study, we set out to replicate the 'man-made' bias found by Hussain Ismail et al. (2019). Since it was found that the bias was not influenced by the spatial frequency content of component images, we predict that the same bias would occur even when hybrids made of unfiltered component images are shown to viewers. Since this study will test a fresh group of participants with a new set of components images, it will allow us to confirm whether the man-made bias is indeed a reliably replicable effect. We used a method fairly similar to that used by Hussain Ismail et al. (2019) to create hybrids that are ambiguous at the categorical level, with a few minor changes to suit the nature of our study (see below). Here, images are created by combining images of animals (for the natural component) and vehicles (for the man-made component).

Since this study was conducted during a time where face-to-face participant recruitment was not possible due to the COVID-19 pandemic, the method in which we measured the man-made bias deviates from that used by Hussain Ismail et al. (2019). They measured the bias by presenting a range of hybrids that vary in the ratio of visibility between the two (man-made and

natural) components of the hybrids. However, when running the experiment online, an accurate measure and manipulation of visibility is not possible as images are presented across different computer displays that vary in dimensions, resolution, and luminance. Therefore, given that Hussain Ismail et al. (2019) already showed that the man-made bias is not affected by differences in the visibility of spatial frequency content, we predicted that the bias should still persist even if we do not account for differences in low-level visibility between component images. Accordingly, we combined unfiltered greyscale component images to create a large set of hybrids and measured the proportion of times these hybrids were classified as man-made. If there is no bias, the proportion should be close to 50%.

Further, Hussain Ismail et al. (2019) only used a constant duration of 100 ms for presenting the hybrids. Following Bayesian assumptions, we can expect the magnitude of the perceptual bias to increase as the hybrid stimulus becomes more uncertain, since the likelihood function becomes broader and thus the reliance on prior expectations increase in such cases (Mamassian et al., 2002). However, seemingly anti-Bayesian models of perceptual biases also indicate that whether perception is biased towards the prior or away from the prior depends on the source of uncertainty about the perceived stimulus, i.e., whether it arises from stimulus noise or sensory noise (Wei & Stocker, 2015). Irrespective of the models used in the past to explain perceptual biases, it is quite evident that perceptual biases are susceptible to stimulus parameters, such as the contrast of the stimulus (e.g., Stocker & Simoncelli, 2006). Following this line of research, if the magnitude of the man-made bias would change as a function of the stimulus duration (which would inevitably affect the width of likelihood function), it will further strengthen the speculation that the man-made arises due to reliance on prior expectations. Therefore, in this study, we used two presentation durations, 50 milliseconds (ms) and 150 ms, to

investigate the effect of a varied presentation duration on the persistence and magnitude of the man-made bias. We assumed that 50 ms would not be too short to recognise the component images of the hybrids because past studies have demonstrated that object categorisation can occur reliably even at durations shorter than 50 ms (Thorpe et al., 1996; VanRullen & Thorpe, 2001).

2.2. Methods – Experiment 1

Participants

An a-priori estimate of sample size using the G-Power software (Faul et al., 2007) indicated that to obtain a large effect size (Cohen's d = 0.8, alpha = 0.05, power = 95%), a sample size of 19 participants is required for a one-tailed one-sample *t*-test analysis. We estimated the sample size using a large effect size based on the large effect sizes obtained by Hussain Ismail et al. (2019). We also aimed to compare the size of the man-made bias between the presentation durations of hybrids with a two-tailed paired-sampled *t*-test. A power analysis conducted for this test, with a statistical power of 95% and an alpha value of 0.05, estimated that 23 participants are required to obtain an effect size of Cohen's d = 0.8. Accordingly, 23 participants were recruited and 14 of them were females. Their age ranged between 20 and 25 years (M = 22.6, SD = 1.46). Through a brief preliminary questionnaire, participants self-reported having normal or corrected-to-normal vision. Furthermore, they were asked to report all the names of the locations (e.g., cities and towns) that they have lived in in the past 10 years of their lives to ensure that all our participants have been living in predominantly urban environments for at least 10 years. They were given a small monetary reward of 4 Malaysian Ringgits for participation. The procedures were approved by University of Nottingham Malaysia's Science and Engineering Research Ethics Committee (SEREC ID: HS280521).

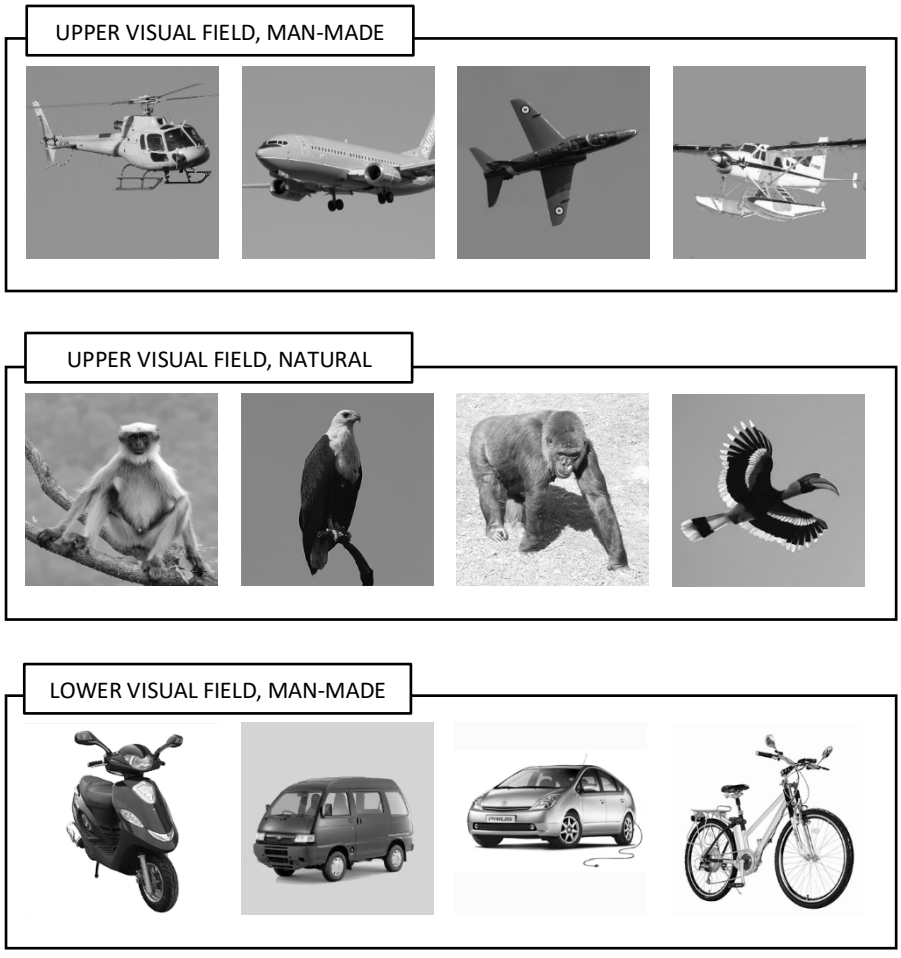
Stimuli

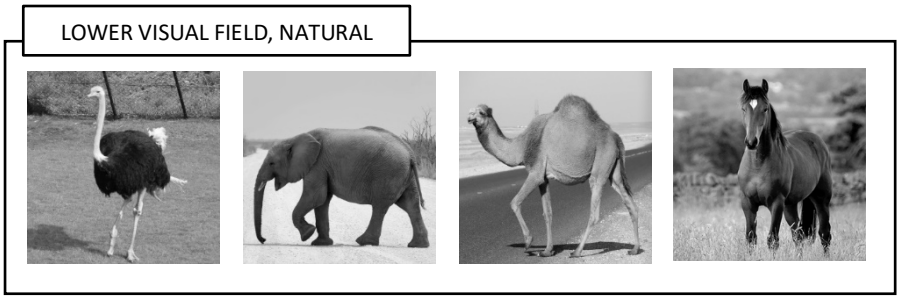
Two-hundred images of man-made objects and 200 images of natural objects were collected to create the hybrid stimuli for the experiments. The images were collected from online databases such as ImageNet (Deng et al., 2009) and other copyright-free online image databases such as Pixabay. The images obtained were man-made (i.e., objects manufactured with human or machinery input) and naturally occurring objects (hereafter "natural objects") that we would commonly encounter in outdoor environments. Images from each superordinate category (manmade and natural) were further divided into two subcategories according to the visual field in which they can be commonly found. Two experimenters (HS and MH) made this allocation based on their subjective judgements. For example, an aeroplane (man-made) and a bird (natural) were categorized as upper visual field objects as they are commonly found in the upper half of the visual field. Similarly, a car (man-made) and a horse (natural) were classified as lower visual field objects (Figure 2.1). One of Biederman's (1981) postulations of relational violations of objects within scenes is the violation of positions of objects in scenes. Biederman (1981) showed that objects undergoing such violations impaired the detection of other objects within the scene. Therefore, to avoid any such confounds within our study, we ensured that only objects from the same visual field were combined in the hybrids, depending on where the hybrid is presented, e.g., a horse-car hybrid or a plane-bird hybrid. A total of 100 images were collected per visual field category and per superordinate category. For the man-made upper visual field category, images of passenger planes, helicopters, fighter jet planes, seaplanes and other light aircrafts were used. For the man-made lower visual field category, images consisted of bicycles, cars, vans, busses, and bikes. For the upper visual field natural category, images of chimpanzees,

eagles, hawks, hornbills, vultures, and monkeys were used. For the lower visual field natural category, images of rhinoceros, horses, camels, ostriches, and elephants were used (figure 2.1). Images were selected such that the target object was not obstructed by any other object present. The background of images with such interfering objects was removed using the software GIMP. The object images were also cropped to fit within a square that is 300 pixels in width and 300 pixels in height.

FIGURE 2.1

Sample stimuli for each category





Note. Examples of images used as stimuli for each visual field category (upper and lower) and superordinate category (man-made or natural).

Once all the object images were compiled, the images were transformed to grayscale by forming a weighted average of the red, green and blue components ($0.299R + 0.587G + 0.114B$; Hughes et al., 2014). Next, a validation study was conducted on six participants (2 males); age range between 22 & 29 years (M = 24.5, SD = 2.58). This was done to ensure that all the object images were unambiguously recognizable as man-made or natural objects when they were presented. Each grayscale image was presented for a duration of 150ms at the centre of a grayscale screen of dimension 1024 x 768 pixels. The distance between the participant's eyes and the monitor was fixed at 60cm using a headrest. Participants were instructed to press the 'm' if they recognise the object as a man-made object and 'z' if it is a natural object. The order of the images presented was random. The validation study yielded an average categorization accuracy of 93.75% (SD = 4.22) for man-made objects, 92.83% (SD= 6.93) for natural objects and an overall categorisation accuracy of 92.76% (SD = 5.77). Accordingly, we decided to use the compiled images as stimuli to create hybrids.

All hybrids were created on MATLAB (Version R2020b). The image pairs consisting of one natural and one man-made object image were randomly selected from the image sets according to the visual field category. Henceforth, the man-made and natural images that make up the hybrid image will be termed component images. Pixel values (between 0 and 255) of all component images were altered to have fixed root-mean square (RMS) contrast. For naturalistic images, contrast sensitivity in humans is found to be best predicted by RMS contrast compared to its Michelson Contrast (Bex & Makous, 2002). Equating component images for RMS contrast only assured that the relative contrast sensitivity between the two component images will be similar. However, it must be noted that, since the hybrids are going to be shown in different computers, their dimensions in pixels will be scaled, and it may alter the RMS contrast of the hybrids as a whole. That aside, a total of 300 unique hybrid images were created for each participant. Each hybrid was created such that no two hybrids contained the same image from each superordinate image category (Figure 2.2). The 300 images were split to have 150 images for each presentation time – 50 ms and 150 ms. The 150 images within each presentation time were further divided to have an equal number of hybrids per visual field (75 each).

