Bangor University

DOCTOR OF PHILOSOPHY

Neurocognitive investigation of objectin-scene representations

Chandler, Hannah

Award date:
2017

Awarding institution:
Bangor University

Link to publication

Neurocognitive investigation of object-in-scene representations

Hannah Louise Chandler

A Thesis submitted for the degree of Doctor of Philosophy (2017)

Bangor University, School of Psychology, UK
Affiliated with Cardiff University, School of Psychology, UK

Acknowledgements

First, I would like to thank my primary supervisor Professor Paul Downing who has been a continued support throughout. Second, I would like to thank Professor Emily Cross for her valuable feedback and her guidance. I have been affiliated with Bangor University for the past 7.5 years. As this Chapter is about to end, I cannot thank the University and Psychology department enough. Without your guidance and passion for science, I would not be here now. Thank you.

I would also like to thank Professor Richard Wise at Cardiff University, for taking me on in the next steps of my career as a research associate. I am very excited for this next chapter and applying all I have learned to examine quantitative methods of MRI and PET in health and disease.

Most importantly, I would like to thank my family. Without you, I would not be here. You are simply the best! Specifically, I would like to thank my parents Kathryn and Simon. I will forever be thankful to you for your continued support, love and guidance throughout my life. It means everything to be able to share this with you. I also thank my younger (but much taller!!) brother Matthew for always being there for me, I am so truly proud of all you are achieving. To my partner Tom, I am thankful for your support, not only throughout the PhD, but also during every other area of my life. I look forward to our life together, exploring the world and having exciting careers in neuroscience.

Finally, I would like to thank my grandparents. You have taught me determination, strength and commitment, which I will take with me throughout every aspect of my life. Thank you.

Thesis Summary

Organisms are required to perceive, process and remember a wealth of visual information from the environment to guide behaviour during spatial navigation. However, our knowledge is limited regarding how the brain encodes and reconstructs in memory, spatial and non-spatial properties of objects in scenes. For example, how are object locations, arrangements and identities encoded and represented across core scene-selective regions? How does the identity of a focal entity influence memory for the spatial extent of a scene? This thesis used functional magnetic resonance imaging (fMRI) and behavioural approaches across 4 independent experiments to investigate these themes. Empirical Chapters 1 and 2 employed fMRI repetition suppression (RS) to examine how activation across scene-selective regions differed in response to spatial (object Locations and Arrangements) and non-spatial (object Identities) conditions. Results revealed no effect of RS in any ROI, considered to reflect the type of task used (inversion detection). The second fMRI experiment employed a novel task, where participants responded to multiple changes between scenes. Results showed a significant effect of RS in two regions, but no dissociable effects between conditions. In two behavioural chapters, we extended these themes by using boundary extension (BE), to investigate whether memory for the spatial extent of a scene is influenced by the type of entity (object/person). Results revealed that BE was observed in both experiments for objects in scenes (in line with previous research), but not for people in scenes. Further analyses demonstrated that this effect might reflect the increase in attention assigned to people compared to objects, possibly to predict their future actions/behaviours. Together, this body of research provides insight into the mechanisms that drive RS during scene encoding, and identifies that possible differences in saliency associated with people and objects may mediate how the spatial extent of a scene is encoded and subsequently remembered.

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Preface

The human visual system provides a window for observing a rich and dynamic world. When glancing out of the window in my office I am able to recognise a row of old Georgian houses, a path, on-street parking and various objects such as cars, people, rubbish bins, a bench, and a post box. On a daily basis, the structural parts of my view remain constant (houses, path, parking spaces) but the people sat on the bench are different, the cars parked on the street are different, and the rubbish bin is usually in a different location. Thus, whilst the scene structure itself stays constant in its global configuration the entities within the scene change in their identities, locations and arrangements.

If I were to walk along this street I would encounter objects as well as people on my journey. I am required to compute the locations of others, predict their actions, and identify/locate surrounding inanimate objects. Comprehension of the visual world demands a high degree of spatial awareness and attention, which allows an individual to navigate from point A to point B without crossing paths with people or objects within the environment.

As these examples demonstrate, the visual world is complex and dynamic, requiring accurate computation and encoding of each spatial and non-spatial component. We encounter these types of experiences every day, in multiple instances. Yet, scientifically, our knowledge about how the brain encodes this type of information is far from absolute. The current thesis follows two lines of investigation across four experiments to determine how variations in scene content are encoded and remembered. First, how does the brain encode information about the locations arrangements and identities of multiple objects in natural scenes? Second, how do variations in scene content influence VSTM for the space around a scene?

1.1. Setting the scene

When viewing the world, organisms are presented with a wealth of visual information. This typically includes spatial elements such as geometric properties

(Oliva & Torralba, 2001) and non-spatial elements such as people and objects (Bar, 2004; Biederman, Mezzanotte & Rabinowitz, 1982). Such information is subsequently integrated across time, space and memory, providing an internal mental representation of the world (Tolman, 1947).

The computation and extraction of such visual information largely influences one's ability to interact within the wider environment (Andrew Browning, Grossberg & Mingolla, 2009). As such, a key motivation for researchers is to determine the contribution of sensory information in facilitating visual recognition and memory for space. For example, what is the importance of objects in our environment? How do we account for obstacles when moving through the world? What kind of spatial information supports navigation? Asking these questions enables researchers to infer how humans utilize different forms of visual information in order to guide their behavior in the world. However, reproducing such a vast and dynamic visual world in laboratory settings presents researchers with practical challenges. One method to overcome these issues is to study how the brain encodes small-scaled static images of single viewpoints known as scenes (Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1999).

One of the most complex and debated questions in spatial cognition is: what is a scene? Some have described a scene as a representation of a real-world view that includes geometric elements, textures, surfaces, and objects (Henderson & Hollingworth, 1999). Critically, the definition of a scene can depend on the spatial scale and viewpoint. Henderson and Hollingworth describe an instance where the scene content becomes a part of the scene background. In one scene, a table may be an object featuring within a room. However, if the photo is taken from a closer vantage such as the top of the table it then becomes part of the global scene structure (the floor). Thus, the point at which a scene is defined must be contingent upon the spatial scale of the view.

1.2. Scene *gist* encoding

Scenes are inherently complex but can be very readily categorised. The meaning, or *gist* of a scene can be encoded in fractions of a second (Biederman, Rabinowitz, Glass & Stacy, 1974; Greene, Botros, Beck & Fei-Fei, 2015; Greene & Oliva, 2009a; Joubert, Potter, 1975; Rousselet, Fize & Fabre-Thorpe, 2007;

VanRullen & Thorpe, 2001). This ability relies upon the computation of both conceptual and perceptual information (Aminoff & Tarr, 2015; Oliva & Torralba, 2001) and can be observed for both intact scenes and scenes that appear visually degraded (Torralba, 2009; Schyns & Oliva, 1994).

What forms of information are required to compute the meaning of a scene? The notion that object identification is necessary for accurate scene categorisation is still heavily debated (Bar, 2004; Greene & Oliva, 2009a; Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1999; Oliva & Torralba, 2001, 2006). However, recent insights suggest that scene categorisation is more accurate when the objects and the scene background are contextually and semantically related (Davenport & Potter, 2004; Greene, Botros, Beck & Fei-Fei, 2015). For example, categorisation of a bathroom will be more accurate if it includes a bath and toilet as opposed to a bicycle and a chair. Conversely, the *gist* of a scene can still be determined from global spatial primitives (Oliva & Torralba, 2006). For example, small rooms are likely to be associated with single bedrooms or home offices whereas, open natural views are more likely to be categorised as parks or beaches (Henderson, Larson & Zhu, 2007; Oliva & Torralba, 2006).

Critically, visual scene encoding requires computation and integration of a wealth of rapidly occurring visual information in order to determine the meaning of a scene. These computations are thought to arise from both structural cues and object based cues.

1.3. Cues within the environment

There are two broader categories of information recognised in the world. First, spatial information provides insight about *where* we are including structural landmarks and the geometry of a given environment (Bar, 2004; Oliva & Torralba, 2001, 2002). Second, non-spatial information informs us about *what* is in our environment, which includes entities such as objects and people (Bar, 2004; Harel, Groen, Kravitz, Deouell & Baker, 2016; Kaiser, Stein & Peelen, 2014; Malcolm & Shomstein, 2015; Mayer, Vuong & Thornton, 2015; McDunn, Siddiqui & Brown, 2014; Sastyin, Niimi & Yokosawa, 2015; van Koningsbruggen, Peelen & Downing, 2013).

1.3.1. Geometric/spatial cues

In the field of scene processing, evidence suggests that processing the gist of a scene relies largely on geometric properties of a view. This may refer to the size/shape of the space (Oliva & Torralba, 2001, 2002, 2006), the perceived distance (Burge, Fowlkes & Banks, 2010; Nasr & Tootell, 2016), or the openness of a space and its degree of naturalness (Greene & Oliva, 2009; Harel, Kravitz & Baker, 2013; Park, Brady, Greene & Oliva, 2011). These forms of cues are considered to provide meaning and context to a scene. Scenes containing similar forms of spatial cues are thought to be categorised as similar. For example, built up scenes containing natural content may be categorised as a view of a busy city street. However, if the scene is built up but contains natural content it may be categorised as a woodland view (Park & Chun, 2014; Park, Brady, Greene & Oliva, 2011). This shows that the geometric structure of a space can provide information about the meaning or category of a scene.

In addition, recent research has investigated the importance of scene layout for supporting scene processing. In particular, it has been evidenced that brain structures involved in processing scene information in the medial temporal lobes (MTL) elicit higher levels of fMRI activation in response to intact rooms as opposed to fractured rooms (room that are displayed with walls fractured into separate pieces) (Kamps, Julian, Kubilius, Kanwisher & Dilks, 2016). These observations demonstrate that the MTL encodes scenes as holistic geometric units as opposed to fragmented scene parts, which might reflect the role of these regions in encoding information relevant for spatial navigation.

How do geometric cues help organisms encode their environment? Early animal models illustrate the importance of scene geometry in facilitating facing direction and orientation within a space during spatial navigation (Cheng, 1982). In this study rodents were primed to receive food in two locations of the room. Their observations demonstrated that the rodents were able to reorient to these locations using geometric cues including the structure of the space as a frame of reference. This suggests that the spatial structure and geometric properties of a space guide spatial navigation and spatial memory for object locations.

Geometric cues may inform an organism of their precise location, heading direction (the direction that the individual is heading/facing in their environment), and the distance they would need to travel from point A to point B (Baumann, Chan

& Mattingley, 2010; Epstein & Vass, 2014; Maguire et al., 1998). For example, in a built-up scene, the layout of the structures will be encoded in memory to provide an overall spatial framework to ensure the organism can effectively navigate through the space.

1.3.2. Landmark cues

The definition of what constitutes an object or a spatial landmark remains controversial (Aguirre & D'Esposito, 1999; Chan, Baumann, Bellgrove & Mattingley, 2012; Tommasi et al., 2012). Some objects are considered structural landmarks if they afford a high measure of spatial permanence (their degree of 'fixedness' to a location), appear highly space defining (objects that encompass a large degree of space), or are large in physical size (Troiani, Stigliani, Smith & Epstein, 2014). Here, I describe landmarks as navigationally relevant space defining units that provide information about one's location within the environment (Epstein, Harris, Stanley & Kanwisher, 1999; Marchette, Vass, Ryan & Epstein, 2015; Vass & Epstein, 2013). Critically, prior evidence has identified landmarks as important units for guiding action and behaviour during navigation (Ganel & Goodale, 2003; Grezes, Tucker, Armony, Ellis & Passingham, 2003; Janzen & van Turennout, 2004; Kim, & Nevatia, 1994).

There are typically two types of landmark cues: global and local landmarks (Steck, 2000). Global landmarks are structures used to guide a person's navigational route and infer their heading direction (the direction they face relative to other external elements) (Marchette, Vass, Ryan & Epstein, 2014). For example, an individual might use a landmark as a guide for their travelling direction (i.e. to travel north, the landmark needs to be on the observes left side). Patients with lesions in brain structures thought to be scene-selective (higher level of fMRI signal to scenes compared to other visual stimulus categories), showed intact scene and landmark identification, but were impaired in using landmarks to reorient themselves and navigate to a given location (Aguirre & D'Esposito, 1999; Maguire, 2001). Therefore, landmarks are critical spatial cues aiding navigation and orientation within the world.

Conversely, local landmarks are considered smaller entities that influence immediate goal directed behaviour (Steck, 2000). These types of landmarks include objects such as monuments, buildings of reference, and post box like entities.

However, the degree of variance between local landmarks in their navigational prominence is less clear. For example, in order to post a letter, a person would be required to find the nearest post box. This type of object is relatively permanent, guides goal directed behaviour, and could be used as a reference to guide someone to a particular house on a given street. However, the same post box might not be considered prominent enough to be a landmark, used to describe a particular location in a city, as would a particular museum. Thus, while the use of landmarks is critical in guiding behaviour, it is important to consider that there are varying types of landmark cues, such as objects with varying degrees of permanence.

1.3.3. Object Cues

Unlike space, which an organism is able to navigate *within*, objects are units that we interact *with* and in many instances, serve a particular function (Epstein, Higgins & Thompson-Schill, 2005). Thus, the ability to accurately identify and locate objects allows organisms to interact with them within the environment. Non-landmark objects include items that may serve a particular function such as bags, shoes, kettles, mugs, utensils and laptops/mobile devices. Given that these types of objects are unlikely to remain in a permanent location, they may not inform us of our relative location within an environment. However, they may inform us of our context (Davenport, 2007).

Non-landmark objects, when linked to a particular context, are bound to episodic memory representations that facilitate scene recognition and categorisation (Hollingworth & Henderson, 2002; Hollingworth & Henderson, 2004; Hollingworth, Williams & Henderson, 2001). While some objects hold strong contextual associations (lawnmower in a garden), others may be more strongly associated with particular location within a scene (kettle on a kitchen work surface) (Hollingworth, 2006). In this instance, syntactic knowledge about object locations may increase the efficiency of visual search for an object, which in turn facilitates goal directed behaviour (Biederman, Mezzanotte, Rabinowitz, 1982), and may be derived from prior experience, object-context relations and amodal perception (Intraub, 2010; Vo & Henderson, 2011). Together, these findings demonstrate the importance of objects in scenes, and the information that they provide in guiding our understanding of the visual environment.

1.3.4. People/animals

Besides inanimate objects, we also encounter animals and people in the environment. Humans extract several forms of information from others; including identity, body language, facial expressions, and predict their future actions (Adolphs, 2003a, 2003b; Gallese, Keysers & Rizzolatti, 2004). Whilst a plethora of work has examined both face perception (Archer, Hay & Young, 1994; Bruce & Young, 1986; Kanwisher, McDermott & Chun, 1997; Vuilleumier, 2000; Xu, Liu & Kanwisher, 2005; Yovel & Kanwisher, 2004) and body perception (Downing, Bray, Rogers & Childs, 2004; Downing, Jiang, Shuman & Kanwisher, 2001; Peelen & Downing, 2005), the focus has been to determine how these entities are encoded based on their individual features. Thus, our knowledge regarding how animate entities are encoded within natural scene contexts and their influence on spatial encoding remains elusive.

Research examining how people are processed in scenes has shown that eye-fixations are initially drawn to the centre of a scene, and are then drawn to people (Bindemann, Scheepers, Ferguson & Burton, 2010). In their study, Bindemann and colleagues used a detection task to infer how individuals encode people in scenes. They identified that participants' reaction times (RT) to scenes containing bodies (with no face) and faces (with no body) were comparable, however, participants favourably fixated on individual faces. The authors conclude that while faces are predominantly important for recognition of people in scenes, both the body and face play an important role in person detection.

Interestingly, (Mayer, Vuong & Thornton, 2015) developed a study to examine whether people 'pop-out' in scenes. Participants were exposed to scenes containing people or machines and were required to search for a scene containing the target entity (person or machine) from an array of scenes. Results showed that RT's were longer when the number of stimuli (load) increased for both conditions, suggesting that there was no difference between search time for scenes containing people or objects. However, their eye tracking data revealed that participants initially fixated more to people in scenes than to the machines. Whilst the results of this particular study show that search and detection for people in scenes is more efficient compared to other objects such as machines.

Together, this research suggests that the occurrence of a person influences the locations of eye-fixations across a scene, however, it does not determine whether encoding of global spatial information within a scene changes due to the presence of a person. More broadly, these studies suggest that spatial attention mechanisms may differ when people are presented in scenes compared to objects.

1.4. Key accounts of scene perception

Various accounts of scene representation have attempted to explain how scenes are processed. Many initial models debate whether scenes are encoded as a function of their objects and textures (content), or their global spatial properties (context) (Alvarez, 2011; Bar, 2004; Bettencourt & Xu, 2013; Biederman, Mezzanotte & Rabinowitz, 1982; Greene & Oliva, 2009b; Hollingworth & Henderson, 1999; Oliva & Torralba, 2001; Renninger & Malik, 2004; Serences, Schwarzbach, Courtney, Golay & Yantis, 2004).

