What are the neuronal correlates and dynamics of gist processing, and what is its role in the interplay of two modes of visual attention?

Lucy Spencer

Doctor of Philosophy

University of York

Psychology

January 2019

Abstract

The ability for humans to extract information from their environment with no more than brief glimpse is well-establish in vision science. This is known as 'gist extraction', and is the product of the first feed-forward sweep of information from the eye, through the visual cortex, and into higher-order cognitive control areas in the frontal lobes. This process results in 'sparse' attention, which contains multiple elements including gist.

The first aim of this thesis was to more clearly define gist in terms of its neuronal locations and time signatures, including observing the effects of multiple gists destructively colliding (in which the presence of more than one relevant image causes participants to be unable to identify the presence of a cued target).

The second aim of the thesis was to explore the relationship between sparse attention and focused attention. As established by the data, sparse attention represents a broad 'vision at a glance' understanding of the visual environment. Focused attention represents a narrow mode of attention.

The following novel findings were identified in this thesis:

1. Processing of target category gist can be observed neuronally using fMRI techniques in category-selective cortex; these selective areas show different % BOLD signal responses to their preferred targets. Place-selective areas show a greater % BOLD signal change in a perceptually-driven manner, and face-selective regions show a greater BOLD change in a manner requiring both the actual presence and the subjective perception of the target.
2. Destructive interference is a decisional, not a perceptual process; it is associated with the N300 and D220 ERPs, which are indicative of greater resources being allocated to a difficult task, conflict resolution, and decision-making.
3. Sparse and focused attention operate in a serial manner, with the first feed-forward sweep (including gist) producing the conscious percept of the world.

Table of contents

List of figures

List of tables

Acknowledgements

I would like to express my deepest thanks, appreciation and gratitude for the direction and support of my primary supervisor, Dr. Karla K. Evans, without whom this thesis would not have been possible. Karla provided unending generosity with her time, patience, and mentorship, for which I am profoundly grateful. I would also like to thank my secondary supervisor, Professor Alex Wade, who was quick to offer help at every stage of my Ph.D., and whose guidance, insight, and cheerfully zen approach helped shape the structure of this thesis as it is today. Additional thanks go to Dr. Heidi Baseler for her insight and suggestions during my TAP meetings, and who was graciously willing to offer her time and support over the course of my Ph.D. and Masters' program. I would also like to extend my thanks to Dr. Sylvia Gennari for her thoughtful input and discussions during my TAP meetings.

The projects within this thesis would not have gotten off the ground without the endless, selfless, and enthusiastic support from the staff at YNiC and the Department of Psychology. Special thanks go to Dr. André Gouws, who provided technical and programming assistance, and who was unbelievably generous with his time in debugging the code for the fMRI project in chapter 3. Thanks to Dr. David Watson, who taught me BASH script from scratch and saved me literally months of analysis time. Thanks also go to Dr. Mark Hymers for his truly excellent grounding in Python programming, and for giving me opportunities to work within different modalities and techniques, which is deeply appreciated. I would also like to extend thanks to Professor Dan Baker, who was very generous with his time in analysing and providing support for the EEG data in this thesis.

Thanks go to Andrea Woodward in the administration team, who was fiercely, relentlessly, and unrepentantly in my corner throughout every single moment of this Ph.D; for her kindness, her willingness to listen, and for her willingness to fight for me.

I would like to thank Kirsty, Anika, Jon, Bella, and Alistair for their humour and wonderful compassion, and for reminding me to take a break once in a while. To all of my friends within the department, thank you for the friendly chats, advice, and

emotional support during the last five years. This Ph.D. would not have been possible without you all. To Val, who has been endlessly kind and supplied me with care packages, I would like to thank her for sharing her space and her dining room table, and for her thoughtfulness. To my family as a whole, thank you for all the joy and love that has made this Ph.D. all worth it. To my parents, my Nana and Granddad, who have supported me ceaselessly and unconditionally, and whose love and encouragement built the foundation for my entire education, I would like to extend all the thanks that it is possible for me to convey.

And finally, I would like to dedicate this thesis to my amazing, wonderful, kind, encouraging, thoughtful fiancé. Craig, I absolutely would not have made it this far without your incredible capacity for love, your kindnesses both big and small, your support even when things seemed tough, your tenacious proofreading skills, and your tendency to make me tea upon request at any time of the day or night. Thank you for everything.

Declaration

I declare that the work in the thesis is my own, and has not been submitted for examination at this or any other institution for another award or degree. It has been carried out with the supervision of Dr. Karla K. Evans and Professor Alex Wade. All sources are acknowledged as references.

Some of the data collection was carried out by undergraduate students in the role of research assistant, whom I co-supervised for data collection: data from chapter 2 were collected jointly by myself, Oliver Hodgkinson and Emma Standley; data from chapter 3 were collected jointly by myself and Oliver Hodgkinson; and data from chapters 4 and 5 were collected jointly by myself, Oliver Hodgkinson, Emma Standley and Patrik Custodio. Additionally, data for the experiment reported in chapter 2 were collected by Annakaisa Ritala. Data analysis for chapters 4 and 5 were coded by Professor Dan Baker. The experiment in chapter 3 was initially designed and coded by Dr. Karla K. Evans. All other reported data analyses, experimental design, creation of the stimuli, and any other data collection were performed by myself, under supervision.

Data from chapter 2 were presented by Dr. Karla K. Evans at the following conference:

VSS, St. Petersberg, FL. (August 2017)

Evans, K.K. & Spencer, L.J. (2017, August), *Allocation of Attention in a Complex Environment.* Talk presentation at the annual VSS meeting. St Petersberg, Florida, USA.

Data presented in Chapter 3 are in preparation for submitting to NeuroImage with co-authors Dr. Karla K. Evans and Professor Alex B. Wade.

Data presented in chapters 4 and 5 were analysed with the help of Professor Dan Baker.

Data from chapter 3 were presented in a poster at the following conference:

YVN, York, UK (July 2016)

Spencer, L.J., Wade, A.B., & Evans, K.K. (2016, July), *Neuronal correlates of Gist Processing*. Poster presentation at the annual Yorkshire Vision Network meeting. York, UK.

Data from chapter 3 were presented in a talk at the following conference:

ECVP, Bareclona, Spain. (August 2016)

Spencer, L.J., Wade, A.B., & Evans, K.K. (2016, August), *Neuronal correlates of Gist Processing*. Talk presentation at the annual European Conference on Visual Perception meeting. Barcelona, Spain.

Data from chapters 3 and 5 were presented together at the following conference:

VSS, St. Petersberg, FL. (August 2017)

Spencer, L.J., Baker, D. H., Wade, A.B., & Evans, K.K. (2017, August), *Neuronal Correlates and Temporal Dynamics of Gist Processing*. Poster presentation at the annual Vision Science Society meeting. St Petersberg, Florida, USA.

Chapter 1 Introduction

1.1. Overview

The phenomenon of visual attention may appear to an observer to be a seamless whole. However, visual attention is the product of more than one mechanism, which operates in relation to one other. Both of the most popular models of these pathways propose two 'extremes' of visual attention, with one being focused and local, and one being distributed and global. This thesis raises two main questions regarding these models:

- What is the relationship between these two types of attention? I.e. Are they opposite ends of a single, continuous resource, or two separate resources that may work serially or in parallel?
- What is the neuronal (spatial and temporal) nature of the distributed, global type of attention?

The questions were assessed using a combination of methodologies and analysis techniques, including psychophysics, functional magnetic resonance imaging (fMRI), electroencephalograms (EEG).

The aim of this chapter is to provide an overview of attention as a whole (with a more detailed look at visual attention specifically), and to establish the models that currently seek to explain the relationship between the two types of visual attention. This will be done by first discussing key concepts behind visual attention, introducing multiple models that seek to explain the relationship between them, and finally by critically evaluating the two dominant models. This will then allow for a more in-depth investigation into the global type of visual attention, which will be covered in subsequent chapters.

1.2. Attention – definition of key terms & themes

1.2.1. What is attention?

Attention is a phenomenon that most people experience as a single, seamless process, which allows a person to choose on what to focus in the world around them. William James makes one of the earliest known attempts to define attention: "Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought." (James, 1890) Whilst more of a philosophical observation than a scientific description, James provides a starting point for the discussion of attention: namely that attention is a subjective perceptual experience of the world, requiring "localization, concentration, of consciousness."

In the decades between 1890 and the present day, there have been many attempts to further define attention in a more scientific manner. Attention can be described from multiple viewpoints, and the following reviews cover these topics: cognitive and perceptual processes, functional features, and anatomical structures.

Chun, Golomb, & Turk-Browne (2011) reviewed attention in terms of the growth of research in the last century, observing its poor definition as a unifying term, and recommending that "we should therefore abandon the view of attention as a unitary construct or mechanism, and consider attention as a characteristic and property of multiple perceptual and cognitive control mechanisms." This is a view supported by other reviews into attention-based literature, and suggests that, despite a common experience of vision as a single, perceptually-seamless unitary system, it's actually the product of multiple interlocking cognitive systems that control different stages of attention (Posner, 1994; Rensink, 2014).

Potential parts of this cognitive network are described by Chun et al. (2011), who attempted a full taxonomy of attention from a functional perspective. In this they identified four core features: limited capacity, selection, modulation and vigilance. 'Limited capacity' describes attention's necessary role in filtering out and refining the perceptual information coming in from the environment to important and/or goal-relevant information. 'Selection' describes the process of biasing competition in favour of important and/or goal-relevant information, of all available information, for example, by directing the eye to targets of interest. 'Modulation'

processes the selected information in the absence of other competing items, controlling factors such as the speed of processing of the object, and speed/accuracy trade-off of any response. Finally, 'vigilance' is the ability to maintain attention over time. These four components of attention also reflect the modular nature of attention, and provide possible examples for the theorised multiple interlocking cognitive systems.

These four components of attention are important, as attention itself is a part of multiple modalities, such as visual and auditory attention (Rapp & Hendel, 2003). However, Chun et al. (2011) establish (in line with previously-discussed findings) that attention is not a unitary system, with mechanisms applying equally across modalities. Visual attention may share capacities with auditory attention, but the actual process of selection, for example, would be different between the two types – for example, the mechanism by which visual attention selects and modulates a target is different to the way in which auditory attention does the same. These four descriptors are labels for processes, but do not attempt to be prescriptive in a functional sense.

Finally, from a more anatomical standpoint, Petersen & Posner (2012) described their view of attention as a system with 3 distinct and basic concepts. Firstly, the attention system is anatomically separate from the processing systems; secondly, the attention system involves multiple anatomical areas; and finally, these distinct areas perform different cognitive functions. Each of these concepts were based on the presence of three selectively-activated networks, the evidence for which suggests that the multiple cognitive mechanisms previously discussed are the product of multiple anatomical areas, rather than a single area with multiple functions.

These reviews mark an important concept in attention: namely, that attention is seemingly not a single mechanism, but comprises of multiple mechanisms or modules, whether perceptual or cognitive. These multiple mechanisms are reflected anatomically in a de-centralised networked system of discreet brain areas. In view of this, it would perhaps be more correct to view the concept of 'attention' as an umbrella term for the variety of functions and networks of which it consists.

1.2.2. Attention & visual attention

It is visual attention that will form the core thesis of this literature review, and it is from this perspective that all information about attention must be received and analysed.

It is important to note that visual attention and visual awareness are *not* the same thing; 'visual awareness', though a contested term, could be defined as a form of consciousness which embodies a subjective experience of seeing visual information (Wyart & Tallon-Baudry, 2008). Visual attention, as the previous discussion might suggest, is a decentralised set of multiple mechanisms, not governed by a single network (Posner, 1990). More functionally, visual attention is the specific application of the features defined by Chun et al. (2011). Specifically, Rensink (2014) defines visual attention as a contingently selective process, which is "controlled on the basis of global considerations". Rensink gives the example of an observer visually tracking a number of vehicles in traffic; the number of vehicles able to be tracked has an upper limit, and once this limit is reached, task performance begins to suffer and eventually fails (limited capacity). However, visual attention is what allows an observer to choose and track an object based on its estimated priority (selection and vigilance), whilst disregarding less-important targets (modulation).

Visual attention has a very close relationship with visual awareness (Wyart & Tallon-Baudry, 2008), and this review aims to assess how visual attention *informs* visual awareness. To do this, this review will compare and contrast two main competing models of visual attention: the Feature Integration Theory (or FIT – AM Treisman & Gelade, 1980; Treisman, 2006), and the updated Guided Search Theory (GST – Wolfe, Võ, Evans, & Greene, 2011). These two key models both attempt to explain visual attention, and although they share some principles, there are a number of key differences. These differences produce alternative predictions about observers' visual awareness, which may allow a conclusion to be reached about the validity of these models.

In debating these models, multiple key topics must be addressed, as they form key underpinnings. Firstly, the basic function of visual attention will be described. This will include top-down and bottom-up processing, spatial and temporal attention, early vs. late selective attention, and a discussion of visual search. These topics are often contentious within the literature, and generally the model of visual attention attempt to reconcile one or more of these points with the others.

Next, specific models of visual attention will be introduced, with a description of early models and the evolution of current models. The debate between the core concepts of the FIT and the GST will be discussed, and mentioned above. Alternative models will also be discussed in terms of their ability to explain visual attention in comparison to the FIT and GST. Finally, conclusions drawn from the literature review will be made, particularly in regards to the predictions offered by both the FIT and the GST.

1.3. Visual attention

1.3.1. What does visual attention do?

Models of visual attention would be very redundant without an understanding of its purpose. In discussing what visual attention does, Evans, Horowitz, Howe, et al. (2011) describe four different purposes. The first purpose is stimulus selection, or 'data reduction', in which visual attention acts as a filter for incoming visual information. This would suppress irrelevant information and prioritise relevant information, with the appropriate brain regions also being inhibited or amplified as a result (Posner, 1994). This may help deal with the 'limited capacity' characteristic of attention by prioritising the processing of some stimuli, ensuring this information makes it through the 'bottleneck' of the capacity limit.

A common example for this first 'purpose' (as used by Chun & Wolfe, 2005), is that of a driver behind the wheel of a car. There is a vast amount of visual information that must not only be processed but actively acted upon, or (just as

importantly) ignored. This requires a variety of mechanisms which allow for the identification and perception of certain prioritised visual stimuli (such as 'give way' signs or road warnings), whilst suppressing or filtering out irrelevant information that might otherwise consume the attentional system's limited resources, and lead to a distracted (and dangerous) driver. Chun and Wolfe described this as an observer only being aware of events or signals which are attended; this prevents the vast amount of visual information overloading perception.

The second purpose of visual attention is described as stimulus enhancement, in which the use of visual attention reduces ambiguity or noise in the visual signal. Itti, Rees, & Tsotsos (2005) described it in mathematical terms as 'internal additive noise reduction', resulting in behavioural performance improvement (such as lower psychophysical thresholds). Evans, Horowitz, Howe et al. suggest this is a result of attention allocation directly to a stimulus, or to part of a stimulus. For our hypothetical driver, this could be demonstrated by faster reaction times to stimuli which are specifically attended, such as braking at red light.

The third purpose is the 'binding' of the features of an object. Traditionally, the 'binding problem' is a multi-part problem that arises when a unitary percept (such as a red square) is created from disparate visual features (the colour red, the lines of the square) which are processed in independent brain areas. In the case of our 'driver', this would involve them binding disparate features of objects (such as shape, colour and movements) to eventually produce the percept of other vehicles, pedestrians and road signs.

There is currently no satisfactory model for the binding problem, though it is suggested that the problem is actually the result of conflating several separate technical problems, covering at least four distinct situations, as per Feldman (2013): co-ordination, the subjective unity of perception, visual feature binding, and variable binding. This would suggest that the 'binding problem' (as referred to above in terms of the vehicle driver) is actually a subset of a larger issue covering several computationally-distinct issues, rather than a single problem to be solved independent of other domains.

Acknowledging that visual feature binding is a part of a larger problem, however, does not negate suggestions on how this specific aspect of the problem

might be resolved. Evans, Horowitz, Howe et al. propose that visual attention plays a role in resolving visual feature binding, potentially in two ways: 1) that it generates stimulus representation, and/or 2), that it resolves ambiguities caused by multiple stimuli appearing within a single receptive field.

In regards to this latter idea, Evans, Horowitz, Howe et al. suggest this may be achieved by "dynamically altering the selectivity or spatial extent of the receptive field of a neuron" (p504). This process is described by Desimone & Duncan (1995) in a single-cell monkey study in which monkeys were primed with the location of a target, and then neuronal response to the target and a distractor were measured. When both target and distractor were within the receptive field for the cell, the response to the distractor was reduced, with the neuronal response being driven almost entirely by the target. Desimone and Duncan reported that "the cells responded as though their receptive fields had shrunk around the target" (p203). This suggests that top-down influences (in this case, spatial bias) such as attention help resolve the visual feature binding problem by attenuating the response of the neurons' receptive fields. Essentially, attention biases the response of the cell to the attended stimuli (Reynolds, Chelazzi, & Desimone, 1999).

These mechanisms may resolve the uncertainty of objects if two or more should share that field. In terms of top-down influences, this might be attention; in terms of bottom-up influences, this might be saliency. In the case of the driver, the features for 'red' and 'circle' may be bound together despite the presence of other features within the receptive field (tree branches, traffic) by the attenuation of the receptive fields in the relevant neurons. This would form the percept of a no-entry sign due both to the saliency of the colour and the attention paid to road signage.

However, it is worth noting that some researchers have argued that the binding problem *itself* is a problem. Di Lollo (2012) argued that binding occurs as a result of a feed-forward sweep through the brain, activating "invariant" (i.e. abstract, platonic) representations throughout the brain and terminating in higher regions (e.g. inferotemporal cortex, or IT). In this proposed model, the IT then generates multiple perceptual hypotheses of the visual object that are broadly consistent with the input stimulus, which in turn produces correlations with low-level (e.g. primary V1) neural representations of the stimulus.

The hypothesis with the highest correlation leads to conscious visual awareness of the object, whilst low correlations are discarded. This utilises both the stimulus-driven and goal-directed features of visual attention to produce a single percept, in which features such as colour and orientation are processed jointly (not independently) in specialised neurons in the visual cortex. This would mean that there is no need to 'reconstruct' a representation of the object.

Bar et al. (2006) combined functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and a behavioural task to investigate how top-down influences can facilitate visual object recognition. In the context of Di Lollo's model, they demonstrated how perceptual hypotheses (recognition networks in the temporal lobe) are primed in advance by the orbitofrontal cortex (top-down facilitation). Bar et al. had argued that top-down processing must occur *prior* to object recognition, and suggested that the orbitofrontal cortex was the site for this top-down processing, due to its presence in object recognition (compared to non-recognition) tasks. They found that low spatial frequencies in an image did activate orbitofrontal pathways (top-down processing), preceding corresponding activity in the temporal lobes (object recognition).

The perceptual hypotheses primed by the feed-forward sweep is then tested through re-entrant processing from the higher areas (IT, OFC) down to the V1, as hypotheses are confirmed or denied through a pattern of correlation. Activation in the early visual cortex (V1 and V2) that correlated with prior higher-level activity (IPS) was found by Weidner, Shah, & Fink (2006), who found that the V1 showed the greatest % BOLD signal change for trials in which the object-substitution network (i.e. the network that helped generate perceptual hypotheses) showed the most activity. Considered as a whole, this implies that the perceptual hypothesis is primed by top-down influences (e.g. the OFC), the hypotheses are generated by an object-substitution network including the IPS, and then these hypotheses are tested via re-entrant processing, so far observed within the early visual cortex.

Finally, the fourth purpose of visual attention, according to Evans, Horowitz, Howe et al., is object recognition and identification. Due to the capacity limit of recognition mechanisms, visual attention serves to subdivide a visual scene to allow for processing small batches of input. This would allow the driver in the analogy to

assign semantic meaning to the precepts produced by the third 'purpose' – to identify another vehicle as such, which in turn allows for the driver to react accordingly.

1.3.2. The mechanisms of visual attention

Mechanisms of visual attention are the building blocks that form part of the larger structure of the models. In terms of the driver in the car, this successful execution of visual attention requires multiple mechanisms to achieve the end result. The modulation (including the suppression) of visual signals may be achieved through a combination of mechanisms, including processing directions, the process of selection, and visual search.

1.3.2.1. Stimulus-driven and goal-directed processing

Visual attention may be considered from two main directions: top-down processing, and bottom-up processing. Stimulus-driven (or 'bottom-up') visual attention is a type of processing in a stimulus' more salient features (compared to competing stimuli) mean it is selected for further processing.. This may be through internal methods, by which a subject's goals or actions are the determining factor; this can include such endogenous cues as symbolic representations of target location, like "up" or "down" (Hommel, Pratt, Colzato, & Godijn, 2001). Stimulus-driven attention may also be driven by external methods, in which the stimulus' external features are the determining factor. These exogenous cues may include features such as luminance or movement (Rosen et al., 1999). This can be observed with reflexive eye-movements, as a stimulus' unusual features 'draw' the eye.

Goal-directed (or 'top-down') visual attention, by contrast, is visual processing that is voluntarily controlled by the subject. This would allow the subject to deliberately switch the field of view to attend to a new area, and be measured by controlled eye movements.

However, these mechanisms do not have to operate on an 'either/or' basis. Fecteau & Munoz (2006) proposed that selection of objects in the visual field is guided by *both* goal-directed and stimulus-driven factors, the combination of which is referred to as 'priority map'. This map is the product of a network, and combines the distinct neural firing properties of both salience (an item's stimulus-driven distinctiveness) and relevance (a goal-directed measure). It is a topographical map of the visual scene, in which objects are spatially represented. In an all-or-nothing competition, the object that maximally combines both salience and relevance is selected.

1.3.2.2. Spatial and temporal processing

Visual attention also operates over spatial and temporal parameters. Spatial attention refers to the direction of attention to a location in the visual environment, and also refers to the attentional system's ability to prioritise certain spatial locations whilst inhibiting others, especially spatial areas that have already been attended ('inhibition of return', Klein, 2000; Chun & Wolfe, 2005). It also refers to the reflexive reorientation of attention to a novel event (Klein, 2000). However, spatial attention is an example of the 'limited capacity' characteristic of attention, as the number of objects that can be attended in the visual field are finite.

Time is perceived as a linear movement from the past to the present, and part of that percept is the way in which incoming visual information changes from moment to moment (temporal attention). It has been studied using an experimental model called rapid serial visual presentation ('RSVP' – Potter, 1976). This involves a fast, continuous presentation of visual objects, which allows for changes in visual attention over time to be studied.

This has revealed a phenomenon known as attentional blink, in which the second of two targets in the same spatial location may not be detected, if the second target is presented less between 180 and 450 milliseconds after the first, depending on the type of target (Raymond, Shapiro, & Arnell, 1992). Attentional blink can also be induced spatially (i.e. simultaneous, neighbouring distractor stimuli), as well as temporally, and Marois, Chun, & Gore (2000) observed the two forms of attentional

blink activate the same neural network. This is telling, as it implies the same mechanism of visual attention is responsible for the distribution of attention across both space and time.

1.3.2.3. Selection

In the description of the driver in the car, one of the mechanisms necessary for a useful visual percept of the environment was the ability to modulate the visual signals according to a goal-directed and/or stimulus-driven system. This would allow certain signals to be perceived, whilst others were suppressed. Attentional selection is this process of prioritising certain signals to be attended, in such a way that the information presented to other cognitive functions (such as working memory) is relevant.

When, and where, do these processes of selection happen to produce awareness? There is some debate over the order in which visual selection and visual processing occurs: whether this occurs 'early' in visual processing, or 'late'. This argument often forms the basis for most models of visual attention.

Broadbent (1958) initially proposed what is now known as an early selection theory of attention, in which perceptual stimuli are selected for attention at a very early stage in processing, before semantic analysis. Stimuli would be filtered for relevance based on low-level features (in the case of visual attention, such as spatial location or line orientation), and the rejected information would not be processed further. Treisman, 1960 proposed an alternative early selection model, known as the Attenuation Theory, which discussed how Broadbent's 'filter' may not be so much of a dichotomous yes/no filter, and may instead selectively raise the thresholds for the rejected source of data.

An alternative to both Broadbent's and Treisman's model was proposed by Deutsch & Deutsch (1963). It was a contrasting, late selection theory of attention, in which information is processed for meaning *after* the low-level features have been analysed. As described by Serences & Kastner (2014), it argued that *all* stimuli are analysed to the semantic level. It is from this point that the stimuli are then separated, with some progressing into further, capacity-limited processing, at which

point an observer becomes perceptually aware of it. (Serences and Kaestner observed that the pre-semantic stage was argued to not be capacity-limited, and would be capable of processing all incoming items in parallel.)

More recent research has attempted to settle the debate between early and late selection models. Chun, Golombe & Turk-Browne discussed the Attenuation Theory in terms of it representing a continuum of selective attention between early selection and late selection, particularly as evidence has been found for both the early (Cherry, 1953; Broadbent, 1958) and late (Moray, 1959; Deutsch & Deutsch, 1963) models. This would mean that attention would fall on a different point on this continuum, depending on task requirements. Lavie (2005) also attempted to resolve the conflict, and claimed to successfully marry the concepts of early and late selection in one unifying model. Called the Load Theory of attentional selection, it argues that the processing of 'distractors' (i.e. non-relevant visual stimuli) varies depending on both the degree *and* type of load when processing goal-relevant information. High perceptual load resulted in a reduction, or even an elimination, of the neural signals associated with non-relevant stimuli. High cognitive and executive load (such as in the working memory) produced the opposite effect, increasing the interference caused by the distractors, measurable in neuroimaging and behavioural experiments. This indicates a late selection system.

1.3.3. Visual search

A person with a typical visual experience of the world may be required to scan their environment for a specific target, whilst ignoring other, non-goal-relevant items. This active deployment of visual attention is often the best way to measure the mechanisms of visual attention, especially when taken in conjunction with the subjective visual awareness it produces. Visual search often forms the backbone of models of visual attention, including for the FIT and the GST; it uses a variety of orienting mechanisms, and involves more than one search type.

1.3.3.1. Overt and covert orienting

Visual search requires the shifting of attention, as indicated by Tsotsos et al. (1995). Exogenous and endogenous cues may result in two different kinds of orienting to a new visual location, and Posner, 1980) observed that this may happen in one of two ways: overt and covert orienting.

Overt orienting is produced by head and eye movements, and are the product of an observer attending to a specific visual location. There are two types of overt eye movements: 1) reflexive, which are quick, and 2) controlled, which are slow and voluntary on the part of the observer. Covert orienting, however, is the product of shifting attention elsewhere in the visual field, but without eye movements.

However, it is arguable that these systems are not independent of one another, as Posner (1980) noted that the central mechanisms that control covert orienting of attention (the parietal lobe is suggested) are also receiving input from subcortical regions involved with overt orienting. This could indicate the presence of a network of visual attention that optimises visual search using both covert and overt orienting systems.

1.3.3.2. Feature and conjunctive search

In terms of *how* visual search operates, Treisman and Gelade (1980) described two different kinds of visual search: 'feature search', and 'conjunction search' and proposed the Feature Integration Theory (FIT) to conceptualize behaviour during visual search.

Treisman and Gelade's feature search involves asking an observer to look for a target with a specific characteristic (for example: colour or letter). The target will have a maximal difference from the distractors (if the target is red, then the distractors will be green). This is also known as 'efficient search', which is defined by a short reaction time in relation to the set size (Hall & Wang, 2001) for correct responses; reaction time also generally does not increase at all or steeply regardless of the number of distractors (Woodman & Luck, 2003). The FIT argues that features are automatically registered in parallel across the visual field, and at an early stage of visual processing.

In conjunction search, the observer is looking for the target stimulus that combines more than one feature (colour, shape, size, etc.). For example, while feature search involves an observer looking for a pink letter amongst purple and brown distractor letters; the conjunction search involves an observer looking for a pink "O" (a conjunction of colour and a letter) amongst distractors of a green "O" and a pink "N".

Conjunction search is also described as 'inefficient search' (Hall & Wang, 2001), as it requires the observer to examine many items in a serial fashion. This can be measured using a reaction time slope; efficient search, regardless of reaction time, produces a flat slope even with the additional of a distractor or a higher set size. Inefficient search, however, produces a steeper slope, as the reaction time increases along with the set size.

The more distractors in the set, the less efficient the search becomes, and the resulting reaction time slope is steeper. Woodman & Luck (2003), observed that a participants' attention would be allocated to one possible target in the visual field for approximately 150ms, before moving on to the next. Treisman & Gelade (1980) describe it as 'top down'-based search, and is therefore also able to draw from contextual information and experience. They also propose that, when observers are unable to focus their attention to the location of the target (due to the capacity limit of top-down processing), 'illusory' (non-existent) conjunctions can be formed from unrelated features.

These illusory conjunctions occur when several features from an object or objects (particularly similar features) in the visual field are incorrectly combined. Treisman (2006) gives the example of a yellow cross object and red heart object producing the illusory conjunction of a red cross, and subsequently creating associative priming for the concept of 'hospital'. These incorrect conjunctions can be resolved with what Treisman terms 'focused' attention, which is the deliberate deployment of a small attentional window to a specific part of the visual field, thus suppressing any features outside of that area.

1.3.3.3. Alternative attributes to visual search

However, much like the early vs. late selection debate, there have been attempts to marry how feature and conjunctive searches work. Duncan & Humphries (1989, 1992) argued that the apparent differences in reaction times did not reflect a distinction between feature and conjunction search. Like Lavie, they proposed a continuum of search efficiency rather than discrete categories, in which the difficulty level of the search task (and thus reaction time) rises due to the increase of similarity between a target and a distractor, and a decrease in similarity between the distractors themselves. Duncan & Humphries discussed this as a part of their proposed Attentional Engagement Theory (AET), an alternative to Treisman & Gelade's FIT.

Other researchers have also sought to identify alternative explanations for the differences in reaction times between 'types' of visual search. Hall & Wang used context-defined, multi-conjunction targets, for which an inefficient search was predicted; they found instead that search was efficient for certain multiple conjunctions (size, shape and topology of the target), and concluded that perceptual organisation could be another factor determining search efficiency. Kristjánsson, Wang, & Nakayama (2002) also argued that priming was a less-recognised aid to efficient visual search, and may in fact play a greater role in conjunctive search tasks than currently acknowledged – and may even account for some of the efficiency otherwise attributed to top-down guidance.

1.4. Models of visual attention

1.4.1. The 'spotlight' and 'zoom-lens' models of visual attention

Many researchers have attempted to construct a single, coherent model that incorporates all of the previously-discussed mechanisms, in assessing the role of attention in vision. One of the earliest 'modern' models was the concept of the attentional spotlight, first proposed by Posner, Snyder, & Davidson (1980). This proposed that attention could be moved spatially around a visual scene much like a spotlight, consisting of a limited-capacity mechanism that operates in a serial manner. (This is in line with Treisman's description processing during conjunctive

search.) It inhibits input from areas outside of the 'spotlight' in a gradient from the focal point (Evans, Horowitz, Howe et al., 2011).

Eriksen & St James (1986) argued that, at one end of a continuum of visual attention, this focused 'spotlight' could be small and tightly-focused on a small area, producing more highly-detailed processing. At the other, the 'spotlight' could be distributed more broadly over a larger visual area, producing a degraded signal. Eriksen & St. James called this the 'zoom lens' model; the analogy proposes that attention can be broad and low-resolution (diffuse), or can 'zoom in' on a single location, producing higher resolution (focused). Their research suggested that, when attending visually to a scene, attention begins in the diffuse state, but is restricted over time in order to limit the effect of non-relevant distractor items.

1.4.2. The Feature Integration Theory (FIT)

At the same time that Posner et al. (1980) were developing the spotlight model, Treisman and Gelade (1980) were refining the FIT. The two-stage FIT proposed in its second stage, like Eriksen & St James (1986), a continuum of visual attention with two extremes.

The first stage is a pre-attentive (unconscious and implicit) level of processing, in which the features of a visual scene are analysed quickly, early on in the perceptual process, automatically, and in parallel. (Examples of these features are colour, shape and orientation, each of which are processed in separate brain regions.) Treisman (2006) clarified this stage as the first feed-forward pass, taking place either when attention is focused at another location, or before attention is fixed on the location in question. This is somewhat analogous to the zoom-lens 'diffuse' state, though it does not require conscious attention and extracts certain broad characteristics such as summary statistics. This is also known as acquiring the 'gist' of a scene. (It is important to note, however, that this 'gist' interpretation of the model is new; whilst discussed by Treisman in 2006, it is not evident in her earlier work – Treisman & Gelade, 1980).

This creates 'maps' of features, which contain information about the feature property (e.g. 'blue'), maintaining some implicit data about the feature's spatial

location. However, like the objects the features 'activate' in the recognition network, this location information is not made explicitly available. All of these feature maps are stored in a single 'master map' (Chan & Hayward, 2009).

The second stage involves focused attention, in which attention is narrowly applied to a single location, and the features outside of this location are suppressed. Kahneman, Treisman, & Gibbs, 1992) propose that the movement of focused attention from location to location in space produces 'object files', which they hypothesise is a response to the binding problem. In these object files, an object's features are bound (to the exclusion of features in other locations), and compared to unconscious representations that have been activated by the features available in the feature map (i.e., prior knowledge). This object file can be updated, which allows the object to be consistently represented even when changed in space or time.

Treisman (2006) gives the example of the representation of 'hospital' being activated via a red cross, which is caused by the presentation of a yellow cross and a red heart. Although a red cross is not itself present, the feature 'red' and the feature 'cross' are primed in this manner because of a top-down expectation. Treisman hypothesised that focused attention was what allowed such illusive conjunctions of features to be eliminated, due to the suppression of features outside of the immediately attended area.

However, much like the zoom-lens model, in the FIT the aperture of attention comes in more than one form. As well as a narrowly-focused window, visual attention can also span over the whole scene. Treisman proposes that the size of the visual aperture can fall at any point along the spectrum in between these two extremes (focused and distributed attention), producing an object file at each level. It is the combination of these object files that produce a visual understanding of the environment at multiple scales; from a global view (a forest), to an intermediate view (a tree), to a localised view (a branch).

A person's ability to store these object files are capacity-limited. Wheeler & Treisman (2002) argued that limited capacity of attention meant that up to four object files could be stored in working memory at any one time. This was based on research by Luck & Vogel (1997), who found participants could nearly perfectly retain feature information for up to three objects, with accuracy declining from four

objects onwards. This 1–4 range of objects was also not limited to features; participants displayed a similar pattern in accuracy levels for objects that combined two features (i.e. one object contained both colour and orientation information), with performance once again dropping beyond four objects.

1.4.3. The Attentional Engagement Theory (AET)

The AET is a late-selection theory of attention by Duncan & Humphries (1989, 1992). Duncan & Humphries acknowledged the FIT, but presented evidence which found a wide variation in search efficiency. This contrasted with the FIT's assertions that feature search would be fast and efficient, where conjunction search was slow and inefficient.

As opposed to the FIT's serial/parallel search dichotomy, the AET is predicated on the notion of a *continuum* of search efficiency (as discussed in the section concerning feature and conjunction search). Duncan & Humphreys (1989) conducted four experiments in letter search, asking participants to identify a target letter amongst distractor letters. In reviewing their findings, Duncan & Humphreys (1992) drew three main conclusions: firstly; that reaction time slopes are flat, assuming that target letter are sufficiently distinct from distractors, even if the distractors themselves are very different from each other (heterogenous). Secondly, that reaction time slopes increase steadily as distractors begin to more closely resemble the target, even when the distractors are heterogenous. Thirdly, the worst reaction-time performance is found when targets and distractors closely resemble one another (more homogenous), but when distractors are heterogenous.

Using this information, Duncan & Humphries contrasted their findings with those of the FIT, and proposed an alternative, late-selection model with three components. Hall & Wang (2001) described the first stage of the AET as a pre-attentive stage, in which all objects in the visual environment are processed in parallel. Here, these object 'descriptions' are turned into individual units of a scene, which form the representation of a scene at several scales.

It is *after* this point in which selective attention is involved, choosing which information is processed into short-term memory. This is in comparison to the FIT, in

which focused attention is required to convert object 'descriptions' into units, rather than this happening in the pre-attentive stage.

In the next stage, search is then 'directed' by comparing these object descriptions to a template, consisting of the required target features. The descriptions in the visual scene are given attentional 'weighting', with those objects that more closely represent the template being weighted more heavily. Critically, Dent, Allen, Braithwaite, & Humphreys (2012) note the phenomenon of weight linkage, in which objects that are similar to one another can also change their attentional weights as a group; Hall & Wang (2001) observes that this allows for the spread of suppression between similar non-target items. In this manner, the target item can be selected based on its weighting to the template, and non-target items can be suppressed, especially if they are homogenous.

1.4.4. The Biased Competition model

The Biased Competition model, developed by Desimone & Duncan (1995), offered a framework for looking at the interaction between bottom-up and top-down processing. At its simplest, it proposed that visual stimuli compete for cognitive resources and for representation on the cortex during visual search. Which objects are selected depend upon the bias operating upon them: whether it is bottom-up, or top-down.

Upon developing the model further, Desimone (1998) described five main tenets. The first is that simultaneous presentations of objects compete for neural representation, and the responses of these two representations will be determined by the way they compete. This will have a mutually-suppressive effect, when averaged, and they predicted this to occur between groups of items, rather than within them. Importantly, this process occurs automatically, and without attention; this is analogous to the pre-attentive stage of the FIT and the AET.

Beck & Kastner (2007) report that this group-based prediction is the result of behavioural studies, in which search times for visual search are not impacted (i.e. there is little competition between target and distractors) when more distractors are added, if those distractors group with other distractors. Further developing the

prediction, Bundesen & Pedersen (1983) found that speed of target identification was influenced by the number of groups of items in the visual field, suggesting suppression occurs between groups, and not within them.

The second tenet is that two items which engage cells in the same part of the receptive field will produce the strongest competition. The third tenet is that 'feedback bias' may bias competition in favour of one of the items, due to a variety of neural mechanisms (rather than a single attentional control mechanism). This can take one of two forms: bottom-up attentional bias, which highlights targets that are distinct from the distractor items (this is the 'pop-out' effect, in which a distinctive target amongst homogenous distractor items appear to draw the eye). This allows for items to be separated from the background (Desimone & Duncan, 1995). There is also top-down bias, in which targets relevant to the observer's goals or behaviour are prioritised.

The fourth tenet, on a relate noted, observes that this feedback bias may consist of features such as colour and shape, and not just its spatial location. This bias can be activated in parallel across a very of locations in the visual field, and is activated when the desired feature is relevant to the observer. The fifth and final tenet is that working memory is responsible for the top-down biases, in that biases generated in the prefrontal cortex are fed into the extrastriate areas.

Neural evidence has been found to support this interaction between bottom-up and top-down processing. Beck & Kastner (2009) discuss how multiple monkey brain single-cell recording studies find that representations of targets in a receptive field are not processed independently at the neural level, but can interact with one another in a way which is mutually suppressive; in particular, Reynolds, Chelazzi, & Desimone (1999) reported that, when presenting two stimuli to a cell's receptive field, their combined response was smaller than that of the stimuli when presented individually. This combined response was a weighted average, supporting the proposal of competitive representation on single cells.

1.4.5. Reverse Hierarchy Theory

Taking a more anatomical approach, Hochstein & Ahissar (2002) proposed a model of visual attention and perception using a specific underlying cortical framework, which they called the Reverse Hierarchy Theory (RHT). The RHT shared some basic similarities with the FIT, in that the RHT was also theory that espoused a continuum (albeit also serial) model of conscious vision, with two extremes. One extreme represents a broad, sparse mode of attention, which they called 'vision at a glance', and which is the product of higher receptive fields within the brain. The other extreme is a narrow, more focused mode, which they termed 'vision with scrutiny', and which represented smaller receptive fields at a lower level in the visual cortex.

Hochstein and Ahissar proposed that explicit visual perception was a result of a reversal of the progression of visual information through the cortical hierarchy. In the first place, the initial feedforward sweep of information through the brain travels from lower-level cortical areas through to higher-level cortical areas; however, access to the information within this feed-forward sweep is not available to conscious perception. Hochstein and Ahissar argue that conscious perception is a result of the reversal of that pattern.

At the very top of the hierarchy, there are large receptive fields that are able to extract information such as basic category, operating in parallel and generalising over space, size, and other aspects of the scene. Hochstein and Ahissar refer to this as 'spread attention', and argue that it detects objects by approximating an initial binding of features, in something of a "guess". This is responsible for the phenomenon of 'vision at a glance'. As attention descends through the hierarchy from higher cortical areas back to lower cortical areas, this re-entry serves two purposes: first, to confirm the initial bindings made during spread attention, and to correct or update erroneous bindings based on the more detailed information available from lower cortical areas. These lower cortical areas make available information such as specific characteristics of the image, allowing further classification of the image from its basic category to a subordinate category by adding more detail to the initial scene information.

At the bottom of the hierarchy the receptive fields are much smaller, and produce narrower, more focused attention; this produces Hochstein and Ahissar's

'vision with scrutiny'. This lower-level, more detailed perception has limitations, much like the bottleneck in focused attention described by the FIT (in which features outside of the immediate zone of attention are suppressed). The RHT describes focused attention as limited and serial in nature, and whilst focused attention is in use, the sensitivity to the characteristics detected by what the authors call 'sparse' attention is reduced.

Essentially, the RHT proposes that attention to a scene begins at the level of 'vision at a glance', and progresses in a reverse direction down the vision hierarchy until the level of detail required is met within the scene. Like the FIT, the RHT proposes that attention is either broad and spread, or it is narrow and focused. However, unlike the FIT, it is required that 'vision at a glance' is the first stage; the progression from one end of the continuum down to the 'vision with scrutiny' end is obligate in its direction.

Critically, where the FIT is an 'early' selection theory, Hochstein & Ahissar (2002) describe the RHT as resembling a 'late' selection theory. The FIT proposes that attention is required for binding the features of an item into a whole perceivable object; the RHT, conversely, argued that binding is 'guessed at' at an early, high-level stage of processing, and the subsequent return to lower levels is what confirms or corrects this binding.

1.4.6. Guided Search Theory (GST)

Most recently, Wolfe, Võ, Evans, et al. (2011) presented an updated version of Guided Search (based on the original models by Wolfe, Cave & Franzel, 1989, and Wolfe, 1994), which eschewed the concept of an FIT-style continuum of visual attention in favour of a two-stage pathway model. This two-stage model favoured a parallel approach, as opposed to the serial information progression hypothesised by Hochstein & Ahissar (2002).

This model uses aspects of a previous version of the Guided Search theory, described by the authors as classic Guided Search. Classic guided search proposes a serial-parallel hybrid model of selection, rather than the FIT's pre-attentive parallel process. There are several forms this might take; one example is the model offered

by eye movement studies, in which the serial eye movements from point to point may parallel-process several items at each location (Zelinsky, 2008). A more common description is that of the carwash analogy, in which items may enter and leave the binding/recognition process in a serial fashion, but inside which multiple items are in the process of being reconstituted and categorised (Wolfe, 2003).

Information in the visual field is fed into a 'selective' pathway, which is somewhat analogous to the FIT's 'focused' attention, in that it is capacity limited and binds features. Access to the bottleneck of the capacity limit is decided by this serial-parallel hybrid model, which allows for a certain set of features to be the guiding point for the search for a target. This could include colour, shape and orientation, and is also informed by episodic and semantic information. For example, if conducting a visual search for a particular book on a bookshelf, episodic memory may inform search as to the location of the book (where was it placed last), and semantic guidance may limit the locations which are searched (the bookshelves; the book is unlikely to be on the floor or on top of the bookcase itself).

It is these combination of factors which arguably 'guide' attention to certain locations when searching for a target. In particular, Wolfe, Võ, Evans, et al. (2011) discuss how semantic guidance is a product of the second of the two processing pathways: the non-selective pathway, which has multiple roles (such as scene gist). They propose that this operates in parallel with the selective pathway, with the output from both forming the visual experience.

One of the roles of the non-selective pathway is gist perception. 'Gist' refers to an observer's ability to extract certain basic global statistics from a scene. This acts on all levels of a scene (Oliva, 2005), and could include average colour, shape, movement direction or size of a set of items, and even more complex information such as whether or not a scene contains an animal (VanRullen, 2009). This information can also be extracted very rapidly; Oliva notes that this can happen within 100 milliseconds of observing a scene, and Green & Oliva (Greene & Oliva, 2009a; Greene & Oliva, 2009b) noted that an experimental observer can correctly identify a scene's basic category (indoor scene, natural scene, lake) in 20 milliseconds.

It is this information, this gist, which Wolfe, Võ, Evans, et al. describe as the source of the semantic guidance in visual search. In particular, this is a key difference between the GST and the FIT, as the FIT does not propose that the initial feed-forward processing of features has any impact on later, conjunctive searches.

1.5. Critical analysis of models of primary interest

Of the discussed models, there are three in particular which offer specific, testable predictions for the relationship between the extremes of visual attention . Those are the Feature Integration Theory, the Reverse Hierarchy Theory, and the Guided Search Theory.

Each of these models proposes two 'kinds' of attention, though the way these relate differs; whilst the FIT argues for a continuum between two extremes, the RHT posits a serial progression controlled by re-entry into earlier visual cortical areas, and the GST proposes two distinct pathways operating primarily in parallel. Each model ascribes slightly different characteristics to their 'kinds' of attention, despite the similarities. For example, though distributed attention (FIT), non-selective attention (GST) and 'vision at a glance'/ broad attention (RHT) all share the characteristic of being applied to the visual field as a whole, they are not synonymous.

To begin with, the FIT talks about distributed attention in terms of it generating an object file for the whole scene, containing subfiles about the separate objects within that scene. Treisman (2006) describes this as distinct from the initial, pre-attentive, feed-forward sweep of information through the brain.

Both the GST and the RHT, by contrast, include the feed-forward sweep as part of their models. For the RHT, this feed-forward sweep is what generates the 'gist' of the scene, and includes the stages to which visual processing can serially re-enter to extract further information. The GST also uses it as the generator of the non-selective pathway, which (like the RHT) includes gist information such as spatial statistics and category information. However, the GST argues that it operates separately from the selective, binding pathway, and that these two systems inform each other and operate together in parallel, providing episodic and semantic

guidance. All of these differences create wholly different hypotheses for how the two extremes ought to interact, and as such, cannot be conflated.

1.5.1. Evaluation of the FIT

The FIT makes several predictions as to how it might be substantiated as a model of visual attention. These arguments include evidence from patients with brain lesions and their subsequent issues with perceptual binding, though this is somewhat disputed. Criticism of the FIT is also possible through object-based theories of visual attention, and through neuroimaging studies. This evidence will now be summarised, so that the relative strengths and weaknesses of the FIT may be understood.

Treisman (2006) argues that evidence from patients with damage to brain regions which control spatial attention, such as bilateral parietal regions, would support the FIT; the argument is that binding errors would result from the loss of spatial information caused by the brain damage.

Patients with this type of brain damage also show problems with binding and object individuation, and display an inability to perform a conjunctive search task. Treisman argues that this is because, in patients who display such damage, spatial attention can no longer be used to individuate items due to the loss of the location map for observed objects. (Access to this location map is controlled through a re-entry process by the parietal areas.) This would mean that observers could not access the location map of an object, and thus could not bind features, leading to perceived illusory conjunctions and incorrect object identification. These patients would also not be able to perform conjunctive search tasks, due to these issues with binding.

With regards to this prediction, Treisman cites Humphreys, Cinel, Wolfe, Olson, & Klempen (2000), who described a 2-stage system of feature-binding. In the first stage, features are bound into shapes, and in the second, those shapes are then combined with surface details to create the percept of the object. Humphreys et al. then discussed a patient with bilateral parietal lesions (GK), who observed illusory conjunctions of colour and form across both hemifields of vision, and left-side visual extinction. Whilst GK showed a greater ability to report accurately on items which

were grouped near to others with the same shape elements, he made several illusory conjunction errors when describing their surface details, such as colour. Humphreys et al. proposed that this was a result of the patient's inability to perform the second stage of binding, a conclusion which aligns with Treisman's predictions for the FIT.

Patients with Bálint's syndrome provide evidence for the FIT in a similar vein. Robertson, Treisman, Friedman-Hill, & Grabowecky (1997) discussed a clinical patient with bilateral parietal lesions (RM), who presented with simultanagnosia. RM displayed similar binding issues as GK, including a large number of illusory conjunctions during visual search, problems with localising a target object, and an inability to correctly identify the number of objects. RM's performance of visual search for objects with a single feature was unimpaired, despite problems with conjunctive search, further suggesting that his issues stemmed from his inability to bind features. This evidence from perceptual binding appears to strongly support the FIT, as the findings fit the predictions.

However, this is not the case for all elements of spatial attention. Roelfsema (2006) studied the cortical mechanisms of perceptual grouping, the phenomenon in which features are correctly grouped into their individual objects post-binding. Roelfsema identified two specific forms of grouping occurring at the neuronal level: base and incremental grouping. For the purposes of discussing the FIT, incremental grouping is the theory of interest; this is a flexible system in which information is exchanged in a feedforward and feedback manner to neurons in higher and lower cortical areas, and also horizontally between more similarly-located neurons.

Roelfsema argues that his findings align with the predictions of the FIT, as the visual cortex produces an enhanced response to incrementally-grouped features, which correlate with visual attention. However, this 'enhanced response' requires a spread of attention through the activated connections, as well as the initial pre-attentive feed-forward stage mentioned in the FIT. This requires time, and suggests that attention is involved in visual perception at a much earlier stage than the FIT would propose.

The FIT also requires focused attention to bind the features in a single location into a single object (and thus suppresses features outside of this location);

conversely, Roelfsema observes that features can be incrementally grouped even if they are spatially separated, proposing that attention is sensitive to objects, rather than specific spatial locations. Unlike the system proposed by FIT's focused attention, this would allow for attention to be directed to objects that share the same space, or overlap. This model is known as 'object-based attention'. Support for this comes from Duncan (1984), who found that attention can be focused specifically on one of two overlapping objects, with prioritised objects resulting in high report accuracy, and the non-prioritised object resulting in low report accuracy. Egly, Driver, & Rafal (1994) also broadly support these findings, though they argue that attention can be object-based *or* location-based, with suggestions for an interaction between them. However, the paper does not suggest a model with comfortably fits the binding mechanisms proposed, such as the attentional engagement theory or (more relevant to this discussion) the GST.

In her overview paper, Treisman (2006) makes mention of multiple studies that support the FIT. However, the evidence discussed here suggests that the FIT is perhaps too simplistic in its handling of visual attention. Like the Posner, Snyder & Davidson (1980) spotlight model of attention, it is limited by its reliance on location-based feature-binding, as object-based theories of attention show.