FIGURE 2.2

Sample hybrid stimuli

A)  B) 

Note. Examples of hybrid image created. A) shows a lower-visual field hybrid where the man-made and natural components are of objects typically found in the lower visual field. B) shows an upper visual field hybrid.

A mask stimulus that is 300 by 300 pixels in dimension was created for each of the 300 unique hybrid images by phase-scrambling the hybrid image itself. To do so, each hybrid image was Fourier transformed to obtain its phase spectrum, to which we added an array of uniformly distributed random values between 0 and 1. The altered phase spectrum was combined with the hybrid's original amplitude spectrum and reverse Fourier transformed to create the mask stimulus. Masks were used to cease the processing of the hybrid stimuli after the stipulated presentation duration and to suppress image aftereffects that could impact performance in subsequent trials.

The screen background was set to be at mid-level grey. The hybrid images and the masks were scaled to have a square width of 6cm, irrespective of the computer displays used by the

participants. This made sure that, when the participant is seated at a distance of 60 cm from the screen, a square with a width of 6 cm subtends 5.7° of visual angle.

Although the images were created externally on MATLAB for each participant, we used Psychopy (v2021.1.4) (Peirce et al., 2019) to create the experimental routine. The experimental routine and the uniquely generated stimuli were uploaded to Pavlovia, which is a platform that allows participants to use their web browsers and run the experiment online.

Procedure

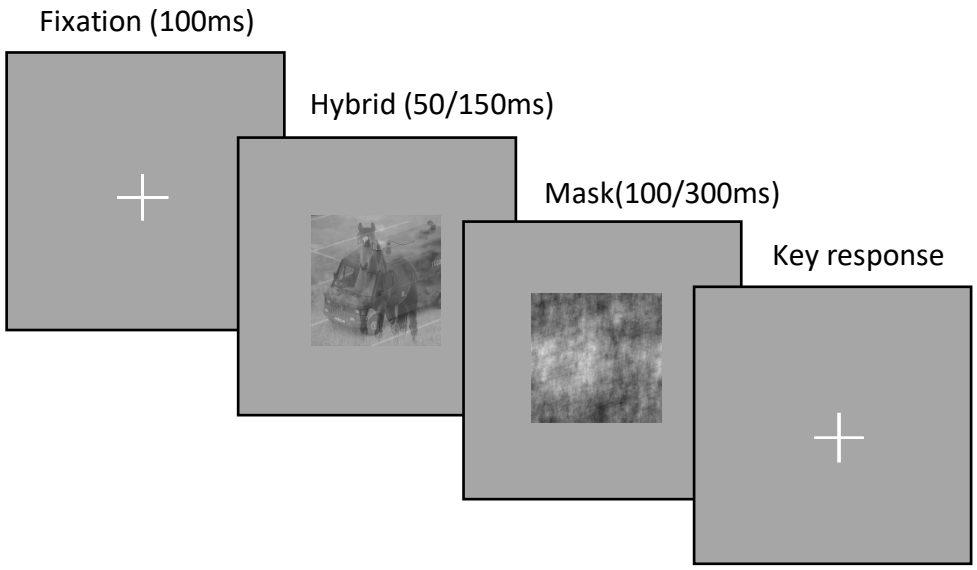
After the observers read and signed the consent form given online, they were sent a hyperlink to the experiment hosted on Pavlovia. To ensure that no more than one participant used the uniquely generated set of stimuli, they were scheduled to a time slot within which they will be able to access the study using the link while the experimenter was available online to troubleshoot any errors that the participant might encounter. Participants were told to sit comfortably with their laptop's or desktop monitor's screen directly ahead of them, approximately 60 cm away from their eyes.

The experiment started with a screen size calibration task where they were required to adjust the size of an image of a credit card on the screen to match that of their standard physical credit or debit card placed on the screen. Participants used arrow keys in the keyboard to adjust the virtual image. To ensure that the calibration was close to perfect, participants were asked to measure a white square shown on their screen and ensure that it was 10 cm in width and height. If their measurements were beyond 10 ± 0.2 cm, they were instructed to alert the experimenter before restarting the calibration task. Once the calibration task was successfully completed, the participants were provided with the instructions for the experimental task. They were told that a series of images will be briefly presented on their screen and that for each image they should

press the “m” key if the man-made object in it was the most visible to them and the “z” key if the natural object was the most visible. Each experimental trial started with a white fixation cross (size = 0.95°) that was presented at the centre of the screen for 100ms. Following this, a unique hybrid image was presented for either 50 or 150ms at the centre of the screen. Immediately after the offset of the hybrid, its respective mask image was presented for twice the duration of the hybrid (100 or 300ms). The subsequent trial commenced only after a key response was provided (Figure 2.3). After 150 trials, halfway through the experiment, they were given a short 5-minutes break. Once the experiment ended, they informed the experimenter to ensure that their data was saved. The experiment took around 30 minutes to complete.

FIGURE 2.3

Timeline of progression of a trial in experiment 1



2.3. Results

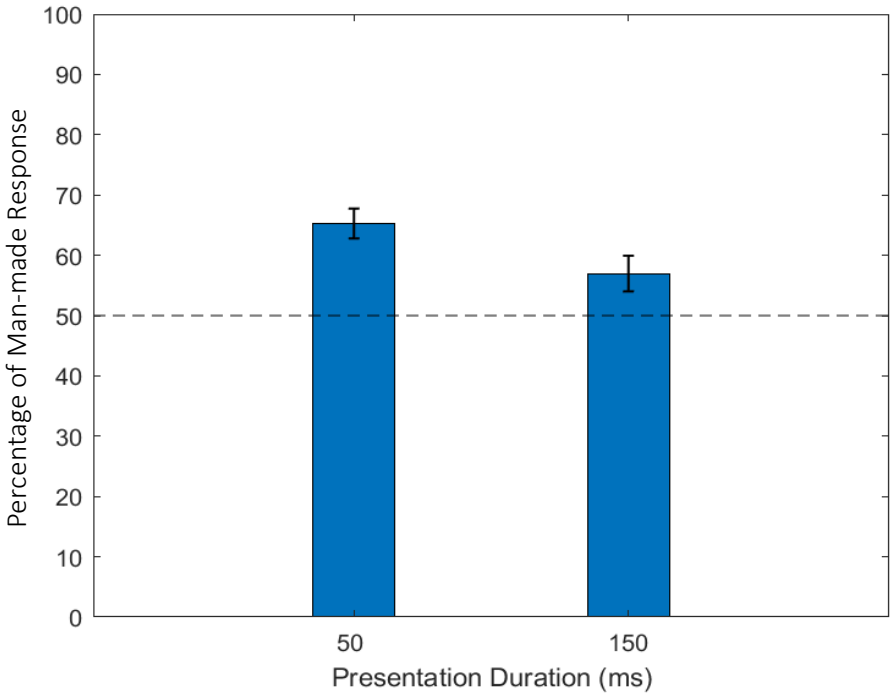
Participants' data were retrieved from Pavlovia's repository. For each participant, the percentage of man-made responses was obtained by dividing the total number of 'man-made' responses (trials in which participants pressed the 'm' key) by the total number of trials per presentation duration. This was done separately for each presentation duration.

A one-tailed one-sample *t*-test was conducted for each presentation duration to check if a man-made bias occurred. The test value was set at 50% with the rationale that if the man-made bias is observed, participants should have classified the hybrid as man-made object on more than 50% of the trials. When the hybrids were presented for 50 ms (M = 65.27, SD = 11.89) a significant man-made bias was observed $t(22) = 6.16$, $p < .001$ (Cohen's d = 1.284) . Similarly, for the 150 ms presentation duration (M = 56.98, SD = 14.22) there was a significant man-made bias too, $t(22) = 2.35$, $p = .014$ (Cohen's d = 0.490).

Next, a two-tailed paired-samples *t*-test was conducted to compare the average percentage of man-made responses between the two presentation durations. The average percentage of man-made responses at 50 ms (M = 65.27, SD = 11.89) was significantly higher than that of 150ms (M = 56.98, SD = 14.22), $t(22) = 4.61$, $p < .001$ (Cohen's d = 0.955). This indicates that the man-made bias is larger when the presentation duration is short (see figure 2.4).

FIGURE 2.4

Percentage of man-made responses across stimulus duration



Note. The graph shows the mean percentage of man-made responses across all participants for each presentation duration (50 and 150ms). Error bars indicate standard error of mean. The dashed line represents the 50% test value used, indicating that for both duration conditions, participants made more man-made responses than natural responses.

2.4. Discussion

Our results support our hypothesis that when observers are presented with ambiguous hybrid images, classification is largely biased towards its man-made objects components instead of its natural object components. This finding successfully replicates the man-made bias found by Hussain Ismail et al. (2019) and indicates that the bias persists despite using a different experimental design and running the experiment online with a totally new set of object images

and a new set of participants. Given that all our participants have spent a significant amount of time living in heavily carpentered urban landscapes, the bias corroborates the possible expectation to see man-made objects within their environment more frequently than natural objects. This finding also goes against the "animate monitoring hypothesis" posited by New et al. (2007), who claim that humans have an ancestral bias that causes a shift of attention towards perceiving animals and other humans than inanimate objects. If this ancestral and evolutionary bias caused the animal component of the hybrid to be preferentially attended towards, participants would have reported classifying the hybrid to the natural category more frequently than the man-made category. One could argue that the animal components in the hybrids were not detectable at the short duration in which we presented our hybrids. However, this argument would not hold true because we have strong evidence from the literature showing us that even when images are presented for less than 50 ms, people are able to reliably detect the presence of animals in images (Delorme et al., 2010; Thorpe et al., 1996).

A key finding of Experiment 1 is that when hybrids are viewed in our central visual field, we find an effect of the hybrid presentation duration on the magnitude of the man-made bias. More specifically, the magnitude of the bias was significantly larger at the 50 ms duration compared to the 150 ms duration. It is quite obvious that at the 50 ms duration, our hybrid stimulus has relatively high uncertainty. Therefore, this finding is in line with previous literature showing that as the uncertainty of the stimulus increases, the magnitude of perceptual biases increases. This effect has not only been demonstrated for perceptual biases associated with fundamental visual features such as orientation (Girshick et al., 2011) and speed of motion (Stocker and Simoncelli, 2006), but also for higher-level, meaningful features of everyday objects such as gaze direction (Mareschal et al., 2013) and gender and age (Watson et al., 2016)

of faces. Extending these finding, we report novel results showing that stimulus uncertainty also increases the magnitude of a perceptual bias demonstrated in classifying the superordinate category of objects. Collectively, these findings favour the Bayesian assumption that as stimulus uncertainty increase, likelihood functions for the stimulus feature would widen and we would rely more on prior expectations. With regards to our study, as stimulus duration decreased, this would have increased the uncertainty associated with the hybrid, which may have then caused our participants to rely more on their prior expectations to see man-made objects more often in their environment.