1.4.1. Object-centred and scene-centred

The *object-centred* approach suggests that objects in scenes are initially encoded, and that this identification provides meaning to the scene (Biederman, 1987; Biederman, Mezzanotte, & Rabinowitz, 1982; Park, Konkle & Oliva, 2015; Schyns & Oliva, 1994) (Figure 1.1). Early studies demonstrated that scene processing is driven by object recognition and spatial relations between objects (De Graef, Christiaens & d'Ydewalle, 1990). For example, objects may be grouped in order to determine a context, which does not rely upon the scene background itself (a kettle, toaster and microwave are going to be linked more strongly to a kitchen than a bathroom).

Evidence from target detection and visual search literature, has revealed that objects influence the guidance of eye movements across scenes (Peelen, Fei-Fei & Kastner, 2009; Guo, Preston, Das, Giesbrecht & Eckstein, 2012). Eye-tracking evidence suggests that attention/knowledge to an objects function can improve reaction times during visual search for an object (Castelhano & Witherspoon, 2016). Specifically, visual search for objects is faster when an object appears in a location congruent with its function, suggesting that knowledge of the object identity and function drives search to relevant locations. In this instance, knowledge of the objects function guides visual search and facilitates the efficiency of subsequent goal directed behaviours.

Moreover, evidence suggests that scene encoding involves processing both objects and spatial properties alongside each other. This is evident from literature suggesting that scenes are categorised more rapidly when the object and the context are related (Davenport & Potter, 2004; Park & Chun, 2014). So, could it be suggested that the spatial-based properties are just as crucial as object-based properties for scene processing?

The *scene-centred* approach suggests that the brain does not process scenes as a function of their individual objects *per se*, but models the shape of the space, extracting multiple geometric, volumetric and object properties from the image (Oliva & Torralba, 2001). In this model, the objects and the spatial properties of the image are computed in parallel but independently. For example, the theory proposes that there are two distinct descriptors of scene information that can be extracted, 1) external features as defined by boundaries and wider spatial components, and 2) internal features such as the type of scene content and layout. This approach suggests that scenes are recognised by their global characteristics; such as openness, natural/manmade, roughness and mean depth (Oliva & Torralba, 2001, 2006; Park et al., 2011; Ross & Oliva, 2010). For example, a cityscape may have large manmade structures, be relatively un-natural and score low on its degree of openness (Park et al., 2011).

Together, these accounts of scene representation bring to light some important points relevant for the current thesis. For example, they suggest that scenes are encoded either by their global image characteristics or the local components. This constricts our ability to extend these models to study how the brain processes spatial properties of objects within scenes that do not hold strong contextual associations. Likewise, these models do not account for more social properties of scenes such as people, and whether the presence of a person within a scene influences how the global spatial structure of a scene is encoded.

However, emerging models have shown that scenes are encoded beyond basic visual input highlighting that prior contextual knowledge is also crucial to form a coherent and contextually rich representation of a space (Intraub, 2010; Linsley & MacEvoy, 2015; MacEvoy & Epstein, 2007, 2011).

Object-centred

Building, rocks, cars, grass/plants and sky.

Scene-centred

Large space, natural scene, open space.

Figure 1.1. Example of the types of information extracted for a scene-centred and object-centred approach to scene processing.

1.4.2. Scene encoding - from visual input to short-term memory reconstruction

While scene-centred and object-centred accounts propose that scenes are categorised/processed as a function of both global spatial properties and object-based properties, it is important to consider how these components contribute to scene encoding and the construction of scenes in short-term memory.

In many instances scene processing is considered the result of a single visual input, or a single-source. This single input relies on information presented in the image, and does not draw upon other modalities to inform what might occur beyond the scene borders. However, there are instances and phenomena such as the *boundary extension* error, for which the single-source model cannot account.

Boundary extension (BE) is a memory error where the brain extrapolates information beyond the occluded borders of a scene, via amodal perception (Gottesman & Intraub, 1999; Intraub & Berkowits, 1996; Intraub, Gottesman, Willey & Zuk, 1996; Intraub & Richardson, 1989). BE occurs when participants falsely report a larger expanse of space than was shown in the original image and is thought to reflect amodal perception of space (see Chapter 2 for methodological discussion of BE). The *multi-source* model accounts for the effects observed in BE, proposing that scenes are processed as a function of multiple inputs (Intraub & Dickinson, 2008; Intraub, 2010) such as visual input, amodal continuation of space, and learned object-context relationships. How does this help humans to view the world in a continuous manner?

During eye movements, vision is briefly suppressed resulting in the world being viewed as an array of continuous snapshots. Evidence suggests that the role of spatial and object VSTM is to help support the completion of a coherent and consistent view of the world (Currie, McConkie, Carlson-Radvansky & Irwin, 2000). This is supported by the multi-source model of scene representation where it is thought that these 'gaps' in vision are filled via a modal perception and knowledge of object-context relationships (Intraub, 2010, 2012). Moreover, the BE error may reflect the process of binding visual representations from VSTM across views to form a coherent and continuous account of the world.

BE has been observed under many conditions, but has critically been shown to only occur for objects within scenes as opposed to objects isolated from a scene (Gottesman & Intraub, 2002). Therefore, suggesting that BE itself relies upon both global contextual scene elements and object information. Moreover, BE is observed across saccades, suggesting that it may facilitate the integration of consecutive scene viewpoints, allowing the world to be viewed as a constant, stable construct.

Further to the lines of investigation discussed here, additional evidence has examined the integration of this information across brain networks thought to be involved in scene encoding.

1.5. Brain based evidence of object-in-scene representations

1.5.1. Ventral versus dorsal processing streams

Traditional theories of visual processing have identified two interconnecting visual processing pathways, referred to as the dorsal visual stream (DVS), and the ventral visual stream (VVS) (Creem & Proffitt, 2001; Ungerleider & Haxby, 1994) (Figure 1.2). The DVS is thought to be involved in processing *where* items are in relation to the user, and is thought to run from the early visual cortex dorsally through to wider occipital and parietal regions. Conversely, the VSS is thought to be involved in processing *what* items are (object processing pathway), running from the early visual cortex, down toward the anterior temporal and medial temporal lobe.

Despite evidence of these processing streams across species (Wang, Sporns & Burkhalter, 2012), it is now considered that they are less distinct than previously proposed (Baldassano, Esteva, Fei-Fei & Beck, 2016; Zachariou, Klatzky & Behrmann, 2014). For example, emerging evidence has suggested that regions involved in processing scenes (traditionally VVS regions) encode a wealth of information, including object identity properties and spatial properties of objects. In the following sections I review this literature with a specific focus on three regions considered critical in scene processing.

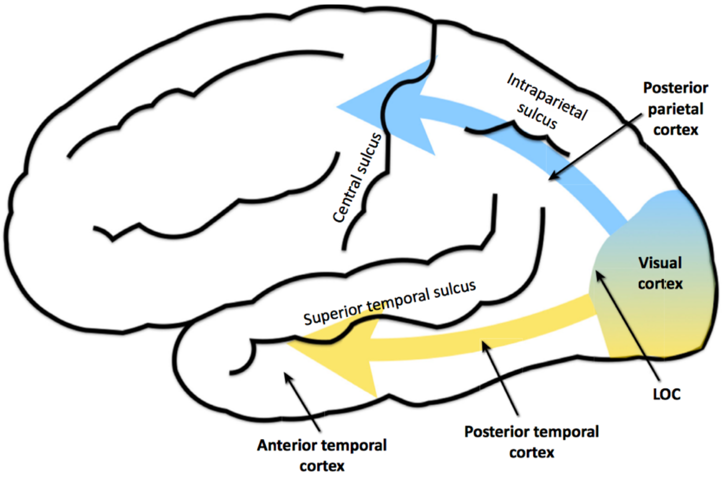


Figure 1.2. Illustration of the ventral visual pathway (yellow) and dorsal visual pathway (blue) in the human brain. Figure from Wilcox & Biondi (2015).

1.5.2. Anatomy of scene-selective network

With the development of brain imaging techniques such as functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation (TMS) and electroencephalography (EEG), it is possible to infer the functional and temporal correlates of scene representation in the human brain. There are three predominant regions considered scene-selective due to their unique hemodynamic properties associated with processing visual images of scenes, in comparison to other visual categories such as faces, bodies and objects. These include the parahippocampal place area (PPA), retrosplenial cortex (RSC) and the occipital place area (OPA, formerly referred to as the transverse occipital sulcus (TOS) (Dilks, Julian, Paunov & Kanwisher, 2013; Epstein & Kanwisher, 1998; Epstein & Higgins, 2007; Maguire, 2001; Maguire et al., 1998; Nasr et al., 2011) (Figure 1.3).

There are two key stances towards the kind of information that these regions represent. Some suggest that these regions are specialised nodes that specifically encode information about a particular category of stimuli (Epstein & Kanwisher, 1998), whereas others suggest that these regions are involved in the computation of information across a number of shared feature dimensions (Harel, Kravitz & Baker, 2013; Kravitz, Peng & Baker, 2011; Troiani, Stigliani, Smith & Epstein, 2014). For example, emerging evidence suggests that the core scene-selective structures also encode information about the spatial properties of objects (Aminoff & Tarr, 2015; Bainbridge & Oliva, 2015; Linsley & MacEvoy, 2015; Park, Konkle & Oliva, 2015; Troiani, Stigliani, Smith & Epstein, 2014). However, our knowledge regarding how spatial and non-spatial information about objects is encoded in these scene-selective structures is still largely unknown.

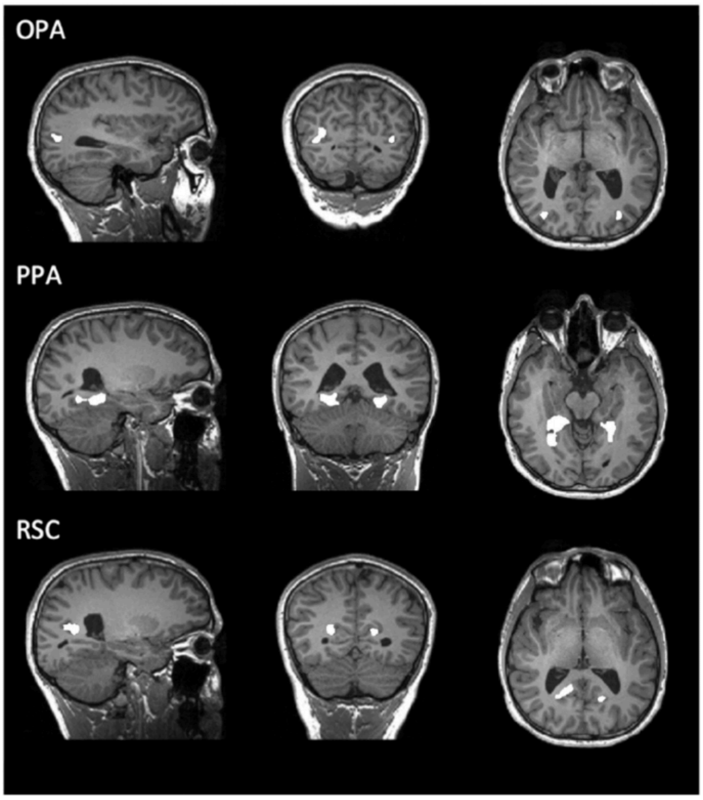


Figure 1.3. Illustration of the core scene-selective regions (OPA; PPA; RSC) represented in the human brain, highlighted as the white areas in the cortical structure. Image from Kamps, Julian, Kubilius, Kanwisher and Dilks (2016).

1.5.2.1. Parahippocampal place area (PPA)

The PPA is a functionally defined region situated in the parahippocampal cortex (PHC), between the hippocampus and the fusiform gyrus. A similar region has been observed in the monkey brain. The macaque homolog of the PPA (mPPA) is situated ventrally and toward the anterior portion of the medial temporal sulcus

(Nasr et al., 2011; Rajimehr et al., 2011) and is also thought to encode visual scene based information.

In an early human fMRI study, Epstein & Kanwisher (1998) identified that the PPA elicited higher levels of activation in response to images of scenes compared to bodies, faces and objects (fMRI is a method that examines the hemodynamic properties of the brain by examining changes in blood oxygen level dependent signal, refer to Chapter 2 for a methodological overview). In addition, the authors found that the PPA responded almost as significantly to images of landmark-based structures such as buildings and houses. Conversely, the PPA only responded weakly to images of small objects such as appliances and furniture.

Moreover, the PPA has been implicated as an important region for processing geometric features, such as the distance between elements within a scene (far/near), the content of the scene (natural versus man-made scenes), and the geometric structure/layout of the scene (open or closed depiction) (Harel, Kravitz & Baker, 2013; Kravitz, Peng & Baker, 2011). Additionally, researchers have also questioned the role of the PPA in encoding viewpoint based information.

In their seminal study, Epstein, Graham and Downing (2003) used fMRI repetition suppression to investigate whether the PPA encodes viewpoint in 15 healthy participants. The authors presented participants with consecutive pairs of stimuli and examined changes in the hemodynamic response across the images (refer to Chapter 2 for detailed description of RS). Participants were presented with changes in the object identity (whilst keeping the scene the same), a place change (keeping the object the same but changing the background scene), viewpoint change (same object from a different viewpoint), and no change (repetition). Results showed a similar level of activation (relative to repetition condition) for viewpoint changes and entirely new scenes changes. Thus, indicating that the PPA holds viewpoint specific representations. Results also revealed that the PPA does not encode information about object identity. Conversely, the lateral occipital cortex (LOC, involved primarily in object processing) showed sensitivity to object based information but not viewpoint information. This dissociation suggests the PPA is involved in encoding spatial based information as opposed to object identity information. However, this evidence is not supported by other research in the field, who have observed that the PPA is involved in processing information beyond just spatial properties of scenes.

Indeed, the PPA’s contribution to processing some forms of object information is largely agreed (Aminoff & Tarr, 2015; Linsley & MacEvoy, 2015; Mullally & Maguire, 2011; Troiani, Stigliani, Smith & Epstein, 2014). For example, evidence has shown that the PPA is able to decode information about object identity to the same accuracy as the lateral occipital complex (LOC, a region involved in processing objects) (MacEvoy & Epstein, 2011), and information about object identity in non-scene arrays (Pihlajamaki et al., 2004). However, others have suggested that that PPA may only encode objects that are navigationally relevant (Spiridon & Kanwisher, 2002). These findings suggest that PPA encodes some information about objects, which can be utilised to potentially aid in place recognition and spatial navigation (Epstein, 2008; Ghaem et al., 1997; Janzen & van Turennout, 2004; Marchette, Vass, Ryan & Epstein, 2015; Vass & Epstein, 2016). Moreover, it is still unclear the extent to which the PPA encodes representations about object identity, requiring further exploration.

The idea that the PPA encodes spatial information about objects is widely supported. For example, evidence suggests that the PPA encodes information about objects that portray a strong sense of spatial presence (Mullally & Maguire, 2011; Troiani, Stigliani, Smith & Epstein, 2014). Mullally and Maguire used fMRI to examine the neural correlates of the PPA in response to two different categories of objects, 1) space-defining objects, and 2) space-ambiguous objects. Results suggested that the PPA was activated significantly more for space-defining objects compared to spatially ambiguous objects. Thus, demonstrating that independent of scene context, the PPA is involved in processing spatially pertinent properties of objects in scenes. Analogous to this finding, Mur et al., (2012) showed that the ranking of the category responses in the PPA showed significantly higher levels of activity for scenes and landmarks compared to faces and other forms of objects. Comparably, neuropsychological research has shown that patients with lesions to the PPA were unable to use landmark information in order to determine a spatial framework during map drawing (Habib & Sirigu, 1987) suggesting the PPA encodes navigationally relevant object information.

In line with this research, others have used fMRI to examine the extent to which the PPA encodes information about navigationally pertinent locations of objects (Janzen & van Turennout, 2004). Some of these objects were located at points considered navigationally relevant, and others were positioned at points not considered navigationally relevant. Participants were familiarised with objects when exploring the virtual museum and were subsequently asked to partake in a task measuring their memory for the objects. Results showed that activation in the PPA was increased for objects previously observed in the museum at navigationally relevant points. In addition, activation was elicited for objects at navigationally relevant locations that participants could not explicitly recall, PPA activation was observed. This observation is in line with the findings of Troiani, Stigliani, Smith and Epstein (2014), identifying that the PPA encodes various spatial qualities of objects which make them navigationally relevant (such as: degree of permenance). These results provide evidence that the PPA is involved in encoding spatial qualities derived from objects and their locations, relevant for navigation.

Despite previous evidence suggesting that the PPA is an important node in the spatial navigation network (Epstein, Harris, Stanley & Kanwisher, 1999), other evidence suggests this is not the case (Dilks, Julian, Kubilius, Spelke & Kanwisher, 2011; Persichetti & Dilks, 2016). Could it be that there are two distinct functional networks involved in scene processing? This question has been examined (Baldassano, Esteva, Fei-Fei & Beck, 2016; Baldassano, Fei-Fei & Beck, 2016) suggesting that there are two distinct scene-based processing streams which the PPA is a part of; one involved in supporting episodic memory for spatial processing and another supporting perceptual processing of scenes. Despite emerging evidence suggesting the PPA's involvement in different aspects of scene perception, our current knowledge about the functional processing properties of the PPA is still relatively unclear. For example, does the PPA encode information about object locations and arrangements in natural scenes?