1.5.2. Evaluation of the RHT

Much like the FIT, the RHT proposes a continuum of conscious visual attention, with an extreme at one end of fast, spread, parallel attention, and another extreme of slow, serial, focused attention. However, unlike the FIT, the RHT seeks to explain the relationship of spread (or distributed) attention and focused attention through a cortical framework, using evidence from both neuroscience and behavioural fields.

The FIT is an 'early selection' model (meaning attention is needed to bind object features into a perceptual whole). In contrast, the RHT argues that the first feedforward sweep of attention results in 'sparse' attention, which produces initial bindings for objects as a 'first guess'. Re-entry down the visual hierarchy (focused attention) then refines those percepts, rejecting incorrect bindings and reinforcing

correct ones. This would mean the RHT resembles a 'late selection' model, in contrast to the FIT.

The RHT also proposes that the progression through the reverse hierarchy goes in a single direction, from higher cortical areas to lower ones. In this regard, the model may be described as a serial progression, with re-entry into progressively lower cortical areas done as necessary. In principle, the concept of a 'reverse hierarchy' has been found neuronally; Peyrin et al. (2010) used fMRI and EEG to find that low spatial frequency information produces an initial increase in activity in prefrontal and temporo-parietal areas, with enhanced responses seen later in early visual cortex areas. This, Peyrin et al. argue, is a retroinjective 'feedback' into areas of 'lower' processing, which allowed guidance based on the information gained from high spatial frequencies. As well as fitting into the model described by the RHT, it also provides an initial suggestion of the mechanisms by which the model may operate.

Perceptual learning is a key part of the argument for RHT, with the authors of the original model producing a separate paper discussing this topic (Ahissar & Hochstein, 2004). Perceptual learning within the RHT, they argue, predicts a specific pattern of learning, in which non-specific, broad learning comes from modifications at neurons at the higher cortical level, with larger receptive fields. More narrow, task-specific learning is a result of modification of neurons at the lower cortical levels, happening later in the RHT model due to the re-entry process. Essentially, high-level learning happens first, and low-level learning has to happen later, as the low-level learning requires guidance from the preceding high-level learning. Ahissar & Hochstein (2004) argue that this is due to visual information in the brain, from high-level to low-level, becoming more usable due to a top-down-guided series of modifications to that information, which prioritises the task-relevant aspects and crops the non-relevant information.

Earlier, Ahissar & Hochstein (1993) demonstrated this by asking participants to firstly identify the global direction of an array of lines, and to secondly identify the presence of an 'odd' array, which was a pop-out search task. Participants received training with one condition (e.g. different target sizes), and were tested on another (e.g. different target locations). Learning effects transferred across conditions only

occasionally. Further analysis (Ahissar & Hochstein, 1997; Ahissar & Hochstein, 2000) found that learning transfer was found with easy task conditions (and presumably modification of higher-level cortical areas), and that learning specificity was found within hard task conditions (presumably at lower cortical levels. They also found that the timeline for this was predictable and unchanging: easy task conditions were learned early and harder task conditions were learned later. Specifically, learning in hard task conditions only occurred after an easier example had been provided.

This order of operations is within the predictions of the RHT, suggesting that re-entry into the visual hierarchy is required for conscious visual attention. Access to higher cortical levels happens early, and allows access to (and learning of) broad, global conditions within the scene. These areas contain receptive fields that are generalised, and so easy task conditions can be generalised over new conditions. Later access to lower-level cortical areas for harder task conditions produces more specific learning, which cannot be generalised.

In particular, the claim from Ahissar and Hochstein's 2004 paper that high-level learning acts as guidance for subsequent lower-level modifications has some empirical evidence. The concept is based on work by Cave & Wolfe (1990), who proposed the initial version of the Guided Search Theory as a response to issues with the Feature Integration Theory in integrating the role of parallel visual search. More recently, work by Wu, Wang, & Pomplun (2014) found that information such as the local spatial layout of an image (spatial dependency) could help semantically guide attention. This suggests support for the RHT, in that information from spread attention (the broad end of the proposed continuum of the RHT, in higher cortical areas) provides guidance for subsequent focused attention (later re-entry into lower cortical areas of the visual hierarchy).

The RHT makes several predictions that can be directly investigated using electrophysiological methods. Ahissar and Hochstein discuss studies which address their hypotheses for neural modifications in perceptual learning at different stages in the reverse hierarchy. Initially, the RHT proposes that neural modifications associated with transferrable perceptual learning would be found in higher cortical areas; primate studies found firstly that changes in the inferior temporal cortex (a higher cortical area) were associated with behavioural context, and were not specific

to the unique task type (i.e. orientation – Vogels & Orban, 1994). Conversely, orientation-specific improvements were seen in the lower cortical area of the V4 (Yang & Maunsell, 2004).

Ahissar and Hochstein propose that these modifications may be found as low as the V1, though they argue that this would not be common. There is some electrophysiological evidence that this is the case; Hua et al. (2010) found that cats who were 'trained' on specific spatial frequencies of grating stimuli showed significantly greater mean contrast sensitivity in the V1 neurons when tested on that specific spatial frequency.

Despite some evidence toward the RHT, there have been some concerns based on conflicting evidence. Watanabe, Náñez, & Sasaki (2001) found evidence that perceptual learning could occur when the stimulus was not consciously perceivable, during a time period during which observers did not consciously attend to this stimulus, and when the stimulus was not task-relevant. In this experiment, participants became more sensitive to the direction of motion of a field of moving dots, despite all of the above factors. This is indicative of a bottom-up process, as opposed to the top-down process argued for by the RHT.

Seitz & Dinse (2007) conducted a review into perceptual learning that also addressed the notion of 'passive' (i.e. in the absence of active deployment of attention) perceptual learning. By assessing multiple studies, they found that repetition was sufficient to produce robust perceptual learning effects, by aiding stimulus signals in rising above a 'learning threshold'. They offered a hypothesis in which co-activation (similar to a Hebbian learning process, in which neural plasticity is directed by synchronous neuron firing) is the common mechanism underlying perceptual learning, rather than a top-down process such as the RHT. This was echoed by Rosenthal & Humphreys (2010), who concluded that subliminal perceptual learning of global contours had a "Hebbian-like" underlying neural plasticity mechanism, which allowed for this learning to occur during repeated, non-perceivable, task-irrelevant stimuli.

It is still debatable as to whether evidence of a bottom-up system contradicts the top-down-dominated RHT, as the RHT discusses only consciously-performed learning and does not explicitly preclude bottom-up learning. That being said, the

RHT argues that implicit processing, i.e. the information produced during the first feedforward sweep of information up the visual hierarchy, is used to develop high-level percepts such as category information. This suggested co-activation system would either argue that this feedforward view is wrong, or – perhaps with more nuance – that the RHT's view of the feedforward sweep is too simplistic; that it does not account for perceptual learning in the absence of attention, or address how this might affect the information that becomes available to conscious perception further up the hierarchy. If this view of the sweep is too simplistic, then questions are raised about the validity of its presentation of 'vision at a glance', and possibly the model as a whole.

The RHT is also built upon the notion of psycho-anatomy (Julesz, 1971), in that information about behavioural outputs of the brain gives information about its anatomy; it assumes a direct link between the function of an area, and the properties of the receptive field it contains. Ahissar and Hochstein address this, noting that there's enough cortical variability that there can be a significant overlap between receptive fields. It's also not as neat a progression from large fields in higher areas to small fields in lower areas, as there exist 'smaller' receptive fields in higher areas. This would produce the argument that learning specificity can occur in higher-level areas.

In terms of evidence for these 'smaller' receptive fields, DiCarlo & Maunsell (2006) indicated that animals trained with precisely-located stimuli show that the receptive fields of the inferotemporal cortex (IT) can be small. Specifically, DiCarlo and Maunsell investigated how monkeys responded (both behaviourally and in terms of 146 anterior IT neurons) to the effects of small differences in presented visual forms (0.6 degrees wide). Monkeys were trained in how to identify these forms. They found that all recorded neurons showed a surprisingly strong response to very small (1.5 degree) changes in position; for example, one neuron showed a sensitivity to 2 of the 4 target forms when presented centrally, but little to no response to any of the target forms when presented 1.5 degrees ipsi- or contralateral to center of gaze. They argued that this might represent a small (~2.5 degrees in diameter) receptive field.

DiCarlo and Maunsell further developed this by hypothesising that the purpose of small receptive fields in the IT might be to limit the intrusion of flanking visual clutter into the neuron's receptive field. They found that the introduction of such flanking distractors had only a mild impact on the neuron's response to best target and best position, and form selectivity. This was in line with the hypothesis that these neurons had a small receptive field.

Whilst Ahissar and Hochsetin indicate this is clear-cut evidence of small receptive fields in higher cortical areas, DiCarlo and Maunsell found no predicted relationship between predicted receptive field size and clutter immunity, suggesting that whilst the IT *may* have smaller receptive fields, this is not completely confirmed nor the whole story. The assumption of higher-level, 'smaller' receptive fields may be sensible, but it is still an assumption, which implies the RHT is not an airtight model (as can be seen in the evidence of bottom-up perceptual learning).

1.5.3. Evaluation of the GST

The limitations of the FIT and the RHT do not automatically suggest that the GST is the superior model. The GST has a separate set of predictions, evidence and criticism, which comes from neuroimaging studies, research into 'gist', and contextual 'guidance' information in vision. It is also important to note what contextual guidance means; the RHT also determines that global attention may be related to top-down guidance, but the GST refers specifically to semantic and episodic guidance, as it is a key part of the model.

According to the FIT, the search for an object consisting of conjunctions of features ought to be slow and inefficient. However, as Wolfe, Võ, Evans, et al. (2011) note, visual search for items in a naturalistic or real-world scene are often extremely fast; in this vein, Peelen & Kastner's (2011) study not only offered a criticism of the FIT, it also suggested possible support for the GST.

The authors found, as part of the previously-mentioned study, that faster and more effective search for objects in natural scenes were associated with preparatory activity in the object-selective cortex (OSC). Peelen & Kastner associated this with a more 'abstract' search strategy, in which participants prepared to search for more

general visual cues, such as spatial locations and category diagnostic criteria. In contrast, less effective strategies were associated with preparatory activity in the V1. This was linked to participants' self-descriptions of searching for lower-level features of the target object, such as orientation cues. Peelen & Kastner described this as a sub-optimal strategy.

The implication of these findings support the GST in that the 'abstract' search strategies appeared to involve multiple sources of guidance, such as scene-based and semantic information. This is very similar to the description of the GST, in which non-selective processing allowed for the recovery of information (such as spatial layout), which allowed for subsequent guidance in search. This evidence could support the assertion that participants were 'priming' their attention for this guidance information, the mechanism for which was the OSC. Without this information (i.e. when participants chose to focus on specific low-level features), visual search of the realistic environment was much less efficient. This conclusion is not drawn explicitly from their data, however, and is merely one post-hoc explanation for the author's findings.

However, this does indicate that 'gist' research can be used to investigate the GST, as Wolfe et al. discuss gist as a product of the non-selective pathway, and therefore a part of the model as a whole. Gist is the phenomenon in which multiple types of information can be extracted rapidly (100 msec – Potter, 1976), as will be discussed in chapters 3, 4 and 5, and is part of the GST's non-selective pathway. Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe (2013) noted that gist cannot infer location information for a specific target; however, Wu, Wick, & Pomplun (2014) observed that gist allows observers to infer the existence and locations of a scene's objects without needing direct gaze fixation.

This implies that semantic information allows for this inference of object location, as suggested by Wolfe et al. (2011); gist provides the semantic information which thus allows an observer to search for a specific item more efficiently. For example, semantic knowledge informs us that a toaster may be found on the countertop of a kitchen, but it unlikely to be found on the floor or in the sink. Visual search for such an item may then begin at waist-level within the kitchen in order to more speedily find the object.

Further evidence comes from Wu et al.'s summary of eye-movement research in visual attention, in that semantic similarity (and the semantic associations between objects) is a cue that guides attention.

There is also research that demonstrates contextual information is an integral part of visual search, fitting with the GST. Torralba, Oliva, Castelhano, & Henderson (2006) developed a Bayseian framework called the Contextual Guidance model to predict observers' fixation points in natural scenes, based on top-down and bottom-up mechanisms combined with scene context. This was further developed by Preston, Guo, Das, Giesbrecht, & Eckstein (2013), who provided evidence to support this scene-context guidance of visual search. Preston et al. (2013) discussed how, in real-world scenes, participants guide their eyes to areas that have global scene properties indicative of the target, as well as toward objects that frequently co-occur with the target. They found that activity in the lateral occipital cortex (LOC) was predictive of both of these functions, indicating strongly that it is important for contextual guidance, and that it represents the importance of objects before search. This indicates that the LOC is a neural correlate for the selective pathway, and suggests neuronal evidence for the GST model of visual attention. This would be strengthened by identifying the neuronal correlates of gist processing, in order to link the two and provide an estimate of the network required for the model.

These studies together represent a system in which gist provides category information, which in itself provides semantic and context information, which then guides attention. This is a close fit to the GST. However, it is important to note that this is an assumption that has not been directly tested. Wu, Wang & Pomplun (2014) directly studied the relationship between scene gist and semantic guidance in visual search, and found that scene gist did not have an effect on semantic-based guidance of attention in real-world and natural scenes; rather, the spatial dependency of the visual objects (such as the layout and co-occurrence of an object) was enough to produce semantic guidance. This serves to highlight the issue that, despite the evidence supporting the GST, the processes by which it explains the relationship between the two modes of attention are assumptions and require more direct evidence.

1.6. Conclusions

While the FIT and the GST are both referred to extensively in visual attention literature, a direct comparison has not yet been made. Both have been criticised and undergone multiple iterations to account for new information, but the proposed core mechanisms remain unchanged. Concurrently, the whilst the RHT has not broadly been revised, it has been used as a framework to explore real-world phenomenon such as visual perceptual learning (Ahissar & Hochstein, 2004) and sensory learning (Ahissar, Nahum, Nelken, & Hochstein, 2009). Each of these models offer different predictions for the relationship between types of visual attention.

Due to the FIT's proposed nature as a continuum of attention, this would predict that a person could perform tasks that required focused attention, but not simultaneously tasks that required distributed attention; any attempt to use both kinds of attention at the same time would result in a deficit in both types of attention, as the observer's attention landed on a point on the continuum between focused and distributed.

The RHT argues that the two kinds of attention operate in serial, with broad attention the result of the feed-forward sweep and narrow attention the result of re-entering the visual processing stream. When loaded, this would result in better accuracy for tasks requiring sparse 'vision at a glance' attention, and reduced accuracy for tasks requiring the narrow 'vision with scrutiny' attention.

Conversely, the GST proposes that the two types of attention – selective and non-selective – occur concurrently, and this predicts that an observer tasked to use both kinds of attention would perform equally well on both measures.

This raises the issue of how to broadly discuss the 'kinds' of attention without choosing the nomenclature of a specific model. The commonalities between distributed/non-selective/broad attention, and focused/selective/narrow attention, are their scope within the visual field. The former operates on what might broadly be determined a global scale, with attention applied to a large visual scene; the latter operates on a local scale, commonly limited to a specific part of the scene and allowing for object binding and identification. Going forward, attention that may be referred to as distributed/non-selective/broad will be termed 'global' attention, and

attention that can be labelled as focused/selective/narrow will be termed 'local' attention. This will allow the 'kinds' of visual attention to be discussed neutrally, and without the assumption of preference for a particular model.

1.7. Outline of the thesis

The thesis objective is addressed in each of the following empirical chapters, in which the relationship between the 'global' and 'local' attentional resource(s), and the topic of global attention more broadly, is investigated. The overview for each chapter will briefly describe the chapter, and the introductions also reference more in-depth literature relevant to the chapter question.

This thesis will adhere to the following structure: chapter 2 introduces the broader, more general topic of visual attention as a whole, investigating the relationship between 'global' and 'local' visual attention through a series of psychophysics experiments. Chapters 3, 4 and 5 then focus on a specific aspect of global visual attention, known as 'gist processing'. Chapter 3 addresses the neuronal correlates of 'gist' processing utilising a GLM-based fMRI design. Chapter 4 investigates the temporal dynamics of gist processing using EEG and an MVPA analysis. Chapter 5 investigates a specific element of gist processing called 'destructive interference' using a GLM-based EEG design. Finally, chapter 6 discusses the conclusions from each chapter, assess the novel contributions to the field, and proposes any further questions raised or directions for future research.

Chapter 2 Perceptual correlates of selective and non-selective attention – testing three separate models

2.1. Overview

How might 'global' and 'local' attention relate to one another, as outlined in chapter 1? The aim of this chapter is to more directly establish this relationship though empirical testing of both modes of attention.

The Feature Integration Theory (FIT), the Reverse Hierarchy Theory (RHT) and the Guided Search Theory (GST) are three of the most robust models for explaining this relationship, and all offer very specific and testable hypotheses. These are addressed in order to see how the two modes of attention operate. Chapter 2 begins by broadly assessing what predictions these models offer in explaining the relationship between the two modes of attention. These predictions are then used to build a hypothesis for how data produced by human participants would look, in line with the continuum model and the binary model of visual attention.

Four experiments were conducted, with experiments 1, 2 and 3 directly addressing the hypothesis. Experiment 4 was an auditory control experiment, in which the experimental design was tested to ensure selective/focused and non-selective/distributed attention were being measured in the experimental conditions, rather than attentional resources as a whole. Experiments 2 and 3 improved on the design of the preceding experiments by altering the characteristics of the stimuli, to better access the two modes of attention and reduce confounds. A 5th experiment is also included based on data later collected by a research assistant. Data from the three experimental tasks (and the later fourth experimental task) were combined to find that visual attention appears to operate using a serial system as indicated by the RHT.

2.2. Background

The three key models of visual attention, as described in Chapter 1, identified two modes of attention. Firstly, there is a broad global mode, possibly involving a rapid, automatic, broad process, responsible for the rapid extraction of global information, including summary statistics and category information. Secondly, there is a narrow local mode: a slower, more detailed mode of attention, allowing for conscious access to the visual scene. Several models of visual attention have applied a framework to this apparent dichotomy of visual attention, asserting a specific relationship between these global and local modes.

Treisman's Feature Integration theory (FIT – Treisman, 2006)posits that these modes represent a single process. This would mean the relationship between focused and distributed attention (as Treisman refers to the local and global modes of attention, respectively) is that of a continuum, with focused and distributed attention representing the extremes of visual attention at the poles of the scale.

Distributed attention, as discussed, provides global information including summary statistics, which allows for the extraction of properties of groups of objects as well as scene layout. This can include the mean size of the objects, irrespective of their number or density (Chong & Treisman, 2005), and represents a 'wide' attentional window which allows the global integration of feature maps. Focused attention, on the other hand, provides observers with the conscious experience of individual objects, including their location and configurations. This represents a 'narrow' attentional window – a bottleneck that interacts with specific aspects of objects and allows for binding of features into a perceptual whole.

Visual attention, therefore, will occupy a position on this continuum. This position may be toward the 'distributed' end, which would allow for the extraction of summary statistics, though not for the individuation of objects. If the position is more towards the 'focused' end, then the opposite is true. In practice, observers can shift from one end of the continuum to the other as the task requires. However, this would mean that good performance in one type of visual attention results in poor performance in the other. When both extremes of visual attention are required in a single task, then the middle position of the continuum is occupied, hypothetically resulting in performance drops on both tasks.

Despite self-describing as a continuum, Hochstein and Ahissar's Reverse Hierachy Theory (RHT – Hochstein & Ahissar, 2002; Ahissar & Hochstein, 2004) does not operate in the same manner as the FIT. Like the FIT, it posits two extremes of attention – broad, sparse attention, analogous to the FIT's distributed attention (the global mode) in that it covers a wider area and allows for the extraction of gist, category information, summary statistics, and so on. However, whilst the FIT proposes that attention sits on a point between or on these two extremes, the RHT argues that sparse attention is always activated first, as it is the result of the first feedforward processing of visual information through the brain.

Access to the more focused attention (the local mode) is then controlled by context, with re-entry into lower cortical areas allowing for the extraction of more details of features and objects. In this manner, visual attention is a serial process, in which sparse attention is used first for broad, general information about the scene, and then focused attention. The RHT does not make any explicit claims as to whether visual attention is a single resource or two, but in the context of the FIT and the GST, it is fair to say that the model argues for two sources of visual attention activated serially.

Wolfe et al.'s updated Guided Search Theory (GST – Wolfe, Võ, Evans & Greene, 2011), in contrast, argues explicitly that visual attention is two distinct processes: a specific resource for selective (local) attention, and another for non-selective (global) attention. This produces two hypotheses. The first is that the system operates in serial, with non-selective attention preceding selective attention, which allows information recruited from the non-selective pathway to inform or 'guide' selective attention to appropriate locations. The second hypothesis is that the two modes of attention operate in parallel, with more 'real-time' guidance information passing from the non-selective to the selective pathway. These hypotheses indicate that, should an observer be asked to respond to a task that requires both kinds of attention, one of the two will outperform the other (as explained in the next section in Figure 2.1).

However, Wolfe et al. argue strongly for the 'parallel' hypothesis as opposed to a serial process; they explicitly state that selective and non-selective pathways operate in parallel, and it is this parallel processing which gives rise to the conscious

perception of a visual scene. Whilst it is possible to fit the GST into a serial process (not dissimilar to the RHT), the authors have made it clear that their model should only be considered within a parallel framework.

In the broader literature, both modes of attention have been studied in isolation, and models have been constructed in an attempt to describe the relationship between them. Chapter 1 gives an overview of the key models from the last 40 years. Many of these models of attention have been developed in response to one another; in particular, the GST and the Attentional Engagement Theory (Duncan & Humphries, 1989, 1992) are explicitly discussed as a response to the perceived shortcomings of the original Feature Integration Theory (Treisman & Gelade, 1980); the updated FIT (Treisman, 2006) also references the research underpinning the RHT, showing that there is some overlap between the three models.

As discussed, the three most robust of these models are the FIT, the RHT and the GST, and it is these models which will be tested in chapter 2. Both the FIT and the GST have undergone multiple iterations, with the latest version of the FIT presented by Treisman (2006), and the latest version of the GST by Wolfe, Võ, Evans, et al. (2011). The RHT was discussed and presented between 2002 and 2004 (Hochstein & Ahissar, 2002; Ahissar & Hochstein, 2004), and has not changed drastically since, though it has been used to model the effects of sensory learning and early perceptual learning (Ahissar et al., 2009). It is these three models that appear to underpin the hypotheses, or are used to explain the findings, of modern research into visual attention.

Each of the models has different strengths and weaknesses. While the FIT has weathered more criticism, and the GST and RHT provide more neuronal and behavioural evidence, it is this key comparison that remains – that of a continuum system contrasted against a serial system, or a binary parallel system. This question remains unanswered to date, and forms the basis for a new investigation into visual attention.

2.3. Aims and hypotheses

2.3.1. Initial experiments

The literature in chapter 1 has indicated that the FIT, the RHT and the GST have evidence to support and to criticise them. However, evidence for and against each hypothesis is indirect. The hypotheses offer specific and testable predictions based on their differing approaches to the two modes of visual attention.

The language convention discussed in chapter 1 for the three hypotheses will be continued throughout this chapter. When multiple possible interpretations of the modes of attention may be applied (such as broad or general descriptions), the term 'global attention' will be used to represent the distributed, sparse or non-selective mode of attention; conversely, 'local attention' will be used to describe the focused/selective mode. These terms will also be used to define the tasks participants will undertake. A task requiring the global mode of visual attention will be referred to as a 'global task', and tasks requiring the local mode of attention will be referred to as a 'local' task.

The aim of this chapter is to directly investigate the relationship of these two modes using a dual-task paradigm. In this dual task, the primary task will have two possible modes: one which taxes local attention, and one which taxes global attention. This task will be done simultaneously with the secondary task, which will only tax global attention in experiment 1–4, and local attention in experiment 5. There will be two possible task types: these 'dual' tasks, and 'solo' tasks, in which participants are asked to respond only to a single task (primary-global, primary-local, and secondary).

Comparing participants' performance on these 'dual' tasks to their performance during 'solo' tasks would allow us to see which, if any, task showed a decrease in performance. This data would allow us to determine which of the models discussed is the closest representation of visual attention. In the examples given in Figure 2.1, the possible responses are cartooned as attention operating characteristics (AOCs). For the dual tasks, performance for the secondary global task is assumed to remain the same, due to participants responding to this task first.

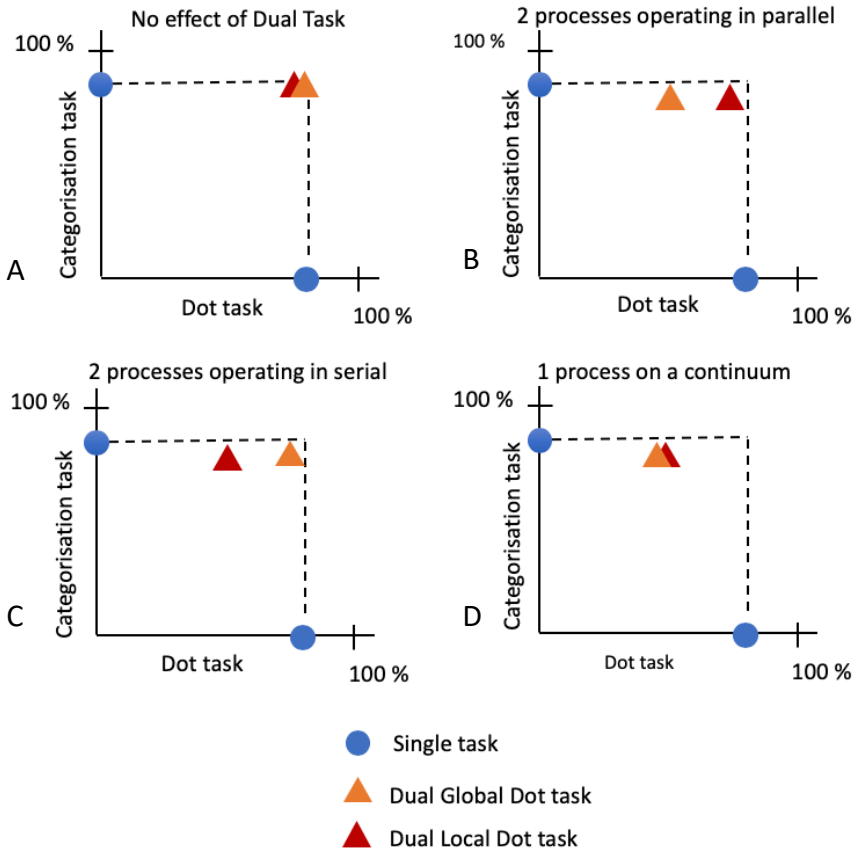

Figure 2.1: Predictions of data for each of the proposed hypotheses in the first 4 experiments, including a control 'no effect' example (A), plotted on attentional operating characteristic graphs. Primary task performance is plotted on the X axis, with each task type indicated using the key. Secondary task performance is plotted on the Y axis. Performance of each task independently is indicated by the dotted line, at 62% threshold, for comparison purposes. The GST hypothesis is represented by the parallel (B) graph, the RHT hypothesis by the serial (C) graph, and the FIT is represented by the continuum (D) graph. A slight dip in performance for the categorisation task during the dual tasks (as compared to the single task) reflects the higher difficulty of the dual task, but this is not predicted to be statistically significant.

AOCs are a type of graph used in the study of attention, in which levels of performance on one task are plotted or the X axis, and levels of performance for another are plotted on the Y axis. This allows the viewer to instantly compare the performance of the two tasks along their axes, and any tasks which combine the two tasks within the graph space.

Figure 2.1 A is the 'baseline' comparison, so to speak; these are the expected results if the null hypothesis were true, and performance accuracy is not impacted by the effects of the dual task. In this example, participants perform equally well on all possible tasks, regardless of whether they are presented as a single task or a dual task. As seen in Figure 2.1 A, performance on the global categorisation task is

recorded along the Y axis, with performance on the single task on the axis itself, and performance on the local 'dot' task recorded along the X axis. The dotted lines indicate the maximum level of performance recorded for these tasks. 'Dual' tasks are recorded with the performance on each task. In Figure 2.1 A, these are proposed to match the performance of the single tasks, as there has been no deficit in performance between single and dual tasks. The markers for the dual tasks are therefore presented at the maximum performance level for the single tasks.

Figure 2.1 B demonstrates the expected results if processing is parallel, as suggested by the GST; in this case, performance in the dual average dot task is worse than for the dual single dot task. This finding would indicate that the global attentional resource is being taxed more heavily, due to both tasks requiring this type of attention. However, as the load is spread more evenly across both resources in the dual single dot task, performance on this task is unimpaired. Conversely, Figure 2.1 C demonstrates the expected findings is processing is serial (as per the RHT); this shows the opposite pattern of expected findings as the parallel model. In this model, performance for the dual average dot task exceeds that for the dual single dot task due to processing only having to enter the first stage of the model (in this case, global processing).

Figure 2.1 D represents the expected findings in visual attention operates using the continuum proposed by the FIT. In this model, both dual dot motion tasks show a reduction in performance compared to their single-task equivalents. This is because the FIT posits that attention is a single resource, and the dual tasks would tax this resource to the same degree.

The ultimate aim of this chapter is to assess each of these possible models of visual attention, with the hypothesis that our data would match one of the models in Figure 2.1 (2.1 B, 2.1 C or 2.1 D). The experiments in this chapter assess this by measuring participants' performance accuracy on tasks requiring global and/or local processing.

2.3.2. Follow-up experiment

The purpose of the four experiments as described above compare observers' responses to experiments that always load global visual processing, and then may or may not load local visual processing. However, this raises a corollary: instead, if it is the local visual process which is always loaded, would we see a double dissociation in the data? In essence, would we be able to confirm the findings of the initial experiments by reversing the design, and producing similar findings? This experiment (henceforth experiment 5) was conducted after the conclusion of data collection for this thesis.

By assigning the secondary task to require local visual attention rather than global visual attention, this has the effect of inverting the predicted responses shown in Figure 2.1. For example, the GST would argue that the data would show the opposite of the prediction in Figure 2.1 B. During dual trials, the local dot task would suffer, as a dual local-local task would tax the single local attentional resource the most heavily.

As before, a serial-based model would predict the opposite of the parallel-based model. The RHT predicts that performance accuracy on the global dot task during the dual local-global task would be reduced. This is because (according to the serial nature of the RHT) dual local trials allowed for re-entry into earlier stages of visual processing, meaning both local tasks would benefit from the engagement of the same resource. Dual global-local tasks, on the other hand, would see the same performance deficit as in experiments 2 and 3, due to the task requiring the recruitment of both the sparse attention mode and the focused attention mode. As participants are not needing to conscious extract information from the spare attention mode, dual local tasks draw from the same resource and take less time to execute.

And finally, the continuum model (FIT) would predict an approximately equal deficit in performance accuracy for both the dual global and local tasks, due to the single posited attentional resource being similarly taxed.

By finding a double-dissociation in the data, this final experiment was intended to confirm the findings of the first four experiments, and bolster the conclusions drawn. Data for this experiment were collected by a research assistant after collection and analyses were done for this thesis.

2.4. Methodology

For the first four experiments, participants took part in one of each of the experiments. Each experiment contained two tasks requiring global attention, and one requiring local attention. The difference between each of the four experiments was the nature of *one* of the tasks. For all experiments, one of the global attentional tasks (i.e. the primary global task) was to identify the average direction of motion of a whole field of dots. For the other (secondary) global attentional task, the first two experiments utilised a 2AFC image task. This involved presenting participants with a category cue word (such as "beach"), followed by an image. Participants were then asked to identify whether the category of the image was the same as the cue word. The local task required the participant to track the direction of dot differentiated either by size (experiment 1) or contrast (experiment 2). Experiment 3 also used dot contrast for the local task, and utilised average dot colour (2AFC) as a global task, in which participants were given a colour cue word (red, pink, green or blue), and asked to identify if the majority colour for the dots onscreen matched the colour cue. Experiment 4 utilised a 2AFC sound task instead of images, in which participants were given headphones and asked to identify the ear to which a sound had been played. The local task required the dot being tracked to be differentiated by contrast.

Experiment 5 differed in that it required participants to respond to a 2AFC colour-bisected disc that recruited local attention, rather than global attention, for this secondary task. They had to identify if the disc was green-red (i.e. green on the left, red on the right), or red-green. Both the primary global and local dot tasks were the same as in experiment 2.

2.4.1. Participants

All participants were volunteers recruited using the University of York Psychology participation website (PEEBS), and were offered either payment or participation

credit as part of their undergraduate Psychology studies. All participants had normal or corrected-to-normal vision. Ethical approval was granted by the University of York Psychology Ethics Committee, and all data were collected prior to analysis.

Over the course of the first four experiments, 145 participants were recruited, with a mean age of 21 years and 2 months, 13 left-handed, 4 ambidextrous, and of which 33 were men. Data for experiment 5 were collected after the body of data collection done for this thesis, and their demographic data are recorded separately below. Participants' data were not used if they failed to equate performance on the two primary dot tasks, i.e. if performance on the two primary dot tasks were significantly different as indicated by a two-tailed t-test. Equated data were important, as they established an equal psychophysical threshold from which to compare their subsequent performance on the two dot tasks. Participants whose data did not indicate significance (i.e. $p > 0.05$) were included in the final analysis.

2.4.1.1. Experiment 1

A total of 45 volunteers were recruited (2 left-handed, 29 paid, 14 men, with a mean age of 20 years 2 months). Of these participants, 16 data sets were used (all right-handed, 10 paid, 7 men, average age of 20 years 2 months). These 16 were selected for further analysis as these participants had successfully equated performance on both dot-motion tasks.

2.4.1.2. Experiment 2

A total of 37 volunteers were recruited (7 left-handed, 3 paid, 4 men, with a mean age of 20 years 3 months). Of these participants, 16 participants were included in the final analysis, as they successfully equated dot-motion task performance (3 left-handed, 1 paid, 2 men, average age of 19 years and 5 months).

2.4.1.3. Experiment 3

28 volunteers were recruited (1 left-handed, 3 ambidextrous, all paid, 8 men, with a mean age of 24 years 3 months), of which 17 data sets were used (2 ambidextrous, 3 men, with an average age of 23 years and 11 months) due to equated performance on the single dot-motion task.

2.4.1.4. Experiment 4

A total of 35 volunteers were recruited (3 left-handed, 19 paid, 7 men, with a mean age of 21 years 5 months). Of these participants, 17 data sets were analysed (1 left-handed, 8 paid, 5 men, average age 19 years 10 months) due to equated single dot-motion data.

2.4.1.5. Experiment 5

A total of 26 volunteers were recruited (3 left-handed, 4 men, with an average age of 20 years and 0 months). All received departmental participation credit. Of this group, 16 participants equated their performance during the staircase dots tasks, and so were included in the final analysis (2 left-handed, 2 men, with an average age of 20 years and 0 months).

2.4.2. Stimuli and apparatus

2.4.2.1. Apparatus

The experiments in this chapter were designed and displayed using MATLAB 2015a (The Mathworks, Natick, 2015), and the MATLAB Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007), on a Dell XPS computer (Intel core i7-4770, with a CPU at 3.4GHz), running Windows 7 Enterprise. Stimuli were presented on a cathode-ray monitor (iiyama Vision Master Pro 514) at 85Hz, at an approximate viewing distance of 60cm. Participants were screened for colour blindness using the Ishihara Color Plate Test (Ishihara, 1917). A Minolta LS-100

photometer was used to measure luminance values for all colours used in the experiment.

2.4.2.2. Image stimuli

All images used were taken from a total of 450 greyscale images, equally distributed 36 'categories' of landscape and scene images. Image landscapes were broad and not defined by a single object. Images were all drawn from the MIT SUN database (http://groups.csail.mit.edu/vision/SUN/). Images were presented centrally, in a square with 20° of visual angle per side.

In order to counteract the visual signature of sudden onset, the image began and ended its presentation at 40% opacity. This was done as sudden changes in the visual field can 'capture' attention; in the context of this experiment, this may mean that local attention is abruptly directed to the image through the exogenous cue of abrupt appearance (Miller, 1989; MacLean et al., 2009), reducing or eliminating the effect of global attention. The opacity of the image, frame-by-frame, was calculated using an exponential curve of opacity over time, based on the number of frames required for image presentation. This allowed the image to fade in, become fully opaque, and fade out again (Figure 2.2).

All trials were presented on a greyscale natural scene texture image, created using the Portilla & Simoncelli (2000) texture algorithm. This texture image was used as a background image, with a mean relative luminance of 93.12 candella/m^2.

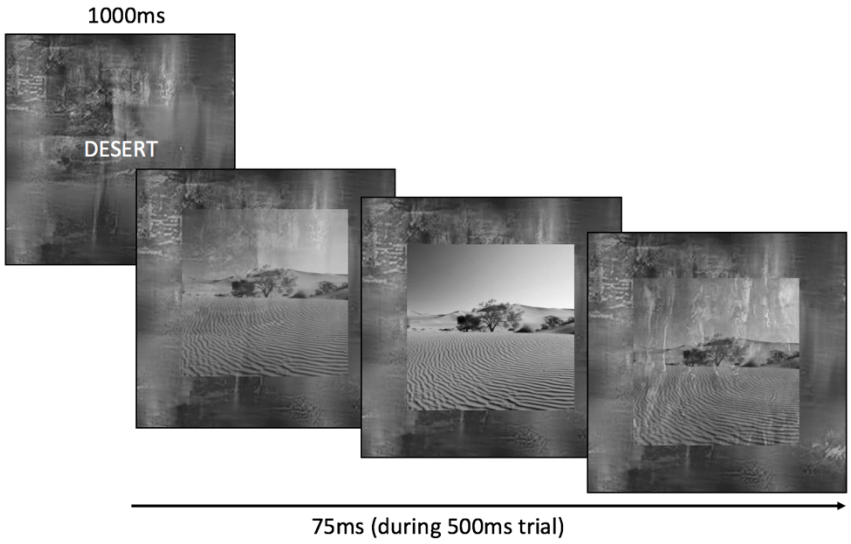

Figure 2.2: Cartoon example of how the images presented faded in and out, from 40% opacity to 100% opacity and back again. This is calculated using an exponential curve, using the total number of frames over which the image is displayed.

2.4.2.3. Sound stimuli

Sounds used were single-tone notes, played for 200ms. Participants wore headphones and manually selected a volume level that was comfortable for them. The sounds were either from a flute, a trumpet, a piano or a violin, and each instrument had 113 possible sounds of varying, randomised notes, ranging from 27.5 Hz to 4186 Hz (sounds downloaded from https://freesound.org. Major notes only from A0 to C8 used).

2.4.2.4. Dot stimuli

For experiments 1, 2 and 3, the dots were white (RGB values = 255 255 255; CIE 1931 XYZ colour values: X= 0.9505, Y = 1, Z = 1.0888). For experiment 4, four colours were chosen based on their equated luminance values using the photometer (an approximate average of 90 cd/m^2 for each colour) . These were red, pink, blue and green (see Figure 2.3).

For all dot-based tasks, 100 dots were generated with random starting positions within the aperture. These 100 dots were split into two 'fields' within a centrally-located square aperture of 20° of visual angle per side. (In trials with images, this was deliberately designed to overlay the image.) Dots were onscreen for

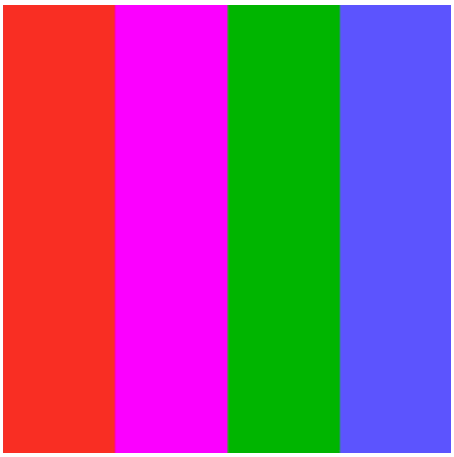

Figure 2.3: The four colours used for the dots. Colours were chosen due to their ability to be equated for luminance values. CIE 1931 XYZ colour values are as follows:

Red:	X = 0.4125	Y = 0.2127	Z = 0.0193
Pink:	X = 0.594	Y = 0.287	Z = 0.97
Green:	X = 0.3576	Y = 0.7152	Z = 0.1192
Blue:	X = 0.1804	Y = 0.0722	Z = 0.9503

500ms, moving 0.002 degrees per second. All dots moved in a straight line. One field contained dots which all moved in the same direction (coherent motion field) while dots in the other field moved randomly (noise motion field). All dots had a randomly-generated start point within the dot aperture, and when dots reached the edge of that aperture, they would be re-introduced from the opposite edge, maintaining speed, direction and with an entry point equivalent to that of the dot's exit point. The direction of the dots in the coherent motion field would be the direction the participants would have to identify (the global task, requiring global attention). The noise dot field contained dots which all moved in randomly-determined, non-coherent directions ('noise' field), but otherwise behaved in an identical manner to the coherent motion dots. All coherent and noise dots were presented with a diameter of 0.263° of visual angle, based on a pixel size of 9.

Additionally, a separate dot was generated for the local task (requiring local attention). This dot was presented centrally in a square aperture 5° of visual angle per side. The dot would begin on the edge of its aperture, and travel the length of the aperture square. Upon arriving at the opposite aperture wall, the dot would 'respawn' at its original starting point. This was to avoid the dot looping along the aperture, which would make the dot harder to track and identify. Depending on the experiment, this dot would be distinguished by being either larger (experiment 1) or of a darker contrast (experiments 2, 3 and 4) than the other dots onscreen. (It is important to note that the dot stimuli presented on screen were identical for both dot motion tasks; participants were simply asked to attend to different aspects of the task.)

For all experiments, the direction of motion of the dot(s) were chosen at random, with some limitations. Dots could be assigned a direction within 360°, with the exception of any direction within 8° of a diagonal (45, 135, 225 and 315 degrees). For example, directions of 37°–53° inclusive would not be generated; see Figure 2.4. This was done to limit the ambiguity of the dots' movements, without making the task too easy. A single direction would be generated for the dots of the coherent field, and individual directions would be generated for each dot in the noise field.

The difficulty of the two dot tasks were determined by a staircase task, which is described in the next section.

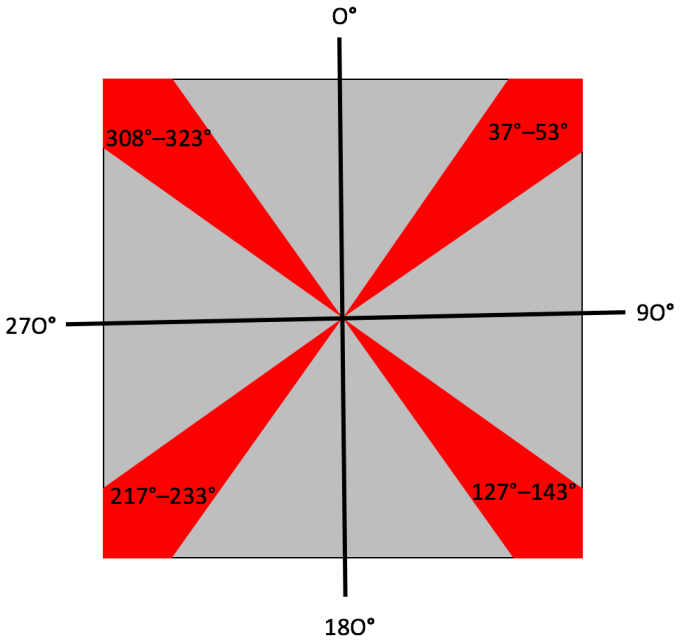

Figure 2.4: A cartoon map depicting the range of dot directions possible. Any direction traveling outwards from the centre within the grey regions was permitted; any direction within the red regions (labelled to include degrees) was excluded.

2.4.2.5. Disc stimuli

For experiment 5, a local-based disc task replaced the secondary global task. This utilised a small, bisected dot in the center of the screen which was split into two halves by colour. The dot would appear either with the left half red and the right half green, or vice versa. It was presented centrally on-screen, and at approximately 0.1 degrees of visual angle.

2.4.3. Design

Within each experiment, all participants performed all three tasks (task types) in both the single and the dual conditions (task conditions) to allow for within-subject comparisons across the experiment. The experiment utilised both 2-alternative-forced-choice (2AFC) and 4-alternative-forced-choice (4AFC) responses to measure participants' performance accuracy during several different task types. After the staircase section was completed, participants were tested on 5 different tasks. Three of these tasks were a single task condition, and two of which were dual task conditions.

Each experiment had two primary tasks (4AFC) reporting on dot motion, with one task recruiting global processing (report on the motion of the coherent field of dots), and the other local processing (report on the motion of the single odd-one-out dot in the centre of the filed). These tasks stayed the same across all experiments, and participants had to respond to the movement direction of the dots (or dot) using the four arrow keys on the keyboard. For the primary global processing task, participants were asked to identify the average direction of motion of the dots (see Figure 2.5). For participants to do so they first needed to identify the coherent field of dots. For the local processing task, participants were asked to identify a single dot in the field of dots based on a given characteristic (size or contrast), and indicate its direction of travel (see Figure 2.6).

For both dot motion tasks, participants indicated dot direction using the four arrow keys on the keyboard. The dot movement direction on each trial, however, was not always a 'cardinal' direction (e.g. 0° for up). Participants responded to trials

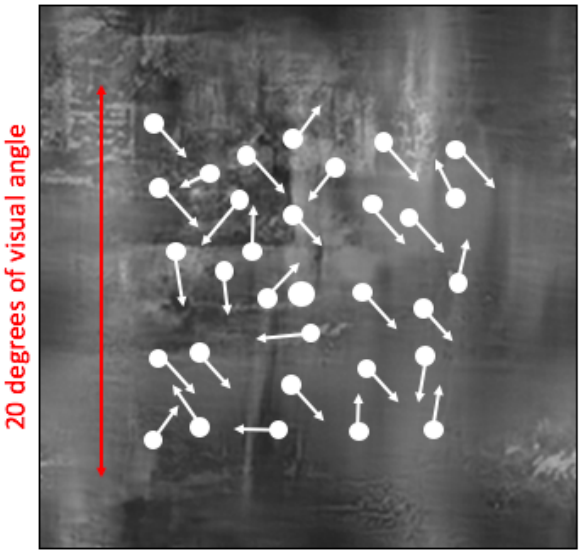

Figure 2.5: Cartoon image of the global dot motion task, with a demonstration coherent motion angle of 135° and coherence threshold of 60%. This angle would not have been used in the experiment to avoid confusion, but is used here to make the coherent angle clear.

with the arrow key that most closely resembled the direction of the dots (see Figure 2.7).

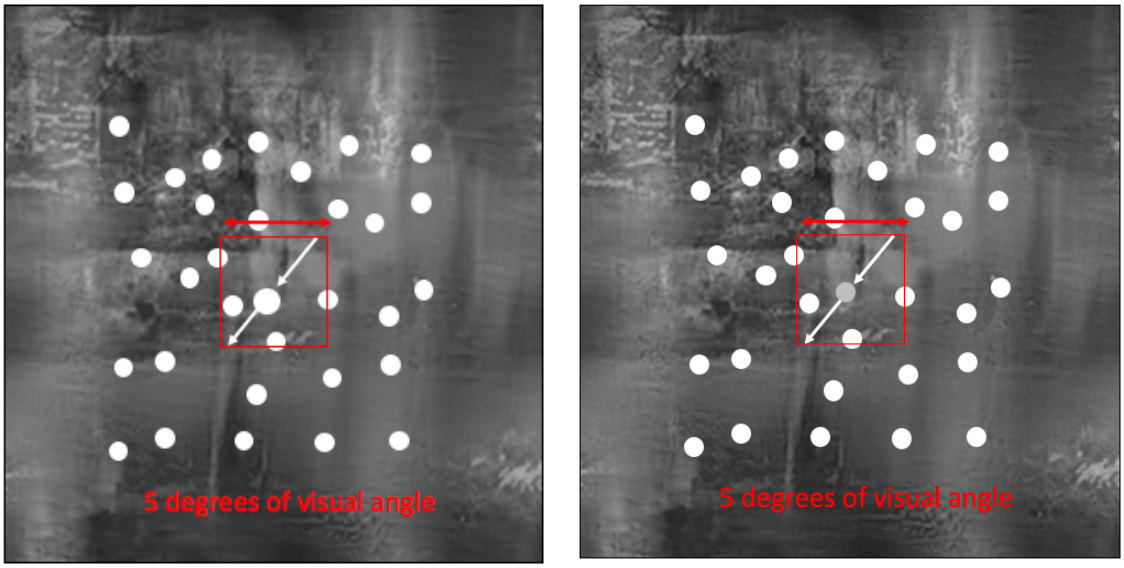

Figure 2.6: Figure A represents a cartoon image of the local motion task. The larger 'signal' local dot is moving at 225°, to make the direction of motion clear in this example. Upon exiting the aperture at the bottom, the dot would begin at its original starting point at the top of the aperture, as indicated. Figure B is identical, except the signal dot is of a lower contrast rather than larger than the surrounding dots.

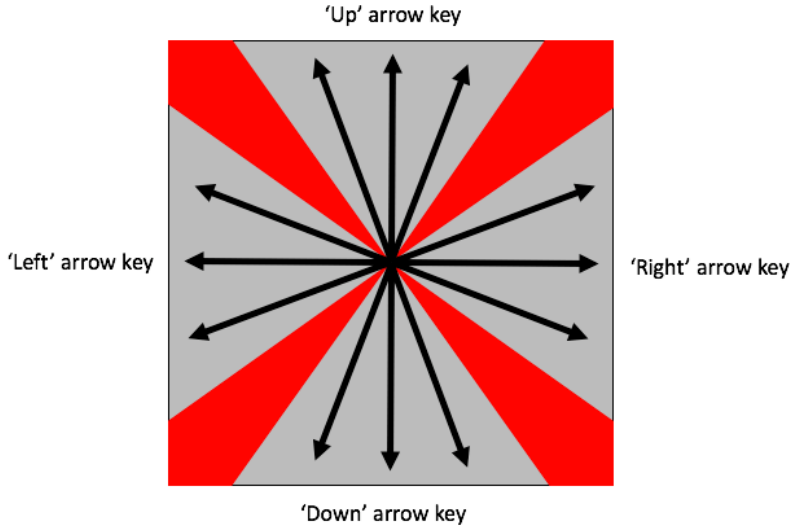

Figure 2.7: A cartoon demonstrating what keyboard button would be the appropriate response for the direction of travel for the dots. No dots would travel within the red diagonal sections, as shown in Figure 26.

2.4.3.1. Staircase

Before participants completed the experimental trials, they were asked to complete 65 trials of each of the single dot motion tasks (and, for experiment 4, of the average dot colour task). They were allowed 5 practice trials for each task, and the results of the remaining 60 trials were calculated using a probit ('cumulative normal') staircase method. The staircase made the task more difficult after three correct responses, and easier after one incorrect response (1-up, 3-down pattern).