Although our findings are clear in suggesting an increase in the magnitude of the bias with increasing stimulus uncertainty, we cannot be certain about the decisional processes that contribute to the man-made bias or the change in its magnitude with shorter stimulus durations. While conventional Bayesian models would consider stimulus uncertainties arising from sensory and stimulus noise alike, recent models, especially ones that aim to explain perceptual biases caused by prior expectations developed with long-term exposure to the environment, distinguish between the different sources of uncertainty. For instance, Wei and Stocker (2015) show that when uncertainty is caused by stimulus noise (e.g., by degradation), perceptions is biased towards the direction of the priors, but when uncertainty arises from sensory noise (e.g., short presentation durations) perception is biased away from the priors. However, this prediction contradicts with our findings, since increase in sensory noise shifted our bias further towards the prior (if there is one). Notably though, while Wei and Stocker (2015) modelled perceptual biases with a single unimodal likelihood function, we could also argue that the likelihood function for our hybrid stimuli is bimodal - for instance, if we assume that man-made and natural categories are two ends of a higher-level feature representing the superordinate category of objects, for any

given hybrid stimulus, we could have one peak near the man-made component and one near the natural component, and this leads to a bimodal likelihood function associated with one single hybrid stimulus. In such a case, how decisional processes would operate within the Bayesian framework is unclear and future studies would need to address this computational issue.

3. Chapter 3 – Effect of visual field on the classification of hybrids

3.1. Introduction

In Experiment 1, we confirmed the presence of a man-made bias when hybrids were shown in the centre of our visual field (foveal and parafoveal regions), when participants were directly fixating at the hybrids. In this experiment, we set out to investigate if the man-made bias is persistent in the peripheral visual field as well. One of the roles of peripheral vision is to monitor and detect changes within the environment while the central vision is more task focused (Li et al., 2021). Therefore, it is of ecological interest that we understand how objects that appear in the peripheral visual field is detected and interpreted, and whether prolonged exposure to specific environmental settings can affect the perception of objects in the periphery.

Peripheral vision processes visual inputs that we receive outside the foveal and parafoveal regions of the retina. Foveal (central) vision subtends approximately 1.7 degrees of visual angle within our visual field (Rosenholtz, 2016). It is known that the fovea is densely packed with cone cells which is sensitive to high spatial frequency information and affords colour vision. On the other hand, peripheral vision, which covers the rest of the visual field, is said to contain mostly rod cells that have a reduced sensitivity to colour and high-spatial frequency information and, therefore, provide a relatively impoverished visual experience (Rosenholtz, 2016). Poor visual acuity in the periphery is often attributed to the reduced density of photoreceptors and retinal ganglion cells in the peripheral vision (Strasburger et al., 2011). Additionally, the cortical area allocated to process 1 degree of visual field within the fovea is much larger compared to the cortical area allocated to process information received from 1 degree of visual field in the periphery (Horton & Hoyt, 1991). However, even if we compensate

for poor resolution in the periphery by enlarging stimuli, recognising shape and form from objects in the periphery is difficult, suggesting that the source of difficulty cannot be limited to poor resolution alone (see Strasburger et al., 2011 for a review).

Due to a number of reasons, including but not limited to reduced spatial resolution in peripheral vision, object recognition performance within the peripheral visual field is thought to be subpar compared to recognition performance in central vision. Investigations of this phenomenon have been done using simple artificial stimuli, such as Gabor patches (Jüttner & Rentschler, 2000; Strasburger et al. 1991) as well as with more naturalistic stimuli such as line drawings and photographs of objects we encounter in real life (Biederman, 1972; Henderson et al., 1989; Thorpe et al., 2001). For example, Thorpe et al. (2001) used images of natural photographs presented at various eccentricities from fixation for an extremely short duration (28ms) to investigate how naturalistic objects are recognised in the peripheral visual field. Using a Go-No-Go task, participants were asked to respond if they saw an animal within the presented image of a natural environment. They found that the participants could successfully categorize objects presented in the peripheral visual field at the superordinate level, although their accuracy decreases and reaction time increased, with increasing eccentricity. Although object recognition performance decreased roughly linearly as a function of eccentricity, even when objects are presented at an eccentricity of 70.5°, people can reliably classify them above chance (Thorpe et al., 2001), demonstrating that object recognition is poor but not absent in the periphery.

Although past studies have demonstrated that classification of unambiguous objects is reliable in the periphery, how we classify ambiguous objects in the periphery is not clear. More specifically, we do not know if the man-made bias we (Experiment 1) and Hussain Ismail et al. (2019) observed in the fovea would persist when ambiguous hybrid stimuli appear in our

periphery. As far as perceptual biases driven by prior expectations are concerned, although there is evidence that we can observe these biases in the periphery (Girshick et al., 2011; Stocker and Simoncelli, 2006), there are also studies showing that biases for certain perceptual judgements driven by prior knowledge may be weaker or even absent in the periphery compared to the fovea. For example, Zhaoping (2017) demonstrated that when participants judge the form of ambiguous dichoptic signals, their judgements were biased by the expectation that signals to both eyes correlate. Although this bias was quite significant in the fovea, it was weak, or rather absent in the periphery. Zhaoping's (2017) findings verified that the bias is caused by top-down feedback, and that weak biases in the periphery are a result of week top-down feedback in the periphery. Not only do such findings imply that the functional role of central and peripheral vision is different, but they also give us a good reason to examine whether the man-made bias, that is seemingly driven by our expectations, would persist in the periphery too.

Accordingly, we presented our participants hybrid images in 6 peripheral locations equidistant from fixation. We randomly presented hybrids at one of 6 locations to ensure that participants do not direct their attention to a predictable location prior to seeing the hybrid, and therefore ruling out any effects of pre-cued attention on hybrid classification. We also presented hybrids at two different durations (50 ms and 150 ms) because we found stimulus duration to influence the magnitude of the bias in Experiment 1 where hybrids were directly fixated at. Both durations at which we presented hybrids were sufficiently shorter to prevent making any eye movements to the hybrids. However, there is still some possibility that participants could make an eye movement to one of the six hybrid locations before the hybrids appear. In that case, if the spatial position they fixate by anticipation overlaps with the hybrid's position, any man-made biases in judgements could be falsely interpreted as a bias that occur in the periphery, when in

fact hybrids were viewed in the central visual field. To minimise this possibility, we gave participants a secondary task that required paying attention to the colour of a fixation cross that appeared immediately before the hybrid, thus preventing eye movements before hybrid onset. Lastly, everyday objects are often best detected when they appear in their typical visual field (e.g., an airplane in the upper visual field, a chair on the lower visual field; see Kaiser et al., 2019 for a review). To ensure that any one component image in a hybrid does not have an advantage in being detected over the other component due to mismatches in typical visual field, we also made sure that only component images typical of the same and specific visual field (upper or lower) are paired together to create hybrids.

3.2. Methods – Experiment 2

Participants

Following the same sample size estimation in Experiment 1, 23 participants were recruited for Experiment 2 as well. Twenty out of the 23 participants from Experiment 1 agreed to take part in this experiment too. The remaining three were new participants. There were 15 females, and the age range was between 20 and 27 years (M = 22.7, SD= 1.75). The new participants were also given the questionnaire to report the locations that they have lived for the past 10 years to ensure that all our participants were living in urban environments. All participants had normal or corrected-to-normal vision and received a compensation of 4 Malaysian Ringgit for participation. The study was approved by the University of Nottingham Malaysia's Science and Engineering Research Ethics Committee (SEREC ID: HS280521).

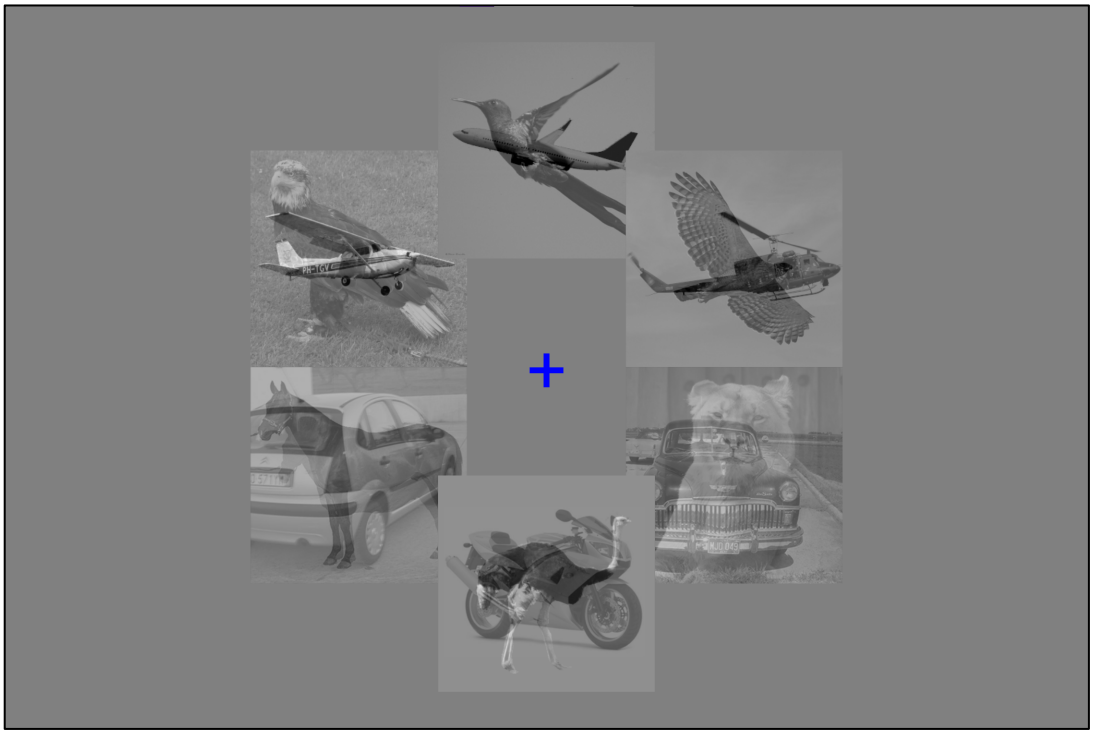
Stimuli and Setup

Images used for this experiment were taken from the database that was created during Experiment 1. To investigate if the man-made bias was present when hybrids were presented in

the periphery, 6 locations equidistant from the centre of fixation were selected (Figure 3.1). The distance from the centre of the fixation to the centre of the hybrid stimuli was approximately 5.58°. Three of the peripheral locations belonged to the upper visual field (i.e., above the horizontal meridian where the fixation point was placed) and the three others belonged to the lower visual field (i.e., below the horizontal meridian where the fixation pint was placed). The hybrids presented at the upper and lower visual field consisted of component images from image sets prepared for the upper and the lower visual field, respectively. Akin to Experiment 1, 300 hybrids were presented in total with the same division of hybrids per presentation duration (50ms and 150ms) and visual field category. In addition, 75 images presented at each visual field category were equally divided among the three locations, which resulted in 25 hybrids per location. The contrasts of the component images were equated the same way using MATLAB as described in chapter 2 and the experiment was setup on PsychoPy and was accessible online through Pavlovia. Each participant was provided with a link to access the experiment on their own computer. Participants were instructed to sit comfortably with the laptop/desktop screen placed straight ahead of them, preferably on a table at a distance of 60 cm between their eye and the screen.

FIGURE 3.1

Peripheral stimuli positions



Note. The 6 possible spatial locations equidistant from fixation at which the hybrid images were presented. The hybrids presented above the fixation were upper visual field hybrids and those presented below the fixation were lower visual field hybrids.