Whilst this has not directly been tested, a study by Pihlajamaki et al., (2004) examined whether different regions of the MTL play a role in encoding familiar object arrangements. Their results showed that the PPA processes information about the arrangements of familiar objects. Critically, however, the authors studied objects in non-scene arrays, thus, it is difficult to deliniate whether this would extend to natural scene contexts and whether this effect would differ across other scene regions such as the RSC and OPA. Likewise, their study has a predominant focus on the hippocampus and entorinal cortext, thus, in order to determine whether the PPA encodes object arrangements in scenes one would need to develop their research to look at variations in object-based informaiton within scenes. For example, does the PPA encode information about the spatial locations or arrangements of object in scenes or object identities?

Together, these findings suggest that the PPA is not just a region involved in processing scenes *per se*, but may also be involved in coding forms of spatial information about geometric/structural properties of scenes, object identity and spatial properties of objects (Epstein, 2008; Ghaem et al., 1997; Janzen & van Turennout, 2004; Marchette, Vass, Ryan & Epstein, 2015; Troiani, Stigliani, Smith & Epstein, 2014; Vass & Epstein, 2016). However, at present, there is still much ambiguity regarding the precise functional role of the PPA. Given that there is evidence linking the PPA with encoding multiple forms of scene and object information, the neural representations of scenes and objects encoded in this region are difficult to delineate. Thus, further work is required to utilise the findings from previous research to determine whether this region encodes relative spatial and non-spatial information about the locations and identities of objects in scenes.

1.5.2.2. Retrosplenial complex (RSC)

The RSC is a region considered important for spatial processing, and is located directly behind the splenium (Broadmann areas 29 and 30) (Maguire et al., 1998; O'Craven & Kanwisher, 2000). In the macaque, the RSC (mRSC) is situated adjacent to the medial parietal occipital sulcus bilaterally (Nasr et al., 2011). In humans, the RSC is implicated in many disease states; where patients present with deficits in spatial memory and an inability to orient themselves within familiar spatial environments (Aguirre & D'Esposito, 1999).

The RSC is widely considered a key 'scene-selective' region but has been critically linked with scene encoding, landmark encoding and spatial memory (Epstein, Parker & Feiler, 2007; Hodgetts, Shine, Lawrence, Downing & Graham, 2016; O'Craven & Kanwisher, 2000). Neuropsychological studies suggest that the RSC is one of the first regions to be affected in the early stages of Alzheimer's disease, where symptoms include deficits in navigation, spatial memory, landmark processing and spatial reorientation (Barrash, 1998; Vlcek & Laczo, 2014). This theory is supported by evidence in healthy controls. For example, it has been found that during rest, temporal coherence in BOLD (functional connectivity) between the hippocampus and the RSC are significantly greater for participants categorised as good navigators compared to those deemed as not as good (Sulpizio, Boccia, Guariglia & Galati, 2016). Together, these observations suggest that the RSC is a key node in networks that support spatial processing, most importantly for supporting cognitive processes such as spatial navigation.

As with the PPA, there have been significant developments in our understanding of RSC contributions to processing spatial landmarks. The RSC may have a role in coding heading direction; crucial for reorientation in a given location, with reference to another landmark or object (Marchette, Vass, Ryan & Epstein, 2014, 2015), and object permanence/landmark-suitability within scene and non-scene contexts (Auger & Maguire, 2013; Troiani, Stigliani, Smith & Epstein, 2014). But how does activation for the PPA and RSC differ?

Research has directly compared the contributions of both the PPA and RSC in scene processing (Epstein, 2008; Epstein & Higgins, 2007; Epstein, Parker & Feiler, 2007; Park & Chun, 2009). Both regions have been largely dissociable in their processing properties. For example, the PPA is not largely considered a region whose primary function is navigation (Epstein, Harris, Stanley & Kanwisher, 1999). However, the RSC is considered a region critical in route learning and spatial navigation (Maguire, 2001; Maguire et al., 1998; O'Craven & Kanwisher, 2000). This is supported by neuropsychological evidence which demonstrates that patients with lesions to the PPA show deficits in recognising scenes and landmarks (Mendez & Cherrier, 2003). However, patients with lesions to the RSC show primary deficits in using landmarks to guide behaviour during navigation (Aguirre & D'Esposito, 1999; Maguire, 2001).

Moreover, evidence suggests that the PPA encodes local features of individual scenes and single viewpoints (Epstein, Graham & Downing, 2003; Epstein, Higgins, Jablonski & Feiler, 2007), whereas the RSC extrapolates information across multiple frames of the scene (Park & Chun, 2009). In their study, Park and Chun used fMRI repetition suppression to examine PPA and RSC responses across panoramic views of scenes. For the panoramic condition, participants were presented with three viewpoints of the same scene. Results showed there were no significant attenuation effects in the PPA, whereas the RSC showed adaptation across panoramic views, suggesting that the PPA holds viewpoint specific representations. This supports the findings of Epstein, Graham & Downing (2003). Conversely, the RSC relates different viewpoints from the same scene to one another.

Together this evidence suggests the RSC plays an important role in processing spatial and non-spatial information relevant for spatial navigation. However, at present, there is limited understanding of how these regions process more basic spatial relationships between the objects and their background.

1.5.2.3. Occipital place area (OPA)

The OPA is a more recently functionally defined cortical region which is situated around the anatomically defined transverse occipital sulcus, TOS, in the occipital lobe (Dilks, Julian, Paunov & Kanwisher, 2013). In macaques, the TOS (mTOS) is thought to be homologous with the human TOS (Nasr et al., 2011) and is also involved in processing visual scene information).

The OPA has been identified as a key node in the scene-processing network (Bettencourt & Xu, 2013; Dilks, Julian, Paunov & Kanwisher, 2013). Dilks, Julian, Paunov & Kanwisher, used transcranial magnetic stimulation (TMS) on human participants, to examine their responses on a scene and face discrimination task. When TMS was delivered to the OPA, participants were unable to correctly categorise scenes, but their performance for faces remained intact. However, when TMS was administered to the occipital face area (OFA), face categorisation was impacted, but not scene categorisation. These results identified the OPA as a region causally involved in scene processing.

Indeed, other research has also identified the OPA as being causally involved in processing scenes (Ganaden, Mullin & Steeves, 2013; Mullin & Steeves, 2011). Using TMS, Ganaden, Mullin and Steeves challenged the object-centred approach, that scenes can largely only be recognised when the objects have been identified. During their experiment, participants were required to partake in a categorisation task (natural/non-natural objects or scenes), while TMS was delivered to the OPA and vertex (control site). When TMS was delivered to the OPA, scene categorisation accuracy was reduced, signifying the OPA is an important node in the scene-selective network.

Whilst the PPA is thought to analyse topographical information about the scenes global layout (Epstein, Deyoe, Press, Rosen & Kanwisher, 2001), the OPA is thought to represent low level spatial information about the layout/local features of scenes (Kamps, Julian, Kubilius, Kanwisher & Dilks, 2016). Using fMRI, Kamps and colleagues examined what kinds of visual information the OPA, RSC and PPA encode. They examined various conditions including intact rooms, fractured rooms, single objects, single furniture and multiple objects. Results showed that the OPA responded similarly for intact and fractured rooms, signifying that it is sensitive to local features, and does not require spatial coherence to process the space. Conversely, the PPA and RSC were significantly more active for spatially intact rooms as opposed to fractured rooms, suggesting that they rely on spatial coherence within a scene to form global scene representations, which provide context. These results are consistent with the theory that more posterior regions such as the OPA represent the local elements and features of scenes (Baldassano, Esteva, Fei-Fei & Beck, 2016; Baumann & Mattingley, 2016).

Given that recent evidence suggests crosstalk between visual processing and memory streams (Baldassano et al., 2016) it could be argued that object and scene processing pathways, two previously assumed parallel streams, cross-talk to support recognition of items in the environment. Indeed, Ganaden, Mullin and Steeves (2013) showed that despite the facilitation of scene processing when TMS was applied to the LOC, object recognition was not facilitated when TMS was delivered to the OPA. Thus, while there remains ambiguity as to how object and scene processing streams utilise object and scene information, these regions may form parts of larger networks that encode spatial and non-spatial information about the environment to support processes such as navigation (Persichetti & Dilks, 2016).

Moreover, these findings collectively suggest that the OPA may encode more basic spatial location and layout information about objects in scenes, relevant for navigation. While evidence has shown that the OPA does not encode object identity per se, further investigation is required to determine the extent to which activation in the OPA differs compared to the PPA and RSC during object-in scene processing.

Although several lines of evidence support a broader role for scene processing in these cortical regions (Epstein & Kanwisher, 1998), little is known about how variation in scene content (spatial and non-spatial properties of objects) affects the representations that are formed in the scene selective network. A key study is that by Troiani, Stigliani, Smith & Epstein, (2014) who examined what kinds of spatial properties of objects the core-scene selective structures are sensitive to, and how this differs across the regions.

Troiani, Stigliani, Smith & Epstein, (2014) used fMRI to study how multiple spatial properties of single objects are processed with and without backgrounds. These spatial properties included *placeness* (the degree to which an item is deemed as having a place in the world), *context* (high or low relationship between the item and its context), *fixedness* (how fixed the item is to its location), *space defining* (how much space the item defines), *distance*, (how close or far the item is) physical size (how large the object is in relation to the real world) and *visual size* (how large the object is visually on the screen). Each condition ranged from 'low' to 'high' in their degree for each property. For example, an object which has a small physical size might be a pen, whereas an object with a large physical size might be a space ship. Each scene selective region (PPA, OPA, RSC) was defined using a functional localiser.

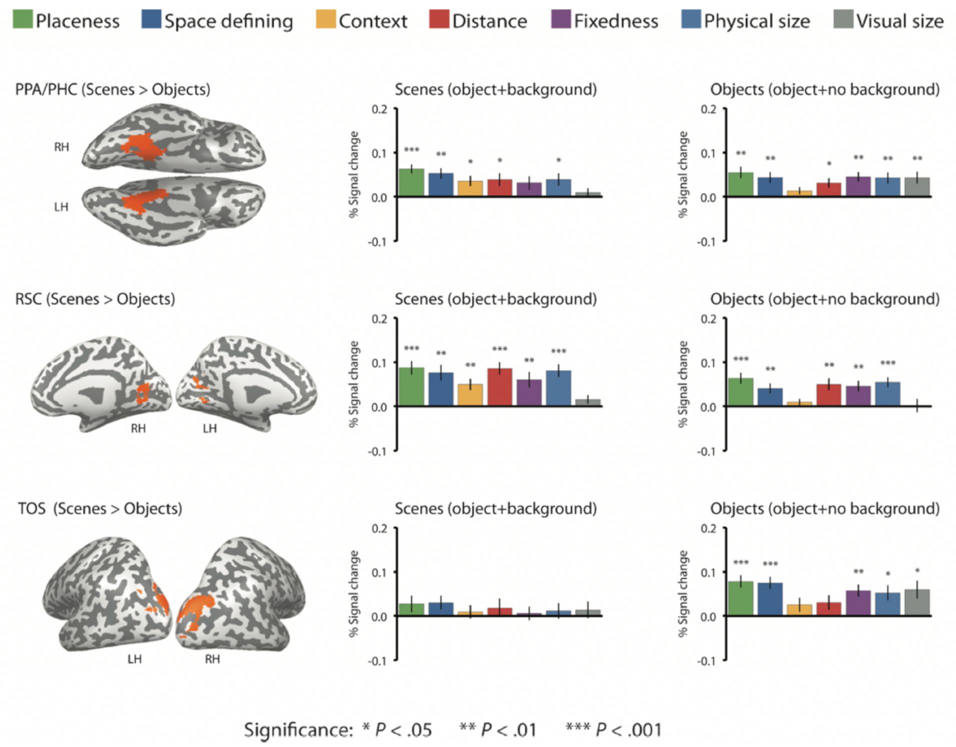


Figure 1.4. Results from the fMRI experiment across all object conditions, with and without a background. Regions include bilateral PPA, RSC and OPA (TOS). Results show the OPA is significantly more activated by spatial properties of objects in non-scene contexts. Image adapted from Troiani, Stigliani, Smith and Epstein, (2014).

Results showed that PPA and RSC showed sensitivity to multiple spatial properties of objects portrayed in a scene context (objects with backgrounds) and isolated from a scene (objects without a background). Conversely, the OPA showed similar responses but predominantly for objects isolated from their background (Figure 1.4). Further analysis revealed that, the PPA and RSC were preferentially recruited for objects that were deemed more navigationally relevant (space defining, place related, degree of fixedness to a location and real world size), but there was some dissociation between their sensitivity to these properties. For example, the PPA was more sensitive to information regarding the objects 'placeness' and its level of space defining qualities. Conversely, the RSC was sensitive to the object distance and its physical size. Therefore, this suggests that the RSC may encode spatial information more related to an objects geometric identity, whereas the PPA encodes information relating an objects spatial qualities to a particular context. These findings demonstrate that it is not scenes *per se* that activate these regions, but rather activation is driven by spatial properties of the objects both with and without a scene background.

Collectively, this evidence suggests that the scene-selective structures discussed here encode a combination of spatial information about objects, which differs across regions. For example, posterior regions such as the OPA integrate information about spatial properties of objects without a background, whereas the more anterior/medial regions such as the RSC and PPA encode spatial information about objects that can be relayed to their individual geometric identity and their spatial relevance for contextual understanding.

1.6. Towards an understanding of object-in-scene representations: converging evidence between brain and behaviour

Repetition suppression is a phenomenon whereby the neural signal and BOLD signal is reduced when a stimulus is repeated. In fMRI, RS is thought to probe attenuation of neural populations and has been identified as a useful tool to establish the function of different brain regions across multiple networks (Bouchon, Nazzi & Gervain, 2015). In the scene-selective network, RS has been used to demonstrate that attenuation is a mechanistic explanation of the behavioural effect boundary extension (BE)

Park, Intraub, Yi, Widders and Chun (2009) demonstrate that when an individual is presented with a pair of scenes where the second scene is a wider-angle view, participants report them as the same (the BE effect). They then model the fMRI signal across multiple pairs of scenes with this trial order and demonstrate that not only do participants behaviourally respond ‘same’ during behavioural testing after the experiment, the effect is mirrored in their neural response as an attenuation like effect similar to that of repetition suppression (figure 1.5). These findings demonstrate that the mechanisms that underpin RS may reflect the behavioural error observed in BE. This seminal paper demonstrates that not only are we able to model the BE error in the fMRI signal patterns, but that we are reliably able to infer attenuation as a mechanistic explanation of BE in the scene-selective network.

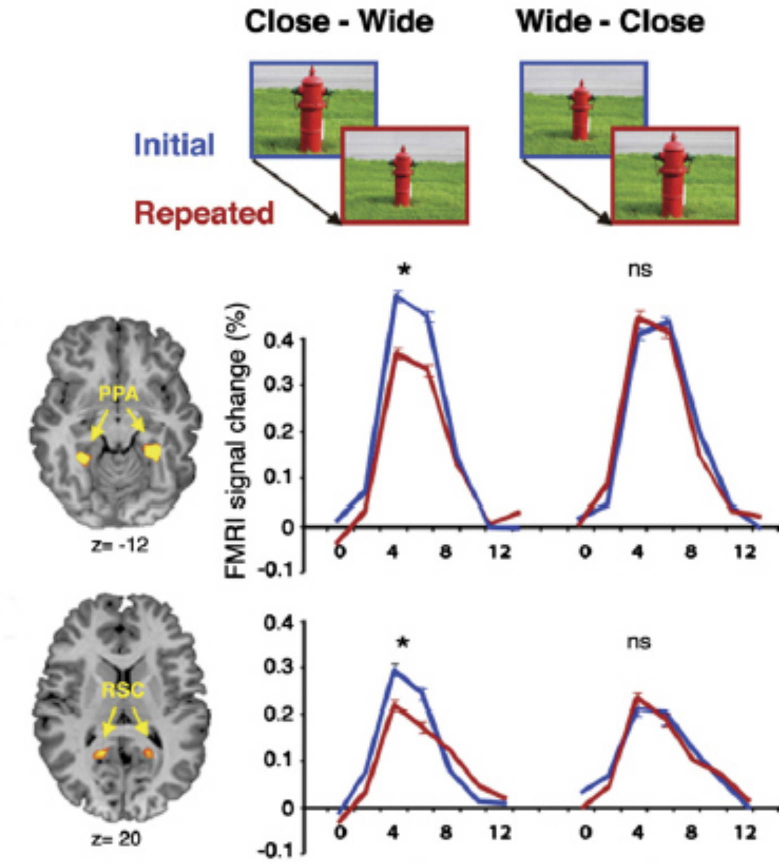


Figure 1.5. Evidence of a repetition suppression like effect observed for a boundary extension paradigm for scenes presented from close-wide, suggesting that the brain encodes this trial as if the two scenes were identical. These findings mirror those observed during behavioural paradigms where participants repose the second wider angle scene as being the same as the initial closer up view. Image

adapted from Park, Intraub, Yi, Widders and Chun (2009).