For all experiments, the difficulty level of the primary global motion task was controlled by changing the ratio of coherent dots to noise dots (as there were 100 dots onscreen, the number of coherent dots were decreased, and the number of noise dots increased by a corresponding amount to make the task harder. For example, with a coherence threshold of 35%, there would be 35 dots in the coherent motion field and 65 dots in the noise field). The threshold for this task was set to 62% accuracy due to the 4AFC nature of the dot tasks, in which 62.5% represented the halfway point between the stimuli being non-discriminable (25% accuracy, or chance performance) and completely discriminable (100% accuracy).

In experiment 1, the difficulty level of the local motion task was controlled by altering the size of the dot. Initially, both tasks used a threshold, calculated using a probit/ 'cumulative normal' staircase method, at 62%. This 62% value was rounded to the nearest integer, in order to be displayed in pixel size. This target was chosen for all staircase tasks as it lies halfway between chance performance on a 4AFC task (25%) and perfect performance (100%).

However, as the size value had to be an integer, this led to situations where the dot was either too large or too small to achieve the desired threshold level. For example, participants whose staircase indicated a value of 12.4 pixels to achieve the desired performance would have this value rounded down to 12, which would be too difficult. This made it hard to equate performance across the two motion tasks. To compensate, the threshold for this task was changed to 82%, as it was easier to calculate correctly for the participants.

In the resulting experiments, this task was changed from looking for a larger dot to looking for a dot of lower contrast. This was to reduce the imprecision in determining a performance threshold for the single dot to 62%, and was then implemented for experiment 2, 3 and 4. Changing the contrast of the dot was

achieved by changing the RGB values for the local dot during the staircase trials. When the task was easy, the dot was clearly a different contrast from the other (white) dots on screen, represented by equally lower RGB values (i.e., where all other dots onscreen had RGB values of 255-255-255, the local dot might have values of 150-150-150). As the task got harder, these values were raised to be closer to that of the white values (Figure 2.6 B). This task also aimed for a performance level of 62%.

2.4.3.2. Secondary tasks

In experiments 1 and 2, the secondary global task used images. Participants were shown a cue word containing a place category such as airport, beach, or lake (1 second). The cue word presented would match the presented image category in 50% of trials. In the other 50%, the cue word would represent a different category to the presented image. The participants were then presented with the background texture mask for 500ms. During this window an image faded in at a variable onset time and remained fully visible for 75ms, after which it faded out. The image start time was jittered so that it would appear in a different 75ms window for each trial. This jittering was limited, so that the image could not begin appearing within the first 100ms of the trial, or finish appearing within the last 100ms. At the end of the trial, the participant would respond by indicating if the image matched the cue word (by pressing the A key), or if it did not (the S key). The next trial would not begin until the participant had made a key press.

In experiment 3, colours were used as an inherent property of the dots, and the colour ratio acted as the secondary global task. In this experiment, participants were given a colour word as a cue (red, pink, blue or green) and had to report whether the average colour of a set of dots on screen matched the cue word. Experiment 3 was the only experiment in which the secondary global task also required a performance threshold to be applied using the staircase tasks. This was achieved by changing the ratio of the 'main' colour compared to the rest of the dots.

To do this, both the coherent motion dot field and the noise dot field had the same proportion of colours assigned to them (red, pink, green and blue). This

proportion was split by determining a 'majority colour'. This colour would then be applied to a set percentage of all dots onscreen, split proportionately between the coherent and noise dot fields. The remaining colours were then distributed equally between the remaining dots, proportionate depending on the size of the coherent and noise fields (Figure 2.8). For example, let's propose that the 'majority colour' for a given trial is red. If the dot field had a coherence of 35%, and 'major colour' proportion of 30%, this would mean 30% of the coherent motion dots (approx. 10) and 30% of the noise dots (approx. 20) would be red, with a total of 30 red dots out of 100. There would be approximately 23 dots of each of the other three colours in this trial, with approximately 8 of each colour in the coherent field, and 15 of each colour in the noise field. The number of 'majority colour' dots were the value which was calculated using the staircase, with a performance threshold of 75%. This number was chosen, like the previously-discussed 62% threshold, as it represents the halfway point between chance and perfect discriminability for a 2AFC task.

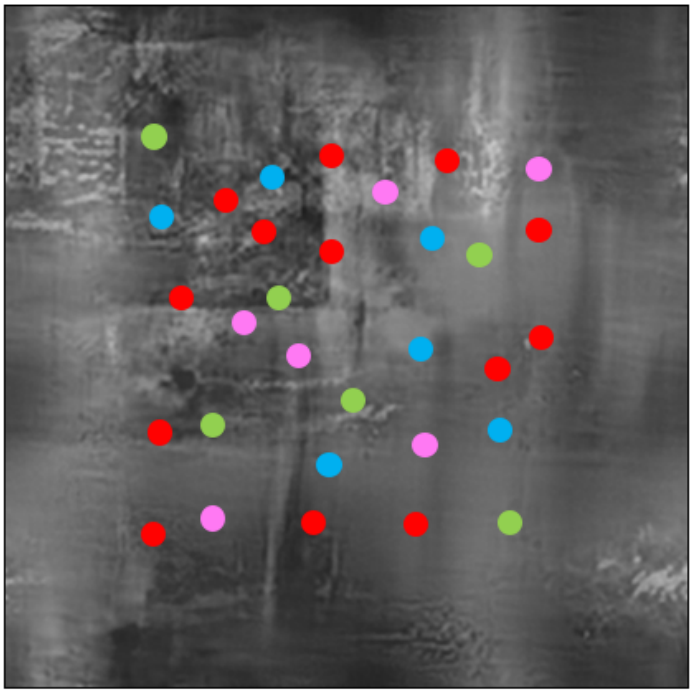

Figure 2.8: Cartoon image of the dots presented in experiment 4. In this example, the 'majority colour' is red, with the other colours distributed equally over the remaining dots onscreen.

Experiment 4 sought to test if previous experimental findings were *not* the result of the attentional requirements of a dual task in general, but rather the interaction of the hypothesised visual attention processes. To this end, experiment 4 procedure was similar to experiment 2 except for substituting the secondary global task with a 2AFC sound task. This task used sound clips, and participants had to indicate to which ear the sound was played (left or right only). The presentation of

the sound, like that of the images, was also jittered. However, due to the longer presentation length and the slight delay on accessing the sound file, the sound could begin at the start of the trial, though (like the image task) it would not finish within the last 100ms of the trial.

Experiment 5, however, did not use a global-based secondary task. Instead, a local-based task was used. This was the disc stimuli as described in the previous section. Participants reported on whether they saw a small red-green dot with one keyboard press, or a small green-red dot with another keyboard press (2AFC). The disc was presented onscreen for 300 msec, and masked for 100 msec before and after the disc presentation by quartered discs (see Figure 2.9). The total trial length remained 500 msec. Participants responded to the disc task first, and then responded to the appropriate dot task. Like experiment 3, the field of moving dots in experiment 5 were a mixture of the four colours shown in Figure 2.3. However, participants were not asked to respond to the dot colours. All other parts of the experiment remained identical to experiment 2.

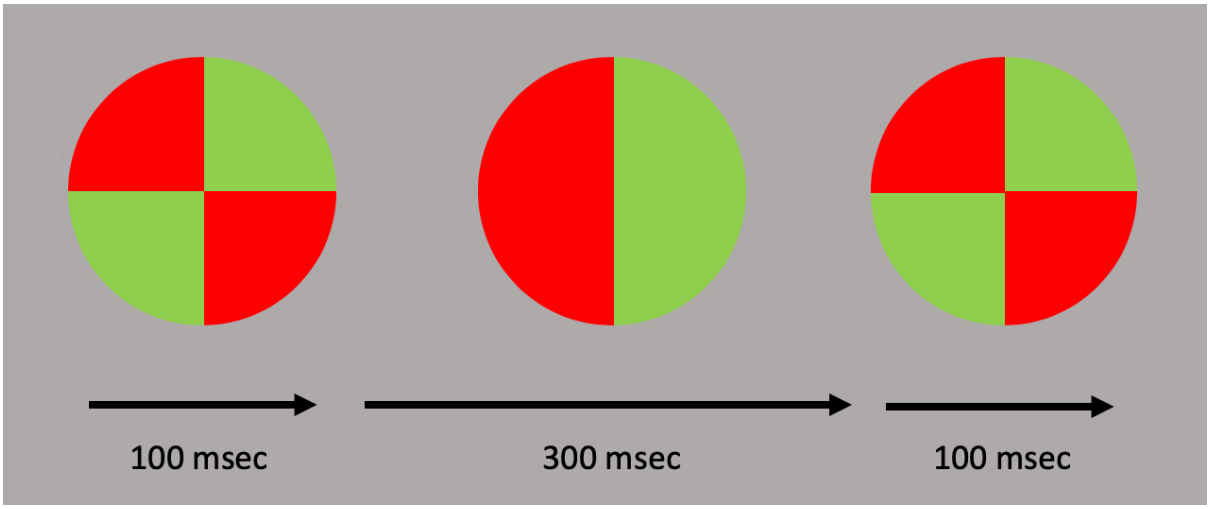

Figure 2.9: A cartoon of the bisected dot to which participants respond, and its corresponding masks. This dot demonstrates a red-green bisected dot. The mask is presented for the first 100 msec of the 500 msec trial; the dot is then shown for 300 msec; and the mask is shown for the last 100 msec.

2.4.3.3. Dual tasks

'Dual tasks' were trials in which participants had to respond to both the secondary global task, as well as either the primary global or local task. These 'dual task' trials would present the cue word for the primary task for one second, followed by a display with a moving field of dots during for 500ms. During the 500 msec trial window, the secondary global task stimulus (e.g. image, or sound) would be presented. For example, during experiment 1, at a pre-determined, jittered period the image would fade in and out behind the dots. Participants were asked to respond to the image or sound task first, before indicating the dot or dots motion (depending on which task they were asked to attend) (Figure 2.10). The same design was used for the colour experiment, though rather than having the global feature fade in and out, the dot colours were present from the onset of the trial. No jittering was used for the disc stimulus in experiment 5; the disc or an appropriate mask was present during the entire trial, as cartooned in Figure 2.9.

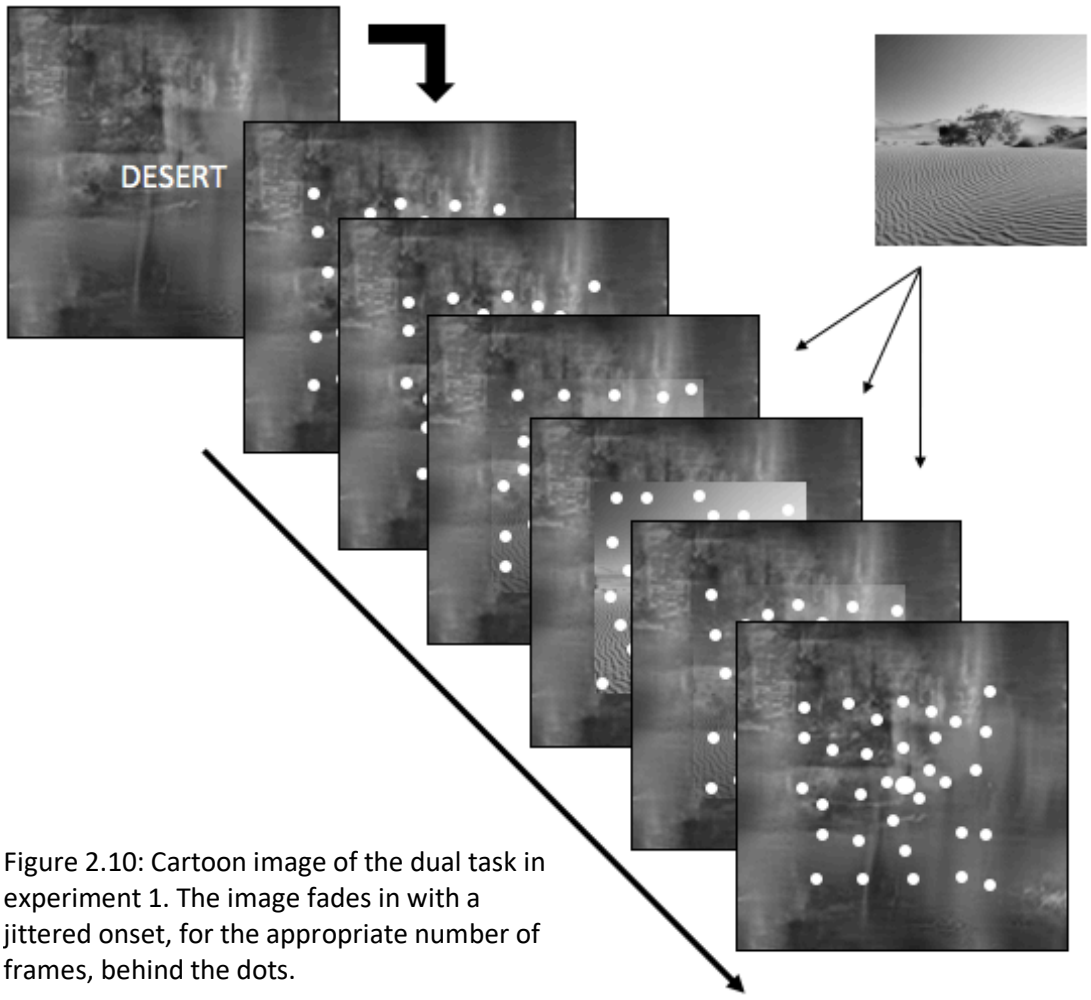

Figure 2.10: Cartoon image of the dual task in experiment 1. The image fades in with a jittered onset, for the appropriate number of frames, behind the dots.

2.4.4. Procedure

Participants completed three different task types in blocks, which were split up into three 'sessions' and completed in one experimental sitting. Participants responded to the secondary global task using the A and S keys on the keyboard, which meant 'match' and 'non match' respectively, and to primary tasks using the keyboard's arrow keys. This allowed the participants to leave both hands on the keyboard during dual-task trials (left hand covering A and S, right hand covering the arrow keys) which helped minimise response errors. Participants would have very short breaks between each session (~1 minute). In experiment 5, participants still used the A and S keys to respond to the secondary task; however, A represented a green-red disc, and S represented a red-green disc.

The first 'session' comprised of two blocks (one block of the global motion task, and one block of the local motion task) of 60 experimental trials and 5 practice trials. In experiment 3, this included a third block of the colour-based secondary global task. The purpose of this session was to determine what threshold the participants needed for each task, in order to meet a specific performance accuracy value. This used the staircase method.

The second session comprised of two blocks (three for experiment 3), during which participants performed 36 trials of each task type (6 practice trials, and 30 experimental trials). These data allowed us to assess if the thresholds provided by the staircase were accurate, i.e. to confirm that participants had equal performance for the two dot motion tasks, and the dot colour task during experiment 3.

At the end of both sessions 1 and 2, participants' output thresholds were manually checked by the experimenter. This was because there was no way to determine whether errors were the result of a mistake in perception or response (i.e. pressing the wrong key than the one intended). This could produce strange thresholds, which needed checking for validity.

The third session was the main experimental phase, and comprised of 5 blocks. The first three blocks had 150 trials of each single task (secondary global task, and both dot motion tasks). The secondary global task had an additional 6 practice

trials, as in experiments 1, 2 and 4 participants had not yet performed this task. This was followed by two blocks of the dual task, in which participants were asked to respond to the secondary global task first, before responding to the appropriate primary task.

2.4.5. Data analysis

Each experiment was analysed independently with a 2X3 repeated measures ANOVA, with the factors being task condition (single and dual task) and task type (image task, local dot task and global dot task). Data were then presented in an Attentional Operating Curve graph (AOC).

In order for the data to be fit into the ANOVA, the performance data for the image task during the dual conditions needed to be collapsed across participants. To see if this was possible, these data were analysed with a paired-samples t-test to determine if they were significantly different. Non-significance meant these data sets could be averaged into a single data set. The two single dot tasks were also analysed using a paired-samples t-test to confirm that they had successfully been equated for difficulty across the group.

Reaction time data were not analysed for this experiment for several reasons, including consistency. It would be hard to measure any difference in reaction time during the dual tasks, as participants were asked to respond to two different tasks in a specific order. There would also be no way to compare the reaction times of the various global and local tasks between their single and dual trial-types, due to the way in which participants responded. Whilst reaction time data could be investigated in the future, perhaps to allow for missed trials to be excluded from analysis, it would require a different experimental design.

2.5. Results

Paired-samples t-tests indicated no significant differences between the secondary task (image, sound, colour or disc task) during dual trials (see table 2.1), so group

data for these tasks were collapsed within each experiment. Paired-samples t-tests also found no significant differences between the group data for the primary, single-task global dot task and local dot task (table 2.2), indicating that the difficulty level had been equated for these tasks within each experiment.

Table 2.1: Means, standard errors, and t-test details for the paired-samples t-tests conducted between the secondary tasks during dual tasks. This included the image, sound, colour and disc tasks to which participants responded when there was also a dot task. Value are for all 5 experiments. None of the t-tests indicated statistical significance.

Secondary global tasks	**Means**	**Standard error**	**Degrees of freedom**	**T value**	**P value**
Experiment 1	Dual global: 79.5% Dual local: 77.3%	Dual global: 1.98% Dual local: 2.86%	15	1.05	> 0.3
Experiment 2	Dual global: 76.33% Dual local: 77.07%	Dual global: 2.3% Dual local: 1.7%	15	-0.35	> 0.7
Experiment 3	Dual global: 63.67% Dual local: 63.94%	Dual global: 2.26% Dual local: 2.59%	14	-0.094	> 0.9
Experiment 4	Dual global: 94.31% Dual local: 92.43%	Dual global: 1.4% Dual local: 1.92%	16	1.06	> 0.3
Experiment 5	Dual global: 86.55% Dual local: 83.92%	Dual global: 3.69% Dual local: 3.57%	15	1.062	> 0.3

Table 2.2: Means, standard errors, and t-test details for the paired-samples t-tests conducted between the primary dot tasks (global and local). Value are for all 5 experiments. None of the t-tests indicated statistical significance.

Single dot tasks	**Means**	**Standard error**	**Degrees of freedom**	**T value**	**P value**
Experiment 1	Global: 75.55% Local: 75.74%	Global: 2.3% Local: 1.86%	15	-0.16	> 0.8
Experiment 2	Global: 62.96% Local: 62.54%	Global: 2.38% Local: 1.53%	15	0.26	> 0.7
Experiment 3	Global: 67.02% Local: 68.05%	Global: 1.78% Local: 1.75%	14	-0.74	> 0.4
Experiment 4	Global: 67.14% Local: 68.55%	Global: 1.65% Local: 2.14%	16	-1.04	> 0.3
Experiment 4	Global: 70.02% Local: 70.81%	Global: 1.54% Local: 1.68%	15	-0.54	> 0.5

2.5.1. Experiment 1

The 2X3 ANOVA was conducted to assess the impact of interaction between tasks. The factors were task type (single or dual task), and the levels were stimulus type (secondary global task, primary global task, and primary local task). Planned contrasts (within-subject contrasts) were implemented for each of the factors, with

task type being a simple comparison and stimulus type a repeated comparison. This was identical for all of the following analyses.

There were main effects of task condition ($F(1,15) = 102.07$, $p < 0.001$), in which participants showed, as expected, a better performance in the single condition (M=78.5%, SE=1.4%) than the dual condition (M=63.3%, SE=2.3%). Participants also showed a significant effect of task type ($F(2,30) = 45.97$, $p < 0.001$), with within-subject contrasts finding that participants performed better for the image task (M=81.2%, SE=1.5%) than the global motion task (M=65.1%, SE=2.4%) ($F(1,15) = 59.47$, $p < 0.001$), though there was no significant difference between the global and local (M=66.4%, SE=2.1%) motion tasks ($p > 0.4$). These findings indicate that, as expected, performance for the image task was greater than that for either of the dot tasks, and performance for the two dot tasks were equated across the experiment.

Investigating these differences further, the ANOVA also indicated a significant interaction effect between task conditions and task types ($F(2,30) = 16.56$, $P < 0.001$). Planned within-subject contrasts revealed a significant interaction effects across the two task conditions, and between the image and global dot tasks (see Figure 2.10). In the single condition, task performance dropped only 8.55% between the two tasks, whereas the drop was 23.72% in the dual condition ($F(1,15) = 36.77$, $p < 0.001$), as demonstrated in figure 2.11. However, there was no significantly

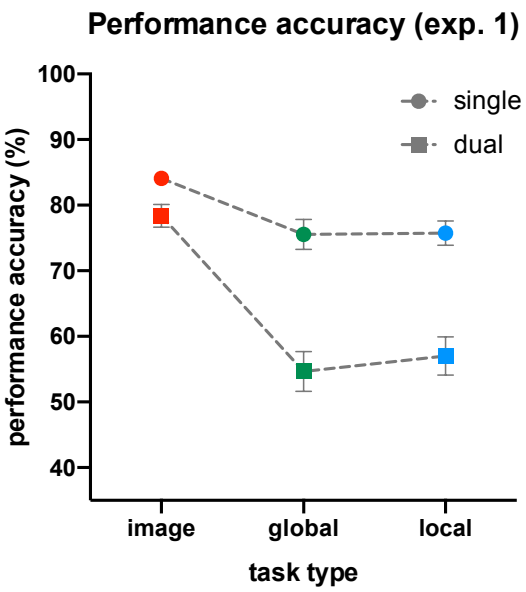

Figure 2.11: Performance accuracy for participants in experiment 1 in the single (*circle*) and the dual (*square*) task condition, across the three task types. Error bars represent standard error of the mean. Where no error bars are visible, the SE was too small to be plotted.

different response profile for the global and local dot tasks between the two task conditions ($p > 0.05$).

These results indicate that participants' performance in the dual condition showed a drop for both the global and local dot tasks, compared to the image task, than in the single task. It was also found that this drop in performance was equal for both the global and local tasks. These data were aggregated into an AOC graph, to help better demonstrate this relationship (Figure 2.12). This graph demonstrates the equivalent drop in performance for the dual-condition global and local dot tasks, and supports Treisman's continuum model of visual attention, as demonstrated in Figure 2.1 D.

However, concerns for this interpretation are raised due to the issues experience in equating participant performance. As this primary local task (dot size) was not suited for staircasing to a participants' unique threshold, participants were equated at 75% performance, rather than the intended 62-63%. The tasks are therefore very easy, which casts doubt on the interpretation of this finding as support for the FIT. The easiness level may instead produce results that mask a finding for a different model.

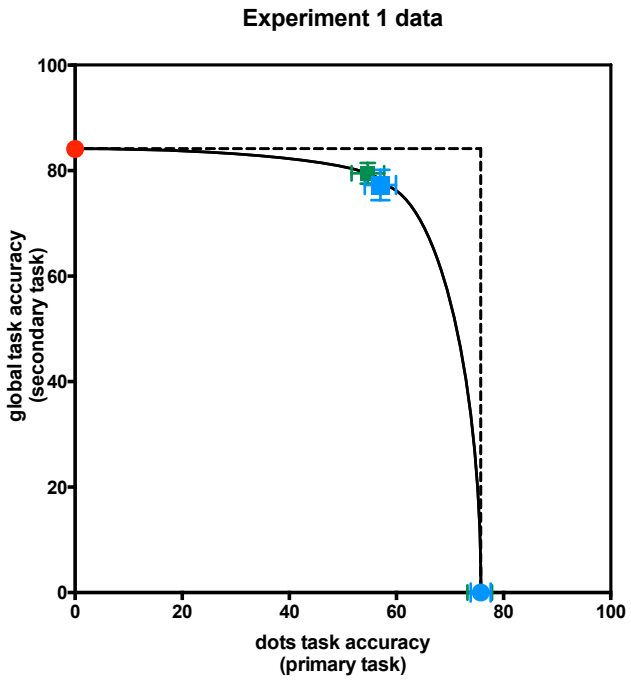

Figure 2.12: An attentional operating curve, demonstrating the relationship between performance accuracy on all three task types (image, global dot and local dot tasks) and both task conditions (single and dual conditions) in experiment 1.

Green represents the global dot task, and blue represents the local dot task. Red represents the image task. Dots represent solo tasks, and squares represent dual tasks.

All error bars represent the standard error of the mean. The attention operating curve has been estimated based on global performance data, in order to better highlight the relationship to local data. Both dual dot tasks show an equal level of performance drop.

2.5.2. Experiment 2

The 2X3 ANOVA found main effects of task condition ($F(1,15) = 156.81$, $p < 0.001$), finding that participants performed better in the single (M=70%, SE=1.5%) condition that the dual (M=55%, SE=1.8%) condition. The ANOVA also found a main effect of task type ($F(2,30) = 77.82$, $p < 0.001$), with a significant difference between the image task (M=81%, SE=1.3%) and the global task (M=56%, SE= 2.9%) ($F(1,15) = 92.82$, $p < 0.001$), though with no significant difference between the global and local (M=51%, SE=1.9%) dot tasks. These data replicate the findings from experiment 1, in that participants performed better for the image task.

The interaction effect of task condition and type, however, showed a different pattern than in experiment 1. The interaction effect was significant ($F(2,30) = 10.49$, $p < 0.001$), and within-subjects contrasts indicated significant differences between the performance profile of the image and global tasks, between the two response conditions (see Figure 2.13). Task performance between the image and global tasks in the single condition dropped 21.5%, but during the dual condition dropped 27.93% ($F(1,15) = 6.02$, $p < 0.05$).

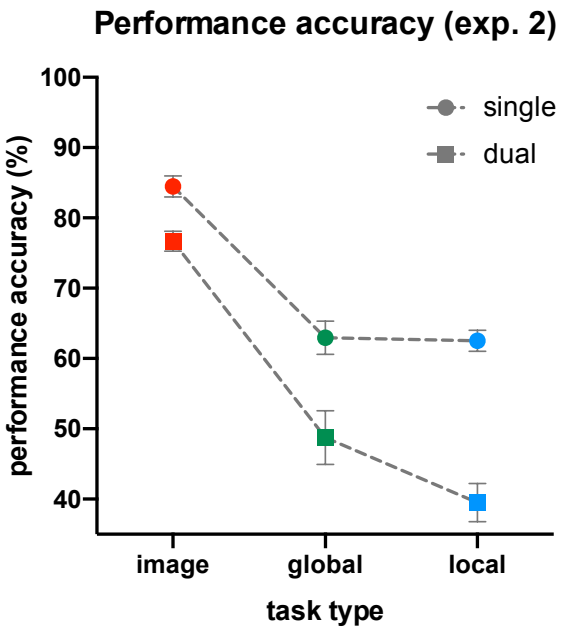

Figure 2.13: Performance accuracy for participants in experiment 2 in the single (circle) and the dual (square) task condition, across the three task types. Error bars represent standard error of the mean. Unlike in figure 2.11, the drop in performance for the global and local tasks in the dual condition is not equated.

However, unlike in experiment 1, there was a significant difference in the performance profile of the global and local dot tasks ($F(1,15) = 5.34$, $p < 0.05$): in the single condition, performance between the global and local dot tasks dropped 0.42%. In the dual condition, it dropped 9.28%, as can be seen in Figure 2.13.

When these data were aggregated into an AOC graph (Figure 2.14), the indicated that performance for the local task was worse in the dual condition than the global task, as predicted by the RHT's serial model of visual attention (Figure 2.1 C). These findings conflict with the findings of experiment 1, which indicated a continuum-based model; this may be because the harder contrast-based local dot task in experiment 2 more accurately represents performance accuracy than the size-based local dot task, as the size-based task was to easy (equated at 82%).

There were some findings replicated in both experiment 1 and 2; they both indicated that participants found the single condition to be easier than the dual condition. However, experiments 1 and 2 produced conflicting findings. Experiment 3, in which the secondary global task is colour (an intrinsic property of the dots rather than an additional stimulus), sought to provide further evidence for this relationship between stimulus types.

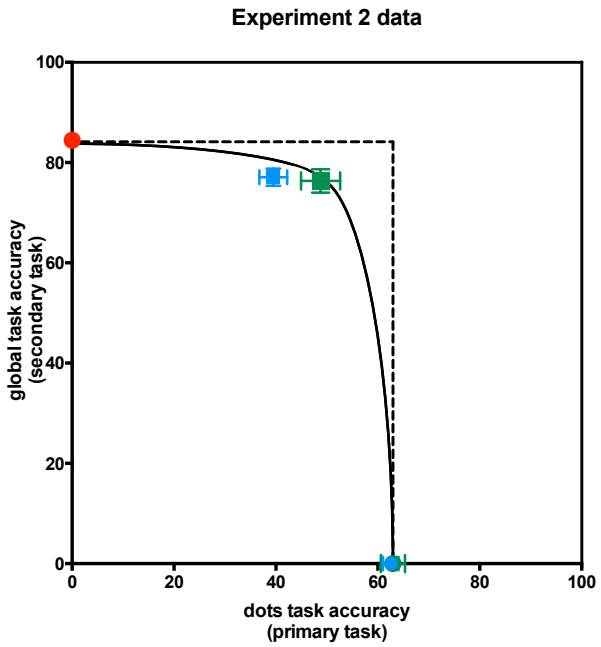

Figure 2.14: An AOC demonstrating the relationship between performance accuracy on all three task types (image, global dot and local dot tasks) and both task conditions (single and dual conditions) in experiment 2.

Green represents the global dot task, and blue represents the local dot task. Red represents the image task. Dots represent solo tasks, and squares represent dual tasks.

All error bars represent the standard error of the mean. Here, the dual local dot task shows a greater drop in performance than the dual global dot task, shown by the blue square being located further to the left of the curve than the green square.

2.5.3. Experiment 3

The 2X3 ANOVA indicated a main effect of task condition ($F(1,14) = 22.188$, $p < 0.001$), in which participants demonstrated greater performance accuracy in the single-condition tasks (M=68.4%, SE=1.5%) than the dual-condition tasks (M=61.4%, SE=1.8%). However, there was no main effect of task type ($F(1.375,19.253) = 1.21$, $p > 0.3$), indicating that participants performed equally well across the colour (M=67%, SE=2.3%), global (M=65.2%, SE=2%) and local (M=62.6%, SE=2.3%) tasks. A significant interaction effect between task condition and task type ($F(2,28) = 4.51$, $p = 0.02$) was found, although within-subjects contrasts indicated no significant differences in response profiles for the colour and global tasks ($F(1,14) = 1.879$, $p > 0.1$). Looking at the previous findings, this suggests that whilst there was a drop in performance between the single and dual conditions for these task types, this drop was the same for both the colour and global tasks.

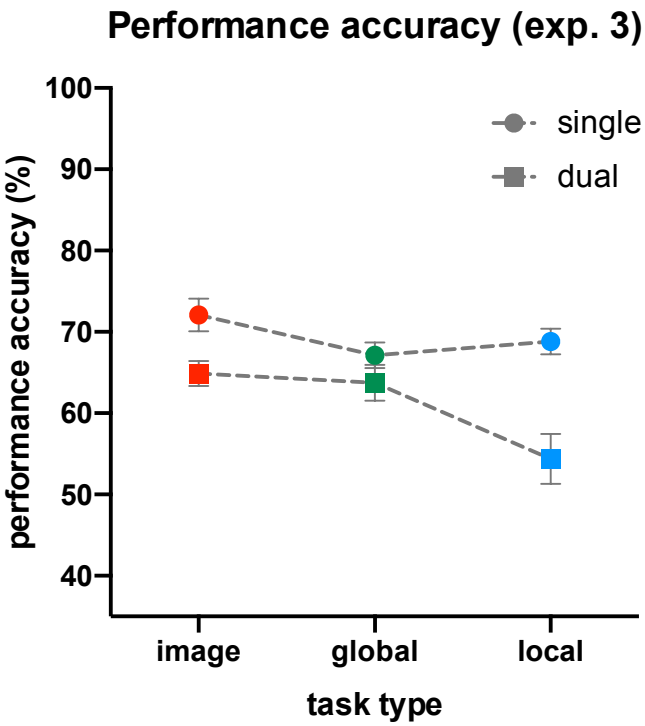

Figure 2.15: Performance accuracy for participants in experiment 3 in the single (circle) and the dual (square) task condition, across the three task types. Error bars represent standard error of the mean. Participants' response profiles showed a similar profile to experiment 2.

There was also a significant difference in response pattern between the global and local tasks ($F(1,14) = 7.939$, $p < 0.02$) between the single and dual conditions (Figure 2.15). As the paired t-test found no statistically significant differences between the global task and the local task in the single condition, this within-contrasts finding suggests that performance for the local task in the dual condition was less accurate (a drop of 6.32%) than in the global task.

As can be seen in the AOC in Figure 2.16, the results for experiment 3 were similar to that of experiment 2, with greater relative drops in performance for the dual-condition local task than the global task. This is further supported by the evidence from aggregating all experimental data into a single AOC graph (Figure 2.17), providing multiple data points along the same attention operating curve. This curve suggests that the data most closely reflect Hochstein and Ahissar's serial model of attention, rather than the parallel model or Treisman's continuum model.

In order to ensure that the data were representative of the mechanisms of visual attention, and not an artefact of taxing attentional resources as a whole, experiment 4 used the non-visual stimuli of sound along with the more-difficult contrast-based dot task from experiments 2 and 3.

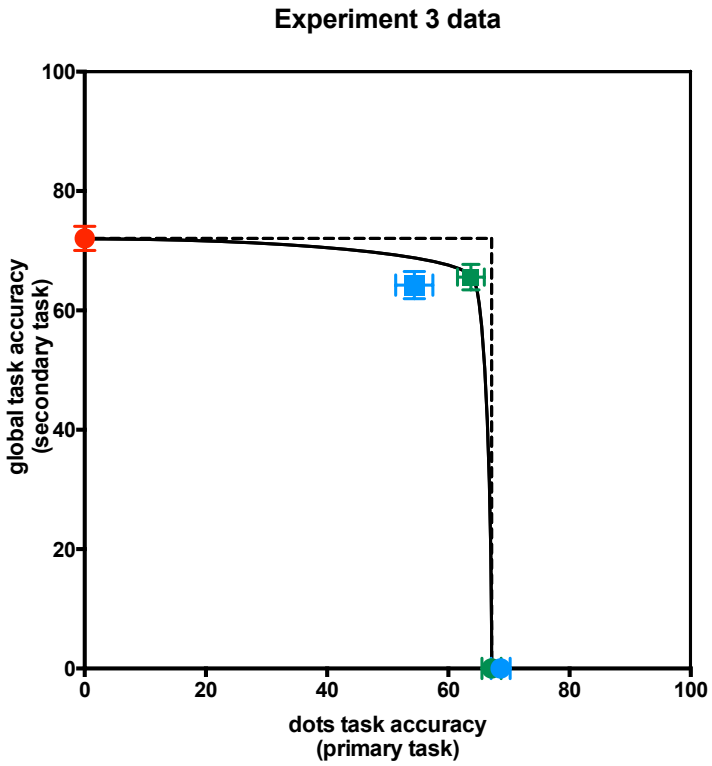

Figure 2.16: An AOC demonstrating the relationship between performance accuracy on all three task types (image, global dot and local dot tasks) and both task conditions (single and dual conditions) in experiment 2.

Green represents the global dot task, and blue represents the local dot task. Red represents the image task. Dots represent solo tasks, and squares represent dual tasks.

All error bars represent the standard error of the mean. Performance on the dual local dot task is reduced compared to that of the dual global dot task, with the data showing a similar relationship to the curve as experiment 2.

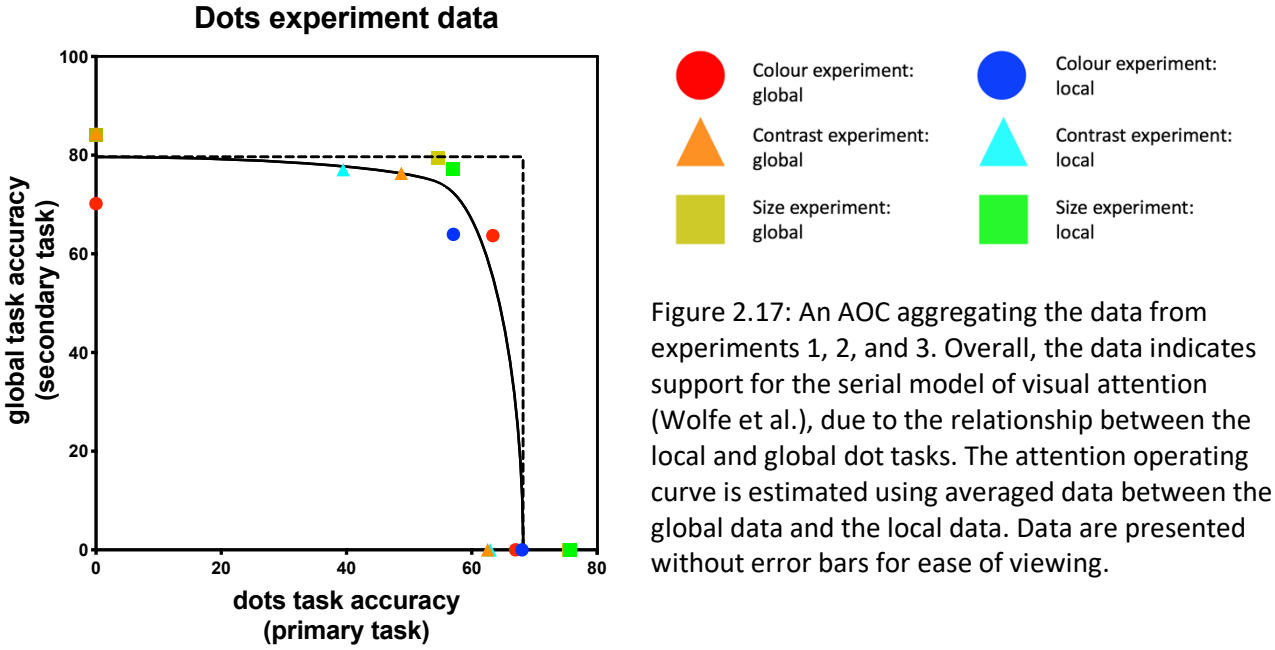

Figure 2.17: An AOC aggregating the data from experiments 1, 2, and 3. Overall, the data indicates support for the serial model of visual attention (Wolfe et al.), due to the relationship between the local and global dot tasks. The attention operating curve is estimated using averaged data between the global data and the local data. Data are presented without error bars for ease of viewing.

2.5.4. Experiment 4

The 2X3 ANOVA found no main effect of trial condition ($F(1,16) = 2.97$, $p > 0.1$), suggesting that there was no difference in performance between the single and dual task conditions. However, the ANOVA did find a main effect of trial type ($F(1.41,22.54) = 134.65$, $p < 0.001$). Within-subject contrasts identified significant differences between participant performance during the sound task (M=96.3%, SE=0.7%) and the global dot task (M=66.7%, SE=1.5%) ($F(1,16) = 493.393$, $p < 0.001$).

However, there were no significant differences between the global and local (M=66.5%, SE=3%) dot tasks ($F(1,16) = 0.006$, $p > 0.9$). This, in combination with the lack of statistically-significant interaction effect ($F(1.44,23.05) = 1.181$, $p > 0.3$), indicates that whilst participants performed more accurately on the sound tasks than the dot tasks, unlike in experiments 1 and 2, participants' accuracy for any of the task types was not impacted by the dual condition (see Figure 2.18). This was suggested further when data were aggregated into an AOC graph (Figure 2.19), which indicates no drop in performance for either dot task in the dual condition.

These findings suggest that experiments 1, 2 and 3 are indeed addressing visual attention, rather than attentional resources as a whole.

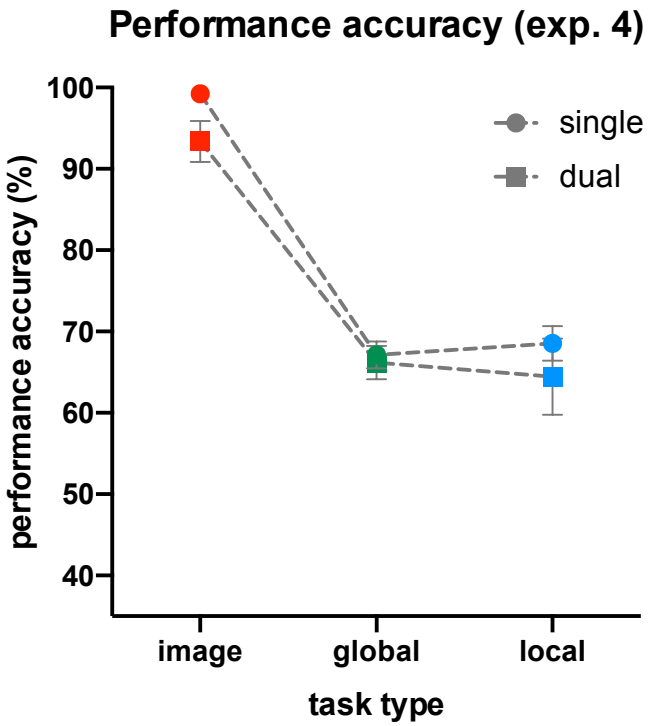

Figure 2.18: Performance accuracy for participants in experiment 4 in the single (dots) and the dual (squares) task condition, across the three task types. Error bars represent standard error of the mean. Participants' response profiles did not differ between the single- and dual-task conditions.

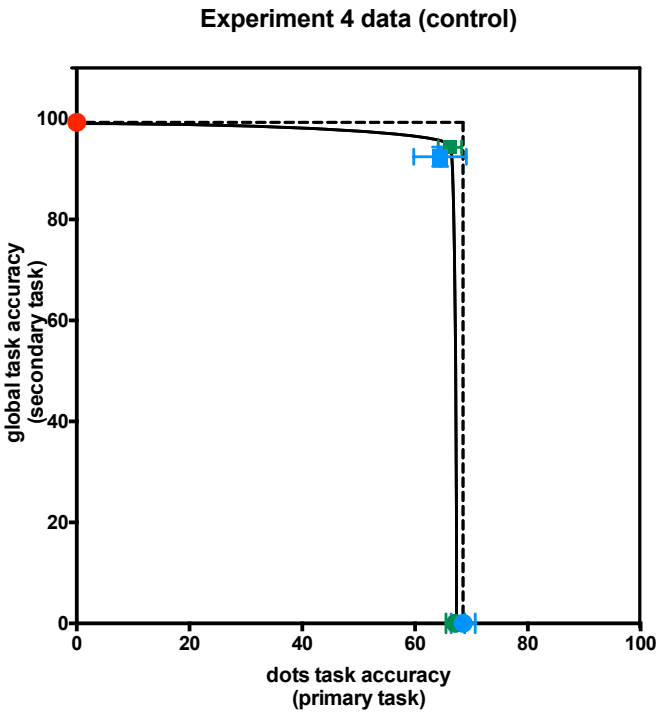

Figure 2.19: An AOC demonstrating the relationship between performance accuracy on all three task types (sound, global dot and local dot tasks) and both task conditions (single and dual conditions). Performance on the dot tasks in the dual condition is not impacted, compared to performance in the single conditions.

2.5.5. Experiment 5

A 2X3 ANOVA was conducted using task type (single or dual task) as the factors and stimulus type (disc, global dot task and local dot task) as the levels. This analysis revealed a main effect of task condition ($F(1,15) = 36.837$, $p < 0.001$), with participants performing better in the single (M=78.2%, SE=1.3%) than the dual condition (M=66.5%, SE=2.2%). There was also a main effect of task type ($F(1.424,21.357) = 53.696$, $p < 0.001$), with significant differences between the disc task (M=89.5%, SE=2.9%) and global dots task (M=61.2%, SE=1.7%) ($F(1,15) = 92.592$, $p < 0.001$), and between the global dots task and local dots task (M=66.4%, SE=2%) ($F(1,15) = 7.364$, $p < 0.02$).

There was an interaction effect ($F(2,30) = 6.744$, $p < 0.01$) between task condition and task type. Both the global and local dot tasks were equated in the single-task condition, but participant accuracy in the dual global-local tasks (M=52.38%, SE= 2.83%) dropped significantly compared to the dual local-local tasks (M=62.01%, SE=2.7%), as shown in Figure 2.20.

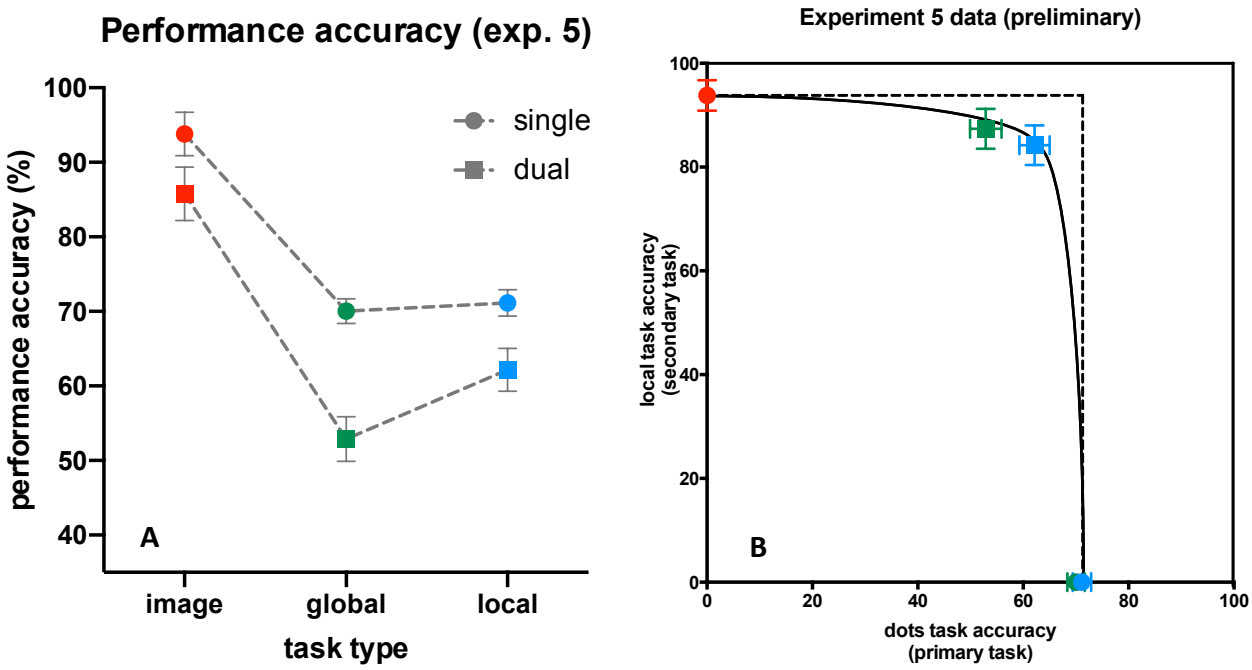

Figure 2.20: A) shows performance accuracy for participants in experiment 5 in the single (circle) and the dual (square) task condition, across the three task types. Error bars represent standard error of the mean. B) shows an AOC demonstrating the relationship between performance accuracy on all three task types (disc, global dot and local dot tasks) and both task conditions (single and dual conditions). Blue represents local data, and green represents global data.

Figure 2.20 A shows a double-dissociation of the data found in experiments 2 and 3; even when the tasks are changed so that the local resource is dual-loaded rather than the global resource, there is still a significant dip in participant performance for the task loading both global and local attentional resources. Figure 2.20 B plots this on an AOC graph, which shows the dip in performance accuracy for the dual global-local task compared to the dual local task.

2.6. Discussion

In chapter 2, we sought to identify what effect stimuli requiring the 'global' and 'local' modes of attention would have on performance accuracy, when presented to an observer simultaneously during a task which required the observer to respond to both. In these experiments, there was a primary global task, whether image- or colour-based, and a control experiment which used a sound task which did not require global attention. For all experiments, the global and local tasks involved identifying the direction of motion of a specified dot (or dots) on screen.

2.6.1. Main findings

In summary: experiment 1 found support for visual attention as a single resource, as described by the FIT; experiments 2 and 3 found support for visual attention as the product of attention operating in serial; and our control experiment, experiment 4, suggested that the previous three experiments had indeed been testing the function and limits of visual attention, rather than attentional resources as a whole.

The rationale behind doing multiple experiments was to identify, and address, potential confounds which impacted our understanding of the relationship between these two types, or extremes, of visual attention. A single experiment would not have allowed us to gain a full picture of this relationship, as the data would have only represented one individual point on the AOC. Adding further data using different stimuli, in the same experimental design, would each add another point to this AOC, and would paint a clearer picture of the relationship between local and global visual attention.

By aggregating the information from each of the 3 experimental conditions on a single AOC, the participants' averaged performance can be compared across experiments using the same logic as the individual AOC graphs (as shown in Figures 2.12, 2.14, 2.16 and 2.19). It can be seen that each experiment forms a separate point of data on an attentional curve (Figure 2.17). When considered together, this evidence strongly suggests that visual attention operates in serial (as described by the RHT), with the limitations in experiment 1 accounted for in experiments 2 and 3 by making the primary local task more difficult.

The final fifth experiment, conducted after data collection and analysis for chapter 2 were completed, also supports this conclusion. As discussed in section 2.3.2., a system that shows a deficit in local performance accuracy for a task requiring global, and then local, attention should *also* show a deficit in global performance accuracy, when asked to perform a a task loading first local and then global visual attention. Figure 2.20 does indeed show this double dissociation.

2.6.2. Critical analysis of experiments

2.6.2.1. Dot size/contrast

Individual design choices within the experiments themselves also allowed us to make each experiment more reflective of visual attention, with each offering some form of improvement over the previous experiment.

As discussed in part 2.4.1., participants' data were excluded if they could not successfully equate their performance (ideally to either 62% or 75%, as required by the task) on both the single global and local dot tasks. Equated performance was required for two reasons: firstly, it ensured that each task was an equivalent level of difficulty for the participant; and secondly, it allowed a consistent baseline between the two tasks, in order to compare any subsequent increase or decrease in dot task performance during the dual tasks.

By equating the single dot task data for participants, we ensure that the analyses done on these data are not dependant on the specifics of the experiments, and allows the data from each experiment to be compared to the others. Whilst this

may skew the results in favour of participants whose task performance can be thresholded within a specific window (i.e. between 62% and 75%, the trade-off is that the statistical analysis of the data across all experiments will be more robust.

Experiment 1 had issues with equating participants' performance across the solo dot tasks; the performance threshold aimed for was 62%, though the final, averaged performance for each task was approximately 75%. This high performance was found to be the result of an inability to titrate dot size accurately during the staircase procedure; dot size was controlled by the number of pixels required to present it, and a difference of 1 pixel could make up to a 15% difference in performance accuracy. A dot could not be presented using fractions of pixels, or scaled due to limitations of distance between screen and observer. Due to this, requirements for the task were changed so that as long as participants equated performance on the two dot tasks, regardless of performance threshold, their data would be included in the final analysis.