Procedure

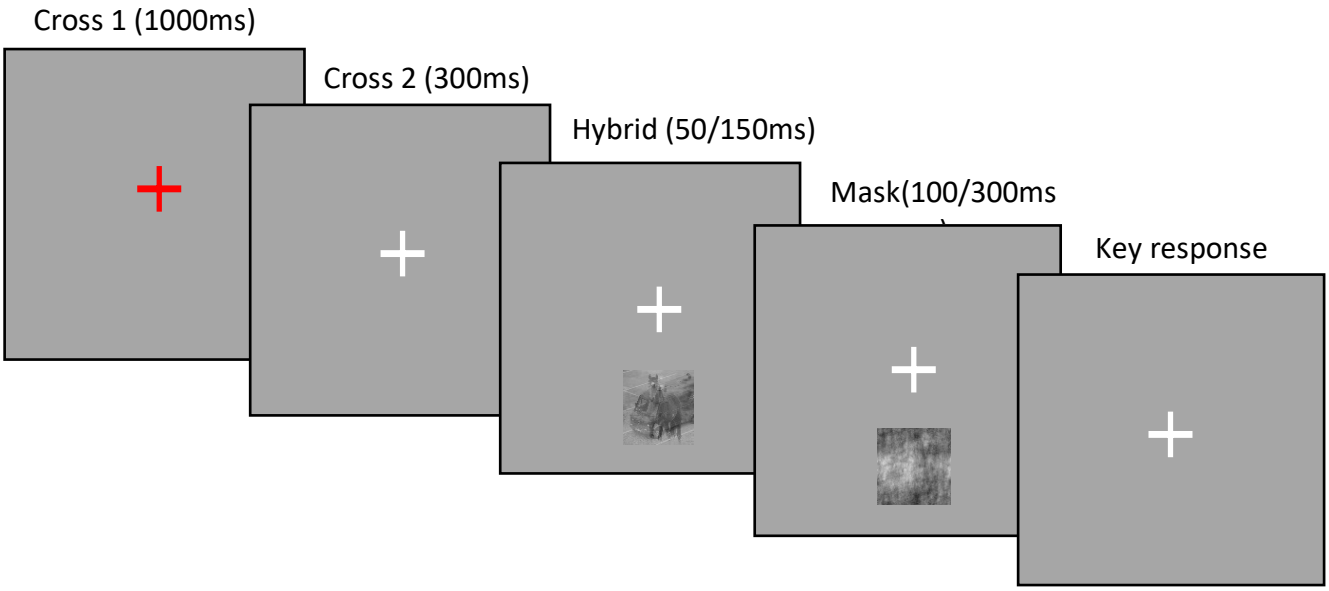
The calibration task from experiment 1 was also used to ensure that the object images were presented at a constant size regardless of the participant's device's screen size. This procedure also ensured that distance to peripheral hybrid locations from fixation were constant across participants. Each one of the 300 experimental trials started with a coloured fixation cross

("cross 1"; size = 0.95°) that was presented at the centre of the screen for 1 second, and this was immediately replaced by a white fixation cross ("cross 2") that was displayed until the participant pressed one of the designated keys. To ensure that the participants' gaze was fixated at the centre of the screen, we randomly varied the colour of cross 1 to be either red or blue across the various experimental trials. The hybrids were presented after a short delay of 300 ms from the offset of cross 2. This ensured that paying attention to the colour of the fixation cross did not make participants miss the hybrid. After the hybrid was shown for 50 ms or 150 ms (depending on the experimental condition), participants had to indicate which component of the hybrid was the most visible to them. cross 1 was blue, they pressed "m" for a "man-made" response and "z" for a "natural" response. If cross 1 was red, they pressed "p" for a "man-made" response and "q" for a "natural" response. Cross 2 remained on the screen when participants were making a response. Once a response was made, the next trial started after a brief delay of 300ms (Figure 3.2).

Participants were given 20 practice trials to familiarize themselves with the different key responses they will use in the main task. The images used in the practice trials were never repeated during the main trials. Similar to experiment 1, participants were given a break at the halfway point during the experimental trials.

FIGURE 3.2

Timeline of a trial in experiment 2



3.3. Results

Each response given by the participants was only considered valid if they pressed one of the keys designated for the colour of the fixation cross (e.g., only "m" or "z" is a valid response for a trial starting with the blue fixation). The total number of valid responses out of the total of 300 trials was calculated for each participant (i.e., both durations combined). The average percentage of valid responses across participants was 92.24% (SD = 5.74), with a minimum of 74.67% and a maximum of 98%. These values indicate that participants performed well in the secondary task that was used to minimise eye movements away from fixation.

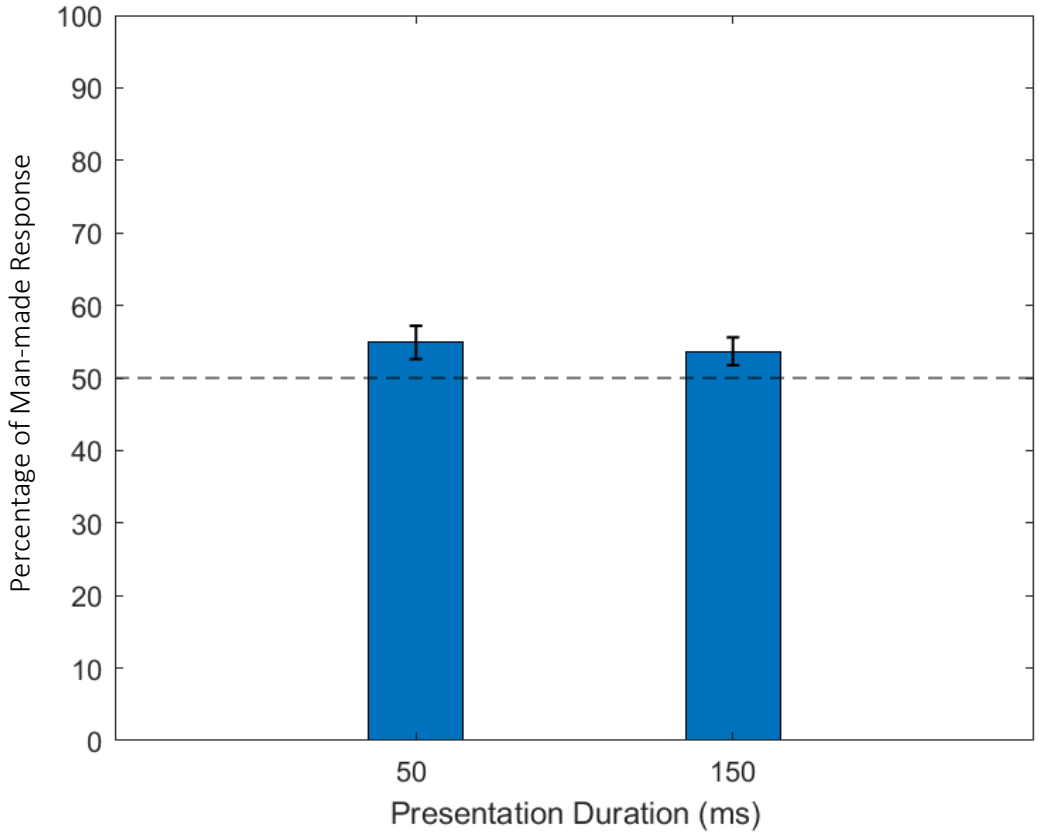
Including only the valid responses, the total number of responses for each presentation time was calculated. Out of that, the total number of man-made responses was calculated. The percentage of man-made responses for each presentation duration was then calculated as the total

number of man-made responses divided by the total number of valid responses. A one-tailed one-sample *t*-test conducted for each presentation time indicated that participants classified the hybrid as man-made more frequently at 50ms (M = 54.79, SD = 11.11), $t(22) = 2.06$, $p = 0.025$ (Cohen's d = 0.431) and at 150ms (M = 53.67, SD = 9.28), $t(22) = 1.89$, $p = 0.035$ (Cohen's d = 0.395) . A paired samples *t*-test comparing the percentage of man-made responses between the two presentation durations showed that categorization performance did not differ significantly between the two durations, $t(22) = 0.46$, $p = 0.646$ (Cohen's d = 0.097) (Figure 3.3).

The analysis was repeated without removing the invalid responses from each participant's data and a one-tailed one-sample *t*-test indicated that the hybrids were classified more frequently as man-made at 50ms (M = 55.04, SD = 11.39), $t(22) = 2.13$, $p = 0.023$ (Cohen's d = 0.442) and at 150ms (M = 53.36, SD = 9.40), $t(22) = 1.72$, $p = 0.050$ (Cohen's d = 0.357). A paired samples *t*-test indicated that the percentage of man-made responses did not differ significantly between the two presentation durations $t(22) = 0.70$, $p = 0.488$ (Cohen's d = 0.147).

FIGURE 3.3

Percentage man-made responses per stimulus duration



Note. Mean percentage of man-made responses across all participants for each presentation duration when the hybrids were presented in the peripheral visual field. Error bars indicate the standard error of mean.

3.4. Discussion

We found that for both presentation durations, participants showed a bias towards classifying the hybrids as man-made. However, the means and the effect sizes indicate that the man-made biases observed in Experiment 2 are relatively smaller compared to those observed in Experiment 1. One could argue that perhaps some proportion of the man-made bias in Experiment 2 may have been lost by excluding data from trials in which an invalid key was

pressed. However, the same analyses performed after including invalid responses produced very similar results. Therefore, it is clear that the overall magnitude of the man-made bias is smaller in the periphery.

According to the animate monitoring hypothesis, attention is preferentially captured by animate than inanimate objects (New et al., 2007). It is possible that preferential attention to animate objects is more pronounced in the periphery (especially when covert attention is employed), large enough to override the man-made bias to some extent. At this point, this is merely a speculation. Moreover, one could easily attribute the reduced bias to a general difficulty in recognising the individual components as well as the hybrid as a whole in the periphery. As discussed in the introduction section of this chapter, object categorization performance tends to be poorer in the periphery compared to the centre (Rosenholtz, 2016; Henderson, 1989; Thorpe, 2001). There are a number of reasons to expect recognition to be difficult in the periphery, such as poor resolution (Strasburger et al. 2011). If recognition is indeed difficult, participants would be forced to guess in most cases, and the man-made bias could have been the result of a few trials where component images were relatively more visible. One way to address this in the future is to account for limitations in the periphery (e.g., poor spatial resolution) when presenting hybrids (e.g., by scaling them to account for poor resolution). This would help us narrow down the source of the reduced bias in the periphery.

One other explanation for the reduced bias is that the task demands (of having to switch between response key sets) might have increased the difficulty of classifying the hybrid. Around 10 participants also verbally reported after completing the experiment that they perceived the task to be strenuous, even though the percentage of valid responses they made were very high. Therefore, it is possible that our participants paid too much attention to the secondary task, to the

extent that it impairs their ability to covertly attend to the hybrid, which would then result in guesses for most trials. Last possible interpretation is that the effect we observe is independent of task difficulty and limitations of peripheral vision. As Zhaoping (2017) demonstrated, the effect of prior expectations on sensory interpretations may be weaker in the periphery, thus leading to a generally smaller bias.

Another finding of Experiment 2 is that there was no difference in the magnitude of the bias between the two presentation durations, unlike Experiment 1. One potential explanation is that biases in both durations were not large enough to capture subtle differences in biases between durations, owing to the general difficulties in performing the task. An alternative explanation is that task difficulty, caused by limitations in peripheral vision or competition from the secondary task, may have influenced the bias in the 50 ms presentation duration than in the 150 ms duration. It is possible that these additional effects reduced a larger bias that would have otherwise been observed with a short presentation duration, consequently resulting in similar magnitudes of biases between the two durations.