Both RS and BE are considered to be driven by similar underlying neural mechanisms of attenuation (Park, Intraub, Yi, Widders and Chun, 2007) where they have both previously been used independently as tools to understand how the brain encodes and remembers visual information. In addition, the temporal effects of BE and RS have been explored in detail showing that they occur in as little as a fraction of a second right up to 30 minutes later. These effects have been shown particularly in RS where adaptation effects are stronger for short lag compared to long lag intervals (Weiner, Sayres, Vinberg & Grill-Spector, 2010). Comparably, BE effects have also been observed, at both long and short intervals, particularly in traditional studies where the interval could be upwards of 30 minutes or rather rapidly within around 500ms (Intraub & Richardson, 1989). In the present thesis, we use BE and RS as tools at short intervals, though to reflect short term memory for scene processing.

Short term memory (STM) is a subcomponent of working memory which underpins and individual's capability to maintain information in mind which is no longer present to them (Baddeley & Hitch, 1974). In the present thesis, we use short interval BE and RS paradigms to assess visual short term memory (VSTM) in response to variations in object-in-scene information. Compared to long term memory (LTM) which holds memory representations for far longer periods of time, VSTM permits an individual to maintain and even manipulate previously seen visual information in mind for a short period of time (seconds). This allows them to accurately recall from this VSTM store whether changes were observed between sequentially presented stimuli. Both RS and BE paradigms used in the present thesis require individuals to hold the scene in memory in order to determine whether there has been a change in the visual properties of the second scene since the first presentation.

Given that the fMRI correlates of RS predict those shown behaviourally during BE, it is reasonable to suggest that these complimentary methods could be used to ask further questions regarding object-in-scene representations in VSTM. For example, very little is currently known about how variations in scene content are represented within the brain, and whether the type of scene content modulates how the scene is remembered. For example, how does they type of object alter the way in which the scene is processed and held in visual short-term memory (VSTM), and does this change how individuals remember global spatial properties of the scene?

1.7. The study of object-in-scene representation in the current thesis.

Collectively, the evidence discussed so far demonstrates a broad and dynamic scope of research in the field of memory and perception for visual scenes. However, there remains a gap in our knowledge regarding how we encode and reconstruct variations in scene content from VSTM. First, how do the scene-selective structures process representations of spatial and non-spatial properties of multiple objects in natural scenes? Second, how does the presence of more social entities in natural scenes influence encoding and reconstruction for the spatial extent of a scene?

1.7.1. Spatial and non-spatial representations in the scene-selective cortex

The core scene-selective structures are thought to encode a magnitude of information beyond basic recognition of scenes, most critically for object-context interactions (Davenport, 2007; Davenport & Potter, 2004; Henderson & Hollingworth, 1999; Hollingworth, 2007; Hollingworth & Henderson, 1998, 1999), spatial properties of individual objects (Mullally & Maguire, 2011; Troiani, Stigliani, Smith & Epstein, 2014) and object grouping regularities (Kaiser, Stein & Peelen, 2014). In addition, some have suggested the PPA encodes novel object identities in non-scene arrays (Pihlajamaki et al., 2004), whereas others do not find this effect in natural scenes (Epstein, Graham & Downing, 2003).

Despite a wealth of studies examining these regions in response to scene-based properties, the literature regarding scene and object-in-scene encoding is overwhelming and in parts contradictory. Aside from the neuroimaging literature studying contextual associations and memory for object arrangements in grid like arrays (Aminoff & Tarr, 2015; Gronau, Neta & Bar, 2008), it remains elusive as to whether these structures represent information about spatial and non-spatial properties of multiple objects in natural scenes. Therefore, we investigate these themes further to determine whether these regions differentially represent object-in-scene information, and how the more recently defined OPA integrates into this functional network of regions.

1.7.2. Social contributions to scene encoding: How the identity of an object influences spatial encoding

Evidence from the scene processing literature has predominantly examined object contributions to scene encoding (Epstein, Graham & Downing, 2003; Harel, Kravitz & Baker, 2013; Intraub, Gottesman, Willey & Zuk, 1996; Troiani, Stigliani, Smith & Epstein, 2014) with very limited consideration for the role that more social entities may play in spatial processing.

People in comparison to objects are considered more salient stimuli (Downing, Bray, Rogers & Childs, 2004; Peelen & Kastner, 2014). While people do not behave like fixed landmarks which can be used for navigation, their presence may influence how people interact within the world, and attend to surrounding space (Adolphs, 2003a, 2003b; Gallese, Keysers & Rizzolatti, 2004). For example, previous evidence suggests that the presence of a person within a scene influences the way in which attention and eye fixations are dispersed across the scene (Bindemann, Scheepers, Ferguson & Burton, 2010). This line of research has predominantly studied how the presence of a person changes the way humans perceive the scene *not* how the presence of a person influences memory for the spatial extent of a scene. For example, do people in scenes influence how the global spatial scope of a scene is encoded? Does this affect short-term memory for the scene?

1.8. Thesis outline

In the preface of this chapter, I describe a situation whereby humans see the same global scene multiple times, but the objects within that view may change in their locations, identities and arrangements. Likewise, not only do we observe objects within the world, but also people. Based on recent evidence reviewed in this thesis, it is clear that variation in scene content (spatial and non-spatial properties of objects in scenes may influence i) the initial neural response to the spatial properties of objects within scenes and ii) reconstruction of scene representations in visual short-term memory.

This thesis employs two different approaches to explore how variations in scene content are encoded and reconstructed in memory. The first approach explores whether activation within the core scene-selective structures differs as a function of scene layout and object identity. Using fMRI repetition suppression (RS) (refer to Chapter 2 for detailed methods), we examine the representational content across the key scene-selective structures (PPA; RSC; OPA). While these regions are important nodes for encoding multiple forms of spatial and object information, research has yet to examine the extent to which these regions differentially encode variations in spatial properties of objects (Locations and Arrangements) and non-spatial properties of objects (Identities) in natural scenes.

The second line of investigation explores whether the type of scene content (person/object) influences how the spatial extent of a scene is reconstructed in memory. Whilst a vast literature of research has examined person perception, this is seldom explored in the context of spatial processing. Here, we examine this by asking, does encoding the extent of space change as a function of the type of entity within a scene? Is this dependent upon on how many people are present? Does this change if the person appears to be dynamic (moving) or sitting in a fixed position? To examine these questions, we use boundary extension (BE) (see Chapter 2 for a detailed description of BE), as a tool to examine whether the type of entity influences memory for the spatial extent of a scene.

Indeed, objects form a substantially large portion of our visual experience, therefore, it is critical to ask these questions in order to better understand human visual spatial cognition. Collectively, these questions will provide a more coherent understanding of the complex nature of natural scene perception, and allow us to better understand the mechanisms by which the wider visual environment is encoded and remembered. Throughout the course of the next chapter, I discuss in detail, the neuroimaging and behavioural methods used explicitly in the current body of work.

In **Chapter 2** I outline the methodology used to examine these lines of investigation. This includes details of the neuroimaging methods (fMRI repetition suppression) and behavioural methods (boundary extension).

Chapter 3 is the first of two fMRI experiments. In this chapter, we investigated differences in the representational content of the PPA, RSC and OPA in response to variations in object Locations, Arrangements and Identities in natural scenes.

Chapter 4 extends the work in Chapter 3 by investigating the same questions, but instead using a task tailored to direct participants' attention to the changes between pairs of scenes.

Chapter 5 is the first of two behavioural studies. Here, BE is used to infer whether the type of entity presented within a scene changes how the spatial extent of the scene is remembered. This experiment directly compared BE effects for people and objects in natural scenes, and included either individual, or multiple entities. The second part of Chapter 5 compared BE for dynamic people photographed in action, and people who appeared sat down and fixed in a location. The third part of Chapter 5 examined BE for scenes containing no focal objects.

Chapter 6 builds upon the findings of Chapter 5 by examining whether saliency differences associated with people and objects influenced BE or boundary contraction.

In **Chapter 7** I discuss the general findings from the four experiments with key links to previous research, limitations and future directions.

Chapter 2

General Methods

Prior to the development of contemporary neuroimaging methods, scientists explored human brain function through the study of brain insult. One of the most notable cases is patient H.M. (Dossani, Missios & Nanda, 2015; Scoville, 1957), who underwent medial temporal lobe (MTL) resection to control seizures. After the procedure, patient H.M. presented with severe memory impairment. This case study provides evidence implicating key brain structures in specific cognitive function. While neuropsychological models have significantly advanced our understanding of brain function, they are limited in that they may not implicate focal brain regions (such as the case with H.M. and in AD). Thus, drawing conclusions about complex cognitive processes and linking them to specific brain structures can be problematic.

With the development of neuroimaging techniques researchers are able to link both anatomy with brain function in both health and disease, using a non-invasive *in vivo* approach. Methods such as fMRI have substantially built on initial patient research. In combination with behavioural cognitive neuroscience approaches, these imaging methods provide neurobiological insight into precise mechanisms of specific cognitive events.

This thesis uses two popular cognitive neuroscience methods for assessing how individuals process subtle variations in scene content during encoding and reconstruction. First, we use fMRI repetition suppression (RS) to examine neural representations of scene-selective structures in response to spatial and non-spatial objects variations. Second, we use boundary extension (BE) to assess whether the type of scene content directly influences scene memory. Methodological considerations and analysis pipelines are discussed for both approaches.

2.1. Functional magnetic resonance imaging (fMRI)

Initial studies used fMRI to assess brain function, comparing regional changes in blood flow during rest and activation (Ogawa, Lee, Kay & Tank, 1990). Here, Ogawa and collegues termed this endogenous contrast method ‘blood oxygen level dependent’ (BOLD). The BOLD signal obtained during fMRI is thought to reflect an increase in neural activity, where metabolic demand (oxygen and glucose consumption) leads to an increase in flow of oxygenated blood to a given brain structure (Attwell et al., 2010; Logothetis, 2010; Logothetis, Guggenberger, Peled & Pauls, 1999). As the oxygenated blood to that structure increases, there is a change in the ratio of oxyhaemoglobin to deoxyhaemoglobin. It is the excess of oxyhaemoglobin that causes a significant decrease in the level of deoxyhaemoglobin, which results in an increase in MR signal, or BOLD.

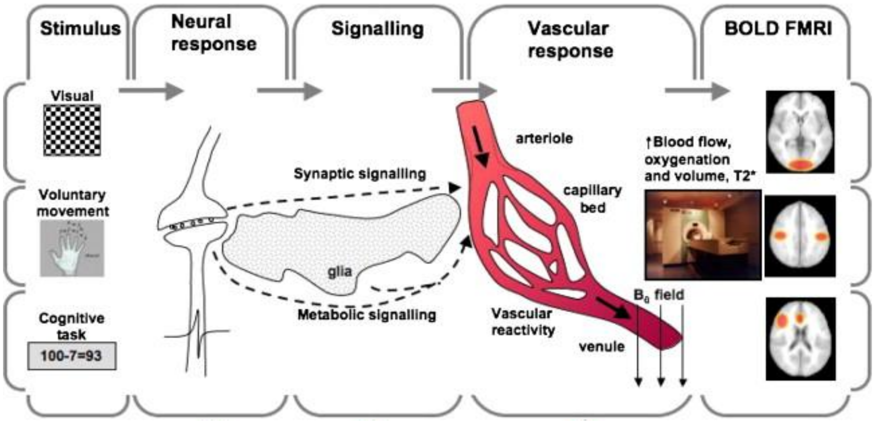


Figure 2.1. An illustration of the origin of the BOLD signal. From the stimulus, to the neural signalling, resulting in a vascular response increasing the demand for oxygenated blood, creating BOLD signal in a given region associated with processing the stimulus. Image adapted from (Iannetti & Wise, 2007).

BOLD signal changes can be examined over time (usually 0-18 seconds), referred to as the hemodynamic response function (HRF) (Logothetis & Wandell, 2004). The HRF typically refers to the time-course of hemodynamic changes (BOLD) across a set of voxels. The HRF usually shows an initial dip in signal in the first few seconds, followed by a positive increase in signal between 4-6 seconds, reaching a peak between 6-10 seconds before returning back to baseline (Buxton, Griffeth, Simon, Moradi & Shmuel, 2014).

However, there is widespread contention regarding the underlying mechanisms that drive changes in BOLD hemodynamics, as the fMRI signal is an indirect measure of neuronal activity (Attwell et al., 2010; Logothetis, Guggenberger, Peled & Pauls, 1999). Nevertheless, the advancement in MRI/fMRI methods has provided advances in our understanding of brain structure and function in both health and disease (Casey, Giedd & Thomas, 2000; Lancaster, Ihssen, Brindley & Linden, 2017).

The development of task based fMRI methods has enabled researchers to investigate differences in BOLD for subtle condition changes. Specifically, it allows researchers to infer the representational content of different brain regions, at the individual region of interest (ROI) level, or across the whole brain. Univariate fMRI methods enable researchers to examine differences in BOLD across the brain for one condition compared to another (Di Salle et al., 1999; Nakada, 1999). For example, in the scene selective cortex, the PPA originally showed significantly higher levels of activation of scenes compared to houses, objects and faces (Epstein & Kanwisher, 1998). Univariate methods such as repetition suppression (RS) enable researchers to probe the representational content of a region by examining responses between novel and repeated stimuli (Barron, Garvert & Behrens, 2016; Desimone, 1996).

2.1.1. Repetition suppression (RS)

Repetition suppression, also referred to as 'fMRI-adaptation' (Grill-Spector & Malach, 2001), is characterised as a reduction in signal/neural firing rates in response to a repeated stimulus. RS has been employed to study single cell firing rates (Lueschow, Miller & Desimone, 1994; Miller & Desimone, 1994) and fluctuations in positive and negative signal amplitudes in EEG (Summerfield, Wyart, Johnen & de Gardelle, 2011). In this thesis, I will discuss mechanisms thought to support RS as well as the strengths and considerations surrounding this method.

In fMRI, RS is used as a tool to probe the representational changes between experimental conditions in pre-defined regions of interest (ROIs) or at the whole-brain level. Typically, RS reflects attenuation in BOLD HRF, in response to repeated stimuli, or stimuli which appear visually similar across multiple feature dimensions. However, we are also able to study RS effects for stimuli which differ in only one feature dimension. If a brain region is sensitive to a particular feature dimension across the two scenes, a release in adaptation (increase in signal) is observed when that feature changes. This is because it has detected the second scene as different, based upon that particular feature manipulation. However, if the region is not sensitive to variations in that feature, an attenuation (adaptation) effect will be observed comparable to the effect observed for the repetition condition. This is because it treats the two images as if they were the same (not sensitive to the feature change). The reader should note that in the present thesis, RS effects refer to the differences between repetition and non-repetition trials.

This thesis uses a paired-stimulus design, which examines HRF responses across pairs of images (Epstein, Graham & Downing, 2003; Kourtzi & Kanwisher, 2001). That is, in one single trial, participants are presented with two images separated by an inter-stimulus interval (ISI). The HRF is then modelled over the two stimuli. In some of these trials the images will be identical (repetition), whereas in others, there will be different in various visual dimensions (object locations, arrangements and identities).

Early studies anticipated that RS in fMRI would provide a bridge between standard univariate contrast models and single cell recordings. However, the mechanisms that drive the repetition suppression effect are still being reviewed at length (Barron, Garvert & Behrens, 2016; Grill-Spector, Henson & Martin, 2006; Henson & Rugg, 2003; Summerfield, Trittschuh, Monti, Mesulam & Egner, 2008). However, through the development of this method, various theories have attempted to explain underlying mechanisms that support RS (Figure 2.2).

The fatigue model describes an overall reduction in neuronal firing rate to the repeated stimulus (Kohn & Movshon, 2003, 2004), where a population of neurons that fired for the initial stimulus show a proportionally reduced firing rate to the second identical stimulus presentation. This is potentially explained by firing-rate adaptation, where the level of neuronal excitability reduces, thus reducing the likelihood of the neurons spiking.

The sharpening model suggests that a population of neurons that fire for the initial stimulus become more refined in response to the second identical repetition (Desimone, 1996; Wiggs & Martin, 1998). This model can be thought of as the group of neurons fine-tuning or pruning their response to the learned stimulus.

The facilitation model proposes that the speed of processing for the repeated stimulus is faster, therefore resulting in shorter neuronal firing rates (James & Gauthier, 2006). Conversely, Summerfield et al., (2008) suggests that attenuation to an unexpected repetition is reduced. Thus, highlighting that RS may be driven by top-down perceptual mechanisms.