This potentially causes issues with the interpretation of the data. In order for the two tasks to be successfully equated, they had to be made easier by indicating a higher target performance threshold during the staircase procedure, and it's possible that this degree of leniency for the tasks masked experimental effects. A dot which was significantly larger than all other dots around it might have a visual pop-out effect, thus biasing observers to attend to its location regardless of their awareness of it (Hsieh, Colas, & Kanwisher, 2011). This would inflate performance accuracy for the local-based dual task, masking the effects of a serial process and presenting a data profile similar to that of the predictions of the FIT. (The GST would argue that no such boost would be present for the global task because it is already drawing from the same attentional resource; the RHT would argue that the same regions in higher cortical areas are being accessed.).

Experiment 2 was therefore intended to be more difficult than experiment 1 by substituting the size-detection task for a contrast-detection task, in order to combat any possible masking. Whilst dot size could only be adjusted within a range of a few pixels, dot contrast was limited only by the numerical values available in the code (i.e. 255). This allowed much more flexibility in fitting the correct dot contrast

for each observer, and also allowed us to more closely approach the target performance threshold of 62% (equated at 62.75%) during the staircase procedure.

2.6.2.2. Images and colour alternative

Following experiments 1 and 2, experiment 3 replaced the image task with a colour task, to address issues relating to category extraction. By changing the task from image categorisation to colour identification, the secondary global task came to more closely resemble the primary global task in that it was to identify a given property of a field of dots. Thus, in dual global-global tasks, observers would need to draw from the same attentional resource for the same stimuli, but attend to two different qualities of that stimulus.

Two separate criticisms were made of the use of images. Firstly, due to participants being encouraged to focus centrally on all presented stimuli, initial category information could be aided using local rather than global attention. This might take the form of identifying the key features of, and binding, contextually scene-specific objects such as a toaster or a refrigerator (for category 'kitchen'), as well as scene statistics. This criticism argued that both local and global attention were used in identifying the target category, meaning the task (and therefore the results) may not be indicative of global visual processing, but of a combination of them both. The second criticism argued that tasks involving more than one distinct task type added an additional layer of complexity to the task, in that observers had to extract the gist of two separate and unconnected tasks – the image category and the dot direction.

The results from experiment 3 produced a similar response profile to experiment 2, in terms of performance in the primary local task dropping further than performance in the global task, so it appears safe to argue that the presence of the image did not impact the overall response to the tasks. That being said, performance did not drop as sharply between single and dual trial types in experiment 3 as it did for experiment 2.

This suggests that whilst image tasks may not have been recruiting both local and global attention as feared, more attentional resources are required to extract

the category information from the image than to identify the average colour from a set of dots. As discussed later in chapter 3, different type of gists require a certain time-frame in order to be extracted; extraction of global scene properties, such as colour, require 19–47 ms, whereas even basic category information (such as city, ocean, etc.) need 30–67 msec (Greene & Oliva, 2009a). It may be that the images required more time to extract, especially if the image were complex or more ambiguous.

Whilst the four colours chosen for the dot task in experiment 3 were equated for luminance using a photometer, some participants reported finding some colours easier to identify, appearing 'brighter'. Whilst we controlled for luminance of colours, there may be a perceptual brightness effect that was not controlled, taking into the account the grey shade used for the background, and the saturation of the colours used (Corney, Haynes, Rees, & Lotto, 2009). This may make some of the colours easier to identify as a majority colour or as the target single dot. This was the advantage to using only greyscale stimuli in experiments 1 and 2, even taking into account any possible issues in using images.

Overall, by considering the results of experiment 2 and 3 together, it appears that the unique advantages and disadvantages of each type of stimuli used in the secondary tasks have allowed us to counterbalance for possible confounds, with participants' performance producing an overall similar response profile for each.

2.6.2.3. Control experiment

In all experiments, the performance for the secondary global task (whether image or dot colour) dropped for the dual tasks compared to the single task. This drop was statistically significant but was also equated across the dual task, meaning performance on the secondary task dropped to the same degree regardless of the dot task also being performed. This reflects the limitations of attentional capacity, in that performing two tasks consecutively is more resource-intensive than a single task; however, the as the degree to which the secondary task drops is equated, it is evidence that the responses to the primary tasks are indicative of the mode and type of visual attention, rather than attentional capacity as a whole.

This was also the purpose of experiment 4, which investigated participants' responses to the primary and secondary tasks when the secondary task was auditory. If experiments 1, 2 and 3 were tapping into general attentional resources, we would have expected to see a similar response profile for the primary tasks; as it turned out, we did not, and participant performance on all secondary tasks showed no significant reduction at all, regardless of single or dual task status.

There was a significant drop in performance accuracy in the auditory task between the single and dual task conditions (approximately 6%), but performance accuracy was still well above 90%, and this may simply be representative of the increased general attentional load of performing a dual task.

2.6.2.4. Bisected disc

The design of experiment 5 allowed us to perform a sanity-check on the data in chapter 2. After 3 experiments (excluding the control), the data strongly suggested support for a serial model of visual attention as proposed by the RHT. Tasks in which participants had been asked to respond to load on both the global and local attentional systems resulted in performance on the local part of the task significantly suffering.

The hypothesis for experiment 5, in line with the RHT, was that participants would produce significantly worse performance accuracy for the global task during dual global-local trials, compared to dual local trials. This would be because (according to the serial nature of the RHT) dual local trials allowed for re-entry into earlier stages of visual processing, meaning both local tasks would benefit from the engagement of the same resource. Dual global-local tasks, on the other hand, would see the same performance deficit as in experiments 2 and 3, due to the task requiring the recruitment of both the sparse attention mode and the focused attention mode. As participants are not needing to conscious extract information from the spare attention mode, dual local tasks draw from the same resource and take less time to execute.

That the double-dissociation is found is is strong further evidence for two things. Firstly, the design of this experiment is successfully loading visual attention, and is not measuring a feature of attention unique to the 'global' task, as these results were predicted from the outcome of the previous experiments. Secondly, this double-dissociation provides further strong evidence for RHT (Hochstein & Ahissar, 2002). As hypothesised, it appears as through the 'double-dipping' of the participant into the focused attentional resource is less demanding and/or requires fewer resources.

Conversely, even when the local task is performed first, performance on the dual global-local task suffers. This suggests that visual attention is serially processed, with the first feed-forward stage resulting in an obligatory 'sparse' attentional mode that must first be processed before focused attention can be brought to bear on the scene.

2.6.3. How does this fit in with existing research?

In contrast to experiment 1, experiment 2 found that participants had a higher performance accuracy in the secondary global task, compared to the local task, during dual trials. Performance declined more significantly ($p < 0.04$) for the local task (an average drop of 23% between single task and dual task trials) than the global task (an average drop of 14%). A similar result was found in experiment three, with a significant difference ($p < 0.001$) between local (14% drop) and global (3% drop) primary task performance in the single and dual-task conditions.

Of the three proposed models, this appears to fit best with Ahissar and Hochstein's (2004) physiological model, which argues for a serial progression of visual attention from a 'sparse' visual window, containing information such as summary statistics and 'gist', to more 'local' or focused attention later in time. Participants' performance on the dual global task generally showed both the primary (dot) and secondary (image/colour) global tasks were equated in performance when compared to their single-task equivalent. This is accounted for by the information becoming consciously available at a high level within the hierarchy, allowing

information about the broad scene as a whole – direction of dot travel, image category and/or colour information – to be available quickly and automatically.

Ahissar and Hochstein's RHT would indicate that this is because all of these elements are processed implicitly and in parallel as they travel up the visual hierarchy, creating generalised information about the scene that contains all of this information. Accessing this information, as indicated by Ahissar and Hochstein (2004), would be the first stage in conscious visual perception. This would mean that in dual global tasks, such as detecting the main dot colour and direction of dot travel, the participant needs to extract 2 pieces of information from a single, high-level percept, which would not have the same time requirements as accessing information from a more focused attentional window.

In contrast, participants' performance in the dual global-local task produced a result in which the performance on the dual dot task was less accurate than performance in the single dot task. In order for this information to be consciously perceived, the RHT argues that attention must be deployed and the visual hierarchy re-entered for local details to become available. This would allow access to details of the scene, even if this access is limited and search is serial. This access is also slow, with observers needing less time to identify a stimulus than to localise it (Atkinson & Braddick, 1989; Saarinen, 1996). Due to the short presentation time of the stimulus, and the task demands (participants were asked to respond to the global task first), this did not allow as much time for re-entry into the visual hierarchy and a slow, serial search for the local stimulus, thus reducing participant performance.

In chapter 1, the topic of 'gist' is mentioned, particularly as it relates to the non-selective pathway as suggested by Wolfe et al. (2011), and the sparse visual processing suggested by Hochstein & Ahissar (2002) and Ahissar and Hochstein (2004). Amongst other features, gist contains the ability to extract the summary statistics of a scene, including the mean colour, size, and direction of motion of a set of objects. This extraction of summary statistics is seen in the participants' ability to correctly judge the average direction of motion of a field of dots. This is in line with findings from Chong & Treisman (2003), who tasked participants with identifying the average size of a group of 12 circles, using circles with heterogeneous and homogeneous sizes. They reported that the mean properties of a set of items (in

their case, size of circles; in our case, our moving dots) could be extracted in as little as 50 ms. In particular, Chong and Treisman note that in the case of motion differences as small as 1-2° of motion can be discriminated.

Chong and Treisman identified this extraction of summary statistics as an automatic and parallel process, the same characteristics of the process which Ahissar and Hochstein (2004) argue produces spread attention (i.e. the first feedforward sweep of information through the visual system, providing access to gist information). However, it is worth noting that Wolfe et al. discuss gist as a product of their non-selective pathway, and also leave open the option that their pathway may operate in a serial manner, despite advocating for a parallel process. So why conclude that our results are a representation of the RHT model rather than a serial-operating GST model?

This is due to the way the GST represents the relationship between scene gist and the non-selective pathway of the model. Wolfe et al. argued that the non-selective pathway, being responsible for the extraction of scene gist, would then use this information to produce semantic and episodic guidance for subsequent visual search (which could feasibly operate in either serial or parallel), directing the selective pathway. Wolfe et al. describe how a visual search for a specified item, such as bread, in a kitchen scene would not begin on the ceiling or the sink, but on the countertops; our knowledge of the scene (i.e. gist-derived semantic knowledge) would direct us to begin searching for the bread in the more likely locations first, thus increasing the efficiency of the search. In this way, the non-selective and the selective pathways work together.

The issue is that later research has indicated that gist is not responsible for semantic guidance during visual search. Wu, Wick, & Pomplun (2014) directly addressed the effects of gist and spatial dependency during real-world visual search, with regards to semantic guidance, and found that scene gist did not affect semantic guidance. Specifically, when Wu et al. removed scene gist (scene background) from visual search tasks, subjects were still able to use spatial dependency (arrangement of stimuli in space) to aid their semantic guidance in visual search. In fact, they argued that scene gist had only a marginal effect on semantic guidance, with spatial dependency required in order to produce the semantic guidance necessary. (During

tasks in which both scene gist and spatial dependency were removed, participants' performance dropped to chance levels.)

Wolfe et al. have primarily and explicitly argued that their model of visual attention predicts a parallel, rather than serial, process for visual attention (despite the remaining possibility for a serial process). This, combined with the requirement of the GST that gist is necessary for semantic guidance of the selective pathway, raises the question of whether the model is an accurate representation of visual processing. When considering the issues with the GST, and the empirical data reported in this chapter, there is strong evidence that the RHT is a better explanation for the relationship between the two types of visual attention.

2.6.4. Further research

The experiments described in chapter 2 covered several angles in addressing the relationship between sparse and focused attention, particularly in regards to 'global' attention as the secondary task. The next logical (to replicate these experiments using a 'local' attentional task as the secondary task, producing local/local and local/global tasks) was conducted in experiment 5.

There are two potential ways to further research into this topic. The first is to consider replicating these findings using a more true-to-life stimulus. By necessity, the experiment designed in this chapter required on a very artificial presentation of stimuli. Fields of dots and their related properties (colour, contrast, size, and direction of motion), bisected discs and brief presentations of tones allowed for precision in the experimental design, but are not reflective of real-world experience.

Experiment 3 was intended to reduce the number of extraneous factors in the design. By replacing the image categorisation task with reporting on dot colour, this allowed participants to respond, in every trial, to some property of the on-screen dots. In this vein, a follow-up experiment is proposed in which participants must respond to some aspect of a briefly-presented visual image. The secondary global task will be to report on the average colour of the image, based on a colour-summariser analysis of a database of scene images, controlled for average colour. The primary global task will be to identify the scene category of the image; the

primary local task will be to locate a specific object within the scene, and report on its location (by dividing the image into quadrants, and asking participants to respond using a key that corresponds to a specific quadrant.

This is, of course, still artificially-presented stimuli. However, by using an image task, this allows us to ask questions more related to real-world visual processing, and (like experiment 3) also limits the participants' responses to those of a single stimulus type. This would allow us to replicate the results of this experiment using solely visual scene imagery, and would help bridge the link between the findings of this chapter and the execution of visual attention in the natural environment.

The second way to consider further research into this topic is to investigate more closely the individual modes of visual attention: the global (sparse) and local (focused) visual processing modes. As discussed in chapter 1, what is termed the local mode of attention has been investigated in some detail; features such as the capacity bottleneck, and visual search are well-known in the literature.

However, what is less understood is the concept of 'gist' processing, a product of the global mode. This is a somewhat amorphous topic that lacks a clear-cut, specific definition; however, as discussed in the introduction to this chapter, some key facts are known about it. It is accessed quickly (Haberman & Whitney, 2012), it allows for the extraction of information such as scene category, and multiple types of information may be extracted (Evans & Treisman, 2005; Evans & Chong, 2012). It also allows for the extraction of information such as the average size of a set of shapes (Chong & Treisman, 2005), and appears to be the mechanism by which participants were able to answer the 'global' tasks posited in this chapter.

How, then, did participants extract the information required to answer these global-based tasks? The rest of this thesis will investigate the neuronal correlates (Chapter 3) and temporal dynamics (Chapters 4 and 5) of gist, and in doing so further illuminate the role and underlying neuronal processes of this global, sparse attentional processing.

2.7. Conclusions

The aim of chapter 2 was to conduct experiments to investigate more directly the relationship between what was termed 'global' and 'local' attention. More specifically, the experiments aimed to understand the relationship by asking participants to perform 2AFC vision-based tasks which taxed either one or both of the attentional resources.

During dual-task trials, it was predicted that there would be one of three outcomes, as cartooned in Figure 2.1: that attention operated in a single continuum, in which performance on both global and local tasks were reduced equally; attention operates in a serial manner, in which participant accuracy on the local task would be reduced compared to that of the global task; or the two attentional modes operated in parallel, in which performance on the global task would be reduced compared to the local task.

Findings from the experiment indicated that attention operated in a serial manner, aligning with the predictions of Ahissar and Hochstein's (2004) Reverse Hierarchy Theory.

Future research should aim to replicate these findings using a paradigm that more closely represents real-world visual scenes, and should also further investigate the mechanisms behind the global mode of visual processing.

Chapter 3 The neuronal correlates of gist processing

3.0. Overview

Gist, as a concept, has been well-established in the scientific literature (examples of gist just for visual scenes are discussed by Oliva, 2005; Loschky & Larson, 2010; Oppermann, Hassler, Jescheniak, & Gruber, 2011; Groen, Ghebreab, Prins, Lamme, & Scholte, 2013; Evans et al., 2013; & Wu, Wang, et al., 2014). However, this is not the case for its neural correlates. Oliva (2005) argues for multiple possibilities as to the location of gist, suggesting dedicated cortical areas, or even a distributed representation across multiple areas and levels of processing. Existing studies have identified gist as part of a process that goes beyond high-level visual cortex (Oppermann, Hassler, Jescheniak & Gruber, 2012), but there is currently no research into the location of category-specific gist extraction.

Chapter 3 features the manuscript for the paper which is currently being submitted to the journal NeuroImage, which represents a study into the location of gist within the human brain.

3.1. Abstract

'Gist' refers to our ability to rapidly extract information from a visual scene. This can include category information, such as whether the scene is a forest or a city. An observer can extract more than one gist at the same time, and different types of gist can interfere, with the presence of one category 'masking' the other, despite them being presented at the same time. Here we ask where in the brain gist processing occurs, and whether we can see evidence for destructive interference in these regions. We find that participants can correctly identify the *presence* of a briefly-presented, pre-cued item in a scene without recalling its position. We find that the neural correlates of this gist processing are seen in category-specific cortex but not retinotopic early visual areas, and that correlates of destructive interference may be seen in cognitive control regions, including the dorsolateral prefrontal cortex.

3.2. Introduction

3.2.1. Gist processing is accessed quickly and early

It takes humans only a couple of hundred milliseconds to extract semantic information about a scene. This rapidly extracted information can include global-property categorisation (for example, is it urban, or is it natural?), and even scene categorization (is it a city or a beach?) (Oliva & Torralba, 2001; Greene & Oliva, 2009a; Potter, 2012). The extraction of the information needed for these judgements is known as *gist*.

Gist processing allows observers to make rapid, computationally 'inexpensive' (Haberman & Whitney, 2012) and reasonably-accurate assessments of a complex environment - providing visual context information that may help guide attention toward specific objects within the scene (Chun, 2000; R. A. Rensink, 2000; Wolfe, Võ, Evans, & Greene, 2011). Gist processing is fast, efficient, and performs the function of forming a 'global first impression' (Groen et al., 2013) of the visual environment.

Gist is a series of characteristics we obtain rapidly to make judgments about the content and characteristics of a novel scene (Wolfe, Võ, Evans et al., 2011). These characteristics include global image statistics, such as the distribution of the scene's basic features or 'spatial envelope' (Oliva & Torralba, 2001; Oliva, 2005), summary statistics, such as the average size of a set of shapes (Chong & Treisman, 2005), and the presence or absence of 'target categories', such as an animals or faces (Evans & Treisman, 2005; Evans & Chong, 2012).

Gist extraction can be achieved even when a scene is presented for only 13–40 msec (Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009b), although different 'types' of gist are accessed at different rates (Loschky & Larson, 2010). At 13 msec, for example, observers are able to detect the presence of a cued target category during an RSVP task, at above chance levels (Potter, 2012; Potter, Wyble, Hagmann, & McCourt, 2014). A 26 msec exposure is needed for both the extraction of scene gist and parallel object processing, which can aid with context categorisation (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). Observers require only 19–47 msec for global-property categorisation (such as mean depth, temperature and navigability), but will need 30–67 msec to determine a scene's basic category (forest, city, etc.) (Greene & Oliva, 2009a). The rapid speed at which these

assessments can be made has led to assertions that gist processing happens during the first feedforward sweep of information through the ventral stream (Serre, Oliva, & Poggio, 2007; Fabre-Thorpe, 2011). It is also argued that it does not require focused attention or re-entrant processing (Li, VanRullen, Koch, & Perona, 2002; Hochstein & Ahissar, 2002).

Gist extraction clearly happens quickly, possibly even preceding object identification (Rousselet, Joubert, & Fabre-Thorpe, 2005). VanRullen & Thorpe (2001) argue that these early time points (as mentioned above) represent a perceptual process, as part of the first feed-forward sweep of visual information through the brain. VanRullen & Thorpe asked participants to identify vehicles or animals within a series of presented images (20 msec per image), whilst measuring neuronal activity using EEG. They found a neuronal component between 75–80 msec post-stimulus onset, which they associated with a perceptual process in which the basic features of the image category were extracted. This occurred regardless of target type, suggesting that this is a reflection of visual processing.

Further studies have also looked into this initial extraction of scene information in the feed-forward sweep. Scholte, Jolij, Fahrenfort, & Lamme (2008), whilst studying the neural and temporal correlates of scene segmentation, found information about texture boundaries was extracted by the occipital areas at around 92 msec post-onset, with activity associated with texture appearing in peri-occipital areas at 104 msec, temporal areas between 104-108 msec, and then parietal areas (104-120 msec). This, Scholte et al. argue, is not the result of object-based attention, but is a part of the first feed-forward sweep of information that can be tracked through the brain.

By assessing this literature together, this produces an argument that the first aspects of gist processing to be seen would be under 100 msec, ranging between 13 msec and 92 msec for extraction of features to produce a percept of gist. It is possible that the C1 component may be identified within this window, which is seen seen in ERP studies between 50 and 90 msec after stimulus onset. It represents the initial response of the visual cortex to visual stimuli, without modulation from attention (Di Russo, Martínez, & Hillyard, 2003; Stolarova, Keil, & Moratti, 2006). Di Russo et al. also noted that this signal identified during the C1 component would be routed from V1 to higher visual cortex areas, and back into the V1, beginning 130 msec post-stimulus onset. This would indicate that the C1 component is a marker of processing during the first feed-forward visual sweep, and also implies the same for time points identified before this ~130 msec cut-off.

In addition to this, visual evoked potentials (VEPs) are known to occur around (100 ms (Sharma, Joshi, Singh, & Kumar, 2015), which is later than some of the observed ERPs as described above.

In fact, these two seemingly disparate facts – that gist can be accessed < 100 ms and VEPs are seen ~100 ms – are related, in that VEPs are an integral part of gist processing. Generally, the ERPS associated with the extraction of whole, single elements of gist information are observed between approximately 100–250 ms (Groen, Ghebreab, Prins et al., 2013) (this will be explored in more detail in chapters 4 and 5). Groen et al. (2013) utilised spatial coherence as a metric for scene gist, and found that it modulated behavioural performance up to 250 ms when assessing scene naturalness. They concluded that the spatial aggregation of responses in early visual areas, such as V1, allowed for the computation of global scene information – such as gist.

This means that VEPs, which are seen around 100 ms, can be used as a part of gist processing. Scholte, Ghebreab, Waldorp, Smeulders, & Lamme (2009) noted that 71% of the VEP found between 90 and 180 ms is explained by the low-level statistics found in natural scenes, to which the visual system is tuned and which is used to quickly classify images (gist).

3.2.2. Where might gist processing be seen?

A common paradigm for measuring gist processing involves presenting the participant with an image, or a series of images, and requiring them to respond yes/no to a specified quality of the image in a two-alternative forced-choice design (for relevant examples, see Rousselet et al., 2005; Evans & Treisman, 2005; Greene & Oliva, 2009a; Groen et al., 2013; and Potter et al., 2014). This, naturally, allows experimenters to calculate ratios of correct and incorrect answers. However, when studying the neuronal correlates of gist processing and visual perception, a more granular distinction is needed in an attempt to link psychophysics with neurophysiology.

For example, Gutchess & Schacter (2012) and Ress & Heeger (2003) examined neuronal activity when responses were categorized as per the participant's *perception* of the stimulus presence rather than as simply 'correct' or 'incorrect' as per the stimuli's physical presence. Whilst these researchers were not attempting to study gist processing

specifically, the signal-detection paradigm they used translates well to the broader question of where gist processing is occurring within the brain. It should be possible to identify neuronal signatures for each of the four response types (hits, false alarms, misses and true negatives), and by using this design in an fMRI scanner, ask which neuronal modulations correlated with the perception of gist.

These examples propose a paradigm to investigate where gist perception might be seen in the brain, but offers no predictions as to what those neuronal areas might be. Work by Peelen & Kastner (2014), however, offers some predictions. Peelen & Kastner investigated focused attention deployed in naturalistic settings, attempting to locate the 'what' and 'where' templates that aid in visual search. Of particular relevance, they found that global scene properties (scene category) were represented within parahippocampal and retrosplenial cortex, and transverse occipital sulcus (i.e. place-selective regions), and that these properties aided in the speed of visual search in natural scenes. (That the parahippocampal cortex and retrosplenial cortex is related to scene perception is well-established – Epstein, 2008; Walther, Caddigan, Fei-Fei, & Beck, 2009; Kravitz, Peng, & Baker, 2011).

Furthermore, Peelen & Kastner (2014) discuss that category-selective neurons activated during the first feed-forward sweep support real-world search in general, noting that body-selective cortex has been implicated in the real-world search for people, but not for other objects such as cars.

All together, this builds a picture that category-selective cortex may be implicated (and potentially observed) in fMRI during gist processing. Gist, as discussed, performs functions such as the extraction of summary statistics and global image statistics; the hypothesis then follows that this processing may occur in cortex that is selective to a specific image type (and therefore sensitive to the specific statistics of the image type). Gist processing for place images may be seen in place-selective cortex, face images in face-selective cortex, and objects within object-selective regions.

3.2.3. Destructive interference

Clearly, gist extraction is clearly not a unitary phenomenon. Rather, it is a flexible computation that responds to task contingencies. In addition, multiple gist extractions can

occur in parallel and interact with each other: Evans et al. (2011) report that, when asked to detect one of two pre-cued categories, observers' performance was improved by the presence of two trial-relevant targets (probability summation). Conversely, when presented with a two-target trial in which both targets had been primed, but in which only one target was trial-relevant, observers' performance suffered (destructive interference). This is both a strength and a limitation of this process, suggesting that different task contingencies can result in perception of different 'gists' at once, even with the same set of images and without prior priming; however, these gists are susceptible to destructive interference, which can impact an observer's visual awareness and accuracy.

Destructive interference has been observed in neural, as well as behavioural, data. Peelen & Kastner (2014) reported that neural responses can be biased by attending to one of two competing stimuli, which can have the effect of filtering out one of the stimuli; in short, attention can bias competition in favor of the task-relevant stimuli. In a similar vein, Seidl, Peelen, & Kastner (2012) found that the processing of objects that were previously task-relevant (but not currently relevant) were actively suppressed, compared to never-relevant objects. Based on their findings, Seidl et al. (2012) argue that this suppression effect happens at the category level.

However, the behavioral findings from Evans et al. (2011) would suggest that even cued targets can be blocked from awareness when there is an interfering (even though non-trial-relevant) category. They observed that, when primed with a task-relevant (i.e. previously trial-relevant, but not currently so) category such as 'animal', performance accuracy for a trial-relevant pre-cued category such as 'beach' would drop significantly when both were presented at the same time (for example, an image of an animal on a beach).

It is important to note, however, that Seidl et al. (2012) and Peelen & Kaestner (2014) reported the results of focused visual selection, whereas the extraction of gist happens without focused attention. Understanding the neural site of interference would shed light on the nature of this process: Would the effects of task contingencies on gist perception be seen in the same areas as those associated with destructive interference, or are these seen in different parts of the brain?

The concept of priming should also be discussed. As described above, Evans et al. (2011) found that priming a category image may have a measurable effect on a

participant's performance during a task cued for a *different* image category. In this case, the priming was achieved through repeated presentations of a particular category type. This priming generally results in decreased neural responses to the stimulus in question (Grill-Spector, 2008). If the hypothesis is that gist processing may be seen in category selective cortex, and potentially frontal cortical regions (destructive interference), where might the hypothetical neural substrate of priming be found?

Generally, paradigms which use pre-cues and priming (i.e. cognitive templates) have not been found to penetrate into early visual processing (Raftopoulos, 2017). Rather, research has tended to suggest temporal and frontal regions. Copland, Zubicaray, McMahon, & Eastburn (2006) investigated semantic priming by asking participants to make quick decisions on presented words (words/non-words). Participants had to identify if a presented word was related to an ambiguous prime (e.g. bank), either in a dominant ('money') or subordinate ('river') way. Copland et al. (2006) used fMRI BOLD responses to investigate the neuronal correlates of this priming, finding that primed dominant targets showed a greater % BOLD signal change in right superior temporal gyrus, right anterior cingulate cortex and the left inferior frontal gyrus. (Copland et al. did not find priming of subordinate meanings.)

Grill-Spector (2008) also notes a consistent, robust finding in the lateral occipital complex, fusiform gyrus and parahippocampal gyrus for a reduced response to repeated stimuli. Grill-Spector also notes that scene and object images (in particular) produce a similar response in frontal regions under repeated presentation.

3.2.4. Aims and hypotheses

For this chapter, there were 2 main questions. The first of these asks: where are brain gist processing computations observed during fMRI? Based on evidence from Peelen & Kastner (2014), the hypothesis was that gist processing would be observed in category-selective cortex. In the case of place images, it would be seen in place-sensitive regions, and the parahippocampal place area and retrosplenial cortex were chosen as the regions of interest for this analysis. This was due to their history of sensitivity to place and scene images during fMRI experiments. To contrast this, face images were chosen, as they also have areas of cortex that selectively responded to this stimulus. The fusiform face area and occipital face

areas were chosen as the regions of interest for these stimuli. These two stimulus types were sufficiently different to allow for direct comparisons in fMRI data.

The second research question asked whether we see two classic signatures of gist processing: changes (specifically destructive interference) due to change in task contingency, and categorical responses without location information.

For this first signature (destructive interference), there was no specific hypothesis as to which areas may be observed. A whole-brain analysis was to be used to determine which regions of the brain responded to a greater degree during events in which destructive interference took place. This analysis was to be conducted by analysing the data using existing behavioural data.

Evans, Horowitz, & Wolfe (2011), as described above, found that participant performance accuracy was reduced during dual-target events, compared to single-target events. Therefore, brain regions which replicate this pattern – i.e. shower a greater % BOLD signal change response to single-target trials, as compared to dual-target trials – may reflect the neuronal correlates of destructive interference.

For the second signature, the hypothesis was that target information was not stored within the early visual cortex (EVC – V1, V2 and V3). Accessing gist information, whilst allowing for the extraction of scene statistics and other data, is not enough to identify the location of a target item within the scene (Evans & Chong, 2012). By splitting the possible on-screen target locations into four positions, this allowed us to investigate whether there was a relationship between the location of a target on-screen, and the % BOLD signal change in each of the four quadrants of the visual field.

3.3. Methods

3.3.1. Participants

Fifteen unpaid volunteers were recruited (11 female; average age 25 years, 13 right-handed), all with normal or corrected-to-normal vision. Informed consent was obtained for each participant. Participants were recruited through an opportunity sample, from within the Psychology and Neuroimaging Departments of the University of York. Ethical approval was

granted by the YNiC (York Neuroimaging Centre) Research Ethics Committee. All data were collected prior to analysis. Two participants' data were excluded from the final analyses, one due to noise from excessive movement in the scanner, and the second for exceptionally poor performance accuracy at the tasks.

3.3.2. Stimuli & Apparatus

The fMRI experiment was designed and presented using Psychopy (v1.82; Pierce, 2007) software for both localiser and experimental scans, on an HP EliteDesk 800 G1 tower (3.4Ghz quad core CPU) running Debian (version 8). Stimuli were presented on an Epson EB-G5900 projector at 60Hz, and projected to a Pro AV Eclipse II rigid screen, which participants observed through a mirror. The participants were at an effective viewing distance of approximately 53 cm (distance from the screen to the mirror of 52 cm, and an average distance from the participants' eye to the mirror of 10 cm).

Data were recorded using Lumina Response Pads using an LSC-400 Smart Controller. All fMRI data were acquired with a General Electric 3T HD Excite MRI scanner at YNiC, utilising an eight-channel high-resolution phased-array gradient insert coil tuned to 127.4 MHz.

All stimuli images (for both the localiser and experimental scans) were created using the MATLAB 2015a software (The Mathworks, Natick, 2015) and the MATLAB SHINE Toolbox extension (Willenbockel, Sadr, Fiset et al., 2010). Images subtended 10 degrees of visual angle and were presented in four quadrants around the central fixation cross. Each image was off-set from the central fixation cross by 4.5 degrees of visual angle laterally, and 4.7 degrees of visual angle vertically.

Two sets of images were used. The first set was used during the initial localiser scan with 72 outdoors place images, such as landscapes, buildings and natural scenes (LabelMe database, MIT Computer Science and Artificial Intelligence Laboratory), and 72 faces (Prof. Tim Andrew's lab database, University of York, 2015). The second set of images were used during the experimental scans consisting of 120 outdoors place images and 120 face images. 216 additional place and face images were analysed and synthesized using the Portilla & Simoncelli (2000) algorithm and acted as 'mask' images (See Figure 3.1). A further 194 place and face images were scrambled to form 'no-target' images.

All images were presented within a Gaussian contrast envelope (FWHM 10 degrees of visual angle), and were processed using the SHINE toolbox to eliminate the frame contours and equate the Fourier spectra and the luminance histograms over the entire image set, eliminating sharp low level feature differences across image categories. (Figure 3.1). Participants saw a total of 180 place and 180 face images over 360 trials. 100 place images were used with an average repetition of 1.8 times, and 47 male and 47 female faces were used with an average repetition of 0.94 times per face).

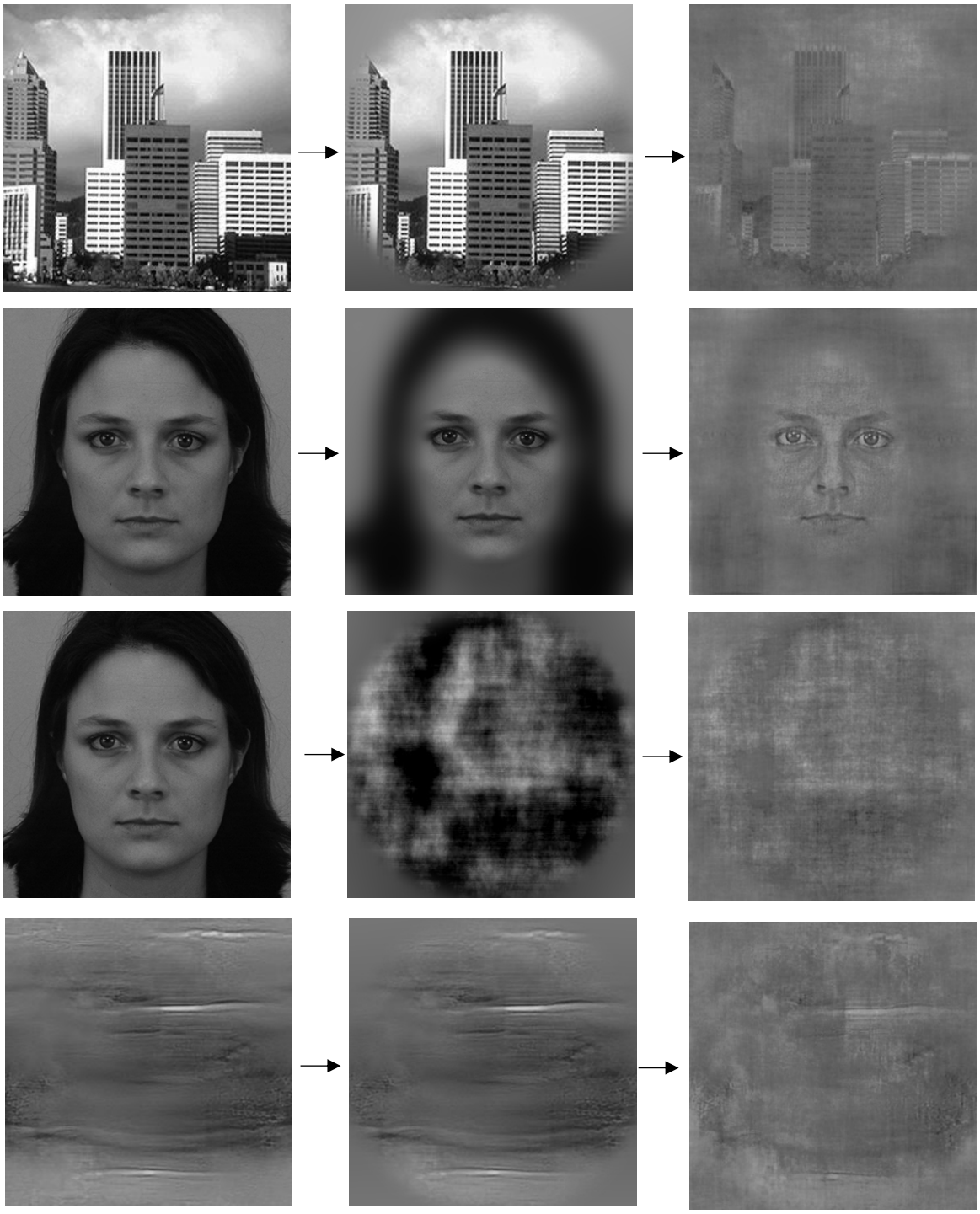

Figure 3.1: Examples of all four kinds of target stimuli (a place image, a face image, a scrambled face image and a mask created using the Portilla & Simoncelli algorithm). Initial images first had a Gaussian filter applied and were then processed using the SHINE toolbox.

This repetition was initially randomized, and then all participants saw the same pattern of randomization. All image stimuli (for both the localiser and experimental scans) were presented on a mid-grey background, processed using the SHINE toolbox scaled to the mean

luminance of the stimuli images. This produced a grey with a consistent luminance to the target images.

3.3.3. fMRI acquisition parameters

A localiser scan and ASSET calibration scan were performed for each participant. Functional fMRI data were acquired using a gradient single-shot echo planar imaging (EPI) sequence with 25 interleaved slices at 3mm slice thickness (TR = 2s, TE = 32.5ms, FOV = 288 x 288mm, matrix size = 128 x 128, voxel size = 2.25 x 2.25mm (axial plane). A FLAIR scan was acquired using 25 interleaved slices at 3mm thickness (matrix size = 512 x 512mm, voxel size = 0.56 x 0.56 mm), and a T1 isotropic volume structural scan was acquired (matrix size = 176 x 256 pixels, slices = 256, voxel size = 1 x 1 x 1mm).

3.3.4. Procedure

Participants took part in one localiser scan (5 mins) and 12 experimental scans (3.3 mins each). The localiser scan was used to determine participants' regions of interest (ROIs) and involved interleaved blocks of place images and face images (see Figure 3.2A) with an initial fixation cross of 12 seconds. The participants were shown blocks of images in alternating locations involving a pair of identical images for 2 seconds at a time, occupying opposite corners of the four onscreen spaces. Images were broken into two blocks of opposing locations, each block consisting of 6 place pairs, and then 6 face pairs. Place images and face images alternated every 12 seconds, and the location of the images (upper left and lower right locations; then upper right and lower left locations) alternated every 24 seconds. Participants saw a total of 72 place images and 72 face images. No task was required except for the maintenance of fixation.

Before the start of each experimental scan, participants were given both verbal and on-screen instructions about their task. They were asked to report whether and where, on screen in the four possible positions, they saw a target image. The experimental scans consisted of 12 fast event-related 'runs', which used three run 'types' repeated four times. The first run type used place images only as a target, and the second run type used face

images only as a target. These run types had a 50% chance of a target being present in a given trial, and all positions that did not show a target used scrambled target images. The third run type also used place images as the target, but in which there were also face images present elsewhere as 'distractors'.

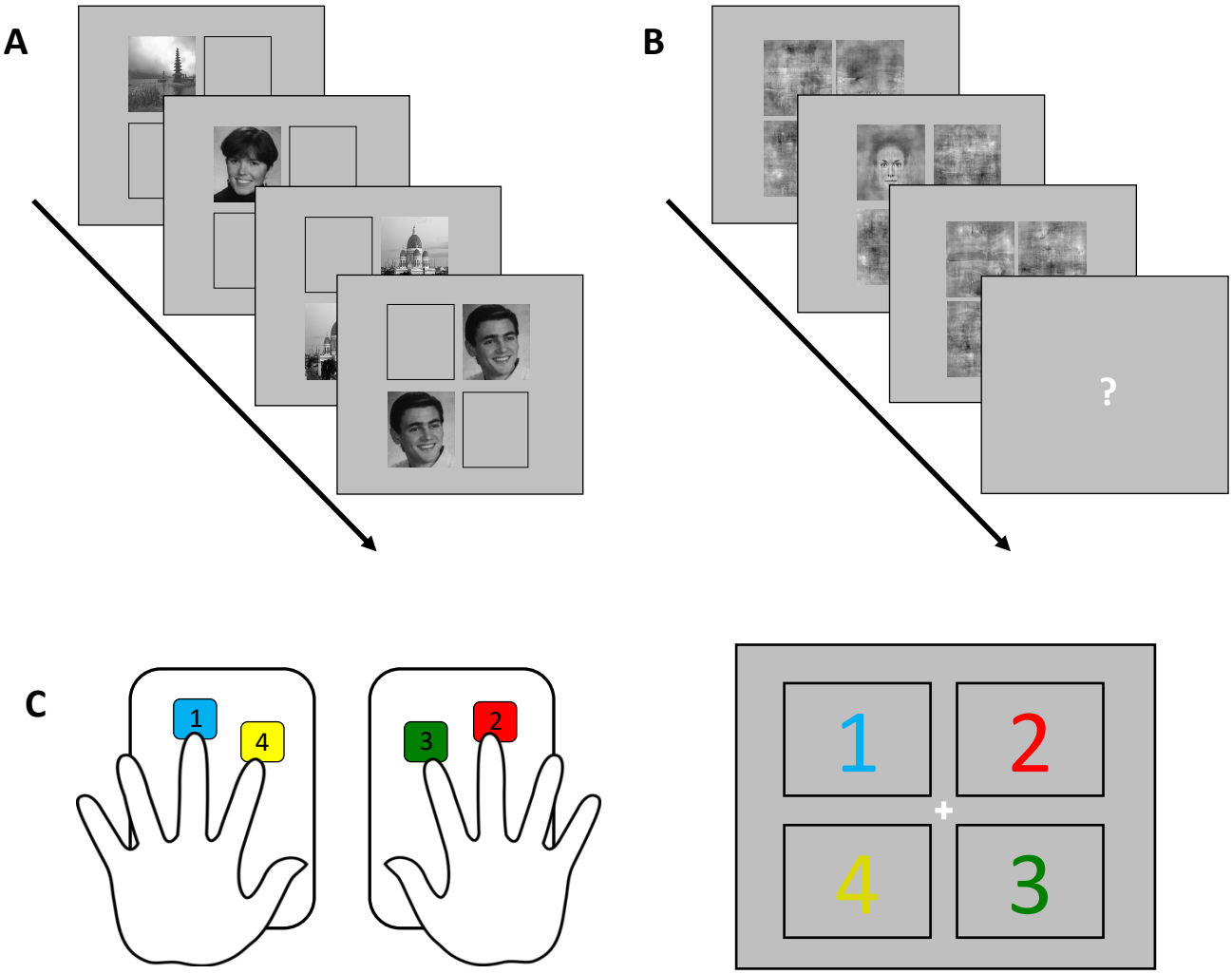

Figure 3.2: Demonstrations of the stimuli types, and how participants respond to them. A shows the localizer stimulus paradigm. B shows the experimental trial structure, in which the face image is the target. The participant is shown 4 texture masks for 200 msec, the four trial images (in this case, a 'target present' trial utilising a face image and 3 scrambled face images) for 200 msec, and four different texture masks for 200 msec. C demonstrates the participants' response button-box set-up, and the corresponding stimulus positions onscreen.

At the beginning of each run, participants were shown a 6-second cue word to indicate the target image. On each presentation, there was a 50% chance of a target, and a 50% chance of a distractor. Each run type contained 30 trials, interspersed with blank temporal 'jitter'

periods in which only a fixation cross was displayed. All experimental trial types and 'jitter' periods were calculated and presented in an order determined by Optseq2 (Dale, 1999; Dale, Greve & Burock, 1999).

As demonstrated in Figure 3.2 B, the participant would be presented with texture mask images, followed by four images with a 50% chance of a target present, and another four mask textures. Participants were shown a question mark to indicate the 1400ms response period. Participants used the response box (Figure 3.2 C) to indicate the location of a target by pressing the button that corresponded to the position on screen. If the participant did not perceive a target, they were instructed to press any two or more buttons as their response. No response feedback was given.

The category image was shown for 200 ms to allow for a clear signal to be seen within the brain, due to limitations of event-related fMRI design. This is a presentation system proposed by Walther, Beck, & Fei Fei (2012), in which they suggested a short presentation time followed by a perceptual mask as the optimum design for natural scene categorisation. This is also a time-frame that allowed for sufficient errors to be made during the experiment, allowing later analyses to contrast hits, misses, false alarms and true negative responses. Shorter presentation times appeared to make the task too difficult, and longer presentation times made the task too easy.

3.3.5. Data analysis

For all scans, the initial 12s of data were removed to allow the fMRI signal to stabilise. All data were analysed using FEAT v5.98 (Woolrich, Jbabdi, Patenaude, et al., 2009). MCFLIRT motion correction (Jenkinson, Bannister, Brady & Smith, 2002) and slice timing correction were applied, followed by spatial smoothing at 6mm FWHM using a Gaussian kernel. Temporal filtering (48s) and high pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 24s) was applied. GLM time-series statistical analyses were carried out using FILM time-series pre-whitening (Woolrich, Ripley, Brady & Smith, 2001) with local autocorrelation correction (Woolrich et al, 2001). Registration to high-resolution structural (T1) and standard brain images (MNI152, 2mm) were carried out using FLIRT (Jenkinson & Smith, 2001; Jenkinson et al., 2002). Individual participant data were analysed

using a fixed-effects design, and then were entered into higher-level group analyses using a mixed-effects design (FLAME 1, FSL), with voxel correction

3.3.6. Localizer

Regions of interest (ROIs) were defined on a subject-by-subject basis from localizer scans. These ROIs were then sub-divided into areas based on face- and place-specific responses thresholded at a z-score of 2.3 (an approximate p-value of 0.02, two-tailed, uncorrected). These consisted of two place-sensitive areas, parahippocampal place area (PPA) and retrosplenial cortex (RSC), and two face-sensitive areas, fusiform face area (FFA) and occipital face area (OFA), as indicated in Figure 3.3. Left- and right hemisphere ROIs were combined to create a bilateral mask.

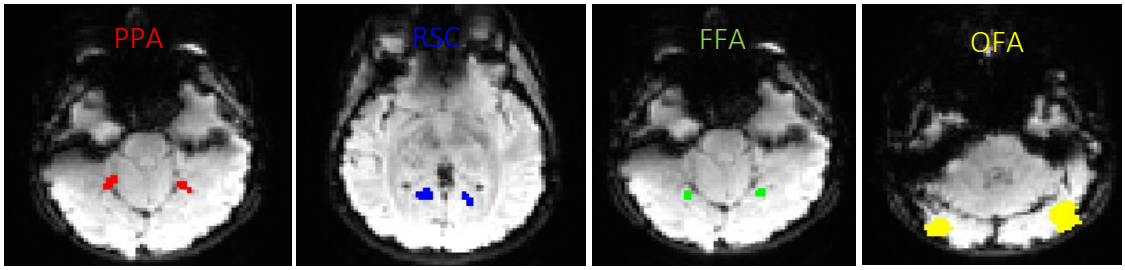

Figure 3.3: Examples of regions of interest from participant R3531. PPA (X = 55, Y = 49, Z = 7); RSC (X = 59, Y = 45, Z = 13); FFA (X = 56, Y = 42, Z = 6); OFA (X = 82, Y = 37, Z = 4).

As the ROI masks had been identified using subjective methods, we needed to ensure that the ROIs we were identifying were correct. The manual masks were therefore multiplied by Harvard-Oxford atlas maps (Harvard Centre for Morphometric Analysis) (PPA = lingual gyrus, RSC = precuneus cortex, FFA = temporal occipital fusiform cortex, OFA = occipital fusiform gyrus). Four early visual cortex (EVC) masks were also created for each participant: one each for the upper and lower calcarine region, for left and right hemispheres. Quadrant masks were masked by hand using a separate localiser analysis. These masks were then multiplied by four V1, V2 and V3 probabilistic maps (Wang, Mruczek, Arcaro & Kastner, 2015) to create four final masks for each participant.

ROI analyses were conducted on each participant's ROIs. Mean voxel activation for every considered condition was calculated for each mask, and participants' data were averaged across the group

3.3.7. Experimental scans

To look at the neuronal correlates of gist perception, participants' response data first were sorted into four distinct categories: hits, misses, false alarms and true negatives. fMRI responses corresponding to each of the four categories were then compared against the baseline activity throughout the scan (determined by average BOLD signal across the length of the run).

In addition to sorting responses by response type, we also asked which areas contained voxels that coded the task contingency. To do this, the responses were sorted into four possible 'event types': 1) target present, in which the target and only the target is shown onscreen; 2) dual target, in which both the pre-cued target and the non-cued target are onscreen; 3) target absent, in which only scrambled images are present onscreen; and 4) distractor-image only, in which only the non-cued target image is onscreen. During runs which included both image types, the target-present and dual-target events were required to localize areas associated with contingency change. In the analysis, we looked for areas which showed a greater signal change for a single target (target-present) compared to a dual-target event. This contrast was chosen as it would highlight regions in which the neuronal response echoed the predicted behavioral responses. Essentially, this contrast would highlight regions which showed destructive collision during the dual-target trials, by comparing activation in these areas between dual-target and target-present trials.

3.4. Results

The goal of the study was twofold: 1) to determine where in the visual processing hierarchy is gist coded, and 2) how and where do contingencies of the task modulate gist processing. As a corolloary to this second question, a sub-goal was to determine if location information of the target is stored on the early visual cortex.

3.4.1. Processing of Gist

To address the first question ("Which regions respond to the perception of gist?"), we performed an initial whole-brain analysis in which we combined data for the separate 'place' target and 'face' target runs. We combined BOLD signal data for hits and false-positive responses and compared them to the BOLD signal data for combined misses and true-negative responses. This created two categories of events: "perceived target-absent", in which participants indicate no target is present and "perceived target-present". We found greater BOLD signal change for a "perceived target-present" event in the left PPA, bilateral RSC, bilateral FFA, and bilateral OFA, overlapping with our predetermined ROIs (see Figure 3.4) and the lateral occipital complex (LOC). Importantly, there was no evidence of significant BOLD signal change between perceived and not-perceived targets in any part of the early visual cortex.

We then asked whether responses in the ROIs corresponded to response type (hit, miss, true negative, false alarm) and target type. The extracted data were analyzed using mixed model ANOVAs with one ANOVA per target type and target selective ROI as a between-subjects variable.