4. Chapter 4 – Effect of background context in classifying hybrids

4.1. Introduction

Objects that we see in our everyday environments do not often appear in isolation. We always see them as an item in a larger meaningful scene. How the scene in which an object is embedded affects our ability to recognise the object itself is a topic that has been highly investigated in the literature. These effects are conventionally referred to as contextual effects on object recognition. One of the factors that cause contextual effects is the semantic congruency between the object and the background scene. An example of this semantic congruency would be a football player (object) in a football field (scene) or a priest in a church. If a priest appears in a football field, or if a footballer appears in a church, they would be example where the object and its scene are semantically incongruent.

Semantic congruency (compared to incongruency) is known to facilitate object recognition by increasing the speed of recognition as well as increasing the accuracy (Beiderman, 1981; Davenport & Potter 2004; Freidman, 1979; Munekke, 2013). It is important to note that the effect of a semantically congruent scene on object recognition is not caused by low-level feature similarity or the overlap of shapes between the object and the scene, and it is not dependent on whether we pay focused attention to the object or the background scene (Munneke et al. 2013). This contextual effect can be explained by the gist processing model and scene schema models. The former indicates that a scene's low-resolution gist is gauged prior to identifying objects within the scene (Biederman, 1981). The low-resolution scene information provides an expectation for the objects found within the scene, and thus facilitating the recognition of objects that match the schema of the given scene (Schyns & Oliva, 1994).

Furthermore, it has been found that this semantic congruency between objects and scenes affects scene recognition too (Davenport & Potter, 2004; Loschky, 2007; Palmer, 1975). Palmer (1975) showed participants line drawings of visual scenes such as a kitchen for 2 seconds and subsequently flashed an image of an object for either 20, 40, 60 or 120ms. The objects presented after the scene were either appropriate to the context (a loaf of bread), inappropriate but was similar in shape to the appropriate object (a mailbox) or a completely different object in terms of shape and semantic congruency (a circular drum). Participants were asked to guess the name of the object they saw. Palmer (1975) found that the probability that the target object was correctly named was highest when the object was appropriate to the context and lowest when the object was inappropriate for the context. Davenport and Potter (2004) showed that when the scene and object were congruent, participants' accuracies in naming the object and naming the scene were better compared to scenarios where the object and scene were semantically incongruent. This indicated that objects and the scene in which they occur mutually influence the ability to recognise each other. Bar and Ullman (1996) suggested that the effect of objects on scene recognition could be explained by the concept of 'context frames'. Context frames are representation of prototypical scenes stored in our memory which consist of objects that typically occur in the scene. During recognition, these objects can activate the prototypical scene in which the object typically occurs in and subsequently enables the recognition of other objects that occur within the scene. Therefore, when Davenport and Potter (2004) showed participants an image of a priest, this activated the prototypical context frame within which a priest usually occurs (i.e. the church), thus facilitating the recognition of a church scene.

While plenty of studies demonstrate the effect of a scene on an object that is unambiguously recognisable when seen in isolation, some studies have also examined the effect of the background scene on recognising objects that are rather ambiguous when seen alone. This effect of context in resolving object ambiguity on a semantic level of perception has been well demonstrated for faces. To quote a real-life example, a photo of a woman crying without the background scene would most commonly be classified as experiencing an adverse event. However, the context might indicate that she was crying out of joy while experiencing a highly positive event, such as meeting her all-time favourite celebrity. Aviezer et al. (2012) demonstrated this using a behavioural experiment. In their study, viewer could not discriminate the valence between intense positive and negative facial expressions extracted from real-life scenarios when they were presented alone. However, they were reliably discriminated when presented along with the expresser's body language.

As far as objects other than faces are concerned, it remains unclear how the background scene would affect their classification. We know from our Experiment 1 and Hussain Ismail et al.'s (2019) study that we have a bias to classify ambiguous objects as man-made when they are directly fixated. In an attempt to study contextual influences on this man-made bias, we decided to examine whether a simultaneously presented background scene (i.e., spatial context) would alter the magnitude and/or the direction of the man-made bias when hybrids are viewed at fixation. To do this, we manipulated the background context in which ambiguous hybrids appeared, at a superordinate level (i.e., using scenes that can be classified as man-made and natural). Here, any given scene would be semantically congruent with one component of the hybrid but not the other. For instance, a natural scene (e.g., a forest) would be categorically consistent with an animal in the hybrid but not a vehicle. In accordance with past studies

demonstrating semantic congruency effects on object recognition and studies on faces showing the effect of context in resolving ambiguous facial expressions, we hypothesised that the man-made bias would be larger when the background is man-made compared to natural, since a man-made scene should facilitate the recognition of the man-made (but not the natural) component in the hybrid. If the influence of the background is strong, we can expect the man-made bias to in fact reverse (i.e., show a bias for the natural component of the hybrid) when the background is natural. Further, we expect effects of the spatial context to be stronger when the hybrid (and the background) is shown for a longer duration (150 ms) compared to a shorter duration (50 ms), since the man-made bias is stronger at shorter durations when hybrids are directly fixated at.

We decided to only test two presentation durations (50 and 150 ms) for a number of reasons. First, we know that scene gist can be recognised at very brief presentations much shorter than 50 ms, even if the scenes are presented in the periphery (Thorpe et al., 2001). Therefore, if there is an effect of the background scene's gist on the hybrid, the durations we present can be sufficient to capture those effects. Second, these are the two durations for which we have already demonstrated a man-made bias. Given that the man-made bias appears to reduce in magnitude with increasing presentation duration (Experiment 1), testing any longer durations may dissipate the bias. We specifically wanted to examine the effect of context when the bias is at play. Lastly, using longer durations would likely cause exploratory eye movements, which may result in the hybrid appearing the participants' periphery when being classified.

4.2. Methods – Experiment 3

Participants

Following sample size estimates for Experiments 1 and 2, we recruited 23 participants (14 females), of which 19 participants participated in both experiments 1 and 2. Furthermore, we aimed to compare the size of the man-made bias between two presentation durations and two background context conditions using a 2x2 repeated measures ANOVA, A power analysis conducted for this test with a statistical power of 95% and an alpha value of 0.05, estimated that 24 participants are required to obtain a large effect size in accordance with Hussain Ismail et al. (2019). The participants age ranged between 19 and 27 years (M = 22.43, SD = 1.76). They had normal or corrected-to-normal vision. In a brief questionnaire, they reported the names of the locations at which they resided in the past 10 years. This ensured that all our participants have been living in urban environments for at least 10 years of their lives. They were given a small monetary compensation of RM 4 for participation. The study was approved by the University of Nottingham Malaysia's Science and Engineering Research Ethics Committee (HS280521).

Stimuli

Eight videos of man-made outdoor environments and eight videos of natural environments were collected from copyright-free sources on the internet. From each video, a 2-minute clip was extracted. The video clips of man-made environments consisted of first-person point of view walking videos of urban streets taken in various countries such as Japan, Malaysia and the UK. We attempted to minimise the number of other human beings walking within the video to ensure minimum animate objects that would be considered as natural are present in the video. The video clips of natural environments consisted of first-person point of view walking videos through different forests and natural areas such as paths alongside rivers and waterfall. Next, all video clips were processed on MATLAB (Version R2020b), where all the video frames were transformed into grayscale.

The hybrid stimuli used in this experiment were created from the component image repositories used for Experiments 1 and 2. The following steps were performed to create sets of hybrid-background composites (i.e., hybrids superimposed on background scenes) that are required for Experiment 3.

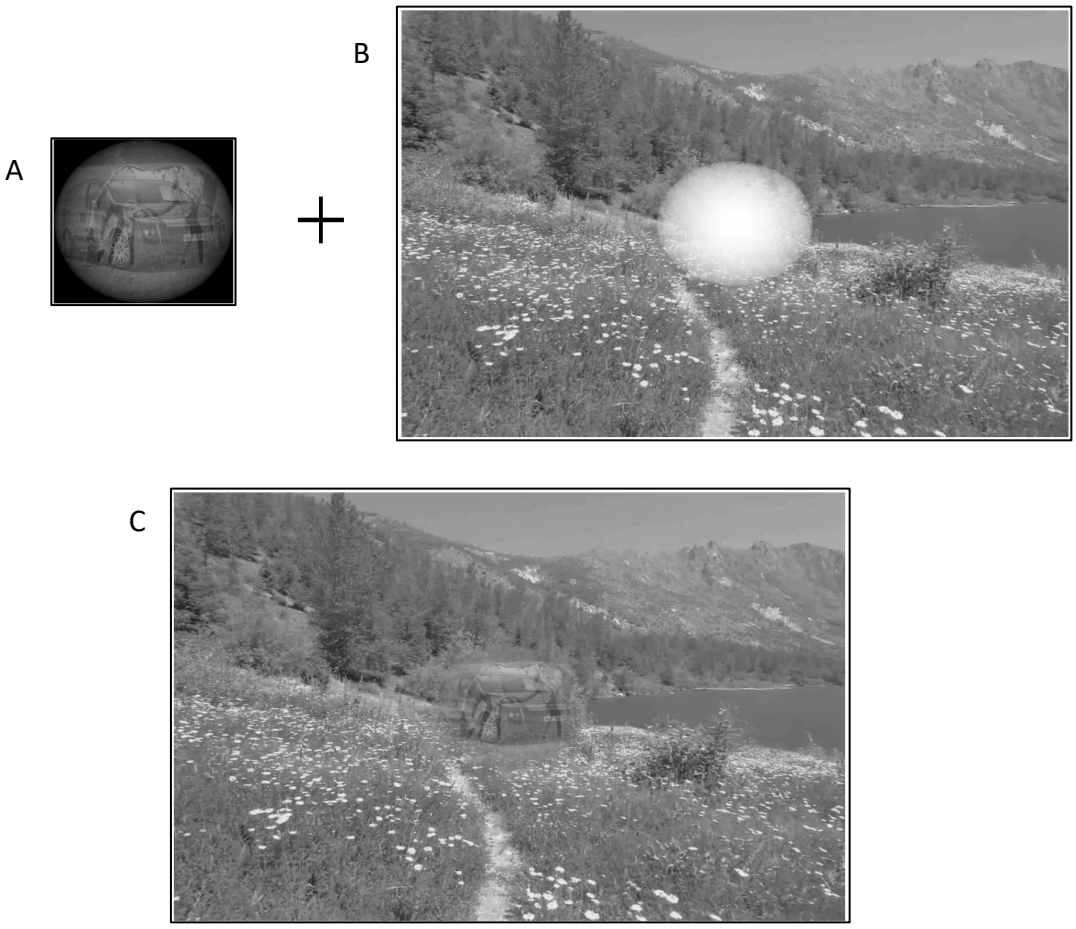
Each greyscale video clip we had comprised of a total of 2400 frames. Accordingly, for the first step of creating hybrid-background composites, we sampled 16 frames from each clip, with mean sampling interval of 150 ($SD = \pm 15$) frames between successive samples. This method was opted instead of a completely random selection of video frames so that consecutive frames will not be chosen, ensuring that background images used are as distinct as possible. Following this method, we sampled background frames for each participant separately. Next, the RMS contrasts of the hybrids' components images were equated as in Experiments 1 and 2, and then superimposed onto each other to create hybrids. Component images for hybrids were selected and combined following the same procedure used in Experiment 1. For example, images from the upper visual field from man-made and natural categories were randomly selected and combined to create hybrids, and similarly, images from the lower visual field categories were also combined to create hybrids.