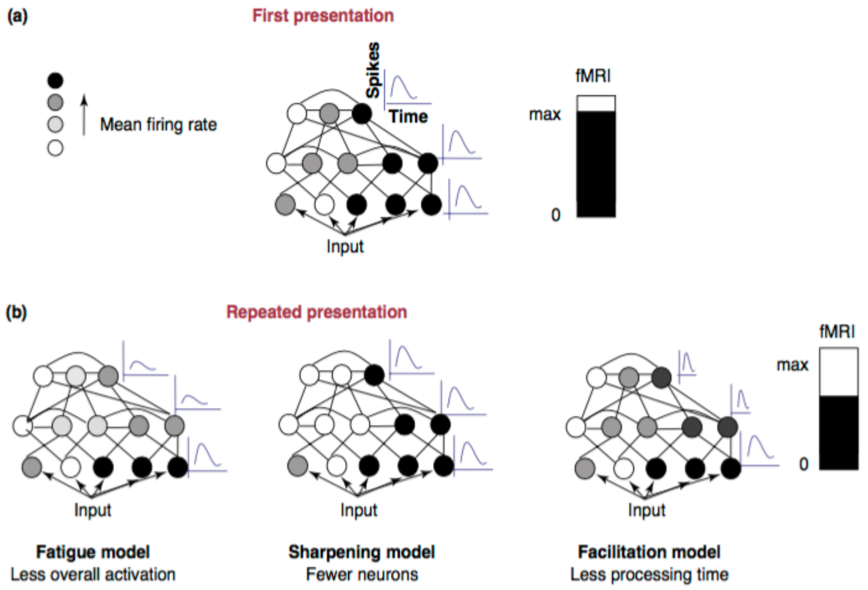


Figure 2.2. An example of the mechanisms that are thought to drive and support RS responses. This particular figure highlights the three most widely described models; Fatigue, Sharpening and Facilitation. Figure adapted from Grill-Spector, Henson and Martin (2006).

These models suggest that repetition suppression may be mechanistically explained by the facilitation of neural signals. This is reflected as a change in the latency of the signal, observed as a reduction in BOLD signal (see Baron et al., 2016 for review). This modal also supports behavioural effects of facilitation where a reduction in reaction time would be observed (Horner & Henson, 2008). Conversely, evidence has more recently shown that repetition enhancement (enhancement of neural signals for repeated stimuli) may be better represented by inhibitory processes where inhibitory signals are disinhibited, resulting in an increase in signal (Wissig & Kohn, 2012). However, the precise mechanisms that drive these effects remain elusive and require further research. While we are unable to directly infer which model best fits BOLD data in the current body of research, one would speculate that facilitation models would most likely explain any reductions observed in BOLD signal to repeated stimuli, as observed in similar RS and scene encoding paradigms (Epstein, Graham and Downing, 2003).

2.1.2. fMRI analysis approaches for the current experiment

2.1.2.1. fMRI region of interest localisation (ROI)

Region of interest (ROI) analysis measures the BOLD and HRF within predefined set of voxels within the brain (Poldrack, 2007). This can be defined at the anatomical (structural) level, functional subject-level, functional group-level, or created with a predefined mask.

Subject-level localisation better accounts for intersubject variability in activation compared to group-level defined ROI's (Nieto-Castanon, Ghosh, Tourville & Guenther, 2003). For example, the precise anatomical location of functional clusters may vary between individuals, however this is accounted for in the subject-level localiser, where activation is taken from the peak location for the individual participant. However, this method of localisation may be problematic if participants do not show active voxels within a given region. Difficulties in localising activation in ROI's have previously been encountered during scene-selective region localisation, most notably for the RSC (Hodgetts, Shine, Lawrence, Downing & Graham, 2016). In this case, functional localised ROIs can be defined at the group-level.

In this thesis, ROI definition was conducted within a separate experiment (within the same scan session), consisting of two runs of a functional localiser. In the present thesis, a run refers to a block of stimuli presented in a set order over a predefined time frame. Here we use two runs to more robustly identify the regions considered selective to processing images of scenes. The fMRI localiser scans involved presenting blocks of different categories of stimuli (faces, bodies, chairs, scenes), to recruit regions considered scene-selective at the individual subject level. To localise regions within the scene processing network (PPA; RSC; OPA), a similar contrast was used to that of Epstein and Kanwisher (1998), of scenes > faces. From here, 6mm sphere shaped masks were created around the local maxima (peak activation) of these regions for each participant. These masks were then used to extract time-courses for the main repetition suppression experiment. A time-course in the present thesis refers to a number of data points modelled across time. In this instance, we use a time-courses to observe the BOLD HRF across 18 seconds.

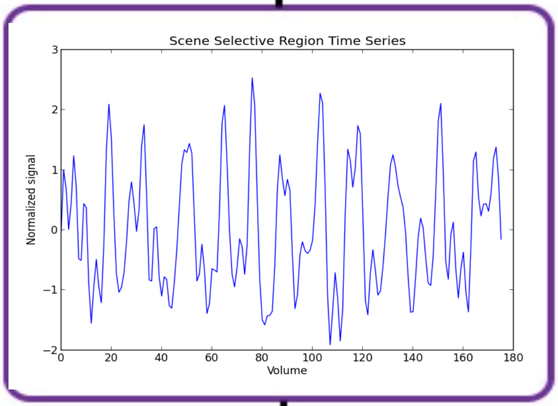
2.1.2.2. Repetition suppression analysis

RS was used in both fMRI experiments in this thesis, as a method of probing the representational content of scene-selective structures (PPA; RSC; OPA), to examine how activation across these structures differs in response to various spatial and non-spatial properties of objects in scenes. These experiments employed an event-related fMRI design, which allows modelling of activation for individual trials at time-locked events.

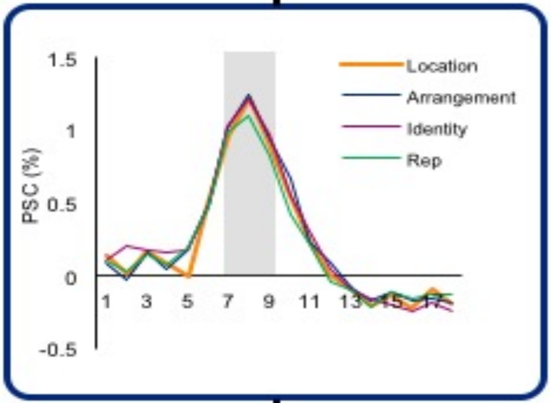
Each trial in the event-related paired-stimulus experiment included a presentation of an initial base image, followed by an inter-stimulus interval (ISI) and then either a repetition, or a scene that changed in one of the following dimensions: objects Locations, object Arrangements, both object Location and Arrangement, or object Identities. Other trials included trials where a fixation cross was presented for 3 seconds, and task trials where the second scene was either upside down (experiment one) or contained two types of changes (experiment 2). Fixation trials are modelled in the analysis for calculation of percent signal change and last for the same duration as scene stimuli trials. During each inter-stimulus interval (ISI, the short interval between each scene within a trial), a fixation was presented to encourage participants to remain focused on the centre of the screen during the experiment.

Time-series were extracted from the individually defined functional ROI masks in both hemispheres for all three scene-selective regions. This was conducted using MarsBaR (http://marsbar.sourceforge.net/), a MATLAB imaging analysis software package. For a given individual, time-courses were extracted for each time-locked event and collated across events, providing a single time-course for each condition, for each run. These time-courses were then inspected (blind to condition), where runs considered to reflect noise were removed from the rest of the analysis. Event-related time-courses were then converted to percent signal change (PSC) and averaged across runs for each participant. The three peak HRF timepoints (7, 8 and 9 seconds) were extracted and averaged. This provided a single numerical value for each condition (both repetition and non-repetition). The value for the repetition condition was subtracted from each experimental condition. This subsequently provided an RS value for each condition (for a review of the analysis pipeline, refer to Figure 2.3). The difference in percent signal change between identical repeats and partial repeats (scenes where object features change) was calculated and used as a measure of RS.

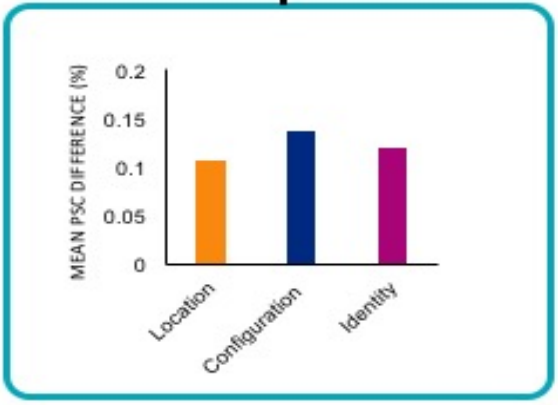
1. ROI definition in native space (individual subject space) from two functional localiser scans.



2. Time-series extraction from each run within each ROI across the number of volumes. Time-series refers to the raw data output collected in volumes across time. This is then transformed into a time-course (below) by sampling and averaging across all of the individual stimulus-locked events (each individual trial for each condition) from the time-series.



3. Time-course analysis of each stimulus-locked event. Quality control of the data and averaging across runs to get a HRF for each condition, within each ROI.



4. Calculation of mean RS score (Experimental condition- Repetition).

Figure 2.3. Pipeline for RS analysis. 1. ROI's are defined in MarsBaR, around the peak coordinate of the functional cluster for each individual participant. (2) Time-series extracted for each run. (3) Time-locked data for each condition was extracted for each run and converted to PSC. Three peak time-points are extracted and collapsed across runs. (4) Three peak time-points are collapsed and the average value for the repetition condition is subtracted from each of the other conditions.

2.2. Behavioural methods

2.2.1. Boundary Extension

BE is a memory phenomenon where participants falsely remember space beyond the occluded boundaries of a previously viewed scene (Intraub & Richardson, 1989). In their original BE study, Intraub and Richardson briefly presented participant with photographs of scenes (Figure 2.4). After a brief interval, subjects were asked to redraw the scene from memory. Participants' drawings typically included additional content beyond the image boundaries, which was not previously shown in the initial photograph. This result reflects inflation in memory of space beyond the borders of the initial photograph. Whilst the drawing method is still widely used, limitations of the method include the time taken to conduct the experiment, and the limited number of trials that can be included.

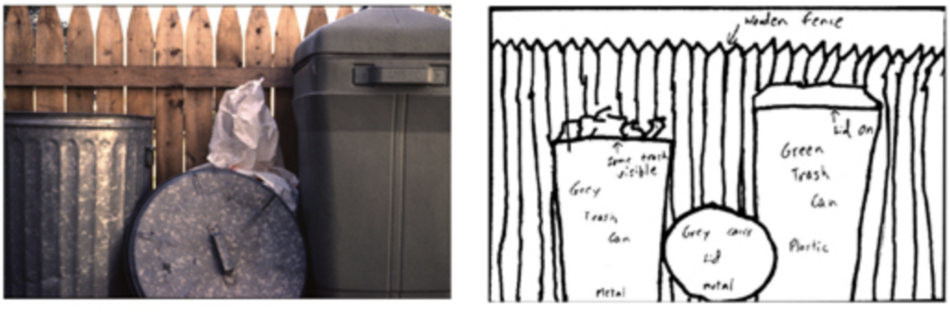
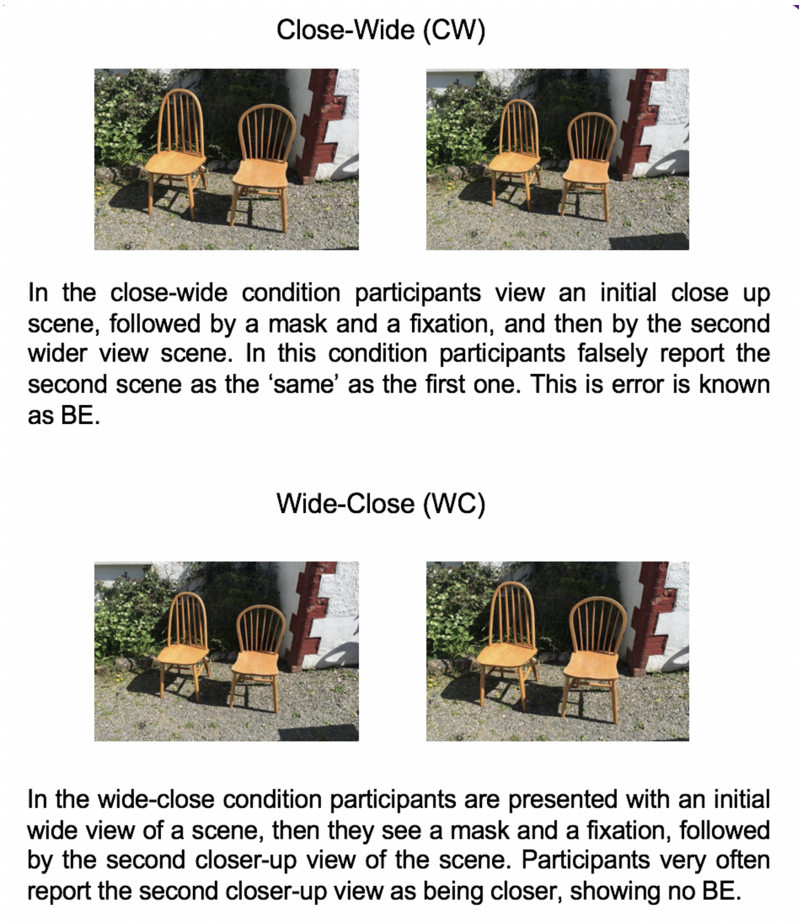


Figure 2.4. Example of the original photograph (left) and the participants drawing (right). This example shows the BE effect as observed in the drawing task. Participants include extra information and complete objects, which were only partially shown previously. Images adapted from Intraub and Richardson (1989).

In BE recognition/rating paradigms, participants are presented with pairs of scenes, separated by a distractor mask and fixation. Scenes can be presented in 2 trial types (Figure 2.5). First, scenes can be presented as close-wide (close-up view of a scene, followed by a wide-view of the same scene). Second, wide-close (initial wide-angle view, followed by a close-up view of the same scene). Participants are then required to rate whether the second picture is farther away, the same, or closer up than the initial scene. This can be done using either a 3-point (-1 farther; 0 same; +1 closer) or a 5-point scale (much farther; farther; same; closer; much closer) (Gottesman & Intraub, 1999; Intraub, Gottesman & Bills, 1998; Intraub, Gottesman, Willey & Zuk, 1996). BE is observed when participants falsely respond 'same' to scenes shown in a close-wide order. For scenes presented in wide-close orders, participants accurately detect that the second scene is closer up than the first. This asymmetry in responses between the two trial types is the measure used to estimate BE.



Close-Wide (CW)

In the close-wide condition participants view an initial close up scene, followed by a mask and a fixation, and then by the second wider view scene. In this condition participants falsely report the second scene as the 'same' as the first one. This is error is known as BE.

Wide-Close (WC)

In the wide-close condition participants are presented with an initial wide view of a scene, then they see a mask and a fixation, followed by the second closer-up view of the scene. Participants very often report the second closer-up view as being closer, showing no BE.

Figure 2.5. Two typical trial types that are implemented in BE. First close-wide (CW) where participants falsely report the second image as the 'same' as the first (defined as the BE error). Whereas for wide-close (WC) participants most often report the second scene as

‘closer’ than the first.

As discussed in Chapter 1, the BE error cannot be explained under the notion that scenes are encoded through a single visual input. The *multi-source* model of scene representation suggests that scenes are encoded through various inputs, building a coherent representation of space based on knowledge of the context, visual input, prior experience and knowledge of object-context relations (Intraub, 2010, 2012). This knowledge together allows parts of the scene, beyond the occluded borders, to be filled in and allows the observer to view the world in a contiguous manner.

There are two approaches when exploring BE. First, researchers have studied the mechanisms that drive such a phenomenon. This includes focusing on aspects such as field of view, type of image (line drawings, photographs), object size, object distance from the boundary and duration between initial image and second image presentation (Bertamini, Jones, Spooner & Hecht, 2005; Hale, Brown, McDunn & Siddiqui, 2015; Intraub & Berkowits, 1996; Intraub & Bodamer, 1993; McDunn, Siddiqui & Brown, 2014; Munger, Owens & Conway, 2005). Second, BE is implemented as a tool to ask questions about wider spatial cognition (Bertamini, Jones, Spooner & Hecht, 2005; Candel, Merckelbach & Zandbergen, 2003; Dickinson & LaCombe, 2014; Hale, Brown & McDunn, 2016; Mathews & Mackintosh, 2004; Menetrier, Didierjean & Vieillard, 2013; Mullally, Intraub & Maguire, 2012). This thesis adopts the latter approach, using BE as a tool to measure whether memory for the extent of space is influenced by the type of scene content.

2.2.2. BE in the current thesis

In the current thesis, BE error was examined in two approaches. First, BE was quantified by averaging the scores for closer (-1), the same (0) and farther (+1) and examined for a significant asymmetry (CW trials reported as closer to ‘same’ compared to WC trials). This is referred to as the mean boundary extension score (Intraub, Gottesman, Willey & Zuk, 1996).

Second, we studied BE by directly examining the proportion of ‘same’ responses for each condition. The actual BE error is observed when participants respond ‘same’ to a close-wide trial. Whilst examining the proportion of ‘same’ responses is not a widely-adopted approach (Intraub, Gottesman, Willey & Zuk, 1996), within the current thesis it provided a means of examining the direction of the BE error between trial types (CW and WC) for people and objects in scenes. In addition, it permitted a more precise investigation regarding how variations in scene content alter the degree to which the extent of space was remembered.

Chapter 3

Investigating object-in-scene representations in the human scene-selective cortex.