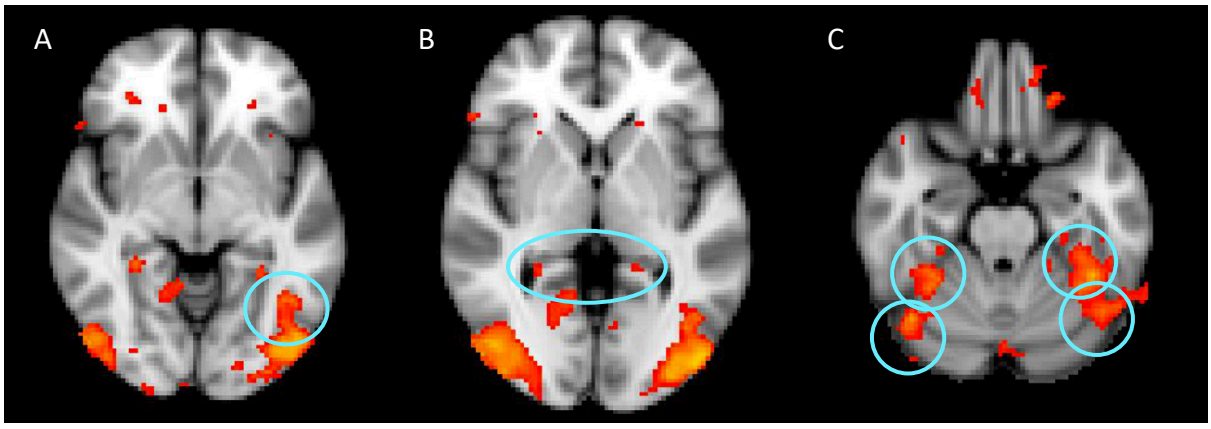

Figure 3.4: Pre-determined ROIs displayed on an MNI152 2mm standard brain. A) PPA, B) RSC, C) FFA (upper) & OFA (lower). Voxel thresholding at $p < 0.05$. Images taken from multiple diagnostic slices.

For place-type targets, data from the place-selective areas (PPA and the RSC) only were analyzed, data (Figure 3.5). There was a significant effect of ROI, $F(1, 12) = 11.667$, $p < 0.01$, indicating that the PPA responded with a greater % BOLD signal change (mean = 0.228, SE = 0.053) than the RSC (mean = -0.008, SE = 0.03). However, there was no significant

interaction between ROI and response type, $F(3, 36) = 1.992$ $p > 0.1$, indicating the pattern of % BOLD signal change for participant responses was the same across both ROIs, despite their difference in mean response amplitude. There was also a significant main effect of response type, $F(1.803, 21.635) = 6.112$, $p < 0.01$. Repeated planned comparisons showed no significant difference in % BOLD signal change between hits and false-positive responses, $F(1, 12) = 2.436$, $p > 0.1$, and a significant difference between false-positive and miss responses, $F(1, 12) = 7.038$, $p < 0.03$. Sidak-corrected pairwise comparisons indicated significant differences in % BOLD signal change between hits and miss responses ($p < 0.05$), hits and true-negative responses ($p < 0.3$), and between false-positive and true-negative responses ($p = 0.04$). This pattern of responses is a signature of an area that responds to the perception of gist, rather than the actual presence of a gist stimulus. These findings suggest these areas are more sensitive to the percept of a target than just the presence of the target.

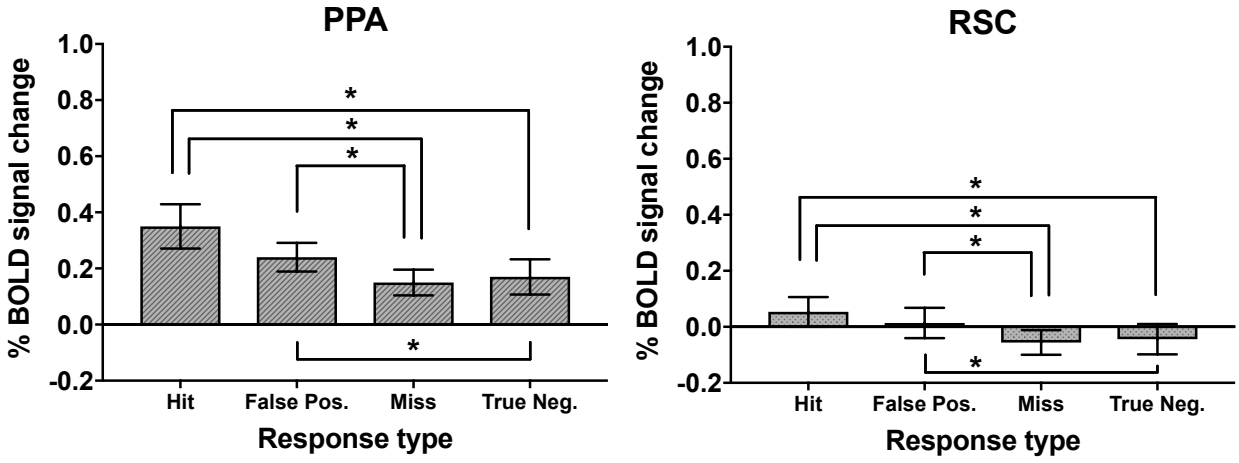

Figure 3.5: % BOLD signal changes in each place-responsive ROI, for each response type. Significant differences are indicated by asterisks. Error bars = standard error of the mean.

Analysis of responses of face-selective regions (FFA and OFA; Figure 3.6) during face target scans showed no significant effect of ROI, $F(1, 12) = 2.077$, $p > 0.1$, and no significant interaction effect between ROI and response type, $F(1.272, 15.267) = 0.1.739$, $p > 0.2$. These results indicate that there was no significant difference in average response amplitude to face targets between the face selective ROIs, nor were the response patterns different.

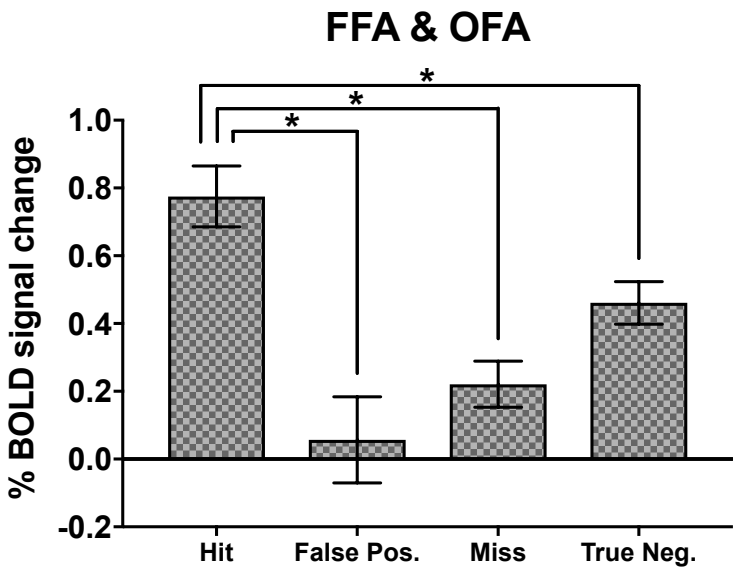

Figure 3.6: % BOLD signal changes for each response type, across collapsed FFA and OFA data. ROIs were collapsed due to the non-significant effect of ROI. Significant differences are indicated by asterisks. Error bars = SE.

Similar to our "place area" analysis, we found a main effect of response type, $F(1.854, 22.248) = 12.148$, $p < 0.001$ for face targets . However the profile was subtly different, with planned contrasts indicating significant differences between hits and false-positive response, $F(1, 12) = 19.174$, $p = 0.001$, hits and miss responses. Sidak-corrected pairwise comparisons revealed additional, statistically significant changes in % BOLD signal increase between hits and miss responses ($p = 0.002$), and between hits and true negative responses ($p = 0.001$). Unlike place-selective regions, for a face target to significantly activate the face–selective areas it is not enough for a target to be only perceived but must also physically present.

This difference in the response profiles for place-sensitive and face-sensitive regions resulted in statistically significant difference when data was combined in a mixed ANOVA, with a significant interaction effect of ROI and response types, $F(1.581, 79.07) = 8.55$, $p = 0.001$.

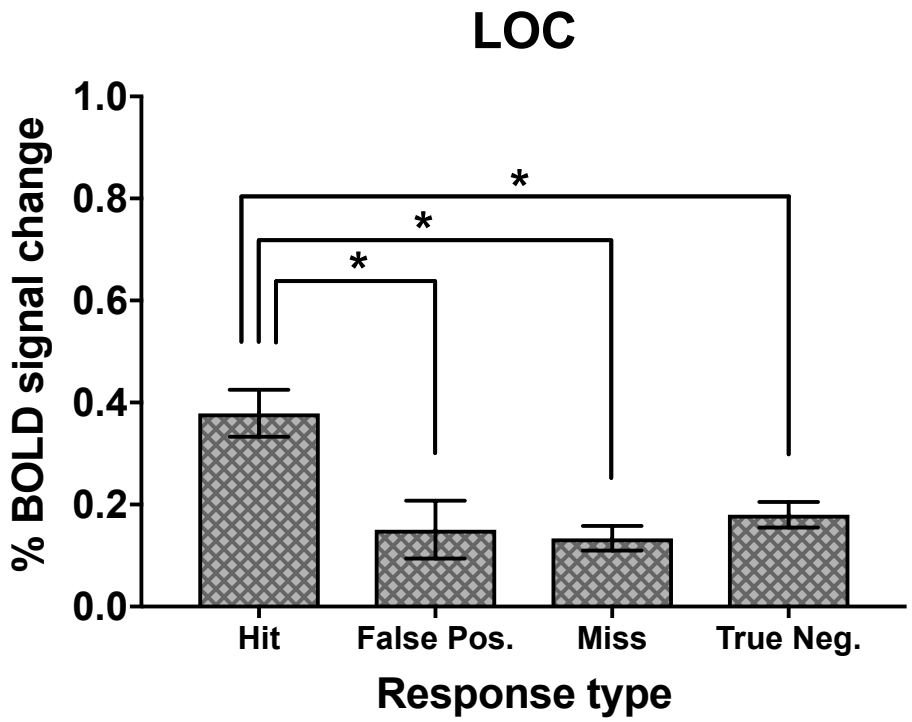

Figure 3.7: Responses of the LOC to the four participant response types, during place-only and face-only trials (averaged). All error bars represent standard error of the mean. Significant differences between categories are indicated with asterisks.

We analyzed the LOC independently (Figure 3.7) due to its presence in the whole-brain analysis even though we did not localize it separately with a functional localizer. The LOC showed a main effect of target type, $F(1,12) = 14.04$, $p < 0.01$, with place-type images producing a greater % BOLD signal change (mean = 0.258%, SE = 0.025%) than face-type images (mean = 0.164%, SE = 0.034%). There was no interaction effect between response and target type, $F(1.712, 20.658) = 2.937$, $p > 0.08$, indicating the % BOLD signal change-related response profile for both targets were not statistically different. There was a significant main effect of response type, $F(1.796, 21.55) = 10.604$, $p < 0.001$. Planned comparisons indicated a significant difference in % BOLD signal change between hits and true-negative responses only, $F(1, 12) = 44.106$, $p < 0.001$. Further Sidak-corrected pairwise comparisons found significant differences between hits and false-positive responses ($p <$

0.05), and between hits and miss responses ($p < 0.001$), in % BOLD signal change. This indicates a response profile similar to that of the face-selective areas, with strong preference for both the presence of and the perception of the target.

Whole-brain analyses did not reveal a category or signal-detection difference in the early visual cortex (EVC) quadrants. To check this more directly, EVC ROI masks were used to analyze the data. We were interested to see if there were any signal-detection-responses to the rapidly-presented targets, specifically when they were perceived in the receptive field areas of the EVC. A repeated-measures ANOVA showed no difference in response between

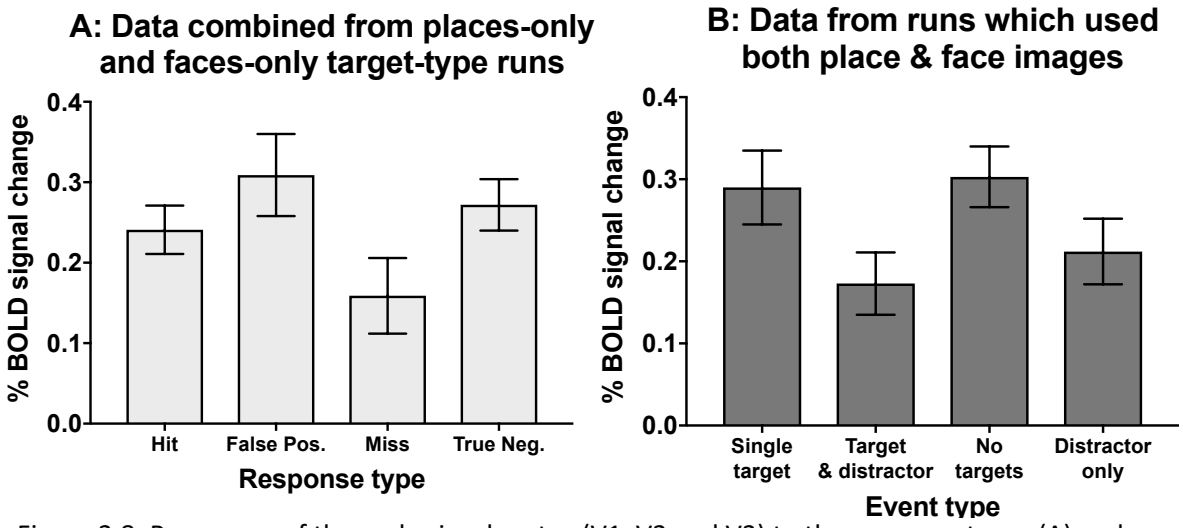

Figure 3.8: Responses of the early visual cortex (V1, V2 and V3) to the response types (A) and event types (B) of the experiment. These graphs are collapsed across EVC quadrants, and across participants. Graph A shows data from places-only and faces-only runs. Graph B shows data only from runs that used both image types.

visual cortex quadrants, so data were collapsed across them. There was no statistical difference between any of the response types (Figure 3.8 A). This indicates that no part of the EVC was responding specifically to the perception of a target; this would mean the gist percept does not depend on feedback connections to the EVC or that observers have access to location information for target or distractors from these areas.

The whole-brain analysis also indicated that the EVC did not show a differential response to single targets compared to dual targets. As in Figure 3.8 A, repeated-measures ANOVA showed no difference in response between visual cortex quadrants, and so data were collapsed across quadrants. None of the target-related event types were significantly different from one another, suggesting that no part of the EVC preferentially responded to the presence or location of a target or distractor images.

3.4.2. Gist Processing with Changing Task Contingency

The second aim of our study was to investigate the neuronal signatures of task contingency effects on gist perception. In terms of the behavioural data, participants successfully extracted gist of target images (places: overall correct = 65%, d' = 0.78; faces: overall correct = 93.78%, d' = 3.11) at above chance levels at a 200 ms exposure. Participants appeared to find it easier to locate face targets (96.03% of presented faces were correctly responded to, of which 2.4% of said hits were mislocalised) than place targets (71.79% of presented targets were correctly responded to, of which 31.79% were mislocalised), possibly due to the high level of saliency.

As in the previous analysis, for the fMRI data we started with a whole brain analysis to identify the regions which showed greater % BOLD signal activation during the presence of a single target (single-image trials), compared to trials that involved the simultaneous presentation of both a target and a distractor (dual-image trials). This analysis used only run types in which places and faces could both appear and was done agnostic of the predetermined ROIs and voxel-corrected ($p < 0.001$). Significantly different activation for

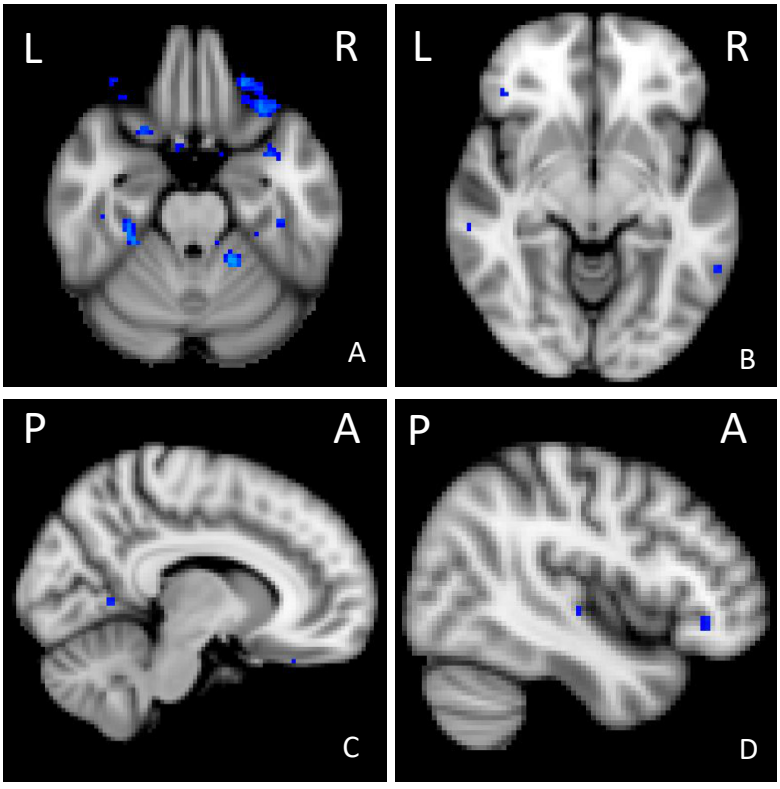

Figure 3.9: Regions of activation in the whole-brain analysis using contrast single target > dual target. Regions of activation include parahippocampal areas (A), dorsolateral pre-frontal cortex and middle temporal areas (B), anterior hippocampal areas (C) and the superior temporal sulcus (D). None of the pre-determined regions of interest appeared to be activated.

Threshold used: z-stat = 2.3, equivalent to uncorrected one-tailed $p < 0.01$.

single-image trials compared to dual-image trials were found in several areas, including the right dorsolateral pre-frontal cortex as can be seen in Figure 3.9. These responses consistently survived both voxel-based and cluster-based correction; however, the region is small (~4 voxels) and, despite surviving multiple analysis tiers, may more accurately be described as a trend.

3.5. Discussion

The aim of this study was to determine where in the brain 'gist' processing for probed image categories was occurring, and further to determine the brain regions demonstrating the neuronal correlates of task contingency change. The central conclusion is that the pattern of neuronal activity for gist processing supports a framework of *feedforward* processing of image gist: Information about target presence and location is not held in the primary visual cortex but rather in category selective regions. The modulating effect when multiple gists are primed but task contingencies change is the result of a decisional rather than a perceptual process and may be correlated with activity in executive control network.

3.5.1. Category selective cortex

Participant successfully (i.e. above chance) extracted target category images at a 200 ms exposure, as seen in section 3.4. This is not surprising, given that humans are able to extract basic image category from exposure duration as rapid as 30-67 miliseconds (Green & Oliva, 2009a). The brain areas that differentiated between perceived target present and perceived absent trials were all situated in the extra striate cortex, in regions which respond preferentially to a type of visual stimulus. Gist processing for place images were seen in the parahippocampal place area (PPA) and retrosplenial cortex (RSC), and for face images were seen in the fusiform face area (FFA) and occipital face area (OFA). Gist processing for both types of image were also seen in the lateral occipital complex (LOC), which was not a pre-identified region of interest. These findings indicate that rapid gist processing (i.e. image categorization) happens at higher levels of the visual hierarchy, in category selective cortex and in associative cortex. However, the primary visual cortex (which surely plays a role in

image processing) does not seem to show activity that differentiates between perceived presence or absence of gist.

Due to the temporal limitations of fMRI, we cannot determine when these regions are showing activity, nor can we determine if this is within the first feed-forward sweep of information through the ventral stream, as Serre, Oliva, & Poggio (2007) and Fabre-Thorpe (2011) have argued. However, observing these regions in the data allies with the findings of other researchers, who have argued that the extraction of characteristics of gist, such as basic category, can be found within these stimulus-selective cortex areas. Specifically, Peelen & Kastner (2014) observe that scene-selective regions such as the PPA and RSC represent global scene properties and the scene category (as found in TMS and fMRI studies – Dilks, Julian, Paunov, & Kanwisher, 2013; Epstein, 2008), with the PPA representing spatial information in particular (Kravitz et al., 2011). As scene image gist includes such elements as spatial information (Oliva & Torralba, 2001; Oliva, 2005) and global properties (Groen et al., 2013), then it is possible that the processing of scene gist is occurring within these regions. In terms of the face targets, the same logic may be applied; Peelen & Kastner argue that these elements of the stimulus (such as global properties) are represented in whichever region of cortex best characterises the target category, so it is arguable that face gist would also be represented within the FFA and OFA, as they are both face-sensitive regions.

Further to this data, we could hypothesise that the gist of objects may be observed within object-selective cortex, which suggests further research into gist extraction: can the gist processing of other targets be observed within their own selectively-sensitive cortices? This might be approached by leveraging the ability to rapidly identify a target object through its disjunctive diagnostic features (Evans & Treisman, 2005; Evans & Chong, 2012), and attempting to observe correlated neuronal activity in object-selective cortex areas. This could replicate the observation that gist processing is seen within category-selective cortex. It is worth noting that these data are, of course, a correlational link, in that we cannot claim that category-selective cortex is the neural basis of gist processing, only that we can see it within these regions.

What is important to note is that these areas of category-selective cortex did not respond to the presented stimuli in the same manner. We further examined the patterns of activity within category-selective regions in order to gain some understanding of how gist operates for certain category information. We observed that the PPA and RSC both showed

a perceptually-driven pattern of activation for place images, with the areas responding to a greater degree during hit and false-positive-type responses; by contrast, the FFA and OFA showed a more hybrid response, indicating that both the on-screen presence *and* the perception of the target were required.

3.5.2. Lateral occipital complex

The LOC was not a predicted region of interest, though its functional presence was consistent across trials, regardless of target type. Initial analyses produced a trend in which the LOC responded to place targets with a response profile similar to that of the place-sensitive regions (PPA and RSC), and to faces targets like a face-sensitive region (FFA and OFA). This trend was not borne out in the final analysis (p = 0.08), however. As the LOC was not pre-identified as a region of interest in gist processing, this raises the question of its presence in our data, namely what functions is it performing, and why are we seeing it?

Answers to these questions could be suggested by research into the anatomical and functional aspects of the LOC. The LOC is commonly associated with object processing (Malach et al., 1995; Kalanit Grill-Spector, Kourtzi, & Kanwisher, 2001), but also is located on or near the transverse occipital sulcus (TOS), which is responsive to scene images (Dilks et al., 2013) and demonstrates a response profile similar to that of the PPA (Bettencourt & Xu, 2013). The area we defined (i.e. masked, using both functional data and the Harvard-Oxford brain atlas) as LOC covered several regions in the occipital cortex, and including the TOS toward the upper part of the mask. This may explain the trend in which the observed LOC region responded to place targets in a manner similar to a place-sensitive region, as the TOS was included within the region of interest analysis. It may also explain the lack of statistical significance, as the response of the TOS was included with the averaging of several regions and thus its response to place images was 'diluted' within the analysis.

In terms of functional research, the LOC has also been found to respond during tasks in which natural scenes are categorised (D. B. Walther et al., 2009), along with the PPA and the RSC, leading Walther et al. to suggest that they are all part of a network which processes (and aids in the categorisation of) scene images. They found information relevant to scene category within these regions, suggesting that the regions may be able to discriminate scene categories. However, Walther et al. note that gist is thought to provide guidance for object

detection (Moshe Bar, 2004), suggesting that the activity seen in the LOC may be the result of a top-down modulation of signal from the PPA and RSC. This would suggest that the PPA and RSC are both engaged with gist extraction, and they then subsequently feed this information back to the LOC, which would explain its appearance in our whole-brain analyses.

The LOC has also been observed in face processing, with connections to both the FFA and OFA (Nagy, Greenlee, & Kovács, 2012). The presence of face stimuli modulates the connection between the LOC and the FFA, and Nagy et al. indicate the LOC aids in discriminating face from non-faces. This connection may explain both the trend for the region to display a response profile similar to that of the face-sensitive regions, and its functional presence even when the participant is viewing non-face targets; Nagy et al. argue the LOC may aid in acting as a gating mechanism for face stimuli, discriminating between face and non-face stimuli. However, the research by Walther et al. suggests the response of the LOC is more fine-grained; they found the LOC contained sufficient information to distinguish between natural scene categories, implying a far less blunt categorisation ability. This may also be due to the fact that, frequently, functionally-defined LOC areas can overlap with the FFA (Kalanit Grill-Spector et al., 2001), which (much like for the TOS and place stimuli) would explain the trend of seeing an FFA-like response profile to face stimuli.

Overall, the appearance of the LOC in the whole-brain analysis is in keeping with the literature as a whole. These findings would suggest that our initial conclusions are correct, with gist extraction being observed in category-selective cortex and activation in the LOC being a result either of modulation of signals from category-selective cortex, or from regions overlapping category-selective cortex being included within the LOC mask. Further to our suggestion that other category-selective cortices are investigated for gist extraction, this also raises the question of where else gist might be observed within the brain, if the category-selective cortex is simply part of a larger gist-processing network; possibly a study utilising multiple further category types, such as objects and animals, may allow us to identify the regions – or identify part or all of a network – common to gist processing across categories. Additionally, our findings indicate the neural correlates of one aspect of gist – 'category gist', if you will. Different aspects of gist are accessed at different time-points within the brain (Loschky & Larson, 2010), and there are different 'types' of gist – such as numerosity, size of a set of items, colour of a set of items – that may also be used in research to develop a larger picture of the gist-processing network of the brain.

3.5.3. ROI response profiles

We have briefly discussed the differing response profiles for category types between different category-selective cortex areas. How does one explain these differences in response?

The PPA may not be a single, homogenous region of interest, but may be divided into two functionally-distinct (though not necessarily independent) subunits (Baldassano, Beck, & Fei-Fei, 2013). These subunits, anterior and posterior, are arguably responsible for different processes: the posterior PPA (linked to the LOC), which processes low-level visual features and object shape, and the anterior PPA (linked to the RSC), which processes memory and scene context. The place mask images, synthesised using the Portilla & Simoncelli (2000) algorithm, maintain some of the same spatial frequencies as the place images from which they are created; some participants reported confusing some mask images as unique place images, frequently as lakes or other place images with a single, identifiable 'horizon'.

Our findings, in which either a hit or a false positive may provoke a significantly larger BOLD response, may be because our PPA mask conflated the two PPA areas. The greater BOLD signal seen in the PPA during false-positive responses may be the result of the posterior PPA responding to the low-level visual features, such as the spatial statistics (Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011) or summary statistics (Cant & Xu, 2012) of the mask images in situations where the mask contains the properties more closely reflective of a scene image (Troiani, Stigliani, Smith, & Epstein, 2014). As our PPA mask covered both regions, anterior and posterior, this more select response would be attributed to the region as a whole.

In terms of the FFA and OFA, our findings replicate those of Rodriguez et al. (2011) who found that face-sensitive regions only appeared to respond to the conscious perception of a masked face, and not when a presented face was genuinely unperceived, in both an fMRI and EEG paradigm. Unlike for scene-sensitive regions, the FFA and OFA did not show a significant BOLD response to false positive identifications, or any other response type; face images had to be both objectively present, and perceived by the viewer, for these regions to significantly respond. This may be due to the manner in which the FFA and OFA help with the initial face vs. non-face distinction when identifying objects in a scene (Maher et al., 2016);

an object first must be present, and must be correctly identified, for these regions to show a significant BOLD response.

3.5.4. Early visual cortex

The early visual cortex showed no evidence of location information for the experience of gist, whether for the perception of a target, or for the actual presence of a target or distractor image. These findings support the suggestion that gist can be attained during the first feed-forward pass of processing through the visual system. However the observers were able to localize the target in one of the 4 quadrants with reasonable accuracy with mean mislocalization for places at 11.4% and face at 1.2% and for runs using both, the mean rate was 8.6%.

So where is location information being accessed? Evidence from research into object perception suggests that location information is stored within category-selective cortex, along with the category information, even if the target is of a type not preferred by the category-selective cortex (Schwarzlose, Swisher, Dang, & Kanwisher, 2008; Cichy, Sterzer, Heinzle et al., 2013). Shwarzlose et al. observed that both category and location information for objects were stored, independently, within object-selective areas; though substantially more location information was found in lateral areas, such as the LOC, TOS and OFA, than in ventral areas, including the FFA and PPA (also found by Radoslaw M. Cichy, Heinzle, & Haynes, 2012).

When contrasting target-present trials against target absent trials, the group analysis revealed regions of increased BOLD response bilaterally within our functionally-defined LOC mask, though the area of increased BOLD respond was greater in the right LOC than the left. Bilateral activation within the LOC mask was also found when contrasting trials in which observers perceived a target, compared to when no target was perceived. Combined with the research discussed above, this raises the possibility that the location of the stimulus was stored within the LOC; we have previously discussed the links between the LOC and our other regions of interest, and it has been found to store location information (Cichy, Chen, & Haynes, 2011).

3.5.5. Destructive interference with multiple gists

Distractors can interfere with the detection of task-relevant stimuli (Evans, Horowitz & Wolfe, 2011). We looked for neuronal correlates of these types of interactions. Specifically, when two things have been primed, but only one of them is task-relevant, we asked if the task-irrelevant stimulus modulate neuronal activity and, if so, where?

Our behavioural findings, in which we observed a statistically-significant drop in performance accuracy for dual-target trials compared to single-target trials, replicate those of Evans et al. (2011) indicating that the participants were experiencing destructive collision.

This same comparison was used to find regions in the brain which responded in a similar manner. The locations of regions that respond in a different manner to single targets than to dual targets was exploratory; there were no specific hypotheses for this analysis, which revealed several voxels of an interference-related BOLD signal change in the DLPFC, amongst other areas. This regions was very small, but resilient, surviving several levels of analysis and voxel correction. It would be more accurate to describe this as a trend, despite its statistical significance, due to the size of the area and its discussion in isolation from other functional data (Cremers, Wager, & Yarkoni, 2017).

The DLPFC has been implicated as playing a role in response/action selection (Procyk & Goldman-Rakic, 2006) as well as in the more well-known fields of response inhibition and competition (Menon, Adleman, White, Glover, & Reiss, 2001), and in conflict monitoring (Taylor, Stern, & Gehring, 2007). This increase in BOLD signal during single-target trials may be representative of the participants' conflict-free response, in which the target signal has not been 'degraded' by the presence of the distractor signal and thus the choice of response is not as ambiguous. This, along with the fact that we see relatively increased activation for single target conditions in cognitive control areas as opposed to perceptual-processing areas, suggest that this destructive interference observed in behavioral results is the result of a decisional process and not a perceptual one.

3.6. Conclusions

To conclude, the extraction of visual category information (gist) is associated with activity in category-selective cortex and the lateral occipital complex. Gist location can be extracted without driving retinotopically specific responses in early visual cortex. Our best understanding, based on existing research, is that this location information is represented within higher-level category-specific areas. Gists can interfere destructively during perception, and we observe correlates of this in cognitive control regions, such as the dorsolateral prefrontal cortex, suggesting that interference occurs in decisional rather than perceptual mechanisms.

Chapter 4 The temporal dynamics of gist processing

4.1. Overview

The literature discussed in chapter 3 suggests that gist information is extracted very quickly from the visual scene, with certain types of gist being extracted a specific time points. Many of these findings are the result of behavioral experiments, with results based on the duration of the image shown and focusing on a single time point. What, then, is the time-course of gist extraction, from onset of a visual stimulus to the point at which the participants makes a decision about that stimulus? When might the different aspects of gist extraction be observed?

Chapter 4 first recaps the key facts about gist extraction, and discusses behavioural and neuroimaging evidence to support a hypothesis that gist extraction happens early on in visual processing. This provides a general temporal window in which gist extraction is hypothesised to happen. There is then further discussion of the existing EEG literature which addresses key event-related potentials (ERPs) and components of gist processing. A diffusion-based model of destructive interference is then discussed, along with the associated ERP.

An experiment was conducted which used electroencephalograms to measure the onset and duration of participants' neuronal responses whilst performing a task. Participants identified whether one of two pre-cued target categories were present in a 2AFC design. It looked for key ERPs that might help explain the time points, and therefore the stages of processing, occurring during gist extraction.

Univariate ERP-based statistics were used to identify the magnitudes of activity across the brain, allowing us to identify statistically-significant ERPs during the time in which participants were acquiring the category information (and so presumably gist) of the images presented to them. We found gist was acquired early, between 30–65 msec after stimulus onset. Relevant targets are differentiated at ~60 msec, and non-relevant targets later at ~130 msec. A 300 ms component, possibly N3 (as opposed to the predicted P3), was found as a correlate of destructive interference when the task contingency has been changed.

4.2. Introduction

To recap from chapter 3, gist is a series of characteristics we obtain rapidly to make judgments about the content and nature of the visual world.

As multiple studies looking into gist processing have noted, its onset is measured in milliseconds, which is a temporal resolution that fMRI – or even behavioural studies, which must allow for reaction time – cannot provide. Experiments that study complex visual scenes and visual processing have often utilised the extremely high temporal resolution provided by electroencephalograms (EEG) to more accurately study visual processing across time. EEG allows us to determine when in the brain a particular cognitive event occurs – or, more accurately, which time points are correlated with a particular event. In this chapter, we used this technique to explore *when* in the brain gist is observed (temporal dynamics of gist), to complement the neural correlates found in chapter 3.

These time points will also be examined in the context of visual ERPs (time signatures locked to certain cognitive, or sensory, events), which are frequently used to track temporal dynamics in neural processing. As ERPs are associated with certain kinds of neural (and in the case of this chapter, specifically visual) processing, this may help explain *why* certain stimuli evoke the observed neural responses, and it also allows for recorded EEG data to be examined in the context of the wider literature.

4.2.2. ERPs associated with further extraction of gist characteristics

The longer the processing time, the more complex the aspects of gist that can be extracted from a visual scene. Moving on from the earliest evidence of gist extraction, the literature suggests that discrimination for global-property categorisation (i.e. natural vs. man-made distinction) is found at around 100ms (Groen, Ghebreab, Prins et al., 2013). The rapid visual categorisation of a pre-cued complex natural scene, such as a scene that contains an animal, is achieved 150ms after stimulus onset (Thorpe, Fize & Marlot, 1996), even when there are two images presented rather than one (Rousselet, Fabre-Thorpe & Thorpe, 2002).

Other research suggests 150 msec is needed for observers to correctly categorise a visual scene (accounting for reaction time) and without requiring the deployment of attention (VanRullen & Thorpe, 2001; Codispoti, Ferrari, Junghöfer, & Schupp, 2006; Hegdé, 2008). Hegdé referred to this time point as allowing for the detection and categorisation of the scene, in that as soon as the category is detected, it is identified (Kalanit Grill-Spector & Kanwisher, 2005). VanRullen & Thorpe (2001) and Hedgé also argued that the 150 msec time-frame represented visual awareness of the stimuli (this also appears to be the earliest point where awareness of gist information has been identified).

Experiments into attentional selection in naturalistic scenes (Kaiser, Oosterhof, & Peelen, 2016) and visual search in naturalistic scenes (Battistoni, Kaiser, Hickey, & Peelen, 2018) identified similar time points, with initial target presence decoded as early as 50 msec post-stimulus onset, and pre-cued targets differentiated from distractor targets between 160–180 msec.

This provides a time-frame between 100–180 msec for various aspects of discrimination of scene types, and with a separation between target and distractor images between 160–180 msec. This research does not necessarily imply the presence of a single ERP, and suggests that the experiment described in this chapter may expect to find several ERPs.

The first of these potential components is the visual N1, which is a negatively-polarised ERP found between 160–200 msec (H. Heinze, Luck, Mangun, & Hillyard, 1990). It is associated with the early allocation of attention, and is seen to be larger when attending to the location of a stimulus. It is also associated with a discrimination process within said attended location, between classes of stimuli (Hillyard, Vogel, & Luck, 1998; Vogel & Luck, 2000). This would align with previous findings that pre-cued scene category (150–160 msec, Thorpe et al., 1996; Kaiser et al., 2016) and perhaps also scene discrimination (220ms, Harel et al., 2016) can be determined in that approximate window.

The second possibility is the P2, though it falls slightly outside of the time frame discussed above. The discrimination of scenes from non-scene images has been observed at approximately 220 msec from stimulus onset, differentiating even between types of scenes (such as open vs. closed) (Harel, Groen, Kravitz et al., 2016). This P2 component appears to be the earliest point at which diagnostic scene information is made available to the

observer, though the authors make clear that this appears to be in relation to scene-specific processing.

Also relevant to this experiment is the N2 component, which peaks between 200 and 350 msec (Folstein & Van Petten, 2008). Of particular relevance are the 'visual' N2 subcomponents. Folstein & Van Petten argue that these are related to novelty detection (or a mismatch from an internally-held template), cognitive control (response inhibition, conflict, and error monitoring), and visual attention. As Folstein & Van Petten argue, this can include deviation of expected stimulus from short-term context (i.e. a mental template).

The N2 component may therefore be associated with the process that determines when a pre-cued gist is not present. This is also seen in the go/no-go Erikson task, in with participants must respond to one type of stimulus ('go'), but withhold a response when presented with another type of stimulus ('no-go'). In this kind of task, the N2 component was associated with inhibitory executive function (i.e. the no-go part of the task – Heil, Osman, Wiegelmann, Rolke, & Hennighausen, 2000). Whilst Heil et al. (2000) caution against treating their findings as a guarantee that N2 components will be found in other response-competition experiments, there is an argument that it may be present during a pre-cued 2AFC task such as the methodology used in chapter 3, and thus may be seen during an EEG experiment utilising a similar paradigm, as here in chapter 4.

4.2.3. The diffusion-based model of gist and destructive interference for ERPs

Given this literature, it is clear that 'gist' information is accumulated over time. However, as in chapter 3, the question of multiple gists arises.

Multiple gists can be extracted from a single scene (for example, multiple categories of image such as 'beach' and 'animal') as found by Evans, Horowitz, & Wolfe (2011), and they discuss the way in which the relationship between multiple categories in an image can impact an observer's perception of the scene. This extraction of more than one category was done simultaneously and with a high degree of flexibility, as the two categories can also interfere with one another, leading to incorrect conclusions as to the nature of the visual scene (destructive interference).

One way of explaining these phenomena is through a diffusion model, which has been used in different tasks to account for multi-choice decision behavior (Philiastides,

Ratcliffe, & Sadja, 2006; Ratcliff & McKoon, 2008; Ratcliff, Smith, Brown, & McKoon, 2016; Tavares, Perona, & Rangel, 2017). As described by Ratcliffe et al. (2016), a diffusion model represents a decisional process (in the case of destructive interference) between two targets. Over time, evidence accumulates for each of the targets, either in favour or against the probability of the target's presence. Thus, if a scene with a beach is presented and the observer is looking for beaches, information about 'beach' accumulates (indicated by the red line in Figure 4.1). If it reaches the upper bound, the observer is sure that beach is present. If the stimulus is cut off and masked (symbolized by the green dashed vertical line, Figure 4.1), the observer must base a decision on the partial accumulated information.

A single fixation of up to 300 msec (Rayner, 1998) is enough to extract semantic information from the scene, to the extent that an exposure of 258 msec is enough to identify an image even with a 'negative' cue (e.g. "not the picture of food"; Intraub, 1981). This suggests that the evidence of the semantic content of the scene accrues quickly, with enough information acquired by 258 msec that even a less-straightforward 'negative' cue can be responded to with accuracy.

What is important to note here is the concept of evidence accruing over time. As

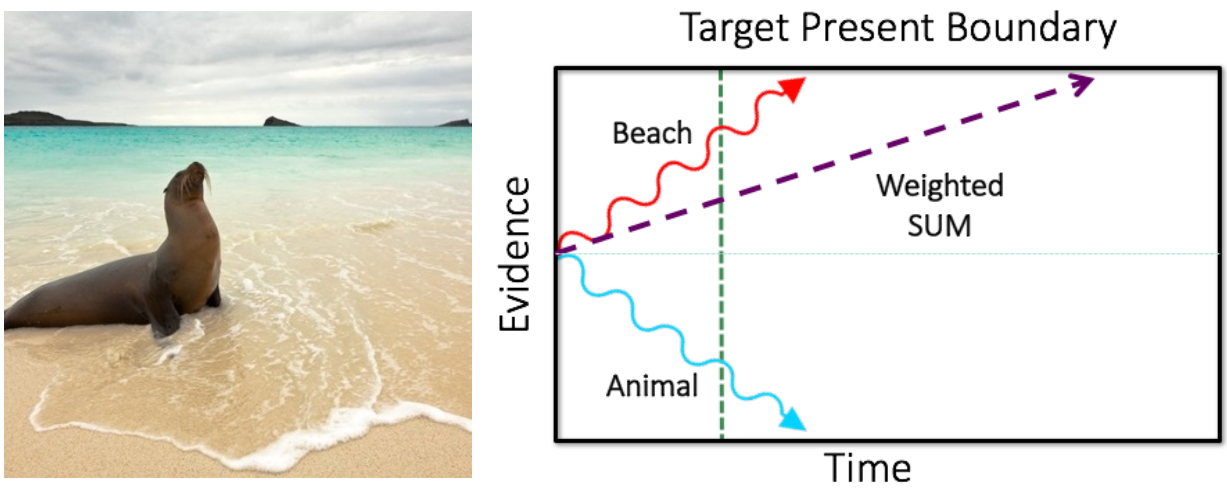

Figure 4.1: A cartoon of a diffusion model. A participant must make a decision about the category information of the presented image (left), after being primed for both the 'beach' and 'animal' categories. The diffusion model (right) displays the hypothesised diffusion process; wiggly lines represent accumulation of evidence for a particular category.

discussed previously, there is evidence that gist information can be extracted from scenes between 13–150 msec, though the rate at which category information is extracted depends on the properties of said category (e.g. global properties vs. basic category). Renninger &

Malik (2004), for example, argued that the processing of scene textures was integral to early scene identification, and found that texture information explained correct responses on 80% of scene categories, with accuracy becoming better with greater exposure duration. This relationship between accuracy and duration fits in with the diffusion model, indicating that evidence accruing over time builds a more accurate representation of the target category.

The diffusion model would explain destructive interference as an attempt to extract information about two competing gist categories; by asking the diffusion model to detect the presence of a pre-cued category in the presence of another (task-relevant but not trial-relevant) primed category image, observer performance is impaired. On trials in which the beach category is cued, for example, animal information is treated as anti-beach evidence (blue line, Figure 4.1), diffusing in the opposite direction. In the condition when both categories are present, summed evidence (purple line, Figure 4.1) will rise more slowly than the beach-alone evidence and, if the stimulus is brief, the accumulated evidence may be inadequate to support an accurate decision. This is what results in the chance performance found by Evans et al. (2011) for trials involving both a task-relevant and trial-relevant target.

The key notion here is the concept of a decision being made between conflicting evidence. Chapter 3 implicated, though did not confirm, the involvement of frontal cognitive control areas in trying to resolve the conflict caused by changing the task contingency. This provides a hypothesis: if these areas are involved in destructive interference, then logically the time signatures associated with frontal areas and decision-making should be seen as observers attempt to resolve the presence of a cued target and an uncued target. In terms of the diffusion model, this should represent an attempt to resolve the pro-target and anti-target evidence. Accumulation of evidence to a decision threshold is associated with late-stage cognitive components (Philiastides & Sajda, 2007), and one of the ERPs identified as a marker for conflict resolution in a diffusion model is the P3 component (Philiastides & Sajda, 2006; Philiastides, Ratcliffe & Sadja, 2006; Ratcliffe, Philiastides, & Sajda, 2009).

The P3 component is a positively-polarised ERP found generally between 250 and 500 msec (Polich, 2007), though a latency window of between 350-450 msec has been found in more vision-specific experiments (Comerchero & Polich, 1999). Like the N2 component, it is also related decision-making, cognitive control and executive control, though Folstein & Van Petten argue it is separate from the N2. Rather, it is a reflection of

neuroinhibition, (Polich, 2007) and can be broken down into two sub-components.

The P3 can be broken down into P3a ('novel') and P3b ('classic'). P3a, seen approximately between 220 and 280 msec (Squires, Squires, & Hillyard, 1975), is associated with novelty but is also associated with target discrimination difficulty (Comerchero & Polich, 1999; Polich, 2007; Philiastides, Ratcliffe & Sadja, 2006). The P3a is observed in frontal areas as a response to physiologically alerting stimuli to which attention is already engaged (Polich, 2007).

P3b is seen anywhere between 250 and 500 ms, and is associated with uncertainty resolution and internal decision-making (Nieuwenhuis, Aston-Jones, & Cohen, 2005). Polich (2007) note that it can be observed during associated attention and memory operations, in temporo-parietal areas.

Because of the P3 component's well-established relationship to these executive functions, it would be logical to expect to see a P3 component during dual-target trials in which destructive interference of gist is seen. As evidence is accumulated (see Figure 4.1) both for and against a trial-relevant target category, a decision must be made in the context of pro-target and anti-target evidence within a short time-frame. The presence of the P3 component, particularly the P3b, would indicate the observer's attempt to resolve the conflicting evidence.

4.3. Aims and hypotheses

For this chapter, we asked two main questions. Firstly, what are the temporal dynamics of gist (when in the brain is gist processing seen, and when are types of gist differentiated)? Due to the existing discussed literature, the hypothesis for this was that initial gist would be seen quickly, before 100 msec after stimulus onset, possibly as quickly as 30-67 msec (Greene & Oliva, 2009a; Groen, Ghebreab, Prins et al., 2013). It was expected the findings from this experiment would mirror this time-frame. In practise, this was the point at which target-present and target-absent trials were differentiated by the participants. It was also expected that a component around 150 msec would be found, due to its relationship to complex natural scenes, differentiation of target types and observer responses (VanRullen & Thorpe, 2001; Codispoti et al., 2006; Hedgé, 2008).

Discrimination between target categories was also hypothesised to be found at two possible temporal locations. The visual N1 ERP (160–200 msec) was one possibility, due to its relationship to the discrimination between classes of stimuli (Hillyard et al., 1998; Vogel & Luck, 2000.); the P2 was additionally a less-likely but plausible option. The N2 component (200–350 msec) was the second possibility, as it was associated with deviation from "mental templates", and therefore may reflect the participants' differentiation between trials that show the pre-cued category image, and trials which do not. This was assessed by comparing the participants' neuronal responses to relevant (primed and pre-cued) target categories.

The second main question asked when destructive interference was seen neuronally, and what did that time point reveal about the nature of the destructive interference? Based on the exploratory analysis from chapter 3, the hypothesis was that destructive interference was a decisional, and not a perceptual process. As such, the prediction for the time signature of destructive interference was the P3 component, particularly the P3b (250–500 msec).

The experiment in this chapter was therefore designed to resemble the methodology of chapter 3 in order to be able to relate the two sets of findings to one another. This was done by asking participants to identify the presence of a rapidly-presented, pre-cued category, in four types of trials: trial in which the cued target was present; trials in which the cued target was absent and a non-task-relevant target was shown; trials in which a task-relevant, but not trial-relevant, target image was shown; and trials in which both a task- and trial-relevant target were shown. This was intended to replicate the conditions in chapter 3, modified for an EEG experiment, whilst still allowing us to assess gist processing and destructive interference between gists.

4.4. Methods

Participants took part in an experiment which required them to identify the presence or absence of a pre-cued target image on a computer screen in a 2AFC design. For the first part of the experiment, the cue word would always be 'PLACE', with possible presented images including places, faces and objects, or a combination of two of them overlapped. During the second part of the experiment, participants could be cued with either 'PLACE' or 'FACE',

with place, face and object images presented either on their own, or as two overlapped images.

4.4.1. Participants

A total of 19 volunteers were recruited (0 left-handed, 1 ambidextrous, all paid, 9 men, with an average age of 27 years 4 months), all with normal or corrected-to-normal vision. All data sets collected were included in the analyses. Informed consent was obtained for each participant, who were recruited through an opportunity sample from the Psychology Department of the University of York participation website (PEEBS). Ethical approval was granted by the Psychology Departmental Ethics Committee, and all data were collected prior to analysis.

4.4.2. Stimuli and apparatus

4.4.2.1. Apparatus

This EEG experiment was designed and presented using MATLAB (2014a – Mathworks, MA, USA) and the PsychToolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007), and using a Mac Pro computer (6-core Intel Xeon E5, 3.5GHz CPU) running Mac OS X (10.9.5). Stimuli were presented on a VPixx 3D Lite with a 120Hz refresh rate, at an approximate viewing distance of 60cm.

Data were collected using ASAlab (version 4.9.2) software (ANT Neuro, Netherlands), with a 1000 Hz sampling rate and a high-speed 64-channel amplifier, on a HP 2230 SFF computer (core i7, 3.4 GHz) running Windows 7 Professional. The EEG caps used had a 64-channel layout according to the 10/20 system (WaveGuard original, ANT Neuro, Netherlands), and were sized individually to participant's skull circumference measurement. The vertical electrooculogram was also recorded using two self-adhesive electrodes, positioned above the left eyebrow and the top of the left cheek. EEG trigger information was sent and recorded using PsychToolbox. Reference electrodes used with the left and right mastoids (behind the ear). The ground electrode was Fz on the mid-line saggital plane.

4.4.2.2. Stimuli

All trials were presented on a greyscale texture image generated by using the Portilla & Simoncelli (2000) algorithm, which was used as a screen background image, with a mean luminance of 93.12 candelas/m^2, at 50% opacity.

Stimuli were greyscale images of places, faces, and objects. A total of 351 place images were used, comprising of a mixture of natural and man-made scenes (lakes, mountains, cities, streets, etcetera). Place images were not defined by a single object; for example, an airport scene would be a broad view encompassing multiple elements of an airport (some aeroplanes, a control tower, gateway buildings, etcetera), as opposed to a single central aeroplane. All place images were drawn from the MIT SUN database (Xiao, Hays, Ehinger et al., 2010). 285 unique face images were drawn from multiple face image databases: Lundqvist, Flykt, & Öhman, 1998; Psychological Image Collection at Stirling (PICS: pics.stir.ac.uk); Weyrauch, Heisele, & Blanz, 2004; Langner, Dotsch, Bijlstra et al., 2010; and Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, http://www.tarrlab.org/). All faces had a neutral affect and were presented facing the viewer. As 372 face images were required, 87 face images were repeated at random, no more than once per image. Additionally, whilst 348 object images were required, 244 unique images were used, with 104 images repeated at random. Both face and object images had a Gaussian filter applied to remove edge contours using the SHINE toolbox. This made the task slightly harder, in that items on screen could not be identified by their outline alone.

All place images were square and were presented centrally, at approximately 13 degrees of visual angle. However, all face and object images (whilst sized to the same dimensions as the place images – see Figure 4.2) had transparent backgrounds, which made presentation size less simple. To solve this, images were sized so that the shortest dimension of the object (vertical length or horizontal length) was matched to the largest dimension, without warping the image. These images were also presented on top of a square texture mask generated using the Portilla & Simoncelli algorithm, to prevent identification of the category image based on the outline of the image itself.

Examples of the types of trials are exhibited in Figure 4.2. 'Target present' trials present a single image which is a match for the cue (no superimposed images; see Figures 4.2 A and 4.2 C). 'Dual target' trials present the cued target image and a non-cued target image at the same time (figure 4.2 B), superimposed on top of one another. 'Target absent' trials displayed one or two superimposed images, in which neither image was of the cued category.