A cosine window (with the same width and height in pixels as the hybrids) was applied to the hybrids as well such that the centre of the window has the maximum weight (i.e., maximum transparency), which gradually decreased towards the edge of the window (i.e., making them opaque) (Figure 4.1). Following this, at the centre of the background scene a circular grayscale cosine window, the same size as the hybrid, was created such that the centre of the window had a 100% mid-grey opacity that decreased towards the edges. The hybrid image was then

superimposed onto this grayscale window in the background scene by adding the pixel values of the hybrid with those of the background frame. When superimposing the hybrid image onto the background frame, the hybrid's outline blends to the background frame without creating an artificial-looking edge between the hybrid image and the background. Lastly, phase randomised masks of the hybrid-background composites were created using the same phase randomising procedure as in experiments 1 and 2.

FIGURE 4.1

The process of creating a hybrid-background composite.



Note. A) shows the image of a hybrid that has been multiplied by a cosine window. B) shows the background scene selected for the composite with a grayscale cosine window at the centre where the hybrid is to be superimposed. C) the resulting hybrid-background composite.

A total of 300 hybrid-background composites were created, with 150 unique composites for each category of background context (man-made and natural) (Figure 4.2). The 150 composites were then divided equally between the two presentation durations (50 and 150 ms). The images were presented in two blocks according to the category of the background scene, such that all the hybrids superimposed onto the natural context were presented in one block of trials and the hybrids superimposed on the man-made context were presented in a separate block. The order of presenting the blocks was randomised between participants. The presentation times were randomised across trials within each block. The hybrid images at the centre had a diameter of 6 cm and subtended 5.7 degrees of visual angle, and the background frame and the mask image had a width of 25.6 cm and a height of 20.8cm which subtended 10 degrees of visual angle. The study follows a 2x2 design where the two independent variables are the context categories (Natural and Man-made) and the presentation duration (50 ms and 150 ms). Similar to the earlier two experiments, the dependent variable is the percentage of man-made responses.

Like the previous two studies, the experiment routine was created on PsychoPy (v2021.1.4) (Peirce et al., 2019). The stimuli for each participant generated on MATLAB along with the experimental routine from PsychoPy was uploaded onto Pavlovia's repository.

FIGURE 4.2

Examples of hybrid images for each context condition.

A)

B)

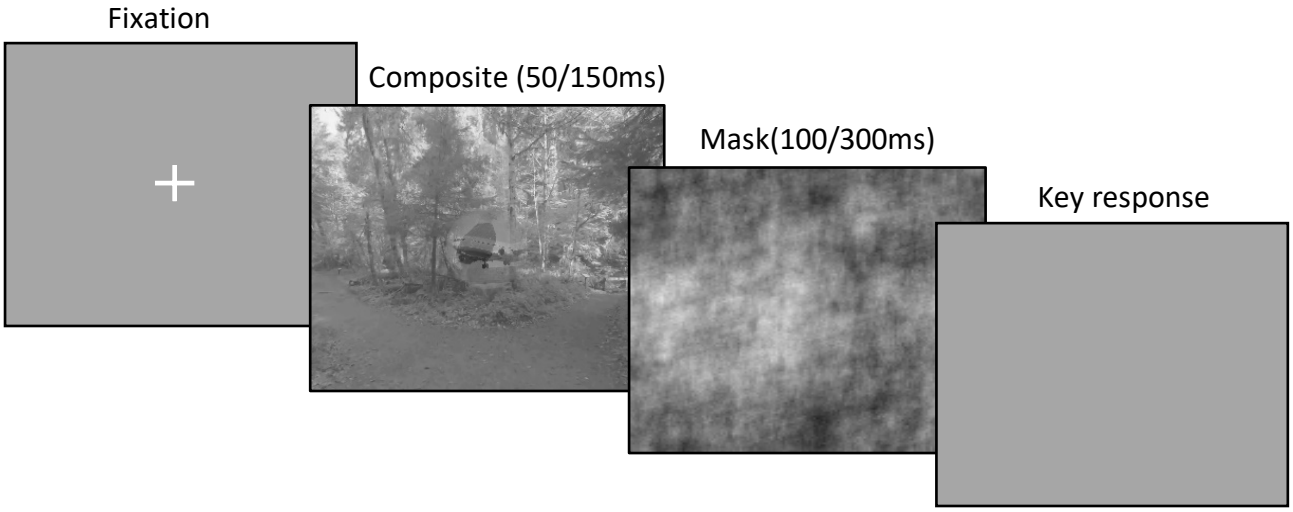
Note. Examples of hybrid images for each context condition. A) a hybrid image embedded onto a man-made background scene. B) a hybrid image embedded onto a natural background scene.

Procedure

After participants gave their informed consent, they were sent a link to the experiment hosted on Pavlovia. Participants were instructed to sit comfortably at a distance of approximately 60 cm between their eyes and their laptop screen. The experiment started with the screen size calibration task where they were required to adjust the size of an image of a credit card on the screen to match that of their standard physical credit or debit card placed on the screen (see Experiment 1 Methods). Once the calibration task was completed, the participants were provided with the instructions for the experimental task. They were given 10 practice trials (5 for each context condition) to familiarise themselves with the procedure. The composites used during the practice trials did not repeat during the experimental trials. A fixation cross appeared at the centre of the screen for 1 second, followed by the hybrid-background composites shown for either 50 or 150ms. Immediately after, the phase-scrambled mask appeared for twice the duration of the hybrid before returning to the blank mid-grey background. Participants were asked to press the key “z’ if the natural object at the centre of the screen was most visible to them and “m” if the man-made object was the most visible. They were allowed after the onset of the mask, and the trial terminated when a response was made. Between the two experimental blocks, participants were encouraged to take a short break for 5 minutes. The experiment took around 30 minutes to complete in total, and they were debriefed and compensated after completion.

FIGURE 4.3

The timeline of a trial in experiment 3



4.3. Results

The percentage of man-made responses for each context condition and each presentation duration was calculated. A One-tailed one-sample *t*-test was done on the percentage of man-made responses for each of the 4 experimental conditions separately (natural context-50 ms, natural context-150 ms, man-made context-50ms, man-made context-150ms), to test whether a man-made bias was present in each condition. If the man-made bias is present, the mean across all participants should be significantly greater than 50%. As shown in Table 4.1, a significant man-made bias was found for all 4 experimental conditions.

TABLE 4.1

One-sample t-test result for every context and stimulus duration

Context/ stimulus duration (s)	M (%)	SD	t - statistic	p - value	Cohen's d
MM/0.05	69.04	12.07	7.56	<.001	1.577
MM/0.15	68.46	11.30	7.83	<.001	1.637
N/0.05	63.82	14.17	4.68	<.001	0.975
N/0.15	64.52	11.46	6.07	<.001	1.267

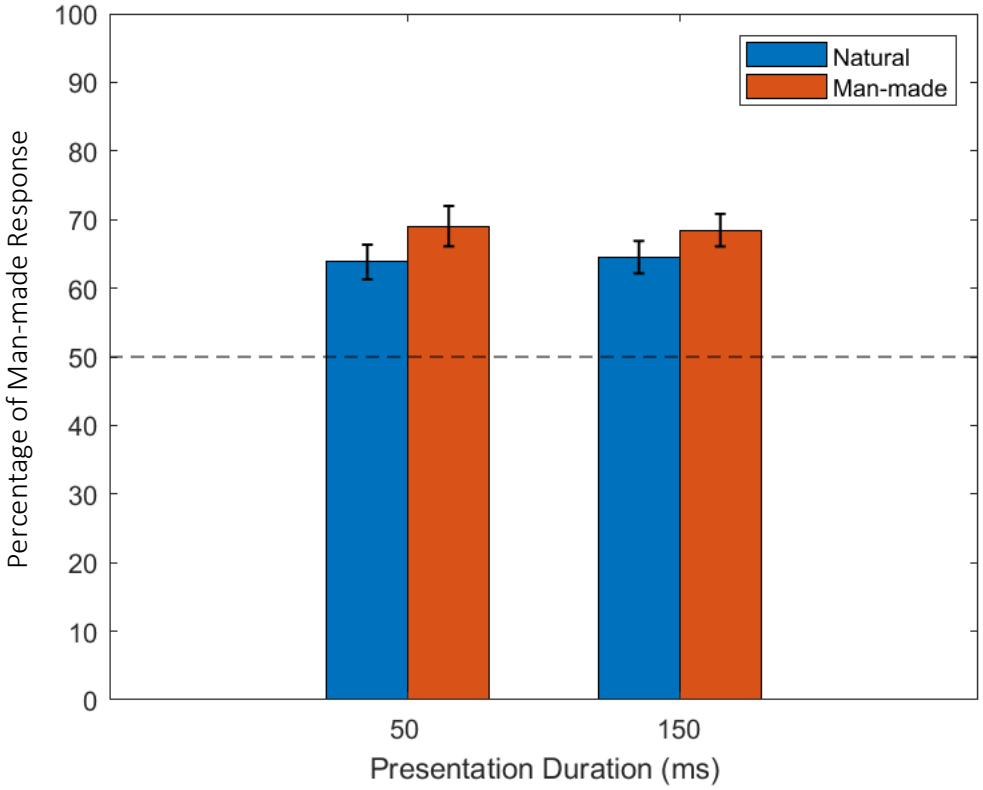
Note. The mean percentage of hybrids classified as man-made, standard deviation and t-statistic for the one sample t-test conducted for each of the four groups. The test value was 50%. MM refers to the man-made context, N refers to the natural context.

Next, a repeated-measures ANOVA was conducted to investigate the effect of background context and stimulus duration on the percentage of man-made responses. There was no significant main effect of background context on the categorisation of hybrids, $F(22) = 3.26$, $p = 0.084$, although the p-value corresponds to an effect that seems to be approaching significance. This indicates that the mean percentage of man-made responses when the background context was a natural environment (M = 64.17, SD = 11.24) did not differ significantly from when the background context was a man-made environment (M = 68.75, SD = 10.47). There was also no significant main effect of stimulus duration, $F(22) = 0.00$ $p = 0.979$, indicating that the mean percentage of man-made responses for hybrids presented for 50ms (M = 66.43 , SD = 13.12) did not differ from the mean percentage of man-made responses of hybrids that were presented for 150ms (M = 66.49 , SD = 11.38). The interaction between the background context and the

stimulus duration did not yield a significant effect, $F(22) = 0.34$, $p = 0.562$. The mean percentage of man-made responses for each condition is visualised in figure 4.4.

FIGURE 4.4

man-made responses per background context and stimulus duration



Note. The mean percentage of hybrids classified as man-made when the background context was either man-made (orange bars) or natural (blue bars) and the presentation durations were either 50 ms or 150 ms. Error bars denote the standard error of the mean.

4.4. Discussion

The results showed that participants were more likely to categorise hybrids as man-made irrespective of the duration of presentation and the background context in which hybrids were superimposed on. Our results did not support the hypothesis that the background context will influence the direction or magnitude of the categorisation bias. Participants tended to classify the hybrid as a man-made object when the background was a man-made environment as well as a natural environment. This main effect seems to be approaching significance, and the means indicate that the effect is in the direction we expected, i.e., smaller man-made bias for hybrids in a natural context compared to those in a man-made context. However, given that there is no statistically significant effect, we prefer to not make any claims based on this effect that is approaching significance. Interestingly, the effect of presentation time is also absent, indicating that participants categorised hybrids similarly at both presentation durations and that this did not affect the persistence of the man-made bias. Although this seemingly contradicts with the findings from Experiment 1 (where more man-made responses were made for the 50ms duration compared to the 150ms duration), it is important to note that hybrids had no context in Experiment 1, making a direct comparison between the experiments difficult. For now, we can only conclude that the presence of a scene in the background eliminates the effect of presentation duration on the magnitude of the man-made bias.