Bangor University, School of Psychology, Wales, UK.

3.1. Abstract

When viewing the world, human observers are presented with unexpected changes to their visual environment. This includes changes to the types of objects and the spatial layout of the scene. Despite extensive research in the field, it remains unclear how activation in response to spatial and non-spatial properties of object in scenes differs across the scene-selective structures (PPA, OPA, RSC). Using a paired-stimulus repetition suppression (RS) fMRI paradigm, we investigate whether RS across these regions differs as a function of the type of spatial or non-spatial condition. This included Locations, Arrangements Locations and Arrangements, and Identities of multiple non-landmark objects in natural scenes. Results showed no distinguishable release from adaptation for any condition in any ROI. These findings reflect the complex nature of scene representation in the human brain, and give rise to discussion about the neural basis of RS, highlighting potential contributing factors that may influence how the effect is observed (such as the modulation of attention to visual changes).

3.2. Introduction

Scenes are rich in detail and represent the local visual environment. However, due to the inherently complex nature of scenes and their lack of uniformity, establishing how they are encoded as a function of their basic visual elements is challenging. Evidence suggests that two broader forms of information are extracted from scenes. These are identified as i) the boundaries, surfaces and geometric/volumetric properties (Harel, Kravitz & Baker, 2013; Kamps, Julian, Kubilius, Kanwisher & Dilks, 2016; Oliva & Torralba, 2001) and ii) features such as the type of objects, local textures and elements that provide identity and context to the scene (Bar, 2004; Bar & Aminoff, 2003; Janzen & van Turennout, 2004).

In a pursuit to determine how these types of visual cues are encoded within the brain, researchers have used methods such as fMRI to identify dominant brain structures that show selectivity for processing images of scenes. These include the PPA, RSC and OPA (Dilks, Julian, Paunov & Kanwisher, 2013; Epstein & Kanwisher, 1998; O'Craven & Kanwisher, 2000). In recent years, there has been a shift in focus to examine how these regions encode spatial properties of objects (Troiani, Stigliani, Smith & Epstein, 2014). However, despite these advances it remains elusive how these scene-selective structures differentially encode object-in-scene representations. More specifically, variations in the locations, arrangements and identities of multiple objects.

3.2.1. Spatial properties of objects in scenes

Objects (particularly landmark objects) play a key role in spatial navigation (Aguirre & D'Esposito, 1999; Epstein & Vass, 2013; Troiani, Stigliani, Smith & Epstein, 2014). However, not all objects we see in the environment are considered landmarks. Janzen and Turennout (2004) devised a study where participants viewed virtual reality museum video clips, and were asked to attend to the objects specifically. Results showed that the PPA elicited higher levels of fMRI activation for objects that were located in areas characterised as critical decision points for navigation. This suggests that the PPA holds representations about spatial locations of non-landmark objects which guide behaviour. This is also supported by more recent research which finds that the PPA and RSC encode spatial properties of individual objects (Auger, Mullally & Maguire, 2012).

Auger, Mullally and Maguire (2012) examined the extent to which different spatial properties of objects are coded in the PPA and RSC, including saliency, the object size, its degree of permanence and navigational use. Their findings indicated that the PPA responded significantly across all conditions, but the RSC primarily coded information about the degree of object permanence. Additionally, participants who had worse navigational ability were less proficient at determining an objects permanence; this was reflected as reduced activation in RSC compared to good navigators. This research suggests that while both regions are critical for spatial processing, the types of information that they encode about objects is largely dissociable, where more anterior regions such as the PPA are thought to process multiple forms of object-in-scene information.

Likewise, Troiani, Stigliani, Smith and Epstein (2014) demonstrated that the core scene-selective structures (PPA; RSC; OPA) encode several spatial properties of objects within scenes (such as permanence, visual size, context, physical size, distance). The PPA and RSC encoded spatial properties of objects both within a scene and isolated from a scene background but mostly for object properties that reflected landmark-suitability (high in permanence, spatially defining/defining a large portion of space, and contextually related). However, the OPA showed increased levels of BOLD signal to these properties, more so, when objects were isolated from a background. Upon an item analysis (of visual dimension condition), the PPA appeared to be more sensitive to contextual driven properties such as the 'placeness' and 'space defining', whereas the RSC was more sensitive to geometric based properties associated with the object identity such as 'size' and 'distance'. This evidence suggests that despite these regions being sensitive to multiple spatial properties of objects, the precise functional processing properties can be dissociated between regions in some instances. Moreover, in comparison to the PPA and RSC, the OPA was found to encode more basic spatial properties of objects, less reliant upon detailed background content. These findings suggest that these regions encode a wealth of information, including spatial properties of objects, and that the nature of these representations is dissociable, where more anterior regions such as PPA and RSC encode more detailed scene context information and the more posterior OPA represents more basic information about the type of object. This evidence gives rise to ask questions regarding the nature of representations in the PPA, RSC and OPA in relation to other spatial and non-spatial dimensions of objects in scenes, such as layout and identity. This is supported by evidence suggesting that there is spatial convergence across the networks that support object and scene processing, where these regions may also encode relative information about object layout and identity in natural scenes (Linsley & MacEvoy, 2014, 2015; MacEvoy & Epstein, 2011).

3.2.2. Scene network contributions to encoding object identity and scene layout

Epstein, Graham and Downing (2003) used a paired-stimulus RS fMRI paradigm to probe the processing properties of the PPA. Here, the authors studied changes in viewpoint, object identity, place change and no change (repetition). Results showed a release in adaptation for changes in viewpoint, which were similar to that of the complete place change condition. These results demonstrate that the PPA encodes each viewpoint as if it were a new scene. Interestingly, results revealed no release in adaptation for changes in object identity. While these findings reveal that the PPA is encodes viewpoint specific representations of the world, others have shown that the PPA is involved in encoding multiple forms of information relating to global spatial properties and object identity.

Harel, Kravitz and Baker (2013) used fMRI to further examine the processing properties of the PPA and RSC. In their study, scenes were manipulated across two dimensions: spatial layout and object content. Results revealed that the PPA was sensitive to both the global spatial layout of the scene and the identity of the objects. However, the RSC was primarily sensitive to the spatial layout and not the type of scene content. These results suggest that these regions have dissociable functional roles in processing these types of visual dimensions, including layout and identity. Does this extend to other types of layout, such as object locations and arrangements? Given these results, one would expect that the PPA would be sensitive to object identity in scenes and the locations/arrangements of objects, whereas the RSC would primarily be invovled in encoding the spatial layout of objects in scenes. While the Harel Kravitz and Baker paper did not examine the OPA, it might be suggested, based on prior research, that the OPA would be sensitive to object layout (Dilks, Julian, Paunov & Kanwisher, 2013). Furthermore, research examining the functional processing properties of the MTL structures have identified a role for the PPA in encoding object identity and the configuration of familiar objects in non-scene grid arrays (Pihlajamaki et al, 2004). Comparably, Fenske, Aminoff, Gronau and Bar (2006) suggest a role for the PPA and RSC in encoding spatial more than non-spatial associations of objects. These objects were again presented in non-scene arrays.

Together, these findings above demonstrate the inconsitencies in our current knowledge regarding the precise role of these regions in encoding spatial and non-spatial information about objects in natural scenes. These inconsistencies in findings may be explained by the differences in research questions and experimental designs used for these papers. For example, the Harel, Kravitz and Baker (2013) experiment used a multivariate approach to study representations in this region, whereas Epstein, Graham and Downing (2003) used repetition suppression with a different research question and analysis approach. Therefore, one must keep in mind that with different approaches and questions, these findings are not directly comparable.

Furthermore, the absolute functional role of the PPA, RSC and OPA in encoding spatial and non-spatial properties of objects in scenes it is largely unclear. Specifically, how they encode non-landmark objects and their locations/arrangements in natural scenes. Furthermore, prior evidence has predominantly examined the PPA and RSC, however it is important to determine the contributions of the more recently defined OPA in processing these forms of object information in natural scenes.

3.2.3. The current experiment

In the preface of the thesis, I describe a situation where an observer examines a view out of their office window. In that example, the structure of the scene and the shape of the space remain constant. However, across time, and multiple exposures, there may be changes to the locations, arrangements and identities of the objects. This experience is one which human observers encounter daily, yet, our knowledge regarding how the scene-selective structures encode these types of spatial and non-spatial variations remains limited. Therefore, the current experiment used fMRI as a tool in a paired-stimulus RS design to examine how spatial and non-spatial information derived from objects in scenes is represented and reflected by changes in BOLD across scene-selective structures (Kourtzi & Kanwisher, 2001). The experiment consisted of 5 primary experimental conditions outlined below:

1. **Location changes**: The locations of objects change across consecutive views of a scene. Therefore, we included this condition to examine how changes to the locations of multiple objects are encoded within the scene-selective cortex. For this condition, the arrangement of the objects was kept constant. That is, whilst they moved locations, their spatial arrangement to one another remained the same.
2. **Arrangement changes**: In the world, observers are often presented with situations whereby the arrangement of objects to one another may change, whilst they remain in the same location. For example, an observer may see the same three cars parked in the same three car parking spaces, however on a daily basis, those cars may be parked in any one of those three spaces. In this situation, the arrangement of the cars may change, but their locations within the car park will remain the same.
3. **Both Location and Arrangement changes**: This condition examined whether any of these regions encode global spatial changes (both changes in the location of the objects, or the local arrangement of the objects).
4. **Identity changes**: Here, we examined whether these regions are primarily sensitive to spatial properties of objects or whether they are sensitive to changes in the identity of objects in scenes. In this condition, the location of the three objects remains constant but each of their identities changes.
5. **Repetition**: The repetition condition here was a necessary component of the design. For example, the repetition condition allows us to measure if there was a significant release from adaptation for any of the experimental conditions relative to the repetition condition.

The present experiment draws its hypotheses from the broad themes identified in the field of visual processing in the occipital and temporal regions which infer that more posterior regions encode object based information (Kamps, Julian, Kubilius, Kanwisher & Dilks, 2016; Kourtzi & Kanwisher, 2000) and more anterior regions encode information about the layout and context of a scene including object location and configuration (Harel, Kravitz & Baker, 2013; Janzen & van Turennout, 2004; Kravitz, Peng & Baker, 2011; Mullally & Maquire, 2011). Therefore, the in the present study we predict that areas such as the OPA will be more sensitive to object identity based information and the more anterior regions such as RSC and PPA might show a larger release in adaptation for object location and configuration changes.

3.3. Methods

3.3.1. Participants

Functional MRI data were collected from 17 healthy participants (Mean age: 29 years; 8 males) from Bangor University and the community. Data for two of the participants could not be analysed due to poor data quality. Subsequently the statistical analysis included *n*=15 participants. The sample size was based upon similar previous studies examining spatial properties of objects (Aminoff, Gronau & Bar, 2007; Troiani, Stigliani, Smith & Epstein, 2014). Participants could fall within any category of handedness and were native-English speaking. No participants reported any neurological disorders, and all had normal or corrected to normal vision. All participants were reimbursed with £10 for their time.

3.3.2. MRI safety screening and Ethics

To assess whether a participant was eligible to take part and be scanned, they were required to complete safety questionnaires administered by Bangor University's Brain Imaging Unit. All participants were provided with full details of the experiment (experimental information and MRI safety information), prior to the day of scanning. The questionnaires screened for multiple pre-existing neurological deficits, hypertension, and metallic implants including dental work. This involved an in-depth interview with a qualified MRI operator who discussed all safety considerations with each participant. Participants were required to remove all metallic objects from their possession and were asked to remove all jewellery and metal from the body prior to entering the scanning room. Participants were subsequently told that they could withdraw from the experiment at any point; regular checks throughout the experiment were put in place to ensure participants were comfortable to carry on after each run. All participants were fully informed of the procedures and gave consent to take part. All participants were given a full debrief after the scan and had the opportunity to ask any questions regarding the MRI procedure and experiment. This experiment was granted full ethical approval from the Bangor University Psychology Ethics Committee and the Bangor Imaging Group. This study was also presented to the Bangor Imaging Group who approved the imaging methods, study justification and analysis approach.

3.3.3 Stimuli and Design

The RS experiment employed an event-related paired-stimulus fMRI design. This design type permitted modelling of the HRF across each trial (Epstein, Graham & Downing, 2003; Kourtzi & Kanwisher, 2001). A total of 7 conditions were included in the experiment. These included changes in object Location, Arrangement, Location and Arrangement, Identity, no change (repetition), fixation and target (task trials: inverted scenes). (Figure 3.1).

Stimuli consisted of 19 families of scenes each consisting of 5 scene photographs. Previous fMRI studies have explored scene representation with single foreground objects (Epstein, Graham & Downing, 2003; Harel, Kravitz & Baker, 2013; Troiani, Stigliani, Smith & Epstein, 2014). However, in this experiment each scene included 3 focal objects which changed in one of the visual dimensions (location, arrangement, identity and location/arrangement). Three objects were chosen as it allowed us to control for any effects of 'load' or number of objects, while still allowing us to change the arrangement of the objects within the scene (see Appendix B). For the identity change condition, the objects were controlled for size to ensure that there were no effects of object size. For the other condition manipulations, all objects within each family remained the same, therefore any effects observed could not be a result of the type of objects included. All scenes within each family were presented from the same viewpoint. Objects within scenes were not considered navigationally relevant. This applies for Chapter 4 where identical stimuli were used. All stimuli were made by the author/experimenter. All rendered images were natural, coloured photographs, including both indoor and outdoor stimuli, and were 1024 x 683 pixels in size.

Figure 3.1. Example of the condition changes in the paired-stimulus RS design. There are a possible of 5 conditions as well as a fixation condition and the task condition.

3.3.4. Experimental Task and procedure.

The task devised for Experiment 1 required participants to press a button on the response pad in the scanner when they detected an inverted scene. This task required participants to attend to each of the stimuli, and ensured they were paying attention throughout the duration of the experiment. Conditions were carry-over counterbalanced. Carry over counterbalancing allows us to confirm that every type of scene is preceded an equal number of times by every other type of scene. It is the best way to ensure that there are no stimulus specific driven effects in the data (muddied results over trials) and also ensures that every participant sees every possible order of stimulus in each trial. This is particularly beneficial for the event-related design and helps when modelling the fMRI time-course data. This was also important for this design as it reduced the likelihood of participants predicting repetition trials, which is thought to influence the magnitude of the RS effect (Summerfield, Trittschuh, Monti, Mesulam & Egner, 2008).

Participants took part in 4 experimental runs each lasting 5 minutes 45 seconds. Each run of the experiment consisted of 109 trials, which included 5 additional fixation scans at the start and 6 at the end of the scan. Each run included one additional fixation trial to allow for the carry-over counterbalancing design. This resulted in 98 trials which were modelled in the analysis. Each trial started with one of the scenes from each family, this was known as the base scene (see Appendix B). To represent a change, the second scene would be one of the other photographs from the same family whereby the objects may have changed in one of the visual dimensions, or was identical (repetition). Each trial lasted 3 seconds. Each trial started with the first stimulus presentation of 600ms, a 400ms fixation and then a second presentation of 400ms. Finally, a 1600ms response period was presented for participants to make a behavioural response using the button box provided in the scanner (Figure 3.2). The experiment was programmed using PsychToolbox (The MathWorks Inc., Natick, MA, 2000).

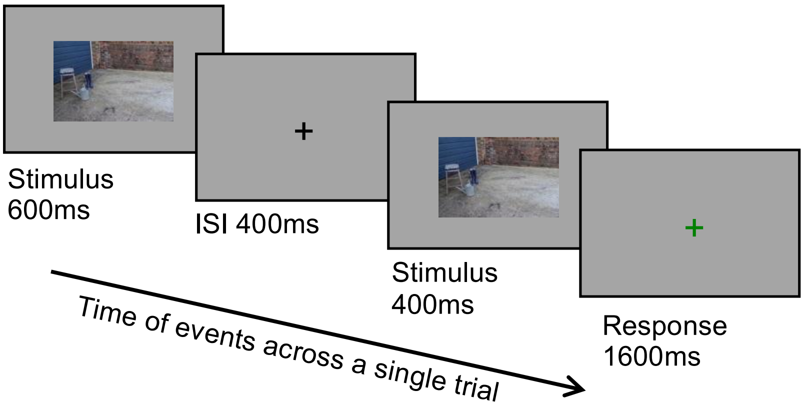


Figure 3.2. This schematic representation shows a typical individual trial from the experiment.

3.3.5. Data Acquisition

3.3.5.1. Anatomical scan

Data were collected on Bangor University's Phillips Achieva 3Tesla MRI scanner with a SENSE-Head-32AH coil. The anatomical scans were collected using standard acquisition measures, a TR of 8.4 seconds, TE of 3.8 seconds and a slice thickness of 1mm. Each voxel on the anatomical scan was 1mm isotropic.

3.3.5.2. Functional ROI localisers

Acquisition of T2* weighted images used a single shot echo Planar Imaging (EPI) sequence. A TR = 2000ms and an echo time (TE) = 35ms was chosen with a flip angle of 90°. A total of 168 volumes (scans) were collected across 29 off-axial slices at 3mm thickness. Voxels were 3mm isotropic.