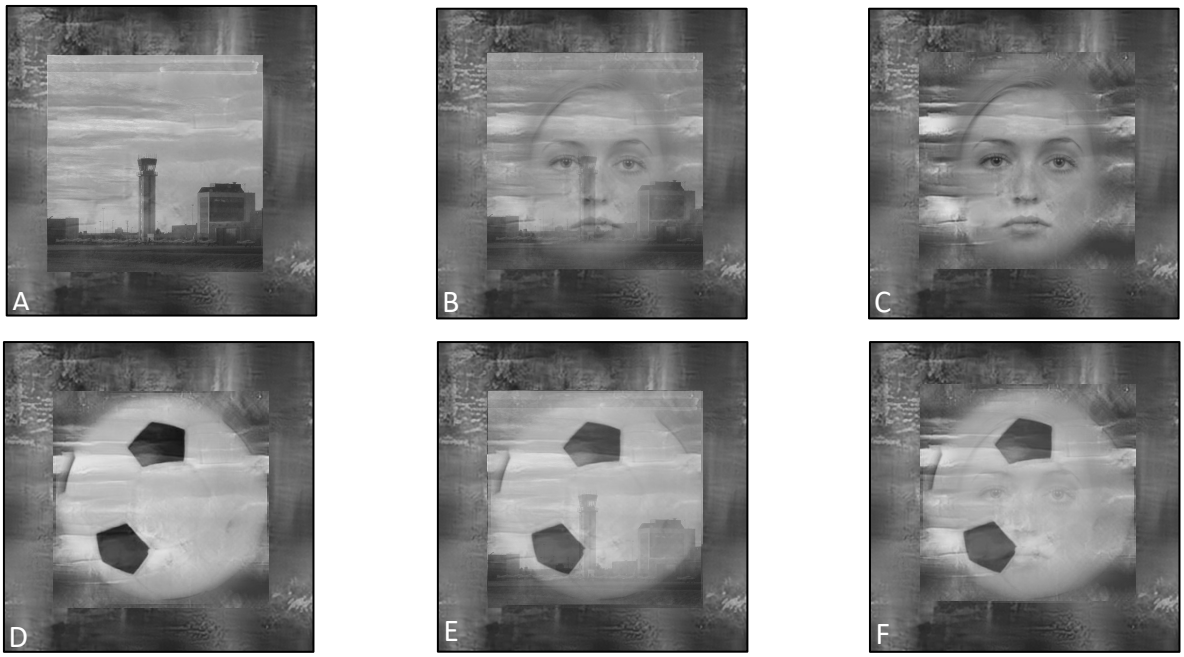

Figure 4.2: Examples of trial types participants may be shown. A shows a trial in which a place only is presented. B shows a 'dual target' trial image, in which the cued target (place or face) and non-cued target is presented simultaneously. C shows a trial in which a face only is presented. D shows a trial in which only an object is shown. E shows both a place image and an object image. F shows both a face and object image.

The ratio of trial types were arranged so that in a total accuracy situation, the participants would respond with a 'match' response in 50% of cases. Trial types were thus distributed so that target-present and dual-target trials each comprised of 25% of trials (72 trials per block), and 'target absent' 50% of trials (split evenly between single and dual 'target absent' trials).

4.4.3. Procedure

Participants were asked to participate in a rapid event-related experiment. They would complete two blocks, each containing 288 trials, in which they were to determine if a presented image 'matched' the cue word given for each trial. This involved giving the participant the cue word (800ms), displaying a fixation cross (300ms) to indicate the area at which participants should direct their gaze, showing the participant an image (25ms), and then waiting for the participants' response (Figure 4.3). The response was 2AFC, with participants pressing the left mouse button to indicate a match, and the right mouse button to indicate a mis-match to the cued word.

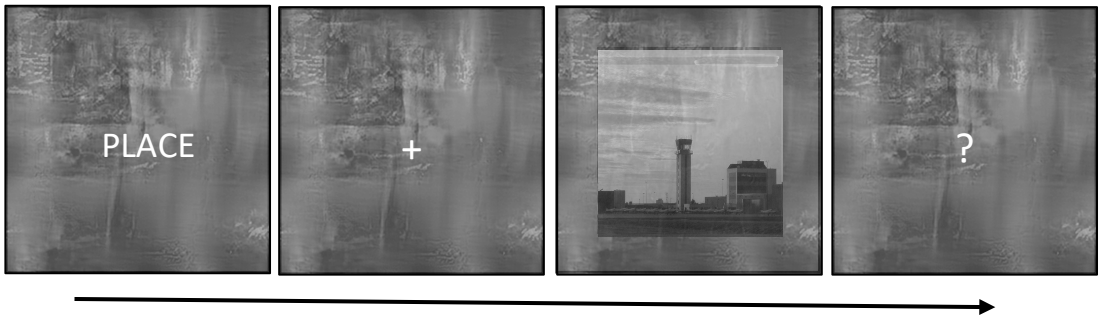

Figure 4.3: Cartoon of the trial structure. Example used here uses a 'PLACE' cue word, and the picture of an airport. This would constitute a 'target present' trial, with no 'distractor' image.

The cue word given would be 'PLACE' or 'FACE'. Whilst all trials used the same structure, there were four possible trial types: target present, in which the cued target was presented solo; target absent, in which a non-cued target image was presented solo; dual target, in which the cued target and a non-cued, though task-relevant, image are presented superimposed together; and dual absent, in which two non-target images are presented superimposed together. Images were either presented alone (for example, a place image only; see Figure 4.2 A), or could be superimposed on top of each other (i.e., an object image presented on top of a place image, such as Figure 4.2 E).

Images were presented on top of a texture to varying degrees of transparency, depending on the trial type. Images presented alone were at 70% opacity (see Figure 4.2 A) in order to make the categorisation task more difficult, and two images presented together were both presented at 50% opacity (as in Figure 4.2 B). For trials using two images, this further reduction in opacity was required for the task to be challenging, and to allow features from both images to be equally visible. There was a need to make the task

challenging in order to take the participants' performance off ceiling level, since the EEG method did not use pre- or post-image masking.

In block 1, participants were always cued with the word 'PLACE'. All image types (places, faces and objects) would be presented, with only places being task-relevant. In block 2, participants were cued with either 'PLACE' or 'FACE', with face images now becoming task-relevant. At no point during the experiment would object images be cued. All cue words, target images and distractor images were distributed equally over each trial type, in each block.

4.4.4. Data analysis

Data were exported to MATLAB and analysed using a univariate cluster-corrected model with $p < 0.05$. ERPs were compared across time and space. Average referencing was used to normalise all waveforms to the mean of all 64 electrodes (at each temporal sample). Each trial was split into eight 50 ms segments, with each segment the result of coherent averaging over that 50 ms period to give a single measure of a phase and amplitude for each trial type, at each electrode. Data were averaged across trials within each observer, and then calculated grand averages and standard errors across observers. The same procedure was used to average signal variances. Eye-movement artefacts were removed in pre-processing.

Figure 4.4. shows the EEG montage with the six highlighted electrodes that were used for univariate analyses (P3, P5, P7, PO3, PO5, PO7).

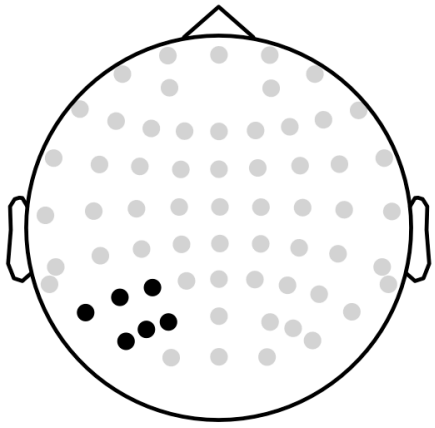

Figure 4.4: Map of electrodes used for univariate analysis. These electrodes are shaded in black.

4.5. Results

4.5.1. Behavioural data results

Participants' behavioural data were analysed using a 2X4 ANOVA, which showed no significant effect of block ($F(1,18) = 0.902$, $p > 0.3$). Data were collapsed across blocks. Data were then compared across trial types, to identify possible differences between them (cued target only present; target absent, in which a non-relevant target were presented; dual target present in which a cued and uncued, though relevant, target were presented superimposed; and dual absent targets, in which an uncued relevant and non-relevant target image were presented superimposed. There was a significant effect of trial type $F(1.638,31.12) = 60.717$, $p < 0.001$), and within-subjects contrasts found that participants performed more accurately in single target present trials (M=92.55%, SE=1.27%) than in any

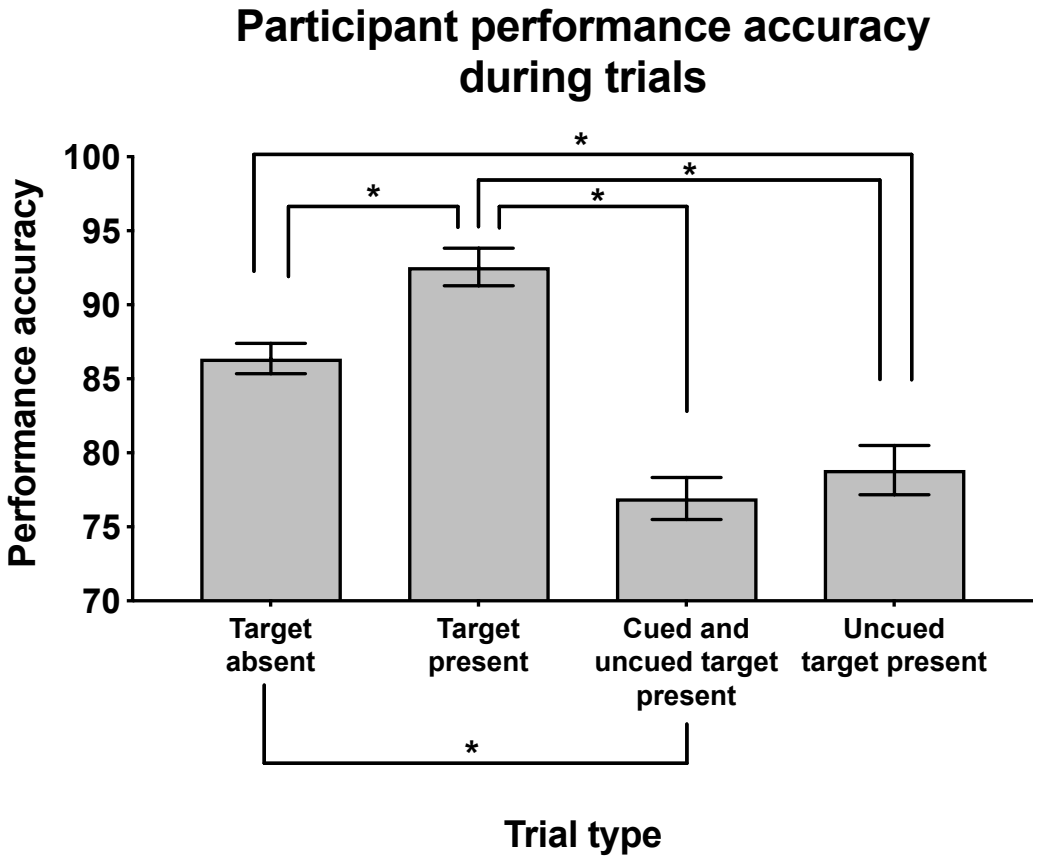

Figure 4.5: Graph of participant performance across the four trial types, with data collapsed across all blocks. Significant differences are indicated by black bars; all asterisks indicate p ≤ 0.001. All error bars represent standard error of the mean.

of the other trial types (dual targets: M=76.92%, SE=1.42%; target absent: M=86.37%, SE=1.03%; uncued target present: M=78.84%, SE=1.66%) ($p \leq 0.001$).

The findings from Evans, Horowitz & Wolfe (2011) were replicated, with participant accuracy during dual-target trial types significantly lower than during single target only present ($F(1,18) = 12.819$, $p < 0.01$), as can be observed in Figure 4.5. Values for significant differences can be found in Table 4.1.

Main effects	**DF**	**F**	***p***
Block	1, 18	0.902	0.355
Trial type	**3, 54**	**31.036**	**< 0.001**
Block * trial type	**2.075, 37.346**	**15.781**	**< 0.001 (Greenhouse-Geisser corrected)**
Trial-type within-subjects contrasts			
Single target vs dual target	**1, 18**	**78.949**	**< 0.001**
Single target vs no target	**1, 18**	**15.767**	**0.001**
Single target vs. uncued target	**1, 18**	**42.71**	**< 0.001**
Trial-type pairwise comparisons			
Dual target vs. no target			**< 0.001**
No target vs. uncued target			**0.001**

Table 4.1: Statistical outcomes for the 2X4 ANOVA. Statistically-significant rows ($p < 0.01$) are indicated in bold.

4.5.2. EEG analysis results

The first analysis for the EEG data intended to identify when a significant difference between target-present trials and target-absent trials, over both blocks, would be found. This would allow a preliminary assessment of gist extraction, and act as a sanity check to ensure target-present and non-target trials were being successfully differentiated.

Topographical data showed an early differentiation between target-present and target-absent trials between 30–100ms post-stimulus onset. The initial response at approximately 30 msec showed activation in the frontal areas, with activation in occipital and parietal areas occurring between 69 and 100 ms. This initial frontal response (30–50 msec) is not seen again until 75 msec after onset (Figure 4.6).

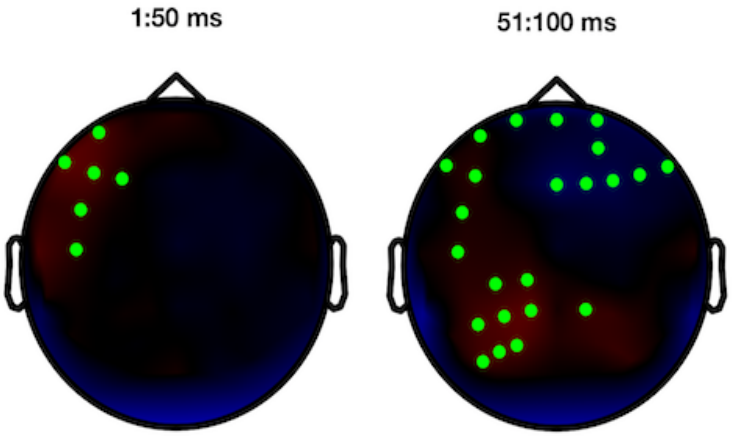

Figure 4.6: Univariate EEG data between 1 and 100ms post-stimulus onset, for target-present only (red) vs. target-absent only (blue) trials. Data are averaged across both blocks 1 and 2.

This can also be seen by comparing the amplitude of the ERP waveforms across the six electrodes used in the analysis. In block 1 (figure 4.7 A), both target present and target absent trial types show an increase in amplitude of approximately 4 µV at 70–100 ms, and again at around 175 ms. By comparing the difference between the amplitudes (figure 4.7 B), target present trials produce a statistically greater µV of between 1 and 2 µV during these time frames.

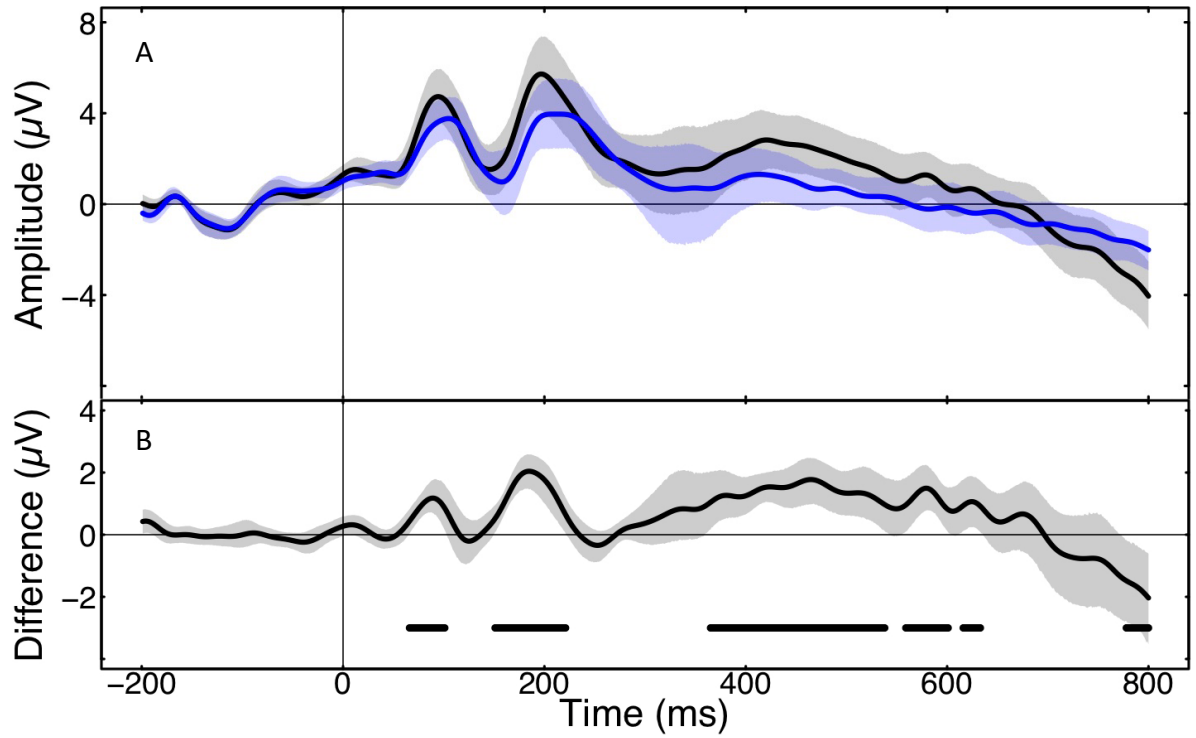

Figure 4.7: Block 1, comparing target present vs. target absent trial types. A) Graph of pairs of ERPs averaged across the six electrodes. Black represents target-present trials, and blue represents target

absent trials. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

In block 2, this differentiation is seen earlier, with target present and target trials both producing positive amplitudes (figure 4.8 A) which nevertheless are significantly differentiated at around 40-60 ms (0.5 µV – figure 4.8 B), and again from 120 ms onwards, with the biggest difference between the two trial types at approximately 2 µV. However, unlike in block 1, the amplitude waves in block 2 show a greater µV for target absent trials than target-present trials, thus leading to a negative difference wave.

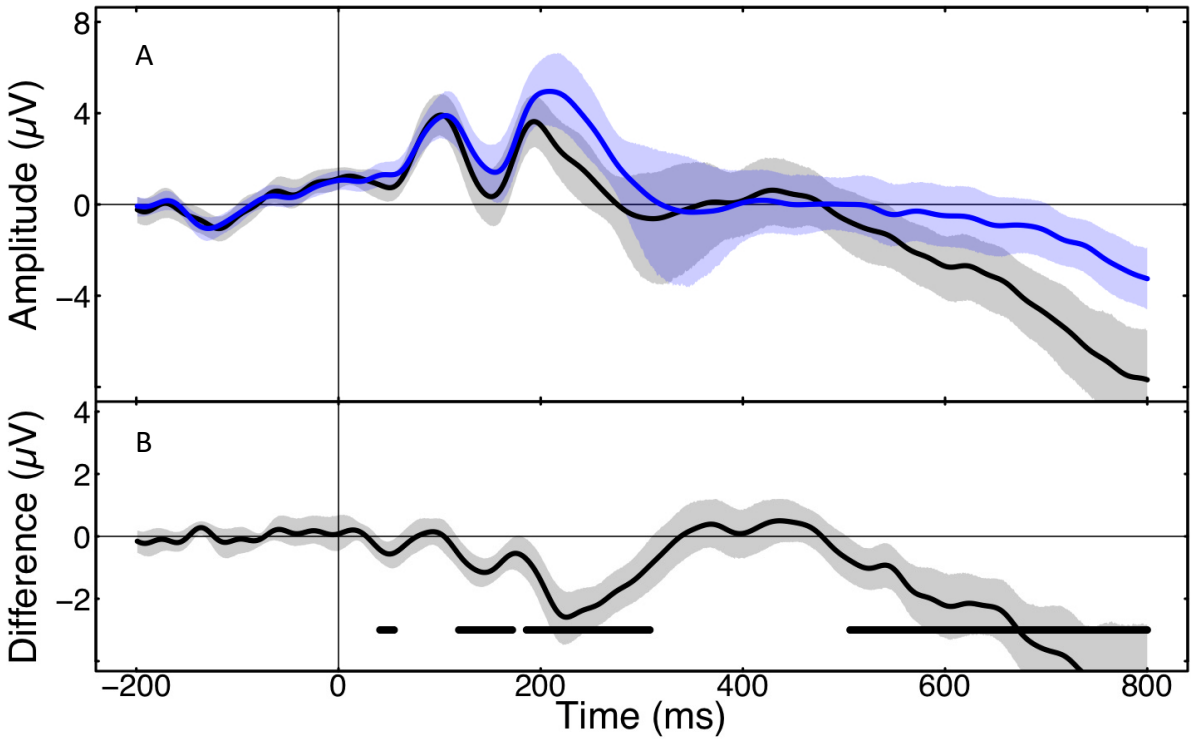

Figure 4.8: Block 2, comparing target present vs. target absent trial types. A) Graph of pairs of ERPs averaged across the six electrodes. Black represents target-present trials, and blue represents target absent trials. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

Following this, the next analysis sought to identify when participants differentiated between different target types. This would indicate a possible time point at category gist was extracted by the brain, and was assessed by comparing the neuronal activity during two

different kinds of task-relevant target trials. Face target-present trials were compared to place target-present trials in block 2, during which both image types were cued targets.

Topographic analysis found that when both image types presented were targets, differentiation happened early; onset of the differentiation occurred at approximately 60 msec, with faces appearing to be processed between 100-150 msec in occipital regions (Figure 4.9). Place images appeared to be processed in frontal and central regions.

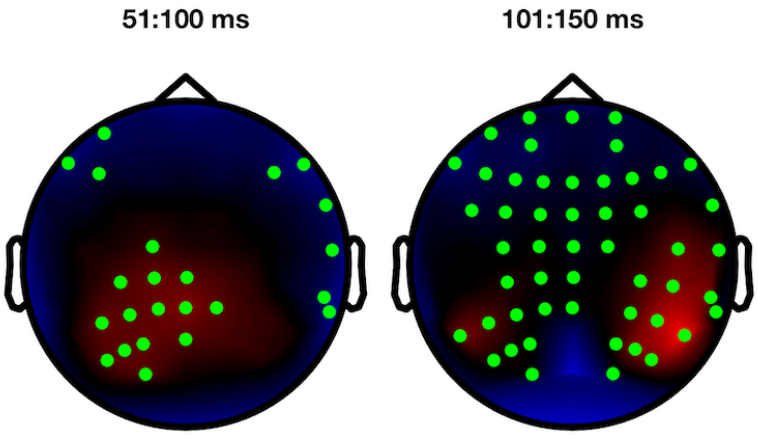

Figure 4.9: Face target-present trials (red) compared to place target-present trials (blue).

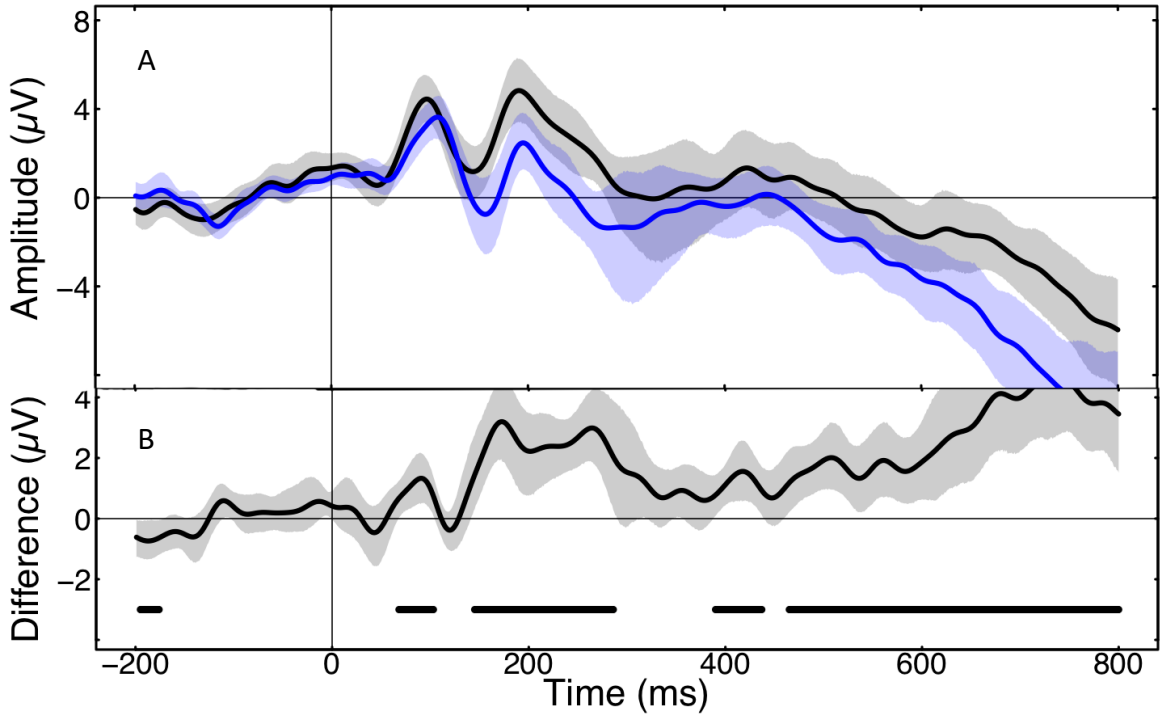

Figure 4.10: Comparing the neuronal response in block 2 to face (black) and place (blue) targets. A) Graph of pairs of ERPs averaged across the six electrodes. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

The ERP data (figure 4.10 A) found that there was a differentiation (figure 4.10 B) between face targets and place targets at approximately 70-110 ms, aligning with the onset of differentiation from the topographic data. In short, when both target types were relevant, differentiation between them happened early.

To complement this finding, neuronal responses to non-relevant images were compared during block 1 only (face and object images). Topographic data revealed onset of

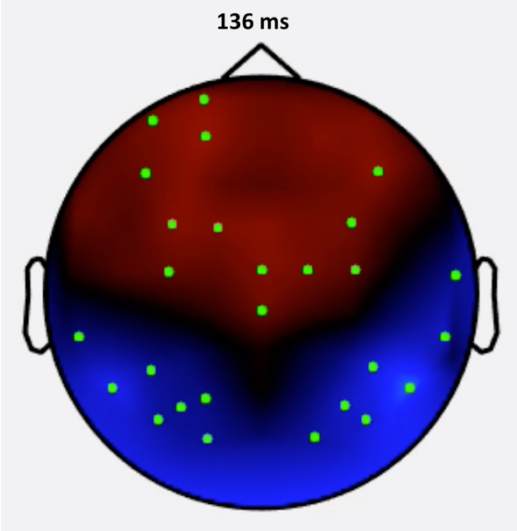

Figure 4.11: Face non-target trials (red) compared to object non-target trials (blue) in block 1. 136 msec time point is chosen to best display the neural activation seen.

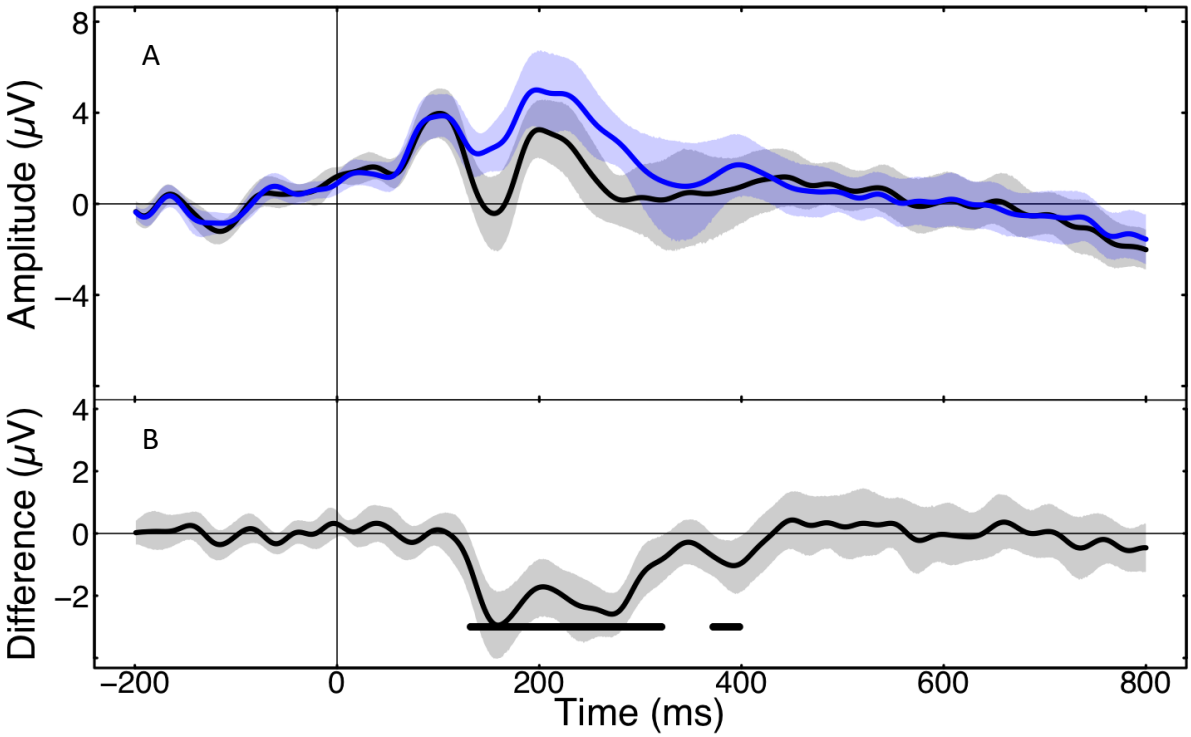

Figure 4.12: Comparing the neuronal response to face (black) and object (blue) images in block 1. A) Graph of pairs of ERPs averaged across the six electrodes. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

differentiation happened at ~130 msec for both face and object images, with faces being processed frontally between 130-140 msec (Figure 4.11). ERP analysis (figure 4.12 A) indicated a similar finding, with objects showing a significantly greater amplitude from 150 ms onwards compared to face images (figure 4.12 B). Compared to the previous analysis, when both target types were non-relevant, differentiation happened later.

By changing the task contingency, destructive interference was found in participants' behavioural responses. In order to find the temporal signature of this event, the neuronal responses in dual-target trials between blocks 1 and 2 were compared. In these trials, the stimuli were the same (both places and faces were presented), but the task was different between blocks – the images went from only one being task-relevant (the pre-cued target only) to both being task-relevant (a cued and uncued target).

In block 1, topographic activation is seen in frontal areas at approximately 65 msec (Figure 4.13). Right frontal activation is then seen as early as 101 msecs, including the left frontal areas between 123–153 ms post-onset. However, in block 2 (Figure 4.14), frontal

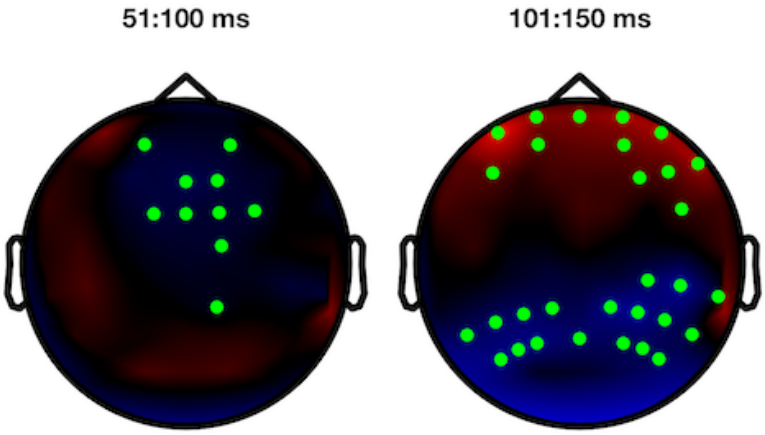

Figure 4.14: Dual-target trials in block 1 (red) vs. block 2 (blue) between 50–150 msec post-stimulus onset.

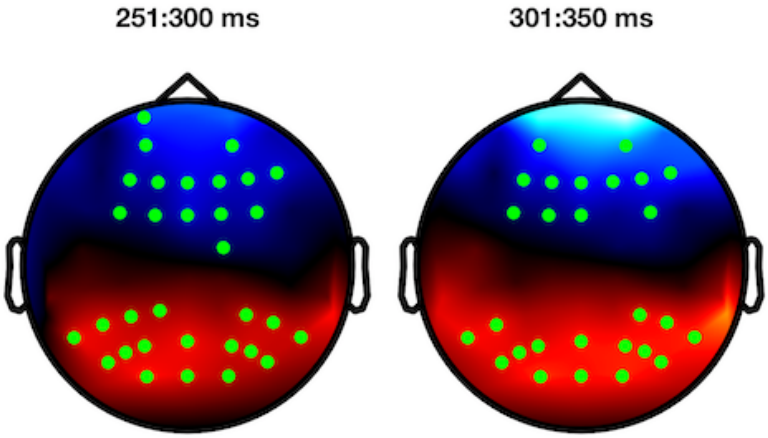

Figure 4.14: Dual-target trials in block 1 (red) and block2 (blue), between 250–350 msec post-stimulus onset.

and central activation is first seen at 222 msec, becoming more consistent at 284 msec and continuing until 364 msec. In particular, there is right frontal activation at the 300 msec mark. Investigating the amplitude data, there are positive-amplitude neuronal responses to dual trials in both block 1 and block 2, at the 100 and 200 ms mark (figure 4.15A), possibly representing the frontal activation seen above. The difference between the amplitude waves is significant at the first peak 120-160 ms, with dual targets in block 2 producing the greater amplitude. This difference becomes more consistent after 200 ms (figure 4.15 B), with dual targets in block 1 producing the greater amplitude. Around 230-300 ms, dual target in block 2 are resulting in significantly less amplitude than in block 1, which may indicate a suppression effect.

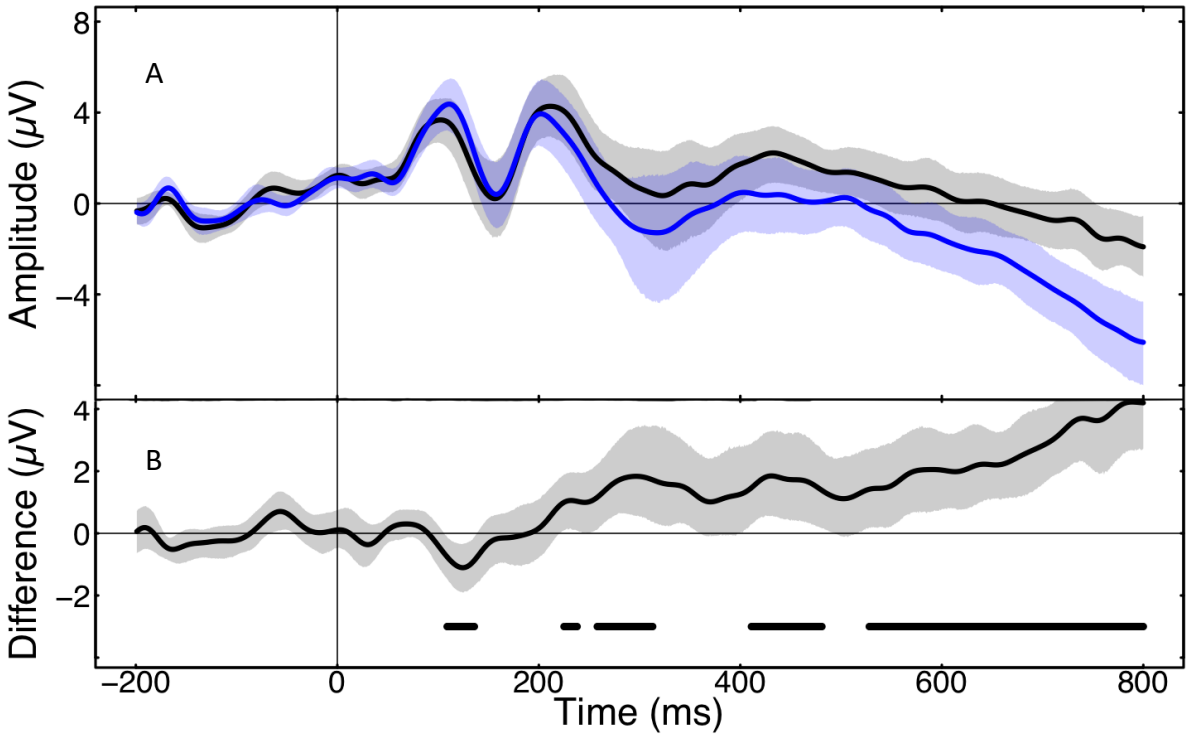

Figure 4.15: Comparing the neuronal response to dual-target trials in block 1 (black) and block 2 (blue). A) Graph of pairs of ERPs averaged across the six electrodes. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

To further investigate this conflict, the next analysis compared target-absent trials in block 1 to block 2. When the target was absent in block 1, participants were being asked about an un-primed target. In block 2, the target was primed, so it needed to be suppressed

differently. Target-absent trials in block 1 (face and object stimuli) showed more activity in the frontal areas for the first 250ms (this may indicate suppression, due to the engagement of frontal cognitive control regions), before becoming more parietal/occipital from 250–500 msec post-stimulus onset (possibly categorisation – see Figure 4.16).

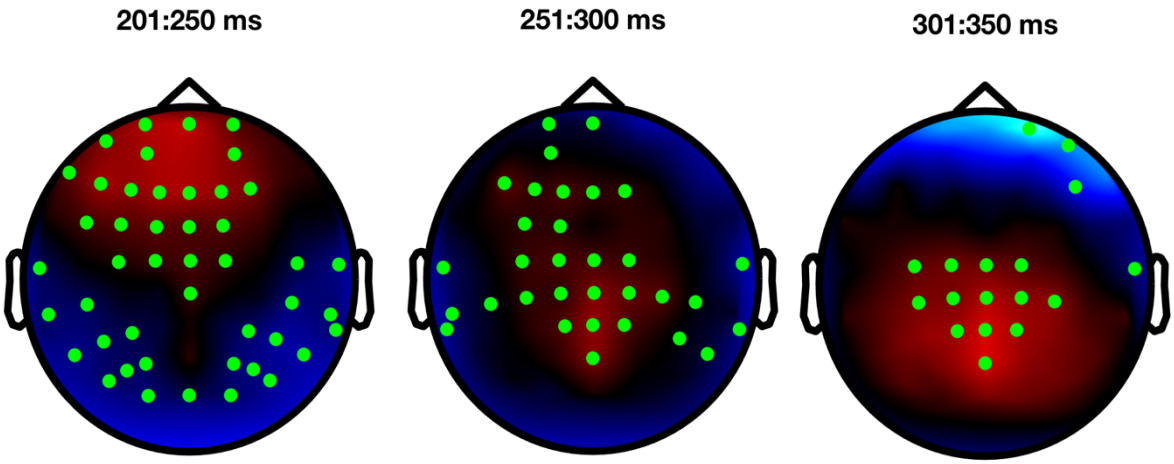

Figure 4.16: Target absent trials in block 1 vs block 2. TA stimuli in block one are never task-relevant; faces become task relevant in block 2. Red = TA in block 1, blue = TA in block 2.

However, target absent trials in block 2 show a different pattern. In this block, when face images have gone from being non-relevant images to task-relevant targets, frontal activation (and possible category suppression) is seen much later, between 300 and 400 msec after stimulus onset (see Figure 4.17). This is similar to the findings of the previous analysis, in which face images have also progressed from non-relevant to relevant targets, and indicates a role in decisional processes in frontal control regions.

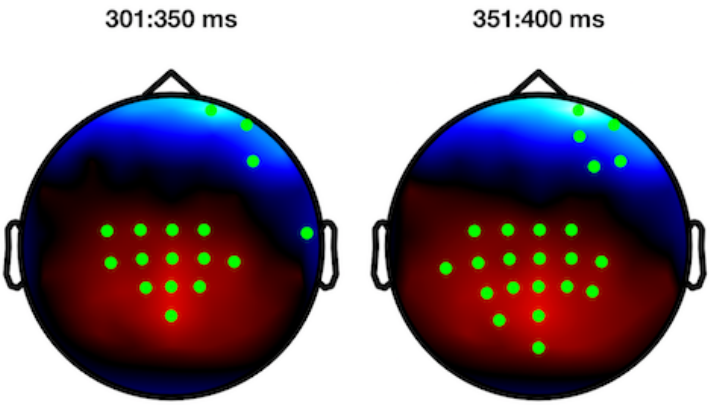

Figure 4.17: Target absent trials in block 1 vs block 2. TA stimuli in block one are never task-relevant; faces become task relevant in block 2. Red = TA in block 1, blue = TA in block 2.

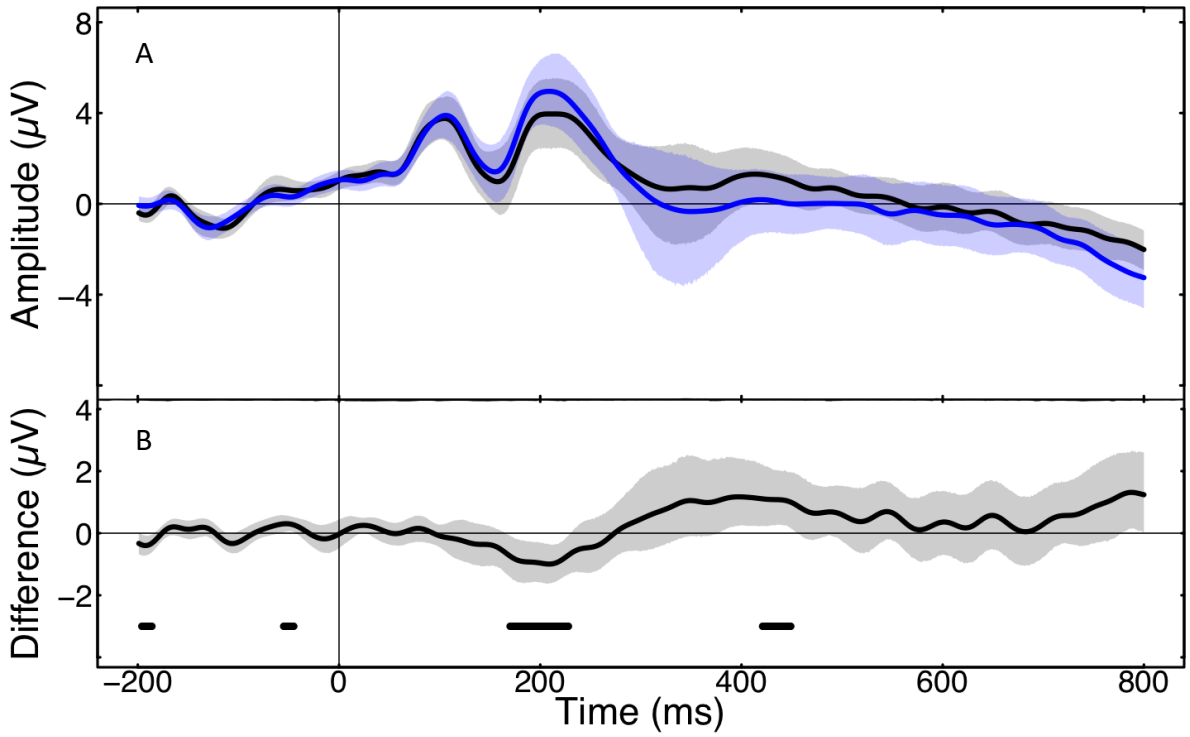

Figure 4.18: Comparing the neuronal response to target-absent trials in block 1 (black) and block 2 (blue). A) Graph of pairs of ERPs averaged across the six electrodes. B) A graph of the difference between the two trial types between the analysis electrodes; black lines near the foot of the graph indicate times that are significant following cluster correction. For both graphs, shaded regions are bootstrapped 95% confidence intervals (10000 resamples across participants).

ERP analyses found that, similar to previous findings, there was a positive amplitude for both blocks peaking at approximately 100 and 200 ms (figure 4.18 A). Significant differentiation between the two target-absent types is shown here around 180–220 ms, with block 2 producing a slightly higher amplitude response than block 1 for target absent trials (figure 4.18 B).

4.6. Discussion

In chapter 4, we aimed to find the time signature (i.e. the temporal dynamics) of gist processing within the brain, and determine when a change in task contingency could be seen. This experiment was designed as an extension to the experiment in chapter 3, so that both the 'where' and 'when' questions with regards to gist processing might be addressed. Univariate statistics were used to identify the magnitudes of activity across the brain during EEG. This allowed for identification of statistically-significant ERPs during the time in which

participants were acquiring the category information, and so presumably gist, of the images presented to them. This allowed for the development of an approximate timeline for the extraction of the target categories, and the point at which destructive interference of gist categories could be temporally observed.

Our key findings were that the gist of a scene category is detected early, with the brain showing differential patterns of activation at around 30 msec after stimulus onset for target-present and target-absent trial conditions. Differentiation of relevant target categories occured earlier (~60 msec) than differentiation between non-relevant images (~130 msec). Upon priming multiple gists and changing the task contingency, conflict was seen between cued and uncued targets between 284–364 msec after stimulus onset. This may indicate that cognitive and decision-making networks are involved in destructive interference, and may involve the P3 component. This evidence will now be discussed in terms of the existing literature.

4.6.1. Accessing gist and differentiating targets

The data presented in chapter 4 found that target-present and target-absent trials could be differentiated as early as 30 msec post-stimulus onset, suggesting that this time point is a neural correlate for the extraction of gist information from a visual scene. This is in line with the data discussed in the introduction from Greene & Oliva (2009a), who found that 30–67 msec were needed for extraction of a scene's basic category. This speed would also suggest that this is the product of the first feed-forward sweep of information through the brain, rather than a top down process.

For this initial frontal activation, it is possible that this is part of the first feed-forward process, initially hypothesising the presence of stimuli. It is found between 30–65 msec post-stimulus onset, with the greatest number of electrode responses seen at 46 msec. This means this is unlikely to represent the C1 component; apart from the component's time signature being delayed in comparison to the findings, this component is also associated with occipital areas, rather than the frontal regions found here.

There is some debate as to the origin of visual evoked potentials (VEPs) seen within this time range. Pratt (2012) discussed research studying early ERPs in vision, observing that there were two main arguments: that VEPs found between 40–70 msecs originated in the

optic nerves, or alternatively that components found between 35–70 msecs were of occipital origin. The findings from this chapter instead found a consistent frontal activation between 30–65 msec. One hypothesis is that this represents the first point at which the first-forward sweep activates frontal areas, and carries a 'hypothesis' of what is being observed (e.g. a relevant target). This in turn would allow some top-down control over subsequent activation, thus allowing for certain patterns of activation for (for example) relevant targets compared to non-relevant images. This time window may mark the brain's earliest guess at identifying the presented image, given partially-accumulated information.

The activation seem in occipital areas around 100 msec may indicate the first indication of target differentiation. As Grill-Spector & Kanwisher (2005) and Hegdé (2008) noted, target detection and categorisation happens concurrently; as soon as a target is detected, it is identified. The data for this observation combines observers' neuronal responses to both place and face targets over both block types, and so is reflective of always-relevant targets (places) and targets which become relevant later in the experiment (faces).

To break down this distinction, Figures 4.9, 4.10, 4.11 and 4.12 indicate the neuronal response to relevant targets and non-relevant images. Relevant targets are identified earlier at 60 msec (in particular, face targets were processed between 56–126 msec in occipito-parietal areas consistent with the location of left face-processing regions found during gist processing in chapter 3). Conversely, non-relevant images are identified later, at 130 msec. These analyses both looked at two targets/images with identical relevancy (block 2 places and faces as relevant targets, and block 1 faces and objects as non-relevant images). The stimuli designated as 'relevant' appear to be prioritised, with neural responses modulated by that priority. This allowed for flexibility of processing.

These findings are, in general, earlier than the predicted ERPs. Categorisation and discrimination of relevant targets (~60 msec) – and even non-relevant images (~130 msec) – appears to occur before the predicted 150 ms. The main difference between the findings in the literature and the findings reported in chapter 4 is that the majority of the previous literature utilises natural scene images, rather than the artificial stimuli used in this experiment (Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2001; Rousselet, Fabre-Thorpe & Thorpe, 2002; Codispoti et al., 2006; Hegdé, 2008). In the case of faces and object targets, whilst natural scene images require an observer to identify a target amongst a

cluttered visual environment (Battistoni et al., 2018), research suggests the type of background does not affect speed of processing (Johnson & Olshausen, 2005). In fact, natural images are predicted to be processed faster and without requiring attentional resources compared to artificial stimuli (Kayser, Körding, & König, 2004). Why then are the findings in chapter 4 seen earlier than the predicted time points?

The answer may lie in the cueing of target categories, and in particular the prioritisation system observers use because of it. The 'priority' of a given target has shown to affect neuronal processing; goal-directed, top-down influences (such as a cue) are sufficient to influence the speed and amplitude of neural processing, acting in a modulatory capacity (Gazzaley et al., 2007; Rutman, Clapp, Chadick, & Gazzaley, 2010). Delorme, Rousselet, Macé, & Fabre-Thorpe (2004) observed that, by pre-cuing observers with a photograph of a target, observes could then utilise low-level cues in the image to increase behavioural performance and reaction time. They also found an decreased latency for onset of neuronal activity. Cues provide a 'template' for the brain to match a potential target to, and targets which meet this 'template' are then prioritised.

In particular, in describing how activity may be seen so early during the processing of overlapped visual stimuli (of particular relevance to this chapter), Rutman et al. (2010) noted that a cue could pre-activate the relevant cortical areas in anticipation of a target type. This would enhance efficiency of subsequent neural processing – if face areas were pre-activated during a 'face' cue, and then an object image were shown, the image might not be processed using the pre-activated networks, and thus would experience a relative delay in processing.

This is seen in the data within this chapter for face images. When a face image is cued and then presented, faces were processed beginning approximately 100 msec in occipital areas. However, when faces are *not* cued and then presented, they are processed beginning 130 msec in frontal regions. This delay may be reflective of the anticipatory gain modulation introduced into processing which pre-activates areas relevant to the cue ("place") and not the presented imaged (face), and may also explain the faster-than-predicted correlates for visual processing seen within the data.