As far as the finding regarding the effect of the background context is concerned, one possible explanation for it is that the man-made bias is caused by a more deeply rooted expectation present in those who have spent a significant portion of their lives in urban environments, such that it supersedes the effect of any simultaneously presented context. If this was the case, it would be interesting to compare performance on this task with a population that has not spent a significant amount of time living in urban environments and instead have lived in

naturalistic environments, such as forest-dwelling indigenous communities (e.g., some Orang Asli tribes in Malaysia).

We could also attribute the lack of a difference in the man-made bias (or a weak difference) between the two context conditions to an interplay between two opposing effects, one low-level and one high-level. A study by Cannon and Fullenkampt (1991) showed that the apparent contrast of a uniformly oriented target sinusoidal grating depended on the orientation of gratings in an immediately surrounding disk. When the two orientations did not differ the apparent of the target was low, whereas when the two orientations differed (e.g., with 60° angular separation) the apparent contrast was relatively high. We can extend this to the context of our study. Although everyday scenes are often dominated by cardinal orientations, this dominance is stronger in man-made scenes (Coppola et al., 1998) and man-made objects (Torralba & Oliva, 2003), compared to natural scenes and natural objects, respectively. When hybrids are presented in man-made scenes, dominant cardinal orientations in the scene may suppress the apparent contrast of the man-made component in the hybrid whose structure is largely defined by cardinal orientations. This would reduce the detectability of the man-made component, and it would make the participants classify the hybrids more often as natural. When this effect occurs in parallel with an opposing high-level effect where man-made scenes would facilitate the detectability of the man-made component in the hybrid due their semantic congruency, they would cancel each other and leave the man-made bias unchanged (compared to a scenario where there is a natural context). One way to rule out low-level effect like the one described above is to equate component images for low-level feature visibility as Hussain Ismail et al. (2019) did. However, due to the online nature of our experiment, we could not achieve this.

More realistically, however, our results can be attributed to experimental limitations as the data suggests a very mild and almost significant effect of context on the bias. As the experiment was done online, we could not control for a variety of factors that would have been otherwise controlled in the lab with in-person experiments. Most importantly, in the absence of a controlled, dimly lit and empty environment of a lab, the presence of other man-made objects and sounds in the location at which the participants performed the task would have been a source of distraction or represented an implicit secondary context. For example, suppose participants performed the experimental task at a home office that overlooks a busy road with loud vehicles and other buildings. In that case, these sounds and peripheral views may have influenced their performance in the task. Such influences can be largely eliminated when experiments are conducted in quiet, windowless and dimly lit indoor spaces, so that the only visual stimuli participants receive is from the computer screen.

5. Chapter 5 - General Discussion

When sensory information received by our visual system is ambiguous, prior knowledge and contextual information can bias perception. For instance, a photo of a woman crying without the background would most commonly be classified as experiencing a negative event; however, the context might indicate that she is crying as she experiences an extremely positive event, such as meeting her celebrity (Martinez, 2019). Some perceptual biases have been found to vary along with the variation in properties and features of the environment, such as the bias towards cardinally oriented lines over oblique lines due to the former's prevalence within the environment (Girshick et al., 2011). The prior knowledge of the regularity of features in the environment has also been found to bias object perception. Hussain Ismail et al. (2019) demonstrated that humans living in the cities are biased in classifying ambiguous hybrid images as man-made objects rather than natural objects. However, as these hybrid images were presented in isolation without context, it is unclear whether this bias is caused due to the long-term exposures to man-made environments and whether this bias can be manipulated in an experimental setting by altering the spatial context within which these ambiguous images appear.

Therefore, in this thesis, we addressed three questions regarding the top-down influences of context on the process of resolving perceptual ambiguity by adapting the technique of creating ambiguous images containing man-made and natural components called hybrids (Hussain Ismail et al., 2019). First, we set out to replicate the man-made bias effect and examine whether altering the duration of the presentation of hybrids affects the magnitude of the bias. Second, we investigated whether the man-made bias persisted when the location of the hybrids varied within the visual field. Lastly, we investigated if the context of a man-made or natural environment can

alter the performance in categorising the ambiguous hybrid such that they're classified more frequently as natural objects when the context is a natural environment and man-made objects when the context is man-made environments.

In the first experiment, we replicated the man-made bias using an online experimental design, a new image set and participants. This established that the man-made bias as a robust phenomenon and provide a basis for the subsequent experiments. Furthermore, we investigated the effect of altering presentation duration of the hybrid on the magnitude of the man-made bias. According to the Bayesian framework, when more ambiguity is induced into the stimuli (in this case by reducing the presentation duration), the perceptual system is known to rely more on prior knowledge whilst making perceptual judgments (Mamassian et al., 2002). We found that participants showed a man-made bias when the hybrids were presented for 50 and 150ms. The magnitude of the bias at 50ms was higher than that at 150ms, suggesting that the magnitude of the perceptual biases increases along with the increase in ambiguity of the stimuli due to the increased reliance on prior expectations of frequently occurring objects in the environment. The results of this experiment also provide against the animate bias that was previously found that states that observers more readily shifted their attention towards animate objects than inanimate objects (New et al., 2007).

In the second experiment, we investigate if the man-made bias in the peripheral visual field. The common notion is that visual information received from the peripheral vision is low in resolution and provides an impoverished visual experience (Rosenholtz, 2016). However, this low-resolution visual information has been found to be beneficial for scene gist processing (Oliva & Schyns, 1997; Larson & Loschky, 2010). Although performance for object recognition in peripheral vision is not perfect, humans are still able to categorise objects into broad

categories, such as animal or non-animal (Thorpe et al., 2001; Juttner & Rentschler,2000; Biederman, 1991). Furthermore, perceptual biases have also been observed in the periphery (Stocker & Simoncelli, 2006), although it remains unclear how and if observers can reliably classify ambiguous objects presented in the periphery. To examine whether the man-made bias would persist in peripheral vision, we presented hybrid images at 6 peripheral locations, equidistant from the centre in the periphery for either 50 or 150ms and found that a man-made bias did occur for both presentation durations. However, the magnitude of the bias was very small. Furthermore, contrary to the first experiment, we observed no effect of presentation time as there was no increase in magnitude observed when the hybrids were presented for 50ms than for 150ms. This indicates that the man-made bias persists within the peripheral visual field, although the effect sizes indicated that the bias observed was relatively smaller compared to the central visual field. In line with Zhaoping's (2017) finding that the bias towards the judgements of ambiguous dichoptic signals were weaker in the periphery than in central vision, one plausible explanation for our findings is that the man-made bias is weaker in the periphery than in the central visual field.

Lastly, we investigated the effect of a spatial context on the classification of hybrid images. The results from Hussain Ismail et al. (2019) and from the first experiment demonstrated that the preference towards categorising ambiguous hybrids as man-made objects occurs possibly due to the expectation to see man-made objects within the heavily carpentered environment that we live in today. However, it is unclear whether the change in environmental context would affect the presence of this bias. The semantic congruency effect and the gist perception theory posits that scenes are recognised rapidly before objects are clearly identified, and this gist can affect subsequent recognition of objects (Biederman, 1981; Friedman, 1979; Davenport & Potter,

2004). Therefore, we investigated whether the presence of a background scene creates a semantic expectation of the objects frequently found in certain environments and if this subsequently affects the classification of hybrid images. We superimposed hybrid images at the centre of a larger background scene that depicted either an outdoor man-made environment, such as a street, or a natural environment, such as a jungle trail. We presented these superimposed hybrid-background composites for either 50 or 150ms. We found that a man-made bias persisted among all four experimental conditions (two background context conditions and two presentation durations), indicating that participants were more likely to categorise the hybrids as man-made irrespective of the duration and context condition. This result did not support our hypothesis that the background context will influence the direction of the categorisation bias. Furthermore, there was no significant main effect of context or presentation duration. This indicates that the man-made bias could be caused a more deeply rooted expectation present in the population living in urban environments such that the alteration of the immediate spatial context does not affect the bias.

Since the experiments were conducted online, numerous extraneous variables could have affected the results that we obtained. Although clear instructions were provided to participants regarding the distance they should maintain between themselves and the computer screens, we cannot control subconscious head tilts and movements towards or away from the screen during the experiment. The placement of their personal computers or laptops at their desk may also be at an angle rather than at exact eye level. Furthermore, since the laptop and computer models vary across participants, it is hard to control or measure the luminance of the screen and the images, which may have affected how participants perceived the hybrid stimuli. Other distractions present within the participant's room may also have introduced significant noise in our data.

Therefore, it would be worthwhile to replicate these experiments in a controlled laboratory setting to confirm if the results found here are robust or confounded by these extraneous variables.

Apart from the limitations posed by the online nature of the experiments, another factor that could have impacted our data is the response bias caused by the use of a 2 alternate-force-choice task. As participants were required to categorise the hybrid they viewed as either man-made or natural objects, they might have been forced to select one of the two options regardless of whether they could reliably recognise the object. Other studies that investigated the semantic congruency effect on non-ambiguous objects used an object recognition paradigm, where participants were required to name the object that they perceived (Davenport & Potter, 2004; Munekke et al., 2013). A similar paradigm could be employed in future variations of this study to ensure that participants are attentive and able to perceive the hybrids accurately.

Nevertheless, the result from this thesis indicates that the man-made bias is a robust effect that is persistent across varying presentation durations, visual field location and context conditions. From this, we infer that the man-made bias is caused by long-term exposures to man-made environments. Such a bias allows us to make perceptual judgements of ambiguous objects more accurately within the surroundings that we live in and provides an ecological advantage. However, to test if this bias is caused due to long-term exposures to certain environments, further studies should be done investigating if a change in temporal context- by varying the length of exposure to the different environments- alters the classification of ambiguous objects. It would also be interesting to see if this classification bias is modified to suit the environment by investigating the performance in populations living in environments that are starkly different to urban environments, such as the forest-dwelling indigenous populations.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, *21*(2), 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406. https://doi.org/10.1038/387401a0

Anna Nikki (n.d.). *Style Classic Vintage Retro Cars Clipart* [Clip art]. Pngitem. https://www.pngitem.com/middle/mhhhTm_style-classic-vintage-retro-cars-car-clipart-vintage/

ArtsyBee (2017). *Building* [Clip art]. Pixabay. https://pixabay.com/illustrations/buildings-clip-art-color-isolated-2933455/

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225–1229.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352. https://doi.org/10.1068/p250343

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A*, *19*(6), 1096–1106. https://doi.org/10.1364/JOSAA.19.001096

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*(4043), 77–80. https://doi.org/10.1126/SCIENCE.177.4043.77

Biederman, I. (1981). On the Semantics of a Glance at a Scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual Organization* (pp. 213–253). Routledge. https://doi.org/10.4324/9781315512372-8

Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, *94*(2), 115–147. https://doi.org/10.1037/0033-295X.94.2.115

Biederman, I., & Cooper, E. E. (1991). Object recognition and laterality: Null effects. *Neuropsychologia*, *29*(7), 685–694. https://doi.org/10.1016/0028-3932(91)90102-E

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177.

Brewer, A. A., Liu, J., Wade, A. R., & Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, 1102–1109. https://doi.org/10.1038/nn1507

Calvillo, D. P., & Hawkins, W. C. (2016). Animate Objects are Detected More Frequently than Inanimate Objects in Inattentional Blindness Tasks Independently of Threat. *Journal of General Psychology*, *143*(2), 101–115. https://doi.org/10.1080/00221309.2016.1163249

Cannon, M. W., & Fullenkampt, S. C. (1991). SPATIAL INTERACTIONS IN APPARENT CONTRAST: INHIBITORY EFFECTS AMONG GRATING PATTERNS OF DIFFERENT SPATIAL FREQUENCIES, SPATIAL POSITIONS AND ORIENTATIONS. *Vision Research*, *31*(11), 1985–1998.