3.3.5.3. Experimental runs

For the functional T2* weighted images, a single shot EPI sequence with a TR=2000ms/ TE=35ms and a flip angle of 90° was used. There were 164 volumes collected at 29 off-axial slices; with a slice thickness of 3mm. Moreover, each voxel size was 3mm isotropic.

3.3.6. Data analysis

3.3.6.1. Image preprocessing

All images were preprocessed using SPM8 a toolbox from MATLAB (MATLAB and Statistics Toolbox Release R2012a, The MathWorks, Inc., Natick, Massachusetts, United States.). All images were slice time corrected, realigned to the mean image, coregistered to each individual subject's anatomical image, and smoothed with a 4mm Gaussian kernel (FWHM). Smoothing was only applied to activation maps for the localiser scans which established the regions of interest. Smoothing was not run on the data used for time-course extraction.

3.3.6.2. ROI definition

Independent functional localisers were run to define ROIs at the individual subject level for the OPA and PPA. For the RSC, the ROI masks were defined at the group-level (in normalised space) due to a number of participants who did not show bilateral RSC activation (*n*=5). This effect has previously been reported in the literature (Hodgetts, Shine, Lawrence, Downing & Graham, 2016). Activation in the RSC does not always fall within the anatomical borders of the region. That is, functional localisation does not always align with the anatomical borders of the RSC. Additionally, it is notably more difficult to localise during a functional scene localiser experiment, in comparison to the PPA and OPA (Hodgetts, Shine, Lawrence, Downing & Graham, 2016), potentially due to its more prominent role in spatial memory for scenes.

To localise these ROI's, an fMRI functional localiser was conducted which consisted of 16 second blocks including an array of either bodies, chairs, faces, scenes or fixation (rest blocks). The task used for the localiser was a 1-back task where participants were required to respond via the fMRI button-box upon detection of an identical image (image presented twice in succession). The localiser was run twice for each participant and lasted for 5 minutes and 50 seconds, and was used for both subject-level definition (OPA/PPA) and group-level definition (RSC).

First-level analysis of localiser data was conducted in SPM8 in order to localise regions of interest (ROIs) bilaterally. Onsets for each block were modelled in SPMs 'model specification' GUI for each run. For each individual, a [scenes > faces] contrast was established with a threshold of $p < .001_{\text{UNCORRECTED}}$, to identify key clusters in the scene selective network. To build the independent ROIs, 6mm spheres were plotted around the local maxima of each cluster to include all voxels that meet the threshold ($p < .001_{\text{UNCORRECTED}}$). The ROIs were built using MarsBaR, a toolbox for MATLAB (Mathworks, USA). The ROIs were saved as nifti images and used as masks to extract BOLD time series from each region for the RS experimental runs. For the RSC, localisers were normalised to MNI space in SPM8 where the coordinates were identified for the RSC bilaterally ($p <. 001_{\text{UNCORRECTED}}$) to create an ROI mask around the local maxima, which included all significant voxels. This group-level mask was used to extract a BOLD time series from the RSC for all participants.

3.3.6.3. Time-course analysis.

The time course extraction was conducted in MarsBaR, a toolbox of SPM (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). For an individual participant, raw time-series were extracted from each ROI (OPA, PPA, RSC), for each run separately. Events were time-locked so that data could be extracted from the time-series for each condition, and plotted over 18 time points (seconds). This allowed the building of separate time-courses for each condition, for each run of the experiment. Quality control of the data included examining the pattern of the HRF's for each run, for of each participant, to ensure the HRF patterns looked representative of a typical HRF (see Appendix A for an example). These were collapsed across conditions to ensure the experimenter was blind to the individual condition responses. Once data were checked for quality, the BOLD time-courses were then converted to percent signal change (PSC), using custom MATLAB scripts. The PSC data for each experimental condition was taken from the three peak time-points of the time-course (7, 8 & 9 seconds), averaged, and then compared against the repetition condition (condition – repetition) in order to gain the measure of suppression. That is, the mean value for a given condition minus the mean value for the repetition condition, resulting in the mean difference PSC. This data was then used for the statistical analysis. Initial analysis examined effects of hemisphere, with no effect of hemisphere, data are collapsed for the remainder of the analysis. A repeated – measures analysis of variance (RM-ANOVA) was run to examine differences in RS between conditions, across ROIs.

3.4. Results

Two participants were excluded from the data analysis due to poor data quality; this included what was thought to be a problem in the scaling of the data extracted from the scanner. No participants had to be removed as a result of head motion (>3mm movement in any direction). Thus, 15 participants were included in the subsequent analysis. The inversion task indicated that participants remained attentive to the task throughout the duration of the experiment (>50% accuracy rate) and thus, no participants needed to be excluded. The low response is thought to reflect participants pressing the button too quickly and thus, their response was not recorded. When the participant was in the scanner, the experimenter ensured they had understood the instructions by observing their responses from the MRI control room.

3.4.1. Region of interest analysis

Individual localisation was successful for all participants bilaterally in the PPA and

OPA; but not in the RSC. Therefore, for the RSC, analysis was run at group level (See localiser methods) (Figure 3.3).

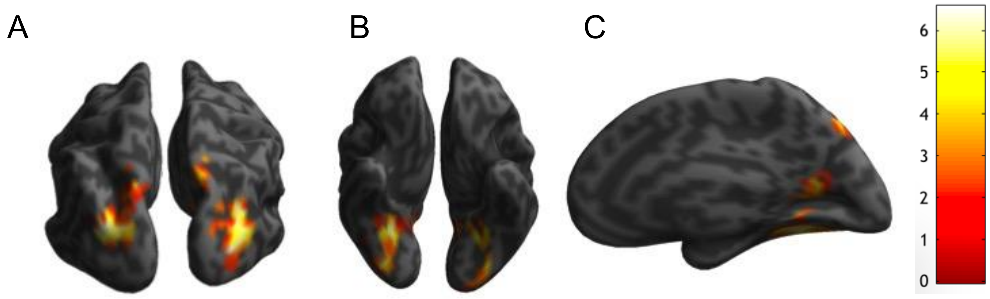


Figure 3.3. Scene-selective regions of interest; A) OPA, B) PPA, and C) RSC. Images show one participant's localiser data concatenated across both runs. The threshold used in order to localise on these regions was $p<.001_{\text{uncorrected}}$. The colour bar represents t-values.

3.4.2. Time-course analysis

The data used in the following analysis was the mean difference for each condition from the repetition condition, collated across the peak time-points (7, 8 & 9 seconds) (figure 3.4).

A RM-ANOVA [2 (Hemisphere) x 2 (ROI: OPA, PPA) x 4 (Condition: Location, Arrangement, Location & Arrangement, Identity)] was first run to explore putative influence of hemisphere. The RM-ANOVA showed no effect of hemisphere [*F*(1,14)=.443, *p*=.516]. Therefore, data from the two hemispheres were averaged for the further analysis. Additionally, the RSC showed no significant effect of hemisphere [*F*(1,14)=.905, *p*=.358] and was also averaged across hemisphere. In addition, there were no significant interactions with hemisphere (*p*>.05).

A RM-ANOVA [2 (ROI: OPA, PPA) x 4 (condition: Location, Arrangement, Location & Arrangement, Identity)] was run to determine the contribution of both object identity and spatial properties of objects between regions in the scene selective network. There was no significant main effect of ROI [*F*(1,14)=1.52, *p*=.238, η2p = .098] and no significant main effect of condition [*F*(3,42)=1.01, *p*=.397, η2p = .067], suggesting no dissociable differences in the degree of RS between conditions. Similarly, there was no significant interaction between ROI and Condition [*F*(3,42)=.40, *p*=.755, η2p =.028]. A RM-ANOVA on the RSC with 4 levels (location, arrangement, location & arrangement, identity), showed no significant main effect of condition [*F*(3,42)=.24, *p*=.870, η2p = .017]. These results show that there were no significant differences in RS between conditions, and no dissociable effects across ROIs (Figure 3.5).

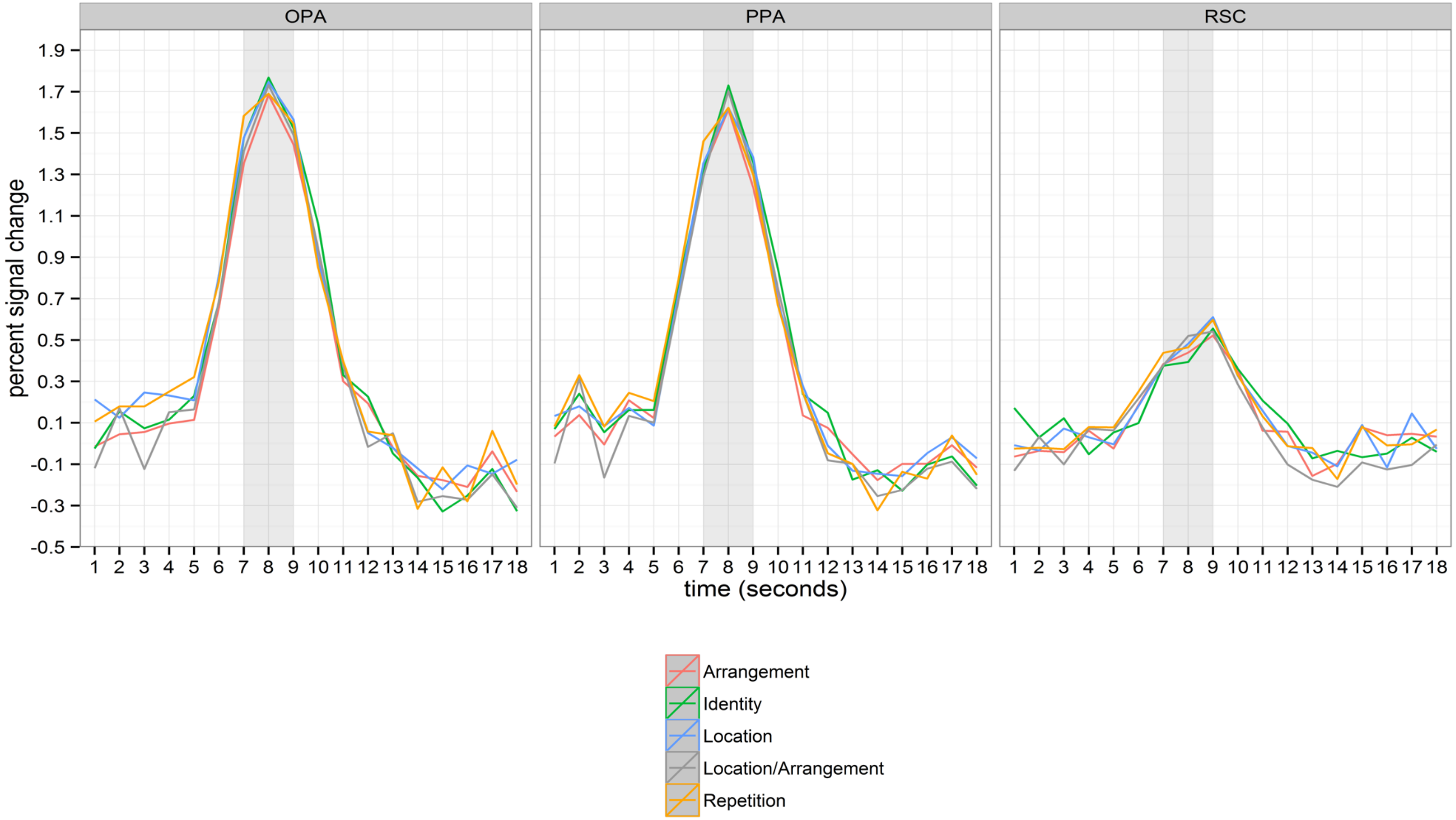


Figure 3.4. HRF time-courses for each ROI, for each condition. Percent signal change for the analysis was sampled during the HRF peak at timepoints 7, 8 and 9 seconds (shaded in grey).

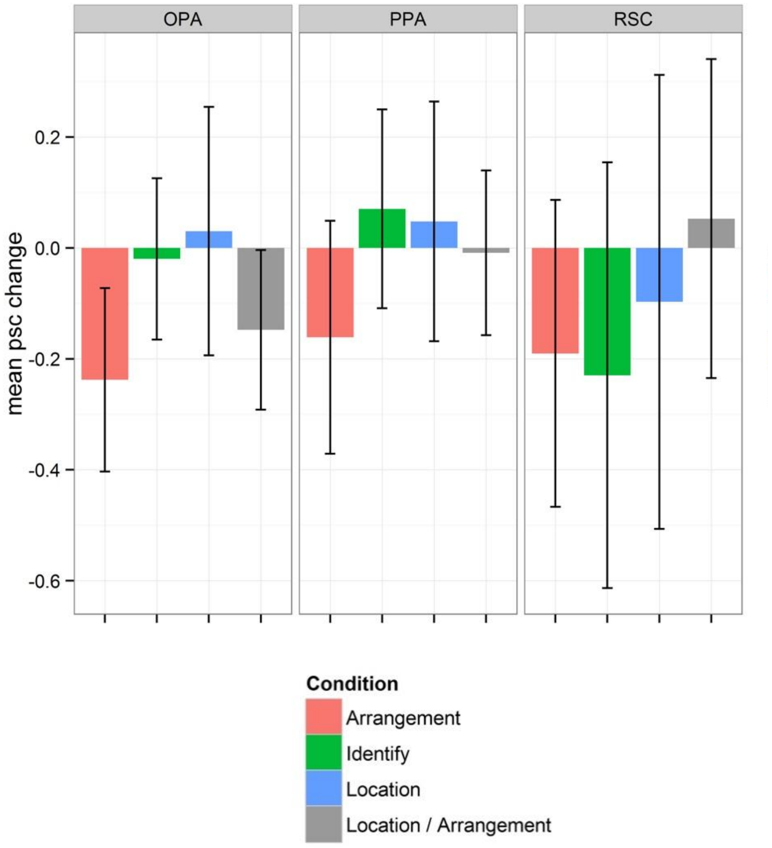


Figure 3.5. Mean percent signal change differences relative to repetition for each experimental condition in each ROI. Error bars are representative of standard error (*S.E*).

3.4.3. Effect of RS

Upon analysis of the time-courses it became apparent that there was no visible effect of RS. To test whether RS occurred within each ROI, paired t-tests were run between the averaged condition values (averaged across Location, Arrangement, Location & Arrangement and Identity) and the repetition values for each subject. Results showed that there was no overall effect of RS within the PPA [*t*(14)= -0.09, *p*=.934] , RSC [*t*(14)= -0.44, *p*=.664], or OPA [*t*(14)= -0.80, *p*=.440]. These results are discussed further in relation to the current experimental design.

3.5. Discussion

The present experiment was designed to assay how core scene-selective regions differentially respond to variations in spatial and non-spatial properties of objects in natural scenes. Specifically, we used a paired-stimulus RS fMRI design to examine whether these regions are sensitive to encoding object Locations, Arrangements and Identities within natural scenes.

In the current experiment, we sought to explore what types of spatial and non-spatial properties of objects these regions are sensitive to. Given that the literature regarding the functional roles of these regions is littered with inconsistencies, we did not have strict hypotheses about the dissociable effects between these regions for the conditions tested. However, based on more recent evidence, we speculated that the PPA may be involved in more global layout based information within scenes such as location (Epstein, 2005), and the RSC would be more sensitive to the arrangement and location of object within scenes. Conversely, with very limited research to date studying the precise functional role of the OPA in scene encoding were did not hold any clear hypotheses about the precise functional role of the OPA in any condition.

Results revealed no dissociable differences between conditions within any of the ROIs. It was further shown that there were no global RS effects in any of the ROIs. Could this suggest that these regions do not encode these forms or information or might these results be explained by a methodological consideration? This paired-stimulus RS design was adapted from that of Kourtzi and Kanwisher (2001) and Epstein, Graham and Downing (2003) who observed significant attenuation effects in the LOC and PPA for stimulus feature changes. Thus, our findings lead us to speculate about the possible methodological considerations associated with RS experiments, such as attention and the types of tasks used.

3.5.1. Task effects, Attention and RS

One explanation for these results is that participants were not directly attending to the changes between each pair of scenes. In the present experiment, participants were instructed to press a button upon detection of an inverted scene, requiring their attention to each scene presentation throughout the experiment. Critically, the inverted scene could have appeared in either the first, or the second scene presentation in each trial and only occurred as often as each other condition.

In the current study our main question aimed to explore how variations in scene content are represented within the brain. This RS study was designed for fMRI and we did not have any prior hypotheses about behaviour. The task was introduced into the design to try and ensure participants remained attentive during scanning while still allow for the carry-over counterbalancing design. However, while participants remained attentive to the scenes, it could be suggested that they may not have noticed the changes between the scenes within each pair, and were instead awaiting a ‘pop-out’ effect from the inverted scene. Specifically, participants may have directed their attention to awaiting the presence of an inverted scene before responding, therefore reducing their attentional focus to less salient changes between the pair of scenes.