4.6.2. Changing task contingency

In chapter 3, further evidence for behavioural evidence of destructive interference (Evans et al., 2011) was found, seen in the form of reduced performance accuracy on dual-target trials compared to single-target trials. Possible neuronal locations included cognitive control areas such as the dorsolateral prefrontal cortex. The findings from chapter 4 once more replicated these results, finding a statistically-significant drop in performance accuracy for dual trials (cued and uncued target present) compared to single trials (cued target present only) (15.63%). In order to identify this phenomenon in the neural data, the EEG data for dual-target trials were compared between blocks 1 and 2. This allowed the comparison of identical task stimuli with the only difference being a change in task contingency between the blocks.

Frontal activation was found for dual-target trials in block 2 between 284–364 msec, with a corresponding difference between the ERPs. However unlike the predicted P3 component, thie amplitude of this finding was negative. Rather than P3, this may represent the N3. Hamm, Johnson, & Kirk (2002) discussed a negative component named the N300 (or d-N300), as a component that is specific to the processing of the semantic properties of picture-based stimuli, marking the differentiation between related and unrelated image pairs. Hamm et al. note that they only found the N300 during between-category mismatches, suggesting strongly that the possible presence of the N300 component is an artefact of the mismatch between cue word and presented image. This may mean that the N300 is part of the process by which a category image is processed prior to identification, contributing to the 'weighted sum' of evidence.

What is also important to note is that this is the time point recognised by Philiastides & Sajda (2006), Philiastides, Ratcliffe & Sadja (2006) and Ratcliffe, Philiastides, & Sajda (2009) as the 'late' attentional component, reflecting the difficulty of the task as well as its decisional nature. They discuss this in terms of the diffusion model, with the participant taking longer to accumulate target vs. non-target evidence until a threshold is reached (as described by the 'weighted sum' of evidence in Figure 4.1).

This notion of additional resources required for resolving conflict is also seen during the target-absent comparison. By comparing target absent trials across blocks, 'target absent' trials (images which were once non-relevant, and are now relevant) in block 2 appear to generate neural signals with approximately a 50–150 ms delay compared to block 1.

This delay is consistent with the above findings of Philiastides, Ratcliffe & Sadja (2006), who identified a delay in the processing of more 'difficult' target images. Philiastides et al. found that there was a significant correlation between the strength of this late component (300 msec) and the drift rates computed within their diffusion model, as well as behavioural accuracy, suggesting that the component is analogous to mean drift rate. The 300 msec marker represents evidence being fed into the diffusion model, which then makes a decision about the stimuli. This process produces a higher performance accuracy correlated to the strength of the 300 msec signal. This is also not an all-purpose decisional mechanism, as it is not found during 'easy' red/green discrimination tasks, and is instead linked to resolving visual representations within a diffusion model.

In light of this research, this 300 msec component appears to represent a decisional process which indicates an attempt by a diffusion model-based system to resolve the presence of 'difficult' target images, either due to the dual presentation or the neuronal management of a now-relevant target. This late component is also seen in frontal areas, which may indicate suppression of non-relevant target categories as the frontal cortex is associated with executive function including inhibition of response (Burgess & Shallice, 1996) and suppression of responses to visual stimuli (Aron, Robbins, & Poldrack, 2014).

Additionally, the frontal N2 component (180 to 325 msec post stimulus onset – Patel & Azzam, 2005) is associated with response-suppression in task-switching (Swainson et al., 2003). The correlates for destructive interference here would fall within the range for this component, and is also both temporally and spatially related to executive control processes. Whilst this experiment did not utilise task-switching in the strictest sense (participants were not asked to attend to two different tasks), participants had to 'switch' their motor set when a relevant stimulus was presented or not presented. Participants also had to operate under the assumption of a task-switching paradigm, in which trials were apparently random, including that each trial may potentially contain the cued stimulus. This frontal activation therefore is potentially a correlate for executive decision-making during what participants perceive as a go/no-go task.

4.6.3. Other findings, and addressing the study's limitations

Additional analyses compared single target trials to dual target trials in block 1 and block 2. Block 1 found differences between single (left occipital regions) and dual (right frontal regions) target types at 60 msec post-onset, though it was unclear if this differentiation related to target type or visual load.

For block 2, trial types were distinguished earlier (40 msec). Single target trials showed more frontal activation throughout the time windows; this aligned with the fMRI data, which indicated greater % BOLD signal change in cognitive control areas for single targets compared to dual targets. There was also a bright spot at electrode F3, though not statistically significant, which is associated with the DLPFC. This is not unquestionable support for the neural correlates of destructive interference identified in chapter 3, but it further indicates a trend. Whilst no definitive statements can be made, the fact that two separate experiments utilising similar (though not identical) stimuli to ask similar questions found activation in frontal regions associated with cognitive control indicates a trend. An experiment designed to further load this process and attempt to more closely monitor the neural response during destructive collision may help to confirm these findings, and such an experiment will be discussed in chapter 5.

There are also limitations specific to univariate analyses; in particular, whilst investigating the temporal dynamics of destructive interference, the univariate data suggested that block 2 dual-target trials showed a frontal pattern of activation from 260–360 msec post-stimulus onset. Whilst this certainly seems to support the hypothesis of a decision-associated region being associated with destructive interference, care must be taken not to over-interpret it. The frontal activation is very strong, and there is a corresponding degree of occipital activation for the block 1 dual-target trials. EEG univariate analysis is limited in the source localisation that can be achieved, and it is possible that these data reflect an echo of the electrodes, especially considering the strong oppositional pattern found. Whilst there may be a clear, legitimate source in frontal areas, this anticorrelation may also indicate an echo.

To further investigate , a natural-scene-based multivariate experiment was designed and implemented in order to find if patterns of activation across the whole brain demonstrated a similar time signature to the ERP found in this chapter. That experiment is the basis for chapter 5.

4.7. Conclusions

The experiment in chapter 4 aimed to identify time signatures of several different aspects of gist processing, include gist processing onset, and the neuronal correlates of destructive interference between gists. Predicted ERPs for these events included a potential 150 msec component, the visual N1/P2, N2, and a 300 ms component that may represent N3. The experiment used a rapid 2AFC design, utilising stimuli (place, face and object images) that could appear solo or superimposed over another image.

The onset of gist processing was seen beginning at 30 msec post-stimulus onset, with relevant targets (all place and some face images) differentiated at 60 msec, and non-relevant images (some face and all object images) differentiated beginning 130 msec. This was earlier than the predicted time points would suggest. Destructive collision was identified at a time point and amplitude associated with the N3 component.

Chapter 5 Multivariate pattern analysis of destructive interference in gist processing

5.1. Overview

Gist is extracted quickly, with initial gist accessed around 30 msec, and differentiation of category images happening between 60–130 msec. These findings have helped build a timeline of gist extraction in the brain, and provide some initial insight into the neuronal time-signature of destructive interference.

Chapter 5 intended to extend the experiment in chapter 4 by further investigating the nature of the destructive interference using multivariate pattern analysis, or MVPA. MVPA allows patterns of information to be 'decoded' across the whole brain, and offers a higher degree of sensitivity for subtle responses that may not be apparent at any single scalp electrode. Using MVPA, when is gist processing detected? Will destructive interference be seen in the same manner for natural scenes as for the artificial stimuli used in chapter 4? Can we replicate the findings of chapter 4 in observing the point in time at which destructive interference is observed?

Chapter 5 begins by discussing firstly the advantages of MVPA and the new factors introduced by using natural scenes. Further literature from EEG studies of the diffusion model of visual processing and decision-making is also discussed. This is then used to develop hypotheses as to which time points in a multivariate difference wave will be identified during an event-related EEG experiment, and how these relate to particular cognitive events. The hypothesis is that gist will, as in chapter 4, be detected early, and that destructive interference will be detected later as a decisional component.

We observed that gist extraction appeared to happen quickly (50–75 msec post-stimulus onset), with category information differentiated for relevant targets at 200 msec. Dual targets were differentiated around 100 msec, and destructive interference was seen in responses to dual-target trials at approximately 300 msec.

5.2. Background

Gist characteristics are extracted quickly as part of a feed-forward, perceptual process. However, if the task contingency is changed, destructive interference between gists is seen

neuronally at around 300 msec after stimulus presentation, indicating a decisional process. This chapter aims to differ from the experiment in chapter 4 in three main ways. Firstly, the experiment will be analysed using MVPA; secondly, the stimuli are natural scenes; and thirdly, the natural scenes contain single or multiple categories (e.g. an animal on a beach, rather than superimposing examples of 'animal' and 'beach' together.

5.2.1. MVPA analysis and natural scene stimuli

MVPA is an alternative to univariate statistics in both fMRI and M/EEG data (Grootswagers, Wardle, & Carlson, 2017), which assesses the relationship between all electrodes, rather than looking for specific ERP signatures (such as peak amplitude, mean amplitude, and onset latency of signals detected at specific electrodes).

MVPA analysis has several advantages over univariate analysis. Because MVPA considers correlations and covariance of voxel activation over the whole scalp, rather than using a voxel-by-voxel approach traditionally associated with univariate analysis, this allows for greater sensitivity by allowing the identification of weak but otherwise consistent differences between experimental conditions (Habeck, 2010; Oosterhof, Connolly, & Haxby, 2016). MVPA also allows for inferences to be made, both within and across individuals, about underlying neural representations (Habeck, 2010). MVPA would allow us to identify statistically significant patterns of activation even when there is no specific time point (event-related potential, or ERP) known for a particular phenomenon.

Because of its advantages, MVPA also allows for the analysis of more subtle differentiations between stimuli, which may be found as a result of using natural scene images. One of the advantages of the artificial nature of the stimuli used in chapter 4 was a high degree of control. The stimuli were equally processed to eliminate edge contours (Gaussian filter), to be of identical size, and were always presented in the centre of the visual field. More generally, artificial stimuli are very stripped down, allowing for the testing of hypotheses without irrelevant or confounding variables (such as colour) (Orhan & Jacobs, 2014).

In chapters 3 and 4, participants were asked to categorise isolated images on texture images. These images were presented without a background or other real-world context, and dual-target trials, in which two target categories were presented, were done so with

images made transparent and layered over the top of one another. The images were undeniably both visible, but it was not a naturalistic setting, and thus were not reflective or representative of the wider visual environment.

The use of artificial stimuli allowed for the neural correlates and temporal dynamics of gist to be explored in a controlled manner. Whilst this allowed us an understanding of when and where gist processing could be seen to happen in the brain, we cannot make specific inferences about the neural or temporal dynamics of gist extraction in natural scenes. Kayser, Körding, & König (2004) noted that assessing the abilities of the human visual system using artificial stimuli often resulted in the underestimation of visual performance.

Importantly, Kayser et al. note that the processing of natural stimuli can be done fast and in parallel, which suggests that the extraction of the gist of the scene may be seen to happen faster for natural scenes than for artificial ones. They also observed that classification of stimuli in natural scenes can be done in the absence of attention, whereas attentional resources may be recruited for classification of 'typical' laboratory, or artificial, stimuli. This implies that the time signature of gist extraction for natural scenes may happen earlier, and with a different neural pattern, than when seen for artificial stimuli.

Like with artificial stimuli, rapidly categorising targets within a natural scene requires the ability to extract the gist of the scene at several levels. However, this extraction requires the brain to identify relevant and non-relevant features within a cluttered visual environment. The location of target and non-target 'distractor' objects can be decoded accurately from MEG data (using multivariate analysis) at 50 msec post-stimulus onset, and a pre-cued target category could be decoded by 180 msec (Battistoni et al., 2018). Further MEG multivariate analysis also indicated that observers can detect pre-cued categories such as cars or people in as little as 160 msec, in spite of other visual clutter and variation within the category target, and only objects relevant to the attentional set (i.e. the relevant target) were processed within the first 200 msec (Kaiser, Oosterhof, & Peelen, 2016).

Battistoni et al. refer to the early 50 msec signal as being part of a stimulus-driven process, which is within the time-frame proposed by VanRullen and Thorpe (2001) and Scholte, Jolij, Fahrenfort, & Lamme (2008) as part of the first feed-forward sweep. Battistoni et al. also identified better decoding of the target location than the distractor location at

240 msec, and argued that both the 180 and 240 msec time-points were indicative of top-down control.

5.2.2. The diffusion-based model of gist and destructive interference for multivariate analysis

As well as replicating the multivariate temporal correlates of 'standard' gist processing, of interest is also the relationship between gists, and the brain's response to gist when the task contingency has been changed. In particular, the time-signature for destructive interference between natural-scene based categories, as decoded by MVPA, is of interest.

Whilst evidence may accrue quickly, the introduction of competing or conflicting sources of information may delay the process. The 'weighted sum' of evidence in Figure 4.1 reflects the conflict of available evidence, and which may be put in terms of task difficulty. Using a diffusion-based decision model, Philiastides, Ratcliffe, & Sadja (2006) specifically looked at perceptual-decision-making. Participants were asked to differentiate either images of faces and cars or the overall colour of an image in a pre-cued 2AFC design. Images were presented for 30 msec, which (Figure 4.1, dashed green line) represents a cut-off point before certainty can reached within the model.

Philiastides et al. found that the colour task (red/green differentiation) represented an 'easy' task, whereas the object task (car/face differentiation) was a 'hard' task. The analysis found a component 170 msec after stimulus onset during the easy task (N170 – Rossion, Joyce, Cottrell, & Tarr, 2003; Philiastides & Sajda, 2006), representing a perceptual process; however, the hard task showed a component at approximately 220 msec after stimulus onset, which they argue represented a decisional process (D220 – 'D' standing for 'difficulty'; onset between 170 and 300 msec). This component had an amplitude correlated to the difficulty of the task.

Philiastides et al. argued that this D220 component was indicative of a top-down decisional process, as it could not be found during simple red-green colour discrimination tasks; the authors also made the argument that D220 is more closely linked to the 300 msec component than N170, due to D220's tendency for its amplitude to reflect the difficulty of the task (as seen in their behavioural findings). Philiastides, Ratcliffe & Sadja suggest that this relationship between amplitude and difficulty reflects how relevant attentional

resources are recruited to make a decision about the presented stimulus. This is supported by later work by Ratcliffe, Philiastides, & Sajda (2009), who found that the later component represented decision-relevant evidence feeding into the diffusion model, in post-sensory processing.

Using this evidence, a diffusion model can be used to contextualise the evidence so far provided into studies of destructive interference (Evans et al., 2011). It provides a timing framework for the relationship between two conflicting pieces of information in visual attention, with a direct supporting framework from neuroimaging and behavioural research even beyond the existing theoretical discussion into diffusion models. By manipulating what a participant sees, we can observe when the conflict between two presented, task-relevant objects is seen in the brain in terms of both the perceptual and decisional processes that accompany it.

5.3. Aims and hypotheses

The literature discussed has used various analysis methods to demonstrate multiple time points at which gist, or conflict between gists, can be said to be differentiated. None of them are conclusive, especially as many of these studies did not specifically set out to measure the time signature of gist processing. However, it provides a timeline that might allow us to predict when certain aspects of gist can be decoded by a multivariate classifier, especially when the findings from chapter 4 are considered. Due to the brain's ability to resolve conflict between targets in a natural scene (Kaiser et al., 2016); the generalisability of results between experiments using natural and artificial stimuli (Peelen & Kastner, 2014); and the generalisability of findings between univariate and multivariate analyses (Battistoni et al., 2018) it was logical to assume a similar timeline for gist processing would be seen in this experiment as in chapter 4.

The aim of chapter 5 was to attempt to replicate the initial findings of chapter 4, in finding the initial neuronal time signature for gist processing. The use of natural scene images was intended to reflect a more realistic, real-world utilisation of gist processing. The use of MVPA would allow any of the more subtle distinctions (by way of patterns presented over the whole brain) to be found, if any, particularly any arising from the use of natural scene images.

As in chapter 4, the first main question asked: when is gist processing seen, and when are targets differentiated? This involved looking at the contrast of target-present trials compared to target-absent trials. An early time signature for this differentiation (i.e. extraction of early gist) was predicted (30–50 msec), based on the findings from chapter 4 and the discussed literature. Initial differentiation between target types was predicted to be approximately 170 msec (comparing trials containing relevant, cued targets against trials containing relevant, uncued targets), in line with the findings of Philiastides, Ratcliffe, & Sadja (2006).

Secondly, when is destructive interference seen neuronally? As in chapter 4, the hypothesis was that destructive interference is decisional rather than perceptual, and a neural pattern variation analogous to the P3 component is expected to be found to reflect this. Additionally, does the pattern of response to different stimuli change based on the task? For this question, the D220 component would be logically expected; its relationship to both task difficulty and the P3 component suggest that it will be found during the harder (dual-target) trials in block 2 and 3, when two task-relevant targets must be processed to determine which is trial-relevant.

The aim of this chapter was to more concisely determine the temporal dynamics of gist processing, using a methodology explicitly designed to load this process. Chapter 3 investigated the neural correlates of gist processing, and the results from chapter 5 were intended to complement and extend the findings of chapter 4, allowing a greater and more cohesive picture of gist processing through natural scene images and multivariate analysis. Despite specific time points being hypothesized for certain processing events, the analysis was exploratory in order to identify as many relevant significant time points for gist processing as possible. This would allow patterns of activation to be considered in the absence of known ERPs.

5.4. Methods

5.4.1. Participants

All participants were volunteers recruited using the University of York Psychology participation website (PEEBS), and were offered either payment or participation credit as part of their undergraduate Psychology studies. All participants had normal or corrected-to-normal vision. Ethical approval was granted by the University of York Psychology Ethics Committee, and all data were collected prior to analysis. A total of 25 participants were recruited, of which 3 data sets were not included in the final analysis due to technical issues (faulty electrodes, EEG triggers not recorded, etc.). Of the remaining 22 participants, the average age was 21 years and 10 months, 11 were female, 21 were right handed (1 ambidextrous), and 15 were paid.

5.4.2. Apparatus and stimuli

5.4.2.1. Apparatus

This experiment was designed and displayed using MATLAB R2014a (Mathworks, MA, USA) and the PsychToolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). Stimuli were presented using a Mac Pro computer (6-core Intel Xeon E5, 3.5GHz CPU) running Mac OS X (10.9.5), and displayed on a VPixx 3D Lite monitor with a 120Hz refresh rate at an approximate viewing distance of 60cm.

EEG data were collected using ASAlab (version 4.9.2) software (ANT Neuro, Netherlands), with a 1000 Hz sampling rate and a high-speed 64-channel amplifier, on a HP 2230 SFF computer (core i7, 3.4 GHz) running Windows 7 Professional. The EEG caps used had a 64-channel layout according to the 10/20 system (WaveGuard original, ANT Neuro, Netherlands), and were sized individually to participant's skull circumference measurement. The vertical electrooculogram was recorded using two self-adhesive electrodes, positioned above the left eyebrow and the top of the left cheek. EEG trigger information was sent and recorded using PsychToolbox. Reference electrodes used with the left and right mastoids (behind the ear). The ground electrode was Fz on the mid-line saggital plane.

5.4.2.2. Stimuli

Stimuli were colour images, representing either a) one or two of six possible categories, which were scenes containing the category image (see Figure 5.1 for examples) or b) non-category filer images (such as forests, cities, etc.). The six pre-determined categories were animals, humans, bridges, mountains, beaches, and vehicles with a total of 1080 images. The number of images for each category were broken down into three groups: for example, in the beach category, there were 60 images in which the category was presented on its own (i.e., pictures of beaches); 60 images in which the category *and* a secondary category were represented in the same image (such as pictures of animals on a beach), and 60 images of the other 5 categories (humans, animals, vehicles, bridges and mountains), equally distributed. This would later allow different trial types to be arranged. There were a total of 270 filler images used.

Figure 5.1: The six categories of images used in the experiment. These constituted images of animals, humans, vehicles, bridges, mountains and beaches. These category images were presented as scenes – for example, the monkey image is that of the animal in its natural habitat, rather than an image of the animal against a black background.

All images were presented in a square and located centrally on the screen with an approximate visual angle of 13 degrees, and were presented on a light grey background (RGB colour space = 192, 192, 192) which filled the screen.

Trial type examples are shown in Figure 5.2. There were four possible trial types, equally distributed so that each trial type represented 25% of all trials. 'Target present' trials contained the cued image (Figure 5.2 A); 'total target absent' trials contained a non-task-relevant 'filler' image (Figure 5.2 B); 'dual target' trials contained an image that represented both the cued and an uncued, task-relevant target image (Figure 5.2 C); and 'cued target absent' trials contained both an uncued, task-relevant target as well as a non-task-relevant filler image (Figure 5.2 D). Due to the equal distribution of these trial types, this meant that 50% of trials were a 'match' for the given cue ('target present' and 'dual target' trial types), and 50% represented a 'non-match' for the cue ('target absent' and 'uncued target' trial types).

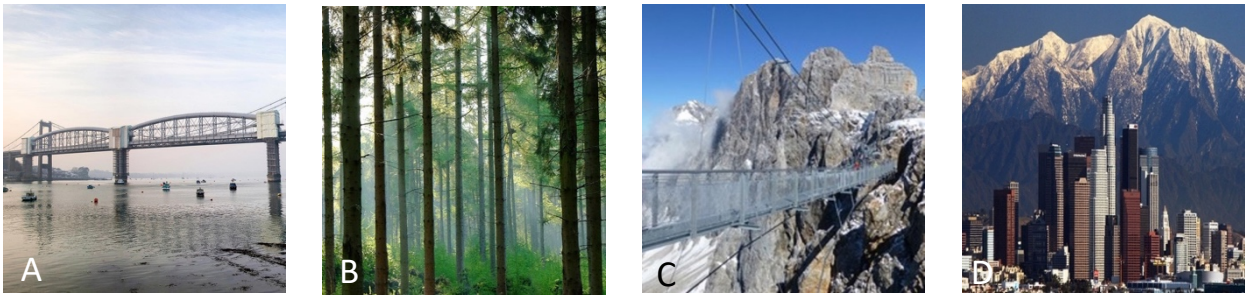

Figure 5.2: Examples of all four trial types. For the purposes of this example, the cue that has been given is 'bridge'. A is a target present only trial, in which the only category given is the one that matches he cue word. B is a target absent trial, in which the image category does not match any of the six possible target categories. C is a dual target trial, in which the cued category (bridge) is present, along with an uncued category (mountain). D is an uncued target present trial, in which the cued target (bridge) is not present, but an uncued image category is present (mountain) as well as a 'filler' image category (city).

5.4.3. Procedure

Participants were asked to take part in a rapid event-related experiment, in which they would complete 3 blocks of 360 trials each. Their task was to respond to presented images by determining if they 'matched' a cue word given at the beginning of the trial.
In each trial, participants were given a cue word relating to one of the six categories (800 ms), a fixation cross to indicate where the participant should focus their gaze (300 ms), and then a category image (25 ms), as in Figure 5.3. Participants would then make their response using the mouse, with the left mouse button indicating a match, and the right button a non-match. The mouse was operated by the participant's dominant hand only.

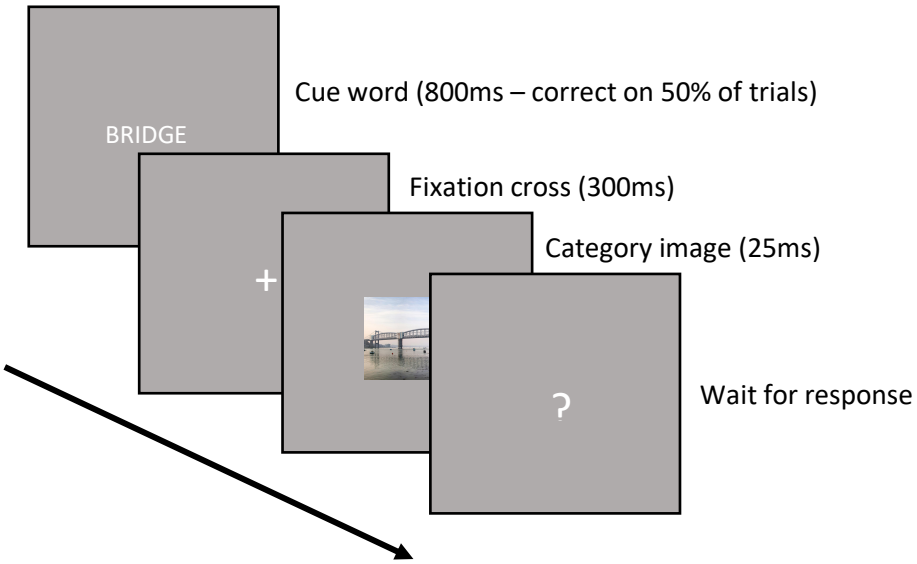

Figure 5.3: Cartoon of the trial structure. This example is of a trial in which the target is present, with no other category or filler images.

During block 1, the code randomly selected 3 of the target categories (for example – human, beach and mountain), and used only these categories as cue words for the 360 trials. The other 3 categories (in this example, animal, bridge and vehicle) became "silent", in that these category images appeared throughout the trials, but would not yet be recognised by the participant as uncued targets. During blocks 2 and 3, participants are cued for all 6 possible categories, and the "silent" categories became task-relevant.

5.4.4. Data analysis

EEG data were exported to MATLAB and analysed offline. EEG pre-processing was performed, which included bandpass-filtering the full time-series data at each electrode, with a low cut of 0.1 Hz (to remove gradual drifts in voltage) and a high cut of 30 Hz (to remove artefacts such as the mains electricity). Eye-movement artefacts were also removed in pre-processing. ERP normalisation was achieved by subtracting the baseline activity (mean voltage data from the 200 msec pre-stimulus onset, at each electrode) from the electrode data at every time point. Normalisation was performed across all electrodes and experimental conditions within the three separate blocks. No trials were rejected from the analysis. This produced a matrix of data that could then be analysed using an MVPA classifier.

EEG MVPA analysis was performed in MATLAB, using the inbuilt support vector machine binary classifier. The classifier was trained to discriminate between the spatial and temporal patterns of EEG responses at each time point and across electrodes, for each participant. It was trained using averages of random subsets of trials (90 per condition, split into blocks of 30). For each time point, the classifier was trained on 2 averages of 30 trials for each condition, and then tested independently at each time point. This process was done with 100 repetitions, using different combinations of trials each time. These data were then averaged across participants, with 95% confidence intervals were generated using bootstrapped resampling. Non-parametric cluster correction was performed (Maris & Oostenveld, 2007) to correct for multiple comparisons, and to determine significant clusters across both time and electrode locations.

The three comparisons used for the MVPA were as follows: target present vs. target absent; target present vs. dual target; and target present vs. uncued target present.

Behavioural data for blocks 2 and 3 were collapsed, as the task was the same for these blocks. All behavioural data were analysed using a 2X4 repeated-measures ANOVA, in which the factors were block type (block 1, and the collapsed blocks 2 and 3) and trial type (target absent, target present, dual target, and uncued target). The comparisons of interest were chosen to match those for the MVPA data.

5.5. Results

5.5.1. Behavioural data

Participants' behavioural data were analysed using the 2X4 ANOVA, which revealed no significant effect of block ($F(1,19) = 0.039$, $p > 0.8$). This allowed data to be collapsed across all blocks. Next, data were compared across trial types, in order to see if there were any behavioural differences within the contrasts specified for the MVPA. The ANOVA indicated there was a significant effect of trial type ($F(1.638,31.12) = 60.717$, $p < 0.001$). Within-subject contrasts found participants performed less accurately during target present trials (M=89.4%, SE=1.1%) than target absent trials (M=92.5%, SE=0.8%) trials ($F(1,19) = 8.661$, $p < 0.01$). The findings from Evans, Horowitz & Wolfe's (2011) were replicated, with participant accuracy during trial types with a cued and uncued targets present (dual target) (M=75.9%, SE=1.9%) significantly lower than during trials with a single cued target present ($F(1,19) = 114.571$, $p < 0.001$), as can be observed in Figure 5.4. Pairwise comparison indicated no significant difference in participant accuracy when there was a single cued target present, and when an uncued target was present (M=89.2%, SE=0.9%) ($p = 1$). Values for significant differences can be found in Table 5.1.

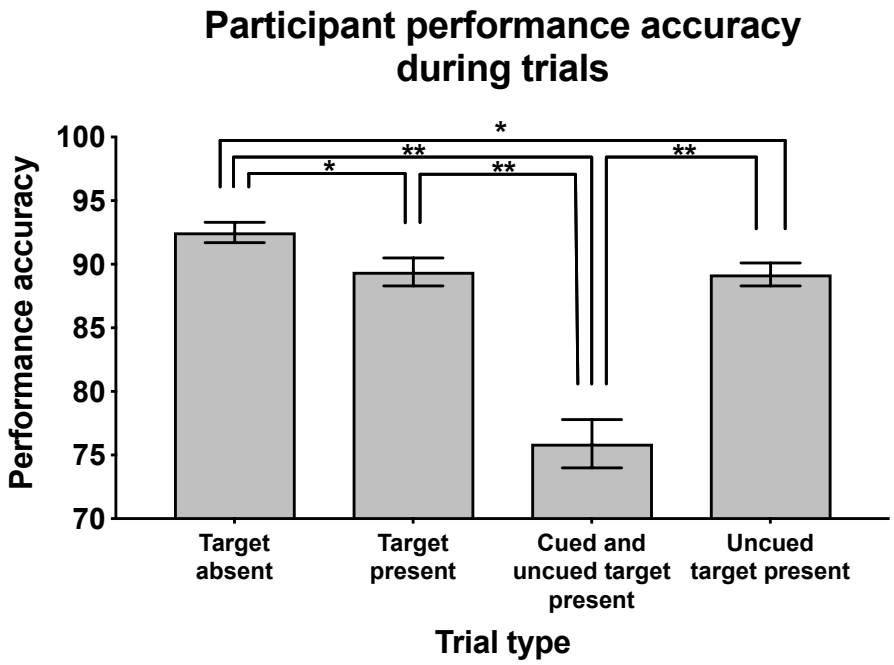

Figure 5.4: Graph of participant performance across the four trial types, with data collapsed across all blocks. Significant differences are indicated by black bars; a single asterisk means $p < 0.01$, and a double asterisk indicates $p < 0.001$. All error bars represent standard error of the mean.

Main effects	**DF**	**F**	***p***
Block	1, 19	0.039	0.846
Trial type	**1.638, 31.12**	**60.717**	**< 0.001 (Greenhouse-Geisser corrected)**
Block * trial type	**3, 57**	**10.507**	**< 0.001**
Within-subjects contrasts			
Target absent vs. target present	**1, 19**	**8.611**	**0.009**
Target present vs. dual target	**1, 19**	**114.571**	**< 0.001**
Dual target vs. uncued target	**1, 19**	**52.629**	**< 0.001**
Pairwise comparisons (Sidak)			
Target absent vs. dual target			**< 0.001**
Target absent vs uncued target			**0.001**
Target present vs. uncued target			1

Table 5.1: Statistical outcomes for the 2X4 ANOVA. Statistically-significant rows ($p < 0.01$) are indicated in bold.

5.5.2. MVPA results

The first analysis run on the EEG data was done to confirm that participants can differentiate between the four different kinds of trials within the block. Voltages for each trial type were averaged across participants and presented for each of the three blocks (see Figure 5.5), as a sanity check. In block 1, as participants do not know that some images are 'silent' targets, the blue (dual target) and orange (uncued target present) lines are not dissimilar to the green lines (target present). However, in block 2 when the 'silent' targets become task-relevant, the microvoltage line for dual targets changes as participants must now respond to previously non-relevant target images. This change appears to be reduced in block 3 as participants adapt to the new paradigm.

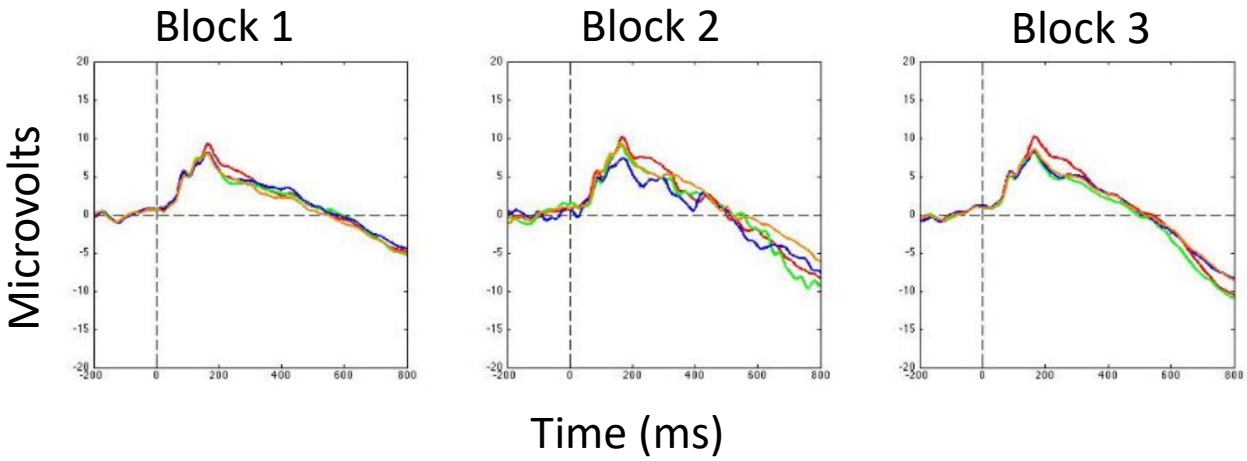

Figure 5.5: Averaged voltages across the four experimental conditions, across blocks, and across all cap electrodes. Green = cued target present, red = cued target absent, blue = dual target, and orange = uncued target present.

The MVPA analysis was run with data from blocks 2 and 3 collapsed, as the task was the same for both blocks. The first contrast was target present vs. target absent (Figure 5.6 A), in which patterns of activation for target present trials were compared to patterns in which target absent trials were presented. In block 1, the difference in patterns were above chance at approximately 50 msec (though it does not become consistently above chance until the 200 msec mark). A similar pattern is observed in the collapsed data for blocks 2 and 3, with the pattern of difference rising above chance at approximately 75 msec and remaining consistently above chance from the point onwards. Both of these blocks indicate an early differentiation between the two trial types, suggesting that gist is accessed/extracted very early on (between 50–75 ms).

The second MVPA analysis compared patterns of activation between the trial types in which the target was present (relevant target), and in which uncued targets were present (non-relevant target) (Figure 5.6 B). In this contrast, the differentiation between the trial types happens later than in the previous contrast for block 1, and the combined blocks 2 and 3. In block 1, the differentiation begins at around 200 ms, whilst blocks 2 and 3 show a pattern in which differentiation occurs at the 200 msec mark, but does not reach consistency until 300 ms. Whilst the previous contrast indicated that gist is extracted early

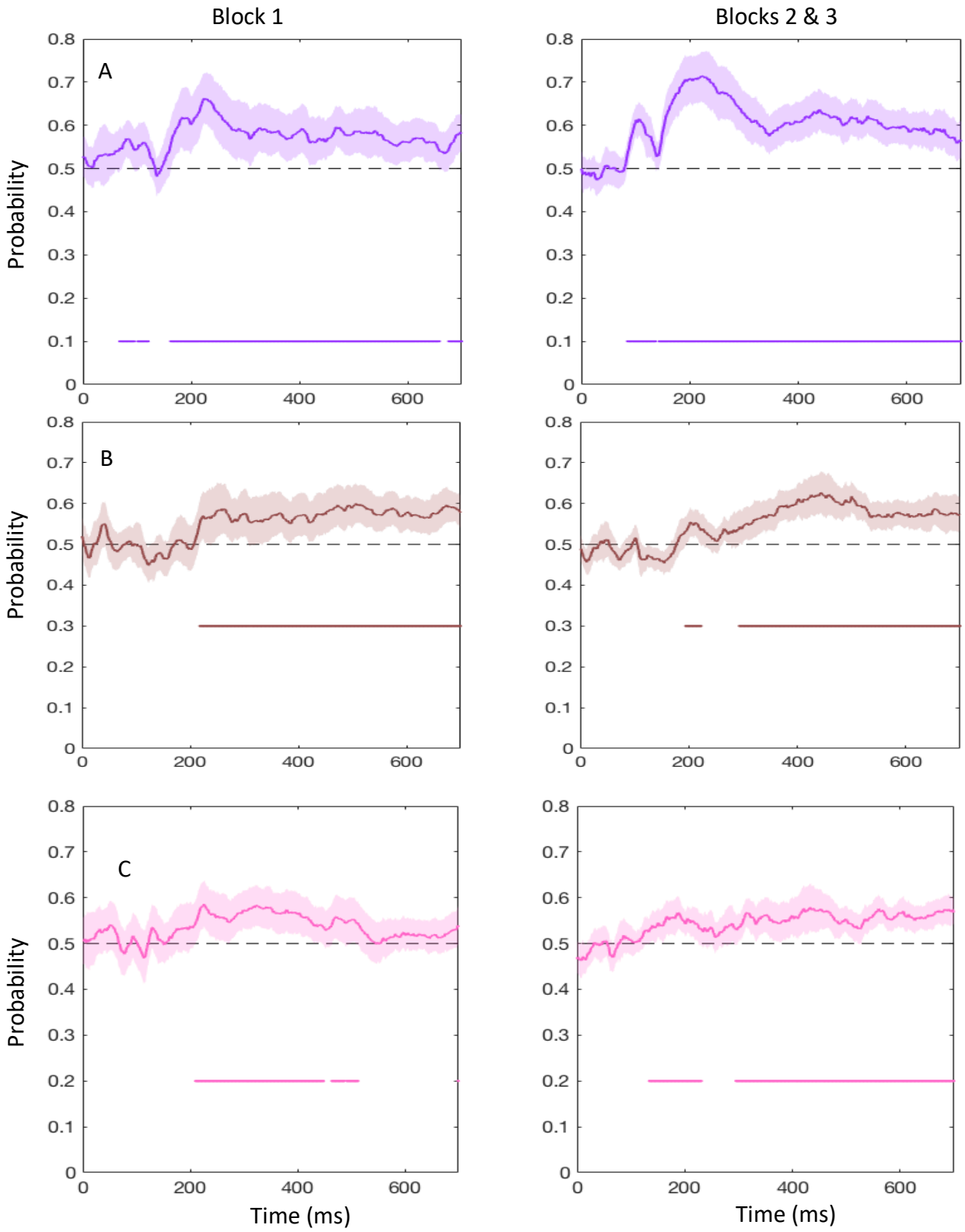

Figure 5.6: MVPA classifier graphs for the three contrasts. A compares cued target present trial to target absent trials. B compares cued target present to uncued target present trials. C compares single target trial to dual target trials. The solid line indicates the classifier; shaded areas are 95% confidence intervals; the straight line at the bottom of the chart indicates when the classifier is above chance (indicated by the dotted line). Data is cluster-corrected and bootstrapped.

(50–75 ms), this contrast suggests that understanding of the relevance of the category

image occurs later on in processing. This occurs even in block 1, during which participants are only aware the target is an uncued category image in 50% of trials.

The third MVPA contrast compares trials in which a single target is presented (cued target present) to trial in which two targets are presented (cued and uncued target present) (Figure 5.6 C). Here, the differences are similar to that in Figure 5.6 B, as the secondary target category is not trial-relevant. Dual targets are differentiated from single targets in 200 msec during block 1, and faster at 100 msec in blocks 2 and 3. This difference may be attributed to participants being unaware in block 1 that dual targets were dual-target events, as half of the image categories were 'silent'. In blocks 2 and 3, these categories are no longer 'silent', and participants are responding to them as task-relevant, though not trial-relevant, categories. Investigating this more in depth, Figure 5.7 compares the difference in brain patterns for dual-target trials in block 1, and the collapsed blocks 2 and 3. The difference wave rises above chance at the 300 msec mark, lasts for approximately 75 ms, and remains at approximately the 55% probability mark for the remainder of the time frame. The 300 msec mark is when the destructive interference can be observed, and due to its relatively late onset in the trial may be part of a decisional, rather than a perceptual, process.

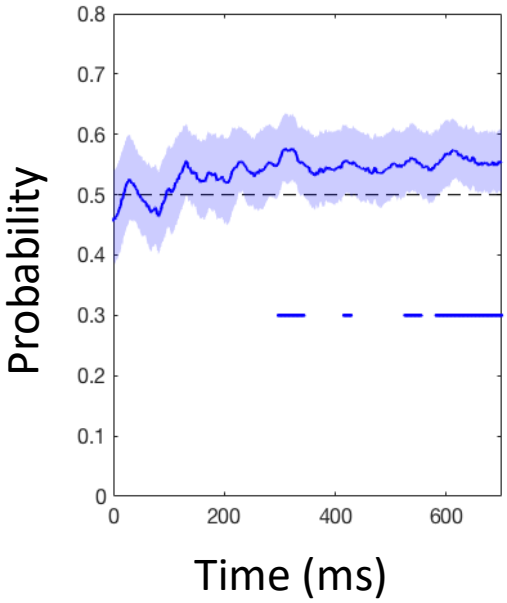

Figure 5.7: Comparisons within the dual-target trial-type, compared between block 1 and blocks 2 and 3. The solid line indicates the classifier; shaded areas are 95% confidence intervals; the straight line at the bottom of the chart indicates when the classifier is above chance (indicated by the dotted line). Data are cluster-corrected and bootstrapped.

5.6. Discussion

The aim of chapter 5 was to replicate the findings of chapter 4, and to further extend the understanding of the neuronal signature of task contingency change. The experiment found gist was accessed rapidly (<75 msec), with target types being differentiated at approximately 200 msec. The effect of changing task contingency, seen behaviourally as destructive interreference, was seen around 300 msec. These findings further support the conclusions from chapter4, in that gist is accessed quickly and as part of a perceptual process; destructive interference, however, is a decisional process seen later.

Tasks which ask participants to identify a single pre-cued target show a differentiation earlier (~200 msec) than tasks which are more demanding (~300 msec), such as differentiating between relevant and non-relevant targets, or processing a relevant and non-relevant target shown in the same scene. This later differentiation appears to delay the time at which decisional processes begin, possibly due to the recruitment of additional attentional resources.

5.6.1. Behavioural performance

In terms of the behavioural data, participants responded with fewer errors during target-absent trials (performance accuracy = ~89%) compared to target-present trials (performance accuracy = ~93%). This is a different finding than the data presented by Evans, Horowitz & Wolfe (2011), who found no significant difference in performance accuracy between these two trial types. However, this experiment did not use rapid serial presentation, as Evans et al. used, which may account for the difference between findings.

Additionally, figure 5.5. shows an increased amplitude for target-absent trials (red line) compared to all other trial types, including target-present trials (green line). This is found between 200–300 ms, possibly reflecting the P3 novelty response.

5.6.2. Gist is accessed quickly

The initial analysis (as seen in Figure 5.5) was intended as a proof-of-concept, or ‘sanity check’; participants could differentiate between the four possible trial types, and in

particular started to differentiate between the presumed target-absent trials in block 1 (with 'silent' categories), and the dual-target trials in block 2 in particular, as shown by the change in the wave for this trial type. This successful discrimination of target types meant that participants evoked consistently, or reliably, different patterns of activation to these trials starting early on.

These data suggested that scene gist was extracted quickly after stimulus onset, as seen in Figure 5.6 A – both block 1 and the collapsed blocks 2 and 3 showed clear differentiation in the MVPA wave between 50 and 75 msec from stimulus onset. As in chapter 4, this supports the notion that gist information is extracted during the first feed-forward sweep of information (Greene & Oliva, 2009a). These findings were also similar to those of Battistoni, Kaiser, Hickey, & Peelen (2018), who found that 50 msec post-stimulus onset was enough for both target and distractor objects (people and cars, in this example) to be accurately decoded in a multivariate analysis.

The particular time-signature of this event would suggest that it may be the C1 ERP component; onset of gist processing can be decoded within 50–90 msec (Di Russo et al., 2003; Stolarova et al., 2006). Stolarova et al. (2006) explored this component further, finding that the C1 component was implicated in differentiation between perceived neutral and unpleasant stimuli, suggesting that the primary visual cortex is involved in the evaluation of learned affective visual stimuli. The authors specifically argue against this being the product of re-entrant processing, noting the short latency of the stimulus. This would fit in with the discussed literature that gist is accessed very quickly as a result of the feed-forward pass.

Our data suggests that the trial types – target-present and target-absent – are being differentiated at this time point. If it is possible for the early visual cortex to learn to distinguish between affective and non-affective stimulus types, as Stolarova et al. suggests, it may also be possible for it to differentiate between stimulus and no stimulus.

However, this is a supposition, and does not have any direct evidence; a more targeted experiment may help unpick the relationship between this early neural response and early visual cortex plasticity. For example, Stolarova et al. argue that classical conditioning is required for this plasticity to be seen. An experiment which removes this element from Stolarova et al.'s design, and instead relies on pre-cue or priming for a target stimulus, may be able to replicate the effect of the plasticity without the conditioning. By

using neutral (not affective) stimuli with a single target category (e.g. urban scenes) amongst 'distractor' categories, it may be possible to see if the early visual cortex can distinguish between a target stimulus and non-target stimulus.

5.6.3. Differentiation of target types occurs later

Whilst initial differentiation of the stimuli is fast (~75ms), later differentiation between relevant and non-relevant targets occurs around 200 msec after stimulus onset (cued trials vs. uncued trials; Figure 5.6 B). Critically, this occurs at 200 msec in block 1, and in the combined blocks 2 and 3, but it does not become consistent in this latter state until 300 ms. This initial response at 200 msec is approximately in line with the response of the visual N1 (160–200 msec) and N2 components (200–350 msec); however, simply declaring this finding as a result of just one (or both) of these components may miss the nuance of the type of scene categorisation being observed.

Our finding of 200 msec is approximately line with other MVPA research, with previous studies finding target differentiation occurring at 180 msec after stimulus onset (Kaiser et al., 2016; Battistoni et al., 2018). Battistoni et al. also used the same stimuli as target and non-target objects, with the relevance of the objects changing via a (symbolic) cue. As Battistoni et al. identified spatial attention occurring at 240 msec post-stimulus onset, we might conclude that this represents the onset of 'local' attention, and gist extraction would be seen before this point (Battistoni et al. found target information was available from 180 msec). As our data suggests a differentiation of target relevance at 200 msec, this appears to be the case.

Kaiser et al. also observed that this 200 msec time point only represents targets that match the current attentional set (a set of rules that aid an observer in distinguishing relevant from non-relevant stimuli – Heisler et al., 2015). In short, if the stimulus does not abide by these rules, then it takes longer to process, which may explain why this differentiation between relevant and non-relevant targets takes longer in the combined blocks 2 and 3 – non-relevant targets becoming relevant does not fit within the attentional set.

The difference between these two findings can be explained by the change in task contingency. In block 1, participants were shown targets which appeared to be non-

relevant, but were actually 'silent' target categories. These silent targets were treated as non-relevant targets, with the differentiation between relevant and non-relevant targets occurring at approximately 200 msec post-onset. However, in blocks 2 and 3, those silent targets became relevant. Participants were cued for one target category, but shown another (which had previously been non-relevant, but now many also be cued during the block). Participants still had to differentiate between relevant and non-relevant targets, but also had to reclassify previously silent, apparently non-relevant targets as relevant. Whilst differences between cued and uncued targets are still seen at 200 msec post-onset, this does not become consistent until 300 msec post-onset.

This is in line with the visual N2 component (Folstein & Van Petten, 2008) which, as discussed in the introduction, may be associated with determining if that a pre-cued gist is not present. As this analysis compared cued and uncued targets, this would fit within that remit; participants are essentially deciding if the presented stimulus is task-relevant within the first 200 msec after stimulus onset, in the perceived absence of other task-relevant stimuli. This is seen in both multivariate and univariate analyses, and through non-natural and natural target images. Conversely, the univariate finding of occipito-parietal activation may reflect our findings from chapter 3, in which category-selective cortex activation is seen during gist processing. This may indicate categorisation of the shown image stimuli.

To dig into this time window in more detail, it aligns with the D220 component (Philiastides, Ratcliffe, & Sadja, 2006). The D220 component is associated more with a later, 300 msec component than the earlier 160–200 msec range given by the visual N1, suggesting a relationship to the N2 as opposed to the N1.

The D220 component is found between what Philiastides, Ratcliffe & Sadja described an 'early' decisional process, found at 170 msec post-onset (N170 – Rossion et al., 2003; Philiastides & Sajda, 2006), and a 'late' decisional process found at 300 msec (possibly P300, though the authors never state this explicitly). Its amplitude is correlated to the task difficulty, which may explain its presence in blocks 2 and 3, and not block 1: as the task has become measurably more difficult, its onset is shifted backwards to account for the additional resources required for processing the stimulus.

Considered in light of the data in this chapter, our data would fit into the framework proposed by Philiastides. Whilst not as quick as 170 msec after stimulus onset (possibly due to the task being more difficult, even in its 'easier' format, that that used by Philiastides et

al.), participants are able to make the decision in block 1 that the presented stimulus is either cued or uncued.

However, in blocks 2 and 3, the contingency of the task has changed; the difficulty level becomes higher, and whilst the ~200 msec component is still seen, it does not become consistent until 300 msec – in line with Philiastides, Ratcliffe & Sadja's 'late' decisional process. As discussed by Philiastides & Sajda (2006), onset for this component can be delayed depending on task difficulty, which would explain the window in which it appears in these data. This would also tie into the findings from Kaiser at al., who argues that the delay in processing relevant and non-relevant stimuli beyond 200 msec represents attention resolving the conflict between competing objects. The findings from this study, Philiastides et al., Philiastides & Sajda, and Kaiser et al. fit together to argue that attention is required to resolve conflicts (as represented by more difficult tasks) which then extends the processing time. This produces the 'weighted sum' element of the diffusion model, as demonstrated in Figure 4.1.

Philiastides et al. also predicted the N170 component as a response to an 'easy' task. However, the blocks in this experiment contained multiple different trial types in this block, with the 'difficult' trials representing a change in task contingency. As the N170 may have occurred in all trial types, we did not see a significant difference in the pattern of activity across the whole brain.

In terms of the univariate ERP experiment, a similar analysis was run on non-target categories. In block 2 of this experiment, participants were cued for either place or face images, with faces images becoming relevant targets (compared to block 1, where they were not; participant were only cued for place images); therefore, in block 2, face images must be suppressed as they are uncued targets.

The data have found corroborating evidence, with the 'easier' portion of the experiment (half of presented targets appearing task-relevant) showing frontal activation around the 200 msec post-onset mark, and once the task contingency has changed (previously 'silent' targets becoming task-relevant), showing frontal activation post-300 msec after stimulus onset. This delay in processing for more difficult tasks ties into the diffusion model of accumulated evidence for destructive interference.

5.6.4. Temporal dynamics of destructive interference

The behavioural effects of destructive interference were once more identified In chapter 5. Participants showed a statistically-significant reduction (13.5%) in performance accuracy during trials in which both a cued and uncued target were present (dual-target trials) compared to trials in which only the cued target was present (single-target trials). The third multivariate analysis compared single target trials (cued target present) to trials with dual targets (both cued and non-cued targets present) in order to identify the time point at which destructive interference can be seen neuronally.