Clipart-Library (n.d). *Tree Cartoon PNG #1835480* [Clip art]. Clipart-Library. http://clipart-library.com/clipart/rcLoj67Xi.htm

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, *95*(7), 4002–4006. https://doi.org/10.1073/PNAS.95.7.4002

Crouzet, S. M. (2011). What are the visual features underlying rapid object recognition? *Frontiers in Psychology*, *2*(NOV), 1–15. https://doi.org/10.3389/fpsyg.2011.00326

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

de Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559. https://doi.org/10.1016/0042-6989(82)90113-4

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Psychology*, *0*(JUN), 21. https://doi.org/10.3389/FPSYG.2010.00021/BIBTEX

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. *IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Deyoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., & Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–2386. https://doi.org/10.1073/pnas.93.6.2382

Eagleman, D. M., Jacobson, J. E., & Sejnowski, T. J. (2004). Perceived luminance depends on temporal context. *Nature*, *428*(6985), 854–856. https://doi.org/10.1038/NATURE02467

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*(2), 181–192. https://doi.org/10.1093/CERCOR/7.2.181

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Felleman, D. J., & van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1

Ferster, D., & Miller, K. D. (2000). NEURAL MECHANISMS OF ORIENTATION SELECTIVITY IN THE VISUAL CORTEX. *Annual Review of Neuroscience*, *23*, 441–471.

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. https://doi.org/10.1037//0096-3445.108.3.316

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. *Journal of Experimental Psychology*, *20*(5), 453–467. https://doi.org/10.1037/H0059826

Gilchrist, A. (2006). Seeing Black and White. *Seeing Black and White*, 1–448. https://doi.org/10.1093/ACPROF:OSO/9780195187168.001.0001

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932. https://doi.org/10.1038/nn.2831

Godden, D. R., & Baddeley, A. D. (1975). CONTEXT-DEPENDENT MEMORY IN TWO NATURAL ENVIRONMENTS: ON LAND AND UNDERWATER. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/J.2044-8295.1975.TB01468.X

Greene, M. R., & Oliva, A. (2009). The Breifest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, *20*(4), 464–472.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Human Brain Mapping*, *6*(4), 328. https://doi.org/10.1002/(sici)1097-0193(1998)6:4<316::aid-hbm9>3.0.co;2-6

He, C., & Cheung, O. S. (2019). Category selectivity for animals and man-made objects: Beyond low- and mid-level visual features. *Journal of Vision*, *19*(12), 22–22. https://doi.org/10.1167/19.12.22

Helmholtz, H. von. (1925). *Treatise on Physiological Optics (Southall, J. P. C., Trans). Electronic edition (2001): University of Pennsylvania.* The Optical Society of America. %0Ahttp//psych.upenn.edu/backuslab/helmholtz

Henderson, J. M. (1992). Object identification in context: the visual processing of natural scenes. *Canadian Journal of Psychology*, *46*(3), 319–341. https://doi.org/10.1037/H0084325

Henderson, J. M., & Hollingworth, A. (1999). HIGH-LEVEL SCENE PERCEPTION. *Annual Review of Psychology*, *50*, 243–271. https://doi.org/10.1146/ANNUREV.PSYCH.50.1.243

Henderson, J. M., Pollatsek, A., & Rayner, K. (1989). Covert visual attention and extrafoveal information use during object identification. *Perception & Psychophysics 1989 45:3*, *45*(3), 196–208. https://doi.org/10.3758/BF03210697

Hess, R. F., Hayes, A., & Field, D. J. (2003). Contour integration and cortical processing. *Journal of Physiology, Paris*, *97*(2–3), 105–119. https://doi.org/10.1016/J.JPHYSPARIS.2003.09.013

Hochstein, S., & Ahissar, M. (2002). View from the Top. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Archives of Ophthalmology (Chicago, Ill. : 1960)*, *109*(6), 816–824. https://doi.org/10.1001/ARCHOPHT.1991.01080060080030

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1961). Integrative action in the cat's lateral geniculate body. *The Journal of Physiology*, *155*(2), 385. https://doi.org/10.1113/JPHYSIOL.1961.SP006635

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/JPHYSIOL.1962.SP006837

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/JPHYSIOL.1968.SP008455

Hughes, J. F., van Dam, A., McGuire, M., Foley, J. D., Sklar, D., Feiner, S. K., & Akeley, K. (2014). *Computer graphics: principles and practice*. Pearson Education.

Hummel, J. E. (2000). Where View-based Theories Break Down: The Role of Structure in Shape Perception and Object Recognition. In E. Deitrich & A. Markman (Eds.), *Cognitive Dynamics: Conceptual Change in Humans and Machines* (pp. 157–185).

Hussain Ismail, A. M., Solomon, J., Hansard, M., & Mareschal, I. (2019). A perceptual bias for man-made objects in humans. *Proceedings of the Royal Society B*.

J4p4n (2014). *Italian Bicycle* [Clip art]. Openclipart. https://openclipart.org/detail/192023/italian-bicycle

Jastrow, J. (1899). The mind's eye. *Popular Science Monthly*, *54*, 299–312.

Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 1–18. https://doi.org/10.1167/8.13.11

Jüttner, M., & Rentschler, I. (2000). Scale-invariant superiority of foveal vision in perceptual categorization. *European Journal of Neuroscience*, *12*(1), 353–359. https://doi.org/10.1046/J.1460-9568.2000.00907.X

Kaas, J. H., & Krubitzer, L. A. (1991). Organisation of Extrastriate Visual Cortex. In *Neuroanatomy of Visual Pathways* (pp. 302–323).

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. v. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/J.TICS.2019.04.013

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Kim, J. G., & Biederman, I. (2011). Where Do Objects Become Scenes? *Cerebral Cortex*, *21*(8), 1738–1746. https://doi.org/10.1093/cercor/bhq240

Laobc (n.d.). *Faceless Woman Walking Clipart* [Clip art]. Creazilla. https://creazilla.com/nodes/3272509-faceless-woman-walking-clipart

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology*, *90*(1), 204–217.

Li, M. S., Abbatecola, C., Petro, L. S., & Muckli, L. (2021). Numerosity Perception in Peripheral Vision. *Frontiers in Human Neuroscience*, *15*, 674. https://doi.org/10.3389/FNHUM.2021.750417/BIBTEX

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513–536. https://doi.org/10.1080/13506280902937606

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The Importance of Information Localization in Scene Gist Recognition. *Journal of Experimental Psychology*, *33*(6), 1431–1450. https://doi.org/10.1037/0096-1523.33.6.1431

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, *92*(18), 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1–B9. https://doi.org/10.1016/S0010-0277(01)00116-0

Mamassian, P., Landy, M. S., & Laurence, M. T. (2002). Bayesian Modelling of Visual Perception -Probabilistic Models of the Brain. In *Perception and Neural Function* (Issue January, pp. 13–36).

Mareschal, I., Calder, A., & Clifford, C. (2013). Humans have an expectation that gaze is directed toward them. *Current Biology*, *23*(8), 717–721.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294. https://doi.org/10.1098/RSPB.1978.0020

Martinez, A. M. (2019). Context may reveal how you feel. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(15), 7169–7171. https://doi.org/10.1073/pnas.1902661116

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Munneke, J., Brentari, V., & Peelen, M. v. (2013). The influence of scene context on object recognition is independent ofattentional focus. *Frontiers in Psychology*, *4*(AUG), 552. https://doi.org/10.3389/FPSYG.2013.00552/BIBTEX

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

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*(1), 72–107. https://doi.org/10.1006/COGP.1997.0667

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision 2001 42:3*, *42*(3), 145–175. https://doi.org/10.1023/A:1011139631724

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. In *Progress in Brain Research* (Vol. 155, pp. 23–36). https://doi.org/10.1016/S0079-6123(06)55002-2

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092. https://doi.org/10.1073/PNAS.0805664105

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. https://doi.org/10.3758/BF03197524

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/S13428-018-01193-Y/FIGURES/3

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: do we know more now than we did 20 years ago? *Annual Review of Psychology*, *58*, 75–96. https://doi.org/10.1146/ANNUREV.PSYCH.58.102904.190114

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rémy, F., Vayssière, N., Pins, D., Boucart, M., & Fabre-Thorpe, M. (2014). Incongruent object/context relationships in visual scenes: Where are they processed in the brain? *Brain and Cognition*, *84*(1), 34–43. https://doi.org/10.1016/j.bandc.2013.10.008

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*, 437–457. https://doi.org/10.1146/ANNUREV-VISION-082114-035733

Rousselet, G. A., Macé, M. J. M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision*, *3*(6), 440–455. https://doi.org/10.1167/3.6.5

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-Frequency Channels in Human Vision. *Journal of the Optical Society of America*, *61*(9), 1176. https://doi.org/10.1364/JOSA.61.001176

Schneider, K. A., Richter, M. C., & Kastner, S. (2004). Retinotopic Organization and Functional Subdivisions of the Human Lateral Geniculate Nucleus: A High-Resolution Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *24*(41), 8975–8985. https://doi.org/10.1523/JNEUROSCI.2413-04.2004

Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews Neuroscience*, *8*(7), 522–535. https://doi.org/10.1038/nrn2155

Schyns, P. G., & Oliva, A. (1994). From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, *5*(4), 195–200. https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

See1,Do1,Teach1 (2017). London Street Scene [Photograph]. Flickr. https://www.flickr.com/photos/mpaulmd/35625021023

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(12), 1182–1190. https://doi.org/10.1093/CERCOR/11.12.1182

Spelke, E. S. (1990). Principles of Object Perception. *Cognitive Science*, *14*(1), 29–56. https://doi.org/10.1207/S15516709COG1401_3

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, *9*(4), 578–585. https://doi.org/10.1038/nn1669

Stone, J. V., Kerrigan, I. S., & Porrill, J. (2009). Where is the light? Bayesian perceptual priors for lighting direction. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1663), 1797–1804. https://doi.org/10.1098/rspb.2008.1635

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*(6), 495–508. https://doi.org/10.3758/BF03212183

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13–13. https://doi.org/10.1167/11.5.13

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, *1*(3), 183–184. https://doi.org/10.1038/630

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, *21*(2), 233–282. https://doi.org/10.1016/0010-0285(89)90009-1

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature* , *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *The European Journal of Neuroscience*, *14*(5), 869–876. https://doi.org/10.1046/J.0953-816X.2001.01717.X

Todorović, D. (2010). Context effects in visual perception and their explanations. *Review of Psychology*, *17*(1), 17–32.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461. https://doi.org/10.1162/08989290152001880

Watson, T. L., Otsuka, Y., & Clifford, C. W. G. (2016). Who are you expecting? Biases in face perception reveal prior expectations for sex and age. *Journal of Vision*, *16*(3), 5. https://doi.org/10.1167/16.3.5

Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. *Nature Neuroscience 2015 18:10*, *18*(10), 1509–1517. https://doi.org/10.1038/nn.4105

Xu, H., Dayan, P., Lipkin, R. M., & Qian, N. (2008). Adaptation across the Cortical Hierarchy: Low-Level Curve Adaptation Affects High-Level Facial-Expression Judgments. *Journal of Neuroscience*, *28*(13), 3374–3383. https://doi.org/10.1523/JNEUROSCI.0182-08.2008

Zhaoping, L. (2017). Feedback from higher to lower visual areas for visual recognition may be weaker in the periphery: Glimpses from the perception of brief dichoptic stimuli. *Vision Research*, *136*, 32–49. https://doi.org/10.1016/J.VISRES.2017.05.002