As in our second comparison, block 1 showed a differentiation between cued and non-cued target types at around the 200 msec post-onset mark; this is not surprising, as participants may not know they are getting a dual-target trial due the 'silent' nature of 50% of possible target categories. However, unlike comparison 2, blocks 2 & 3 indicated that dual targets were differentiated earlier, at around 100 msec after stimulus onset. This difference wave remains significant for approximately 150 ms, and does not reach significance again until 300 msec post-stimulus onset.

The question is then raised: why might we see this early modulation to a greater degree in dual-target trials compared to single-target trials? As discussed in chapter 4, discrimination for global-property categorisation (i.e. natural vs. man-made) is found at around 100ms (Groen, Ghebreab, Prins et al., 2013). Groen et al. more specifically state the spatial coherence and contrast energy of a natural scene modulated ERP amplitudes between 100 and 150 msec post-stimulus onset, as we find in this analysis. They later described this finding as related to the discussed D220 component, with spatial coherence in particular containing information important to global property categorisation. Activation in this time window would be part of the first feed-forward sweep of information through the brain, and it may represent the categorisation of multiple gists, rather than just a single one.

This is also the time point at which the P1 component is found (90–100 msec after stimulus onset) which is associated with early processing for attended targets (Heinze, Luck, Mangun, & Hillyard, 1990). An enhanced P1 response is found during the presence of valid (compared to invalid) targets on-screen. In block 1, differentiation between cued and uncued targets is not a task required consistently of participants; however, blocks 2 and 3

require this for every dual-target trial. It is possible then that this 100 msec component reflects an enhanced P1 response, as participants must differentiate between the validity of the on-screen targets.

It is also important to observe that this differentiation between dual- and single-target trials occurs again, and more consistently, at 300 msec post-stimulus onset. Further, this is seen not only when comparing dual- to single-target trials, but also when looking into dual-target trials independently, as seen in Figure 5.7. When comparing dual-target trials between block 1 and blocks 2 & 3, the pattern of activation in the brain becomes significantly different at 300 msec post-stimulus onset. This ties in with the findings from the previous analysis, indicating strongly that this is the time point at which destructive interference is seen neuronally. It is possible that this time point represents the P3 component, especially the P3b, in line with the findings from chapter 4. Even if this event is not a direct consequence of the P3 component, it is still strong evidence that destructive interference is the outcome of frontal, decisional areas.

In both the experiment by Philiastides, Ratcliffe & Sadja (2006) and our experiment, in the case of higher task difficulty (in our case, specifically the change in task contingency), the presence of 'anti-target evidence' means that the participant's decisional process takes longer. Philiastides et al. identified this as the point at which decision-relevant information was fed into the diffusion model, accumulating evidence both for and against the pre-cued target category. Unlike in comparison two, where we discussed the change in difficulty due to task-irrelevant stimuli becoming task-relevant, this more directly reflects the traditional pro- and anti-evidence diffusion model as discussed in Figure 4.1. Our data provided further evidence for a decisional process at approximately 200 msec (block 1) after stimulus onset, and for more difficult tasks to delay the later decisional component to ~300 msec after stimulus onset (block 2 & 3).

In chapter 3, we performed an exploratory analysis on the fMRI data to see which areas of the participants' brains responded to a greater degree during single targets, compared to dual targets. This suggested that frontal areas may be involved, such as the dorsolateral prefrontal cortex, though this evidence was not strong. Taken altogether, the evidence from the multivariate and univariate EEG studies, and the fMRI study, indicate that destructive interference is a decisional process that may be the result of frontal cognitive control and error-monitoring areas.

5.6.5. Limitations and advantages of the design

MVPA is a now-commonly-used analysis tool for EEG data, with its own set of advantages and disadvantages. As discussed in the introduction, MVPA allows for a map of correlations and co-activations over the whole scalp, rather than comparing electrodes directly as in ERP analysis. This allows for MVPA to detect consistent, though weak, patterns across the scalp (Habeck, 2010; Oosterhof et al., 2016). This allows for the analysis to be exploratory, rather than relying on pre-determined ERP time points.

However, MVPA is not without its flaws. Its single biggest concern is its philosophical underpinning, as it relies upon the concept that information decoded (using analysis methods that 'mimic' the way the brain decodes neural information) from neural activity patterns produces answers , or evidence, about what those patterns indicate. In this case, this would mean that if the classifier used in this analysis can differentiate between the trial conditions, then it is reporting a differentiation between the way neurons in the brain treat those trial conditions. This is known as the "decoder's dictum", as discussed by Ritchie, Kaplan, & Klein (2017). Ritchie et al. note the issues with this assumption, observing that classifiers are something of a 'black box' process with no transparent means of understanding how the classifier exploits the neural data. Classifiers are also 'informationally greedy', in that all neural information is fed into the classifier, not just the neural activity which underlies the specific brain process.

This is offset somewhat by considering the MVPA findings alongside the results of the univariate ERP analysis. It does not counter all of the underlying (and possibly flawed) assumptions inherent in MVPA, but it does allow for a 'sanity check' of the classifier used in the analysis. As similar results are found in both the univariate and multivariate analysis, then (whilst not definitive) this suggests that the classifier is responding to the same neural patterns generating the data found in the univariate analysis. At best, the data are likely representative of neural processing; at worst, the data are consistent. Conversely, whilst the univariate data may show some echo artefacts, the fact that a strong response to the change in task contingency was found between 260–360 msec in frontal areas is supported not only by the findings of the MVPA, but by the initial exploratory analyses found in chapter 3.

This 'sanity check' approach also applies to the choice of stimuli used. The multivariate experiment in this chapter used very naturalistic scene images (see Figure 5.1 for examples), whereas the univariate experiment used artificially-presented stimuli (see Figure 4.2). By utilising these different kind of images, this allowed a broader examination into the nature of gist, rather than limiting understanding to artificial presentations as in chapter 3.

Rapidly categorising targets within a natural scene requires the ability to extract the gist of the scene at several levels, as discussed in chapter 3. By using naturalistic stimuli, the findings can be more consistently related to both the existing literature and reflect more closely the response of the brain to real-life gist processing. Additionally, by combining these findings with those of the artificial, non-natural stimuli in chapter 4, this allows the best of both worlds: the realism of the natural stimuli, with the high degree of stimulus control of the artificial stimuli. This makes the findings more robust, and allows for better interpretation of the neural signal when the results are considered together.

Finally, this chapter predicted 3 specific time points would be decoded within the MVPA data: an early correlate for gist extraction, a 170 msec component, and a component related to the D220. Of these, both the early component and the D220/P300 appeared to be identified. In particular, the N170 was not found as a hypothesised 'easy' marker for trials. This may mean one of two things: either the component appeared consistently in all trial types, and so the MVPA did not identify it as a pattern change, or the task was not easy enough.

Both of these criticisms have something to recommend them, though the latter requires some further explanation. Even though the single-target task is 'easy' in comparison to the dual-target task, Philiastides, Radcliffe and Sadja's (2006) 'easy' task involved the differentiation of red and green targets, with no further challenge, and they subsequently argued that N170 represented an early perceptual process rather than a decisional one. This may be reflected in this chapter in the differentiation component being found later for cued and uncued trial types (~200 msec). This possibly suggests that more attentional or cognitive resources are required for category identification, compared to colour identification; put simply, red/green identification is a simpler process than identifying whole category images, and so the N170 component was not likely to be found.

5.6.6. Future research

The philosophical limitations of MVPA have been discussed, most notably the assumption that the patterns of activation revealed by the MVPA contain key evidence to what those patterns represent. However, as well as criticising MVPA methods, Ritchie et al. (2017) also offer a suggestion to improve the quality of future MVPA. This could be applied to a similar study as the one described in this chapter.

Ritchie et al. argue that MVPA analyses must be supplemented with a psychological model or 'space', in which elements of the stimuli form a structure into which participant responses, such as reaction time, can be modelled based on the relationship between the stimuli. The diffusion model represents such a 'space', with participant response being modelled based on the relationship between stimuli and attention (as cartooned in Figure 4.1).

Representational similarity analysis (RSA) would allow for the analysis focus not just on the overall pattern of activity across the brain during trial types, but would also for an analysis of the neuronal response in relation to the proposed model (i.e. within the structure of the psychological 'space'). This would allow the diffusion model to be applied directly to the neural activation in order to predict observer behaviour.

Ritchie et al. report the most common way of doing this is to use behavioural judgements of a stimulus to create a 'similarity space', and then use these in an RSA matrix directly with the neural information collected (Bracci & Op de Beeck, 2016). By doing this, it would be possible to identify the neural activation spaces that best correlate with behavioural similarity spaces (Kiani, Esteky, Mirpour, & Tanaka, 2007; Kriegeskorte, Mur, & Bandettini, 2008). As this chapter focuses on a phenomenon identified in behavioural research (destructive collision, Evans et al., 2011), it would be logical to analyse it using a method that uses the associated behavioural responses as a framework for the analysis; this would also allow other types of multiple gists (such as constructive interference) to be studied in the same manner.

RSA was not employed for these data, as it was not considered as a possible option until data acquisition and analysis were complete. Particularly as this type of analysis had not been considered when building the design of the experiment, it did not seem

appropriate to immediately re-analyse the data; there were also time constraints due to data collection for other chapters.

5.7. Conclusions

The multivariate experiment in this chapter aimed to expand upon the time signatures of several different aspects of gist processing. This included the onset of gist processing, differentiation between different categories of gist, and destructive interference (or 'category collision') between two primed gist categories. Initially, several time points were predicted at which differences in the pattern of brain activation would be found (N170, D220 and P3). The experiment utilised a 2AFC design, along with rapid presentation of categories of stimuli (beach, animal, human, vehicle, mountain and bridge images).

The findings indicated extremely early responses to initial gist (50–75 msec post-stimulus onset) with the C1 component; later differentiation of the relevancy of targets at 200 ms, and at 300 msec for targets that were previously non-relevant (N2, D220); and at approximately 300 msec post-onset for dual targets (possibly P300, and/or part of a delayed D220 component). Additionally, time points consistent with the P1 (differentiation between validity of onscreen targets) and N2 (determining if a pre-cued gist is present) were found. No component was found within the window suggested by the visual N1 or N170 ERPs. However, this does not mean that they did not occur; only that the design of the experiment may not have allowed them to be observed.

As found by other literature, replication of this study should be possible using representational similarity analysis to account for the behavioural framework of destructive interference. This would allow the observed ERPs and neural patterns to be linked more closely to the behavioural effects of destructive interference, or of other effects of changing the task contingency.

Chapter 6 Conclusions and general discussion

6.1. Overview of the thesis

This thesis had two aims: the first was to assess the neural correlates and temporal dynamics of the visual perception phenomenon known as 'gist'. The second was to assess the relationship between two 'modes' of attention in terms of possible models of visual attention. The experiments described within the chapters of this thesis make novel contributions to the wider understanding of visual neuroscience, in terms of the neuronal correlates and temporal dynamics of 'gist' processing, and by investigating the way in which the two extremes of visual attention interact.

Chapter 2 utilised 4 separate behavioural experiments to identify the nature of the relationship between the two attentional modes. These experiments were the first to directly load these attentional modes, with the aim of clarifying which model of visual attention more accurately reflected the nature of visual attentional processing. In particular, it represented the first time that the testable hypotheses given by the three models (Feature Integration Theory, Reverse Hierarchy Theory, and Guided Search Theory) were assessed explicitly in relation to one another. Overall, the data generated by the first three experiments indicated that visual attention operates in a serial manner (as per the RHT), and not as a flexible continuum, or in parallel; the fourth (control) experiment confirmed that the paradigms used were loading visual attentional resources, rather than cross-modal attentional resources as a whole

Chapter 3 aggregated existing literature on 'gist' processing in order to more clearly define the somewhat amorphous term. It was then demonstrated that gist processing could be observed neuronally in category-selective cortex. The specific, observable neural correlates of gist processing have not otherwise been noted in research; previous literature had focused on single aspects of gist, such as the location of the neural correlates of processing of specific types of category information. This study was also the first to find that gist processing produced different response profiles in category-selective cortex, with predetermined place-selective areas responding in a perceptually-driven manner (greater % BOLD signal change for the perception of a target, regardless of its actual presence on-

screen). The pre-determined face-selective regions, however, responded in a hybrid perceptual-stimulus driven manner, requiring both the presence of the stimulus and the observer's subjective perception of it to produce a greater % BOLD signal change. The findings of chapter 3 also suggested a possible frontal location for destructive gist interference, though this requires further exploration.

Chapters 4 and 5 applied univariate and multivariate pattern analysis respectively to further locate the temporal dynamics of gist processing, confirming the previous literature in that gist is accessed very early in processing (30–50 msec post stimulus onset). The experiments in these chapters were also the first to identify ERPs and significant neural pattern differences during the change in a task contingency (i.e. destructive interference) at ~300 msec post-onset, indicating a decisional process. This also tied into previous findings of a difficulty component at approximately 220 msec post-onset, further suggesting strongly that the time signature of destructive interference is related to conflict resolution and decision-making.

6.2. Further discussion fMRI and EEG data

6.2.1. Spatial and temporal locations of gist processing

Due to the well-known limitations of fMRI and EEG, the three studies were designed to complement each other by using fMRI's superior spatial resolution and EEG's superior temporal resolution. By using similar paradigms, this would allow similar questions to be asked of the fMRI and EEG data.

The fMRI data observed that category-selective cortex showed a greater % BOLD signal change during gist processing. Whilst this does not make a definitive argument that this is the location of gist processing, it is (at minimum) a part of the network or that performs this kind of processing. This raises a question: is this response seen as part of the first feed-forward sweep of information through the brain, or is it a product of re-entrant processing? Due to the poor temporal resolution of fMRI, it is likely that these findings are the result of re-entrant processing. However, this does not preclude their involvement in the initial feed-forward sweep as well.

The initial EEG data discussed in chapters 4 and 5 observed that gist is accessed early, from between 30–50 msec post stimulus onset; previous EEG research into the predetermined regions of interest have observed that the PPA can distinguish place from non-place targets at approximately 80 msec post-onset, with further differentiation (e.g. between building and non-building place images) occurring later, at around 170 msec post-onset (Bastin et al., 2013). The initial response of the FFA to face images is observed as a modulation to the P1 component (90–100 msec after stimulus onset; Olivares, Iglesias, Saavedra, Trujillo-Barreto, & Valdés-Sosa, 2015) as well as the more well-known mapping of the N170 component to both the FFA and the OFA (Oruc et al., 2010; Eimer, 2011).

That both the PPA and the FFA can differentiate their preferred stimuli before 100 msec – even if this is as rudimentary as identifying the stimulus as preferred or not-preferred – suggests that category-selective cortex may be involved in the first feed-forward sweep, acting as the neural location for the quick differentiation of target types observed in previous literature. However, the time signatures for these regions are somewhat beyond that of the 30–75 msec range observed in chapters 4 and 5; along with the poor temporal resolution of fMRI, this implies that the regions observed in chapter 3 are seen due to re-entrant processing, rather than capturing the first feed-forward pass of information through the brain.

Using the fMRI data as a starting point, the EEG data implies that category-selective cortex may be involved for very early differentiation of gist category, but a more targeted study would need to be conducted in order for this to be observed. Use of a combined EEG-fMRI paradigm (Schomer et al., 2000; Mayhew, Ostwald, Porcaro, & Bagshaw, 2013) would allow for cortical excitability in both spatial and temporal dimensions to be correlated, thus providing a better picture of when category-selective cortex is involved within gist processing.

Based on the findings from chapters 4 and 5, these data would hypothesise that category-selective cortex is involved both initially during the feed-forward sweep (as seen by early access of gist, both in chapter 4 and from the literature), as well as in re-entrant processing, as areas like the FFA are associated with multiple ERPs including the later N170. This would potentially explain the delay in differentiating basic category information; 30–67 msec are required to determine a scene's basic category, but 170 msec are needed to make an 'easy' differentiation (Rossion et al., 2003; Philiastides & Sajda, 2006).

6.2.2. Spatial and temporal locations of destructive interference

The findings from the EEG data also aid in understanding the exploratory analysis conducted on the fMRI data. In order to identify possible regions involved with the processing of destructive interference between gist, categories, an analysis was conducted which looked for regions which showed a greater % BOLD signal change for single-target, compared to dual-target, trials. This analysis showed small areas of activation in frontal areas of the brain, including parahippocampal areas, dorsolateral pre-frontal cortex and middle temporal areas, anterior hippocampal areas and the superior temporal sulcus (see Figure 3.9). These areas were small, but survived multiple analysis types, including both cluster-corrected and voxel-corrected analyses.

These areas were too small for a concrete conclusion to be reached regarding their role in the processing of destructive interference. However, they were found consistently, and were generally regions implicated in cognitive control and executive decision-making. A tentative hypothesis was formed that destructive interference was a decisional, not a perceptual process, unlike the findings from the first ("where is gist processing observed in the brain?") fMRI analysis.

This was supported by the EEG findings, with the ERPs associated with destructive processing being D220 and N300. These are both components that represent top-down influences on decision making, with D220 representing task difficulty and greater attentional resources (Philiastides et al., 2006), and N300 representing uncertainty resolution of visual stimuli (Hamm et al., 2002)). What appeared as a possible trend in the fMRI data has further evidence from the nature of the ERPs found in both the ERP analysis and multivariate analysis of the EEG experiments, and indicates that overall, destructive interference is the product of decision-related frontal regions.

This suggests strongly that a follow-up experiment, more specifically to address the neural correlates of destructive interference, would be in order. There are three issues to content with from the study discussed in chapter 3: firstly, the relatively low number of dual-target trials (a total of 30 dual-target trials out of a total of 360 trials) over the course of the experiment. Secondly, the use of face images, which participants frequently reported as appearing far more salient than the place targets; and thirdly, all stimuli used in the

experiment were extremely artificial, including an occasionally-reported issue in which the processed place images appeared, in the 200 msec window participants had to observe the stimulus, to bear more than a passing resemblance to the Simoncelli textures used for masking the images.

This suggests that any follow-up experiment needs to increase the number of dual-target trials. It may also benefit the experiment to use more naturalistic scenes, closer to the stimuli used by Evans et al. (2011) and in chapter 5 of this thesis, where participants were asked to report on the category of real-world images such as beaches, mountains, and animals. This would aid in reducing the high saliency of the secondary target category faces, which participants reported as easier. This was seen in participants' behavioural data, with 'face-only' trials reporting a higher performance accuracy (93.78%, d' = 3.11) compared to 'place-only' trials (65% correct, d' = 0.78).

By boosting the representation of 'destructive interference' trials within a 2AFC experiment, this should hypothetically result in a higher corresponding neural signal for the analysis, and result in a more targeted analysis. The initial hypothesis is that the areas seen in this new experiment for destructive interference would be the frontal areas implicated in chapters 3 and 4, such as the DLPFC and anterior cingulate cortex (ACC), the latter due to its relationship to error and performance monitoring (Carter et al., 1998; Orr & Hester, 2012).

With some minor modifications, this would also allow for the examination of other kinds of gist extractions, such as constructive interference. Evans et al. (2011) discuss constructive interference as a counterpart to destructive interference, as it is also a product of changing the task contingency. Destructive interference is observed when a participant attempts to identify a cued target category in the presence of a relevant, uncued target category (in chapter 3, it was seen when participants were cued for a place image, but were shown both a place and a face image). However, constructive interference occurs when the participant attempts to identify either of two cued targets presented at the same time (for example, the participant would be cued "place OR face", and shown both target types). This results in a statistically significant improvement in performance accuracy. The hypothesis is that there are some shared mechanisms between destructive and constructive interference, and so this experiment would provide a more well-rounded picture of the effect of changing the task contingency in the human brain.

6.3. Conclusions

This thesis defined the neural and temporal corelates of the fast 'gist' processing that is a product of the sparse attentional mode. This attentional mode was also studied in relation to its counterpart, focused attention, in order to better understand the relationship between them. In assessing the possible models to explain this relationship, visual attention was found to operate in a serial manner, with sparse attention engaged as a result of the first feed-forward sweep of information through the brain. Focused attention is then engaged by re-entering parts of the visual processing 'lower down' from cognitive control, exerting a top-down influence on visual attention.

Possible future experiments were discussed to further expand understanding of the nature of gist processing both spatially and temporally, and an extra experiment was discussed that provided further evidence for the serial nature of visual attention.

References

Ahissar, M, & Hochstein, S. (1993). Attentional Control of Early Perceptual Learning. *Proc Nat Acad Sci Usa*, *90*(June), 5718–5722. https://doi.org/perceptual learning, attention control, vision, visual task

Ahissar, M, & Hochstein, S. (1997). Task difficulty and the speciticity of perceptual learning. *Nature*, *387*, 401–406.

Ahissar, M, & Hochstein, S. (2000). The spread of attention and learning in feature search: Effects of target distribution and task difficulty. *Vision Research*, *40*(10–12), 1349–1364. https://doi.org/10.1016/S0042-6989(00)00002-X

Ahissar, Merav, & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Ahissar, Merav, Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 285–299. https://doi.org/10.1098/rstb.2008.0253

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Atkinson, J., & Braddick, O. J. (1989). "Where" and "what" in visual search. *Perception*, *18*, 181–189.

Baldassano, C., Beck, D., & Fei-Fei, L. (2013). Differential Connectivity Within the Parahippocampal Place Area. *Neuroimage*, *15*(75), 228–237. https://doi.org/10.1016/j.neuroimage.2013.02.073

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Halgreen, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Bar, Moshe. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bastin, J., Vidal, J. R., Bouvier, S., Perrone-Bertolotti, M., Benis, D., Kahane, P., … Epstein, R. A. (2013). Temporal Components in the Parahippocampal Place Area Revealed by

Human Intracerebral Recordings. *Journal of Neuroscience*, *33*(24), 10123–10131. https://doi.org/10.1523/JNEUROSCI.4646-12.2013

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. (2018). Spatial attention follows category-based attention during naturalistic visual search: evidence from MEG decoding. *BioRxiv Preprint*. https://doi.org/http://dx.doi.org/10.1101/390807

Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, *7*(2007), 19.1-12. https://doi.org/10.1167/7.2.19.Introduction

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165. https://doi.org/10.1016/j.visres.2008.07.012

Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *Journal of Cognitive Neuroscience*, *25*(10), 1711–1722. https://doi.org/10.1162/jocn_a_00422

Bracci, S., & Op de Beeck, H. (2016). Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways. *Journal of Neuroscience*, *36*(2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press Elmsford.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics*, *33*(5), 487–493. https://doi.org/10.3758/BF03202901

Burgess, P. W., & Shallice, T. (1996). Response suppression, Initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*(4), 263–272. https://doi.org/10.1016/0028-3932(95)00104-2

Cant, J. S., & Xu, Y. (2012). Object Ensemble Processing in Human Anterior-Medial Ventral Visual Cortex. *Journal of Neuroscience*, *32*(22), 7685–7700. https://doi.org/10.1523/JNEUROSCI.3325-11.2012

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Cohen, J. D., Botvinick, M. M., … Cohen, J. D. (1998). Anterior Cingulate Cortex , Error Detection , and of Performance the Online Monitoring. *Science*, *280*, 747–749. https://doi.org/10.1126/science.280.5364.747

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search.

Cognitive Psychology, *22*(2), 225–271. https://doi.org/10.1016/0010-0285(90)90017-X

Chan, L. K. H., & Hayward, W. G. (2009). Feature integration theory revisited: dissociating feature detection and attentional guidance in visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 119–132. https://doi.org/10.1037/0096-1523.35.1.119

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with 2 ears. *Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.1907229

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900. https://doi.org/10.1016/j.visres.2004.10.004

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, *4*(5), 170–178. https://doi.org/10.1016/S1364-6613(00)01476-5

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Chun, M. M., & Wolfe, J. M. (2005). Visual attention. In E. B. Goldstein (Ed.), *Blackwell Handbook of Sensation and Perception* (pp. 272–310). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470753477.ch9

Cichy, R., Sterzer, P., Heinzle, J., Elliott, L., Ramirez, F., & Haynes, J. (2013). Probing principles of large-scale object representation: Category preference and location encoding. *Human Brain Mapping*, *34*(7), 1636–1651. https://doi.org/10.1002/hbm.22020

Cichy, Radoslaw M., Heinzle, J., & Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Cichy, Radoslaw Martin, Chen, Y., & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *NeuroImage*, *54*(3), 2297–2307. https://doi.org/10.1016/j.neuroimage.2010.09.044

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage*, *32*(2), 583–591. https://doi.org/10.1016/j.neuroimage.2006.04.180

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, *110*(1), 24–30. https://doi.org/10.1016/S0168-5597(98)00033-1

Copland, D. A., Zubicaray, G. I. De, McMahon, K., & Eastburn, M. (2007). Neural correlates of semantic priming for ambiguous words : An event-related fMRI study. *Brain Research*, *1131*(1), 163–172. https://doi.org/10.1016/j.brainres.2006.11.016

Corney, D., Haynes, J. D., Rees, G., & Lotto, R. B. (2009). The brightness of colour. *PLoS ONE*, *4*(3). https://doi.org/10.1371/journal.pone.0005091

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLoS ONE*, *12*(11), 1–20. https://doi.org/10.1371/journal.pone.0184923

Delorme, A., Rousselet, G. A., Macé, M. J. M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. https://doi.org/10.1016/j.cogbrainres.2003.11.010

Dent, K., Allen, H. A., Braithwaite, J. J., & Humphreys, G. W. (2012). Parallel distractor rejection as a binding mechanism in search. *Frontiers in Psychology*, *3*(AUG). https://doi.org/10.3389/fpsyg.2012.00278

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *353*(1373), 1245–1255. https://doi.org/10.1098/rstb.1998.0280

Desimone, Robert, & Duncan, J. (1995). Neural Mechanisms Of Selective Visual Attention. *Annual Review of Neuroscience*, *18*, 193–222. Retrieved from http://www.annualreviews.org/doi/pdf/10.1146/annurev.ne.18.030195.001205

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

Di Lollo, V. (2012). The feature-binding problem is an ill-posed problem. *Trends in Cognitive Sciences*, *16*(6), 317–321. https://doi.org/10.1016/j.tics.2012.04.007

Di Russo, F., Martínez, A., & Hillyard, S. A. (2003). Source Analysis of Event-related Cortical Activity during Visuo-spatial Attention. *Cerebral Cortex*, *13*(5), 486–499. https://doi.org/https://doi.org/10.1093/cercor/13.5.486

DiCarlo, J. J., & Maunsell, J. H. R. (2006). Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position.

Journal of Neurophysiology, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The Occipital Place Area Is Causally and Selectively Involved in Scene Perception. *Journal of Neuroscience*, *33*(4), 1331–1336. https://doi.org/10.1523/JNEUROSCI.4081-12.2013

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, *113*(4), 501–517. https://doi.org/10.1037/0096-3445.113.4.501

Duncan, J, & Humphreys, G. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology. Human Perception and Performance*, *18*(2), 578–588; discussion 589-593. https://doi.org/10.1037/0096-1523.18.2.578

Duncan, John, & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037//0033-295X.96.3.433

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology. General*, *123*(2), 161–177. https://doi.org/10.1037/0096-3445.123.2.161

Eimer, M. (2011). The Face-Sensitive N170 Component of the Event-Related Brain Potential. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *The Oxford Handbook of Face Perception* (pp. 329–344). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0017

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, *12*(10), 388–396. https://doi.org/10.1016/j.tics.2008.07.004

Eriksen, C. W., & St James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Perception & Psychophysics*, *40*(4), 225–240. https://doi.org/10.3758/BF03211502

Evans, K. K., & Chong, S. C. (2012). Distributed attention and its implication for visual perception. In J. Wolfe & L. Robertson (Eds.), *From Perception to Consciousness: Searching with Anne Treisman* (pp. 288–296). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734337.003.0026

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychonomic Bulletin & Review*, *20*(6), 1170–1175. https://doi.org/10.3758/s13423-

013-0459-3

Evans, K. K., Horowitz, T. S., Howe, P., Pedersini, R., Reijnen, E., Pinto, Y., … Wolfe, J. M. (2011). Visual attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 503–514. https://doi.org/10.1002/wcs.127

Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When categories collide: accumulation of information about multiple categories in rapid scene perception. *Psychological Science*, *22*(6), 739–746. https://doi.org/10.1177/0956797611407930

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1476–1492. https://doi.org/10.1037/0096-1523.31.6.1476

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*(OCT), 1–12. https://doi.org/10.3389/fpsyg.2011.00243

Fecteau, J. H., & Munoz, D. P. (2006). Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences*, *10*(8), 382–390. https://doi.org/10.1016/j.tics.2006.06.011

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10. https://doi.org/10.1167/7.1.10

Feldman, J. (2013). The neural binding problem(s). *Cognitive Neurodynamics*, *7*(1), 1–11. https://doi.org/10.1007/s11571-012-9219-8

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Gazzaley, A., Rissman, J., Cooney, J., Rutman, A., Seibert, T., Clapp, W., & D'Esposito, M. (2007). Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. *Cerebral Cortex*, *17*(SUPPL. 1), 125–135. https://doi.org/10.1093/cercor/bhm113

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Grill-Spector, K. (2008). Visual Priming. In H. Eichenbaum & J. H. Byrne (Eds.), *Learning and Memory: A Comprehensive Reference* (1st ed., pp. 219–236). San Diego, CA: Academic Press (Elsevier). Retrieved from http://vpnl.stanford.edu/documents/LEME-visualpriming.pdf

Grill-Spector, Kalanit, & Kanwisher, N. (2005). Visual Recognition. *Psychological Science*, *16*(2), 152–160. https://doi.org/10.1111/j.0956-7976.2005.00796.x

Grill-Spector, Kalanit, Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Groen, I. I. a, Ghebreab, S., Prins, H., Lamme, V. a F., & Scholte, H. S. (2013). From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(48), 18814–18824. https://doi.org/10.1523/JNEUROSCI.3128-13.2013

Grootswagers, T., Wardle, S., & Carlson, T. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Gutchess, A. H., & Schacter, D. L. (2012). The neural correlates of gist-based true and false recognition. *NeuroImage*, *59*(4), 3418–3426. https://doi.org/10.1016/j.neuroimage.2011.11.078

Habeck, C. G. (2010). Basics of Multivariate Analysis in Neuroimaging Data. *Journal of Visualized Experiments*, (41), 1–6. https://doi.org/10.3791/1988

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J Wolfe & L. Robertson (Eds.), *From Perception to Consciousness: Searching with Anne Treisman* (pp. 339–349). Oxford University Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=Kw9pAgAAQBAJ&oi=fnd&pg=PA339&dq=Ensemble+Perception:+summarizing+the+scene+and+broadening+the+limits+of+visual+processing&ots=a86oVl7Xzv&sig=zidTJ8yrkovWCEoXf5ecyh3Q4tk

Hall, P., & Wang, D. (2001). *Efficient Visual Search without Top-down or Bottom-up Guidance : A Putative Role for Perceptual Organization*. Retrieved from

https://notendur.hi.is//~ak/TechReport.pdf

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*(8), 1339–1350. https://doi.org/10.1016/S1388-2457(02)00161-X

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439. https://doi.org/10.1016/j.pneurobio.2007.09.001

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-task: Inhibitory executive processes? *Journal of Psychophysiology*, *14*(4), 218–225. https://doi.org/10.1027//0269-8803.14.4.218

Heinze, H. J., Luck, S. J., Mangun, G. R., & Hillyard, S. A. (1990). Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalography and Clinical Neurophysiology*, *75*(6), 511–527. https://doi.org/10.1016/0013-4694(90)90138-A

Heinze, H., Luck, S., Mangun, G., & Hillyard, S. (1990). Visual event-related potentials index focussed attention with bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalography and Clinical Neurophy Siology*, *75*, 511/527. https://doi.org/10.1016/0013-4694(90)90138-A

Heisler, J. M., Morales, J., Donegan, J. J., Jett, J. D., Redus, L., & O'Connor, J. C. (2015). The Attentional Set Shifting Task: A Measure of Cognitive Flexibility in Mice. *Journal of Visualized Experiments*, (96), 2–7. https://doi.org/10.3791/51944

Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1257–1270. https://doi.org/10.1098/rstb.1998.0281

Hochstein, S., & Ahissar, M. (2002). Hierarchies and reverse hierarchies in visual system, *36*(3), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic Control of Visual Attention. *Psychological Science*, *12*(5), 360–365. https://doi.org/10.1111/1467-9280.00367

Hsieh, P., Colas, J., & Kanwisher, N. (2011). Unconscious pop-out: Attentional capture by unseen feature singletons only when top-down attention is available. *Psychological Science*, *22*(9), 1220–1226. https://doi.org/10.1177/0956797611419302.Unconscious

Hua, T., Bao, P., Huang, C. B., Wang, Z., Xu, J., Zhou, Y., & Lu, Z. L. (2010). Perceptual Learning Improves Contrast Sensitivity of V1 Neurons in Cats. *Current Biology*, *20*(10), 887–894. https://doi.org/10.1016/j.cub.2010.03.066

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A., & Klempen, N. (2000). Fractionating the binding process: Neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Research*, *40*(10–12), 1569–1596. https://doi.org/10.1016/S0042-6989(00)00042-0

Intraub, H. (1981). Rapid Conceptual Identification of Sequentially Presented Pictures. *Human Perception and Performance*, *7*(3), 604–610.

Itti, L., Rees, G., & Tsotsos, J. K. (Eds.). (2005). *Neurobiology of Attention*. Elsevier Science/JAI Press.

James, W. (1890). *Principles of Psychology*. Dover Publications.

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, *5*(4), 2. https://doi.org/10.1063/1.4810390

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. https://doi.org/10.1016/j.visres.2007.09.013

Julesz, B. (1971). *Foundations of cyclopean perception*. Oxford, England: U. Chicago Press.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). the Reviewing of Object Files - Object-Specific Integration of Information. *Cognitive Psychology*, *24*(2), 175–219. https://doi.org/10.1016/0010-0285(92)90007-o

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *Journal of Neuroscience*, *36*(41), 10522–10528. https://doi.org/10.1523/JNEUROSCI.1385-16.2016

Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology*, *14*(4), 468–473. https://doi.org/10.1016/j.conb.2004.06.002

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*, *97*, 4296–4309. https://doi.org/10.1152/jn.00024.2007.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

https://doi.org/10.1016/S1364-6613(00)01452-2

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-World Scene Representations in High-Level Visual Cortex: It's the Spaces More Than the Places. *Journal of Neuroscience*, *31*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(4). https://doi.org/10.3389/neuro.06.004.2008

Kristjánsson, Á., Wang, D., & Nakayama, K. (2002). The role of priming in conjunctive visual search. *Cognition*, *85*(1), 37–52. https://doi.org/10.1016/S0010-0277(02)00074-4

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

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

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

MacLean, K. A., Aichele, S. R., Bridwell, D. A., Mangun, G. R., Wojciulik, E. W., & Saron, C. D. (2009). Interactions between Endogenous and Exogenous Attention during Vigilance. *Atten Percept Psychophys*, *71*(5), 1042–1058. https://doi.org/10.3758/APP.71.5.1042.Interactions

Maher, S., Ekstrom, T., Yong, T., Nickerson, L., Frederick, B., & Chen, Y. (2016). Greater sensitivity of the cortical face processing system to perceptually-equated face detection. *Brain Res.*, *15*(1631), 13–21. https://doi.org/10.1117/12.2008529.Image-based

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

Marois, R., Chun, M. M., & Gore, J. C. (2000). Neural correlates of the attentional blink.

Neuron, *28*(1), 299–308. https://doi.org/10.1016/S0896-6273(00)00104-5

Mayhew, S. D., Ostwald, D., Porcaro, C., & Bagshaw, A. P. (2013). Spontaneous EEG alpha oscillation interacts with positive and negative BOLD responses in the visual-auditory cortices and default-mode network. *NeuroImage*, *76*, 362–372. https://doi.org/10.1016/j.neuroimage.2013.02.070

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, *12*(3), 131–143. https://doi.org/10.1002/1097-0193(200103)12:3<131::AID-HBM1010>3.0.CO;2-C

Miller, J. (1989). The control of attention by abrupt visual offsets and onsets. *Perception & Psychophysics*, *45*(6), 567–571.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Nagy, K., Greenlee, M. W., & Kovács, G. (2012). The lateral occipital cortex in the face perception network: An effective connectivity study. *Frontiers in Psychology*, *3*(MAY), 1–12. https://doi.org/10.3389/fpsyg.2012.00141

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (1st ed., pp. 251–256). Academic Press. https://doi.org/10.1016/B978-012375731-9/50045-8

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175. https://doi.org/10.1023/A:1011139631724

Olivares, E. I., Iglesias, J., Saavedra, C., Trujillo-Barreto, N. J., & Valdés-Sosa, M. (2015). Brain Signals of Face Processing as Revealed by Event-Related Potentials. *Behavioural Neurology*. https://doi.org/10.1155/2015/514361

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*(July), 1–27. https://doi.org/10.3389/fninf.2016.00027

Oppermann, F., Hassler, U., Jescheniak, J. D., & Gruber, T. (2011). The Rapid Extraction of Gist—Early Neural Correlates of High-level Visual Processing. *Journal of Cognitive Neuroscience*, *24*(2), 521–529. https://doi.org/10.1162/jocn_a_00100

Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, and Psychophysics*, *76*(7), 2158–2170. https://doi.org/10.3758/s13414-014-0649-8

Orr, C., & Hester, R. (2012). Error-related anterior cingulate cortex activity and the prediction of conscious error awareness. *Frontiers in Human Neuroscience*, *6*(June), 1–12. https://doi.org/10.3389/fnhum.2012.00177

Oruc, I., Cheung, T., Dalrymple, K., Fox, C., Iaria, G., Handy, T., & Barton, J. (2010). Residual face-selectivity of the N170 and M170 is related to the status of the occipital and fusiform face areas in acquired prosopagnosia. *Journal of Vision*, *10*(7), 585. https://doi.org/10.1167/10.7.585.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the Event-Related Potential. *International Journal of Medical Sciences*, *2*(4), 147–154. https://doi.org/10.7150/ijms.2.147

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125–12130. https://doi.org/10.1073/pnas.1101042108

Peelen, M. V, & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/j.tics.2014.02.004

Petersen, S., & Posner, M. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, (35), 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525.The

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., … Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: A combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(12), 2768–2780. https://doi.org/10.1162/jocn.2010.21424

Philiastides, M. G., & Sajda, P. (2007). EEG-Informed fMRI Reveals Spatiotemporal Characteristics of Perceptual Decision Making. *Journal of Neuroscience*, *27*(48), 13082–

13091. https://doi.org/10.1523/JNEUROSCI.3540-07.2007

Philiastides, M., Ratcliffe, R., & Sadja, P. (2006). Neural Representation of Task Difficulty and Decision Making during Perceptual Categorization: A Timing Diagram. *Journal of Neuroscience*, *26*(35), 8965–8975. https://doi.org/10.1523/JNEUROSCI.1655-06.2006

Philiastides, Marios G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*(4), 509–518. https://doi.org/10.1093/cercor/bhi130

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clin Neurophysiol.*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Portilla, J., & Simoncelli, E. P. (2000). Parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49–71. https://doi.org/10.1023/A:1026553619983

Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, M. (1994). Attention: The mechanisms of consciousness. *Proceedings of the National Academy of Sciences*, *91*(August), 7398–7403. Retrieved from http://www.pnas.org/content/91/16/7398.full.pdf

Posner, M I. (1990). Hierarchical distributed networks in the neuropsychology of selective attention. In A. Caramazza (Ed.), *Cognitive Neuropsychology and Neurolinguistics: Advances in Models of Cognitive Function and Impairment* (pp. 187–210). Psychology Press.

Posner, Michael I, Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

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

Potter, Mary C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology*, *3*(32). https://doi.org/10.3389/fpsyg.2012.00032

Potter, Mary C, Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception & Psychophysics*, *76*(2), 270–279. https://doi.org/10.3758/s13414-013-0605-z

Pratt, H. (2012). Sensory ERP Components. In E. Kappenman & S. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (Online edi, pp. 89–114). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195374148.013.0050

Preston, T. J., Guo, F., Das, K., Giesbrecht, B., & Eckstein, M. P. (2013). Neural representations of contextual guidance in visual search of real-world scenes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(18), 7846–7855. https://doi.org/10.1523/JNEUROSCI.5840-12.2013

Procyk, E., & Goldman-Rakic, P. S. (2006). Modulation of Dorsolateral Prefrontal Delay Activity during Self-Organized Behavior. *Journal of Neuroscience*, *26*(44), 11313–11323. https://doi.org/10.1523/JNEUROSCI.2157-06.2006

Raftopoulos, A. (2017). Pre-cueing, the epistemic role of early vision, and the cognitive impenetrability of early vision. *Frontiers in Psychology*, *8*(JUL), 1–12. https://doi.org/10.3389/fpsyg.2017.01156

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology*, *9*(4). https://doi.org/10.1371/journal.pbio.1000608

Rapp, B., & Hendel, S. K. (2003). Principles of cross-modal competition: evidence from deficits of attention. *Psychonomic Bulletin & Review*, *10*(1), 210–219. https://doi.org/10.3758/BF03196487

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliffe, R., Philiastides, M., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG Roger. *PNAS*, *106*(16), 6539–6544. https://doi.org/www.pnas.org/cgi/doi/10.1073/pnas.0812589106

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-

1523.18.3.849

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychological Bulletin*, *124*(3), 372–422.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301–2311. https://doi.org/10.1016/j.visres.2004.04.006

Rensink, R. A. (2000). Scene Perception. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology* (Vol. 7, pp. 151–155). New York: Oxford University Press.

Rensink, Ronald A. (2014). A Function-Centered Taxonomy of Visual Attention. In P. Coates & S. Coleman (Eds.), *Phenomenal Qualities: Sense, Perception, and Consciousness* (pp. 347–375). Oxford University Press.

Ress, D., & Heeger, D. J. (2003). Neuronal correlates of perception in early visual cortex. *Nature Neuroscience*, *6*(4), 414–420. https://doi.org/10.1038/nn1024

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *19*(5), 1736–1753. Retrieved from http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10024360&retmode=ref&cmd=prlinks%5Cnpapers2://publication/uuid/072C6D73-06FE-4988-9AD6-C4D0C608AE77

Ritchie, J. B., Kaplan, D. M., & Klein, C. (2017). Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience. *The British Journal for the Philosophy of Science*, *00*(July), 1–27. https://doi.org/10.1093/bjps/axx023

Robertson, L., Treisman, A., Friedman-Hill, S., & Grabowecky, M. (1997). The Interaction of Spatial and Object Pathways: Evidence from Balint's Syndrome. *Journal of Cognitive Neuroscience*, *9*(3), 295–317. https://doi.org/10.1162/jocn.1997.9.3.295

Rodriguez, V., Thompson, R., Stokes, M., Brett, M., Alvarez, I., Valdes-Sosa, M., & Duncan, J. (2011). Absence of Face-specific Cortical Activity in the Complete Absence of Awareness: Converging Evidence from Functional Magnetic Resonance Imaging and Event-related Potentials. *Journal of Cognitive Neuroscience*, *24*(2), 396–415. https://doi.org/10.1162/jocn

Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, *29*(March), 203–227.

https://doi.org/10.1146/annurev.neuro.29.051605.112939

Rosen, A. C., Rao, S. M., Caffarra, P., Scaglioni, A., Bobholz, J. A., Woodley, S. J., … Binder, J. R. (1999). Neural basis of endogenous and exogenous spatial orienting. A functional MRI study. *Journal of Cognitive Neuroscience*, *11*(2), 135–152. https://doi.org/10.1162/089892999563283

Rosenthal, O., & Humphreys, G. W. (2010). Perceptual organization without perception: The subliminal learning of global contour. *Psychological Science*, *21*(12), 1751–1758. https://doi.org/10.1177/0956797610389188

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

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

Rutman, A. M., Clapp, W. C., Chadick, J. Z., & Gazzaley, A. (2010). Early top-down control of visual processing predicts working memory performance. *Journal of Cognitive Neuroscience*, *22*(6), 1224–1234. https://doi.org/10.1162/jocn.2009.21257.

Saarinen, J. (1996). Localization and discrimination of "pop-out" targets. *Vision Research*, *36*(2), 313–316. https://doi.org/10.1016/0042-6989(95)00093-F

Scholte, H. S., Ghebreab, S., Waldorp, L., Smeulders, A. W. M., & Lamme, V. A. F. (2009). Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, *9*(4), 29–29. https://doi.org/10.1167/9.4.29

Scholte, H Steven, Jolij, J., Fahrenfort, J. J., & Lamme, V. A. F. (2008). Feedforward and recurrent processing in scene segmentation : EEG and fMRI ., 2097–2109.

Schomer, D., Bonmassar, G., Lazeyras, F., Seeck, M., Blum, A., Anami, K., … Ives, J. (2000). EEG-Linked Functional Magnetic Resonance Imaging in Epilepsy and Cognitive Neurophysiology. *Journal of Clinical Neurophysiology*, *17*(1), 43–58.

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences*, *105*(11), 4447–4452. https://doi.org/10.1073/pnas.0800431105

Seidl, K. N., Peelen, M. V, & Kastner, S. (2012). Neural evidence for distracter suppression

during visual search in real-world scenes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(34), 11812–11819. https://doi.org/10.1523/JNEUROSCI.1693-12.2012

Seitz, A. R., & Dinse, H. R. (2007). A common framework for perceptual learning. *Current Opinion in Neurobiology*, *17*(2), 148–153. https://doi.org/10.1016/j.conb.2007.02.004

Serences, J. T., & Kastner, S. (2014). A multi-level account of selective attention. *The Oxford Handbook of Attention*, 76–103. https://doi.org/10.1093/oxfordhb/9780199675111.013.022

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

Sharma, R., Joshi, S., Singh, K. D., & Kumar, A. (2015). Visual evoked potentials: Normative values and gender differences. *Journal of Clinical and Diagnostic Research*, *9*(7), 12–15. https://doi.org/10.7860/JCDR/2015/12764.6181

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency wavesevoked by unpredictable auditory stimuli in man. *Electroencephalography & Clinical Neurophysiology*, *38*, 387–401.

Stolarova, M., Keil, A., & Moratti, S. (2006). Modulation of the C1 visual event-related component by conditioned stimuli: Evidence for sensory plasticity in early affective perception. *Cerebral Cortex*, *16*(6), 876–887. https://doi.org/10.1093/cercor/bhj031

Swainson, R., Cunnington, R., Jackson, G. M., Rorden, C., Peters, A. M., Morris, P. G., & Jackson, S. R. (2003). Cognitive control mechanisms revealed by ERP and fMRI: Evidence from repeated task-switching. *Journal of Cognitive Neuroscience*, *15*(6), 785–799. https://doi.org/10.1162/089892903322370717

Tavares, G., Perona, P., & Rangel, A. (2017). The attentional Drift Diffusion Model of simple perceptual decision-making. *Frontiers in Neuroscience*, *11*(AUG), 1–16. https://doi.org/10.3389/fnins.2017.00468

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural Systems for Error Monitoring. *The Neuroscientist*, *13*(2), 160–172. https://doi.org/10.1177/1073858406298184

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-

295X.113.4.766

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, *12*(4), 242–248. https://doi.org/10.1080/17470216008416732

Treisman, AM, & Gelade, G. (1980). A Feature-Integration Theory of Attention. *Cognitive Psychology*, *12*, 97–136. Retrieved from http://www.sciencedirect.com/science/article/pii/0010028580900055

Treisman, Anne. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex*, *24*(4), 883–897. https://doi.org/10.1093/cercor/bhs364

Tsotsos, J. K., Culhane, S. M., Wai, W. Y. K., Lai, Y. H., Davis, N., & Nuflo, F. (1995). Modeling Visual-Attention Via Selective Tuning. *Artificial Intelligence*, *78*(1–2), 507–545. https://doi.org/10.1016/0004-3702(95)00025-9

VanRullen, R. (2009). Binding hardwired versus on-demand feature conjunctions. *Visual Cogn*, *17*(1–2), 103–119. https://doi.org/10.1080/13506280802196451

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.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203. https://doi.org/10.1017/S0048577200981265

Vogels, R., & Orban, G. A. (1994). Does Practice in Orientation Discrimination Lead to Changes in the Response Properties of Macaque Inferior Temporal Neurons? *European Journal of Neuroscience*, *6*(11), 1680–1690. https://doi.org/10.1111/j.1460-9568.1994.tb00560.x

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009

Walther, Dirk B, Beck, D. M., & Fei Fei, L. (2012). To Err Is Human: Correlating fMRI Decoding and Behavioral Errors to Probe the Neural Representation of Natural Scene Categories. In N. Kriegeskorte & G. Kreiman (Eds.), *Visual Population Codes: Toward a Common*

Multivariate Framework for Cell Recording and Functional Imaging (pp. 391–416). London, England: Massachusetts Institute of Technology. https://doi.org/10.7551/mitpress/8404.003.0018

Watanabe, T., Náñez, J., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, *413*, 844–848.

Weidner, R., Shah, N. J., & Fink, G. R. (2006). The neural basis of perceptual hypothesis generation and testing. *Journal of Cognitive Neuroscience*, *18*(2), 258–266. https://doi.org/10.1162/jocn.2006.18.2.258

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64. https://doi.org/10.1037/0096-3445.131.1.48

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*(2), 70–76. https://doi.org/10.1016/S1364-6613(02)00024-4

Wolfe, JM, Võ, M., Evans, K., & Greene, M. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2). https://doi.org/10.1016/j.tics.2010.12.001

Woodman, G. F., & Luck, S. J. (2003). Serial deployment of attention during visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(1), 121–138. https://doi.org/10.1167/1.3.103

Wu, C. C., Wang, H. C., & Pomplun, M. (2014). The roles of scene gist and spatial dependency among objects in the semantic guidance of attention in real-world scenes. *Vision Research*, *105*, 10–20. https://doi.org/10.1016/j.visres.2014.08.019

Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*(FEB), 1–13. https://doi.org/10.3389/fpsyg.2014.00054

Wyart, V., & Tallon-Baudry, C. (2008). Neural dissociation between visual awareness and spatial attention. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(10), 2667–2679. https://doi.org/10.1523/JNEUROSCI.4748-07.2008

Yang, T., & Maunsell, J. H. (2004). The effect of perceptual learning on neuronal responses in monkey visual area V4. *J.Neurosci.*, *24*(1529–2401), 1617–1626. https://doi.org/10.1523/JNEUROSCI.4442-03.2004

Zelinsky, G. J. (2008). A Theory of Eye Movements during Target Acquisition. *Psychological Review*, *115*(4), 787–835. https://doi.org/10.1037/a0013118.A