Eger, Henson and Dolan (2004) studied the effects of attention in observing BOLD during RS. In their study they identified that attention to images of objects was critical for evoking RS effects, whereas when attention was directed away from the objects, the RS effect disappeared. In the case of the present study, the absense of RS was not modulated by explicitly ‘ignored’ stimuli, but it could be suggested that the task inadvertently modulated participants attention towards detecting the inverted scene and away from the subtle scene variations.

Prior evidence has suggested that RS/adaptation should be automatically observed across the cortex as it is driven by various underlying neural mechanisms which also support repetition priming (Wiggs & Martin, 1998). Conversely, others have demonstrated the importance of attention in RS paradigms, reporting that varying task types can influence how RS is observed (Eger, Henson, Driver & Dolan, 2004; Henson, Shallice, Gorno-Tempini & Dolan, 2002). Henson, Shallice, Gorno-Tempini and Dolan recognised that RS is not an automatic response to repeated stimuli, but is in fact driven by top-down attention mechanisms. Indeed, this does not explain instances where RS is observed during passive viewing (Jessen et al., 2002).

The mechanisms that drive changes in BOLD during RS are still undetermined (Barron, Garvert & Behrens, 2016; Feuerriegel, 2016; Hsu, Hamalainen & Waszak, 2014; Kovacs & Vogels, 2014). Some evidence suggests that the type of stimuli, the regions of interest examined, and the type of task needs to be accounted for in any RS experiment (Epstein, Parker & Feiler, 2008). Thus, with such variation across fMRI results it is challenging to know exactly what was driving the absence of RS in the present experiment. However, based on the evidence outlined above we consider this to reflect a methodological concern such as the type of task used.

Nonetheless, these findings suggest that direct attention to the changes between scenes in this particular paradigm may be critical for achieving the desired attenuation effect, and that the type of task used could have modulated participants' attention specifically toward inverted scenes and subsequently reducing their attention to observing variations in scene content between pairs of scenes. To the best of our knowledge, this has not been explicitly examined for paired-stimulus designs in response to complex scene encoding.

3.5.2. Considerations and future research

A major strength of the paired-stimulus fMRI paradigm is that we can infer how variations in visual content are represented within specific brain regions. Given that the visual world is dynamic, having a paradigm that can infer how the brain encodes subtle variations in scene content is hugely advantageous. However, in the present study we suggest that the type of task used may explain the inconsistencies in findings. A main limitation in the field is the inconsistency in our knowledge regarding how the type of task modulates RS, specifically for a paired-stimulus design.

Therefore, moving forward, it is important to consider how attention focus during the task changes the way in which RS is observed. For example, will developing a task that requires participants to directly attend to the changes between image pairs be strong enough to evoke RS? Examining whether the type of task changes how RS is observed, will provide valuable contribution to other researchers considering using RS as a method to study how variations in visual content are represented across various cortical structures.

3.5.3. Concluding remarks

To conclude, more research is required in order to determine the mechanisms that influence RS in the context of the present experiment, and other similar RS paradigms. The following chapter examines these inconsistencies and delivers a new task directing participants' attention directly to the changes between scenes, to examine how variations in scene content are represented in the scene-selective cortex.

Chapter 4

Examining object-in-scene representations: Task differences in fMRI repetition suppression

Bangor University, School of Psychology, Wales, UK.

4.1 Abstract

The experiment in Chapter 3 used fMRI in a paired-stimulus RS design to examine how variations in spatial properties (Locations and Arrangements) and non-spatial properties (Identities) of objects are represented in core scene-selective structures (PPA; RSC; OPA). The previous study revealed no effect of RS in any ROI, considered a reflection of the type of task used (detection of inverted scenes). The present study built upon this observation by using the same paradigm to re-examine the representational content of these brain structures, while using a task that directed participants' attention to the changes between paired scenes. Results indicated a significant effect of RS in the OPA and RSC, however no dissociable differences in RS were observed between experimental conditions. No effects of RS were observed in the PPA. Results of the whole-brain analysis revealed bilateral clusters in the inferior parietal lobe (IPL) for changes in object Arrangement compared to changes in object Location. These findings are discussed in relation to our current research questions. While we are unable to draw any further conclusions about the processing properties of the scene-selective structures, our results indicate the importance of direct attention to changes between paired scenes, for observing RS.

4.2 Introduction

Research studying the functional architecture of the human brain have identified regions which show specialised hemodynamic properties in response to certain categories of stimuli. In the domain of visual scene processing there are three prominent structures, which are thought to encode information about the visual environment (PPA; RSC; OPA). Despite recent evidence suggesting these structures encode spatial properties of objects in scenes (Troiani, Stigliani, Smith & Epstein, 2014), our knowledge regarding how these regions further encode variations in spatial layout and object identity is far from clear. Given that objects form a large portion of our visual experience; asking these questions is critical, in order to better understand human visual spatial cognition.

In the previous Chapter, participants took part in a paired-stimulus RS paradigm, where they were required to press a button upon detection of an inverted scene. While the task was only included to ensure participants remained attentive throughout the duration of the experiment, results revealed no effect of RS within any of the ROIs; and no dissociable differences in RS between conditions. Thus, we considered that the type of task, and possible attention mechanisms associated, might have influenced these results.

4.2.1. Considerations for findings in Chapter 3

When considering the results from Chapter 3, three possible explanations arise. First, these particular scene-selective structures are not sensitive to the spatial and non-spatial properties of objects, which were examined in Chapter 3. Second, the paired adaptation method may not have been sufficient for assessing subtle changes in scene content. Third, the task used (inversion detection) may have distracted participants from noticing the changes between stimulus pairs and thus, a lack of attention may have prevented a release in adaptation.

When considering the first instance it is plausible that these regions may not be sensitive to these particular kinds of features/manipulations. However, prior evidence has suggested that these regions, particularly the PPA, are sensitive to object arrangements and some suggest even identity in non-scene arrays (Aminoff & Tarr, 2015; MacEvoy & Epstein, 2011; Pihlajamaki et al., 2004). In other instances, these regions have been shown to elicit BOLD activation in response to basic spatial properties of objects in scenes (Troiani, Stigliani, Smith & Epstein, 2014). Thus, based on these previous findings, it is plausable to study these regions further in response to spatial and non-spatial properties of objects in scenes.

Second, we must consider whether the paired-stimulus design was sufficient for detecting subtle changes between scenes. This design was chosen specifically to examine how variations in scene content are represented within these regions. Similar research has used the same design and successfully observed effects of RS (Epstein, Graham & Downing, 2003; Kourtzi & Kanwisher, 2001). Epstein, Graham and Downing used this particular RS method to examine the PPA's response to processing small manipulations of scene content/context. Results demonstrated that this version of the RS paradigm was sensitive enough to evoke attenuation effects for subtle scene-feature manipulations in natural scenes. Thus, when evaluating these studies, it is unlikely that the style of RS paradigm was an influencing factor.

Our third consideration and the one that we examine in the present Chapter, is whether the type of task influenced the likelihood of observing RS. Here, we propose that the type of task in the previous chapter may have distracted participants from detecting changes between the scenes within each pair, and therefore, instead of attending to the scene variations, participants may have overtly awaited the inverted scene to appear. Therefore, while they may not have 'ignored' the changes per se, the task may have directed their attention away from more subtle scene changes.

4.2.2. Towards an understanding of attention in repetition paradigms

Research examining the underlying mechanisms that influence the BOLD response have identified that independently of RS, attention is critical for evoking the BOLD signal. Comparably, if stimuli are unattended, the BOLD signal reduces (Kastner & Ungerleider, 2000; Mack & Rock, 1998). Prior evidence has discussed the notable effect of attention in modulating the degree of RS (Henson, Shallice, Gorni-Tempini & Dolan, 2002; Eger, Hendon, Driver & Dolan, 2004). However, these findings contradict those showing that, the adaptation effect can be observed during passive viewing, with no attention-demanding task (Jessen et al., 2002).

Indeed, there has been large discussion in the literature regarding the effects of attention on RS. Specifically, RS is thought to be reduced when the repetition trials are predictable (Summerfield et al., 2008). However, in the case of the present study, the carry-over counterbalancing design reduced the predictability of events. Here, Summerfield et al suggest that attention directed towards one condition more than another will increase the likelihood of seeing a release in adaptation (relative to repetition). Interestingly, this might suggest that attention directed away from the experimental conditions, might subsequently reduce the chance of seeing any RS effects.

In light of our first experiment, if participants were not attending to the changes between pairs of scenes, and direct attention to the changes is critical to evoke RS, then this would explain why no significant differences were observed between repetition and non-repetition conditions. However, it is critical to note that there may be differences in the underlying mechanisms that drive attention during repetition suppression, specifically for the paired-stimulus design.

4.2.3. Possible influence of task in RS

A number of discussions have arisen about the importance of experimental design in fMRI studies, and more recently this has included the type of tasks used in RS experiments. The inversion task included in our fMRI experiment in Chapter 3 was previously used by Summerfield et al (2008), however, the effect of this type of task has more recently been reviewed by Larsson and Smith (2012) where it is now thought that to maximise subtle RS effects, the task should be designed to focus participants' attention directly to the scene content (some examples include judgement of the stimulus identity or location and whether there have been changes to the stimuli). In their paper, Larsson and Smith suggest that in order to maximise the effects of RS, a highly attention demanding task should be considered whereby attention is directed to particular features of the stimulus.

Evidence suggests that for a paired-stimulus fMRI paradigm, attention to both stimulus presentations is crucial to evoke RS (Moore, Yi & Chun, 2013). In their study, Moore, Yi and Chun identified that attention to scenes in both the initial presentation and the second presentation was fundamental in producing a significant RS effect in the PPA. In the design of Chapter 3, the inverted scene could have been either the first or second scene of a trial. Therefore, participants should have been attending to both scene images. However, it could be suggested that the type of task did not encourage participants to attend directly to variations in the type of scene content and instead they were awaiting an inverted scene.

4.2.4. The current experiment

Here we re-examined whether RS is observed for each of these spatial and non-spatial conditions and how this differs within and across ROIs. We held the same predictions as Chapter 3, with the addition of examining a new attention focusing task.

Crucially, this experiment employed a new task to ensure that participants' attention was directly drawn to the changes between pairs of scenes. We predicted that absence of RS in Chapter 3 was driven by an absence of direct attention to the changes between the scenes within each pair. We anticipate that RS will be observed with a more attention-demanding task, where participants are encouraged to attend to the changes directly between the two scenes within each trial.

In addition to the time-course analysis, whole-brain analyses were run. Given that prior research has suggested cross talk across dorsal and ventral processing streams (Zachariou, Klatzky & Behrmann, 2014), it could be suggested that regions beyond the scene-selective structures encode spatial properties of objects in scenes. This exploratory analysis is examined with no prior hypotheses.

4.3. Methods

4.3.1. Participants

Seventeen participants (8 males, mean age: 27.35) recruited from Bangor University and the local community took part in the fMRI experiment. Two participants had to be removed due to motion artefact (movement of more than one voxel size 3mm). All participants had normal or corrected-to-normal vision and were native English speaking. Participants could fall within any category of handedness. Participants were reimbursed with £10 for their time.

4.3.2. MRI safety screening and Ethics

Participants were screened to ensure they were eligible to take part. All participants were knowledgeable of the MRI procedures, and were fully aware of all experimental procedures prior to taking part. Each participant was sent all the MRI safety information and experimental information prior to the day of the MRI scan. All participants provided consent to take part in the MRI experiment, and were informed that they were able to withdraw at any point throughout the experiment. Participants were given a full debrief post scan, and given the opportunity to ask any questions. The experiment was granted full ethical approval by Bangor University Psychology ethics panel.

4.3.3. Design and stimuli

Stimuli used in this experiment were identical to the first experiment (Chapter 3). However, the change in task resulted in one less condition. Therefore, the conditions included for this experiment were Location changes, Arrangement changes, object Identity changes, repetition, fixation and multiple changes (task condition). The task in this experiment is again included to keep participants attentive to the stimuli and is not modelled in the analysis as a condition of interest (Figure 4.1).

Figure 4.1. Examples of stimuli as used to represent changes within each paired trial.

4.3.4. Experimental task and procedure

The task was designed to direct participants' attention to changes between the stimuli in each paired trial. Participants were required to press one of the buttons on the button-box when they detected *multiple changes* within a scene (for example, scenes where more than one change occurred: both Location and Identity changes, or Location and Arrangement changes). This task allowed us to direct participants' attention to the changes between the scenes, while also ensuring that they remain attentive throughout the duration of the experiment. Within this design, either of the images within each trial could be inverted, however participants were instructed to only press at the end of the trial, after the second scene had disappeared.

Participants took part in 4 experimental runs, each lasting 4 minutes and 15 seconds. Each run of the experiment consisted of 83 trials. Each run included 5 fixation scans at the beginning and at the end of the run. Likewise, there was an additional fixation trial included to account for the carry-over counterbalance design. Each trial lasted 3 seconds and was made up of the first scene presentation for 500ms, an ISI lasting 500ms, followed by the second scene presentation lasting 500ms and a response period of 1500ms (Figure 4.2). The design timings for this experiment were changed from Chapter 3 to allow participants slightly longer to attend to the stimulus, in order to ensure they had time to detect the multiple changes. Each condition was presented an equal number of times within each run. The experiment was programmed using PsychToolbox (http://psychtoolbox.org) in MATLAB.

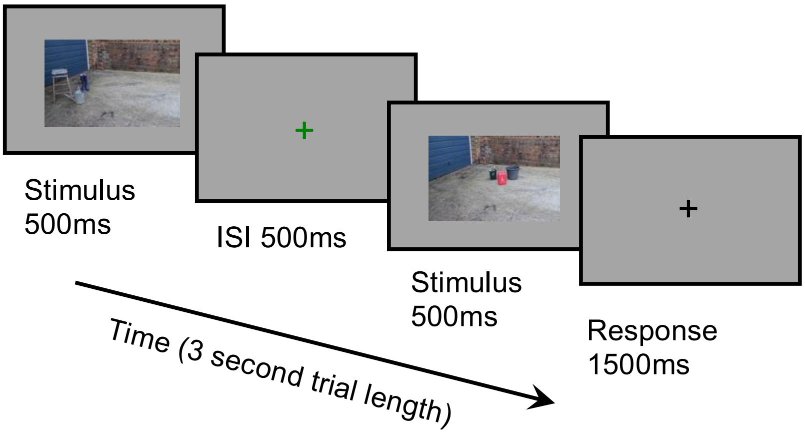


Figure 4.2. A schematic representation of a typical trial. This particular trial is an example of a target (task) trial, where participants are instructed to press a button in response to multiple changes. This figure shows both a Location and Identity change.

4.3.5. Data Acquisition

4.3.5.1. Anatomical scan

Scanning parameters for the T1 weighted anatomical scan were identical to that of Chapter 3.

4.3.5.2. Functional ROI localiser

The imaging parameters for the functional localiser runs were identical to the localisers run in Chapter 3.

4.3.5.3. Experimental runs

The data were collected on a Philips 3T MRI scanner at Bangor University using a SENSE-Head-32AH coil. A total of 31 slices were collected for a total of 125 volumes in each experimental run. Voxels were 3mm isotropic. All other parameters were identical to those in Chapter 3.

4.3.6. Data Analysis

4.3.6.1. Image preprocessing

Preprocessing was conducted in an identical method to data in Chapter 3. All preprocessing was conducted in SPM8 (in MATLAB).

4.3.6.2. ROI definition

The PPA, OPA and RSC were all defined in a procedure identical to the one outlined in Chapter 3, where two localiser runs were conducted to identify functional clusters at the individual subject level (native space). Due to difficulties in localising the left RSC, this region was excluded from the analysis. The right RSC was localised successfully at the individual subject level in all participants. Therefore, the right RSC was included but analysed separately. All regions were analysed with 6mm spheres and included significant voxels ($p<.001_{\text{UNCORRECTED}}$).

4.3.6.3. Time-course analysis

All data for this experiment were analysed using the same pipeline as Chapter 3 with the addition of changes in the number of conditions and the number of brain volumes collected.

4.3.6.4. Whole-Brain Analysis

The whole-brain analysis was performed in SPM8 where key contrasts were defined prior to the analysis (Arrangement > Location and vice versa). These conditions were of particular interest as it was considered that there may be regions beyond our hypothesised ROIs which encode information about the relative spatial relationships between objects and/or the wider scene context. Multiple comparisons across the whole brain were corrected using cluster-wise false discovery rate (FDR) (Genovese, Lazar & Nichols, 2002). FDR is a cluster-wise alternative to Family-wise Error (FWE) correction (based on Gaussian Random Field Theory) which reduces the instances of false positives associated with non-uniform Gaussian noise (Eklund, Nichols & Knutsson, 2015).

4.4. Results

Individual subject motion was assessed and participants who moved more than one voxel size (3mm) in any direction within the scanner were subsequently removed. Statistical analysis was performed on 15 participants (two participants were removed due to motion artefacts). Assessment of the behavioural task showed that on average participants were accurate >60% of the time within a window of 1500ms. No individuals were removed from the dataset.

4.4.1. Localiser analysis

Data collected from the two functional localiser runs were used to localise on the OPA, PPA and RSC (Figure *4.3*). Due to problems with localising the left RSC at both individual subject level and at the group level, data for the RSC were analysed separately throughout the course of this analysis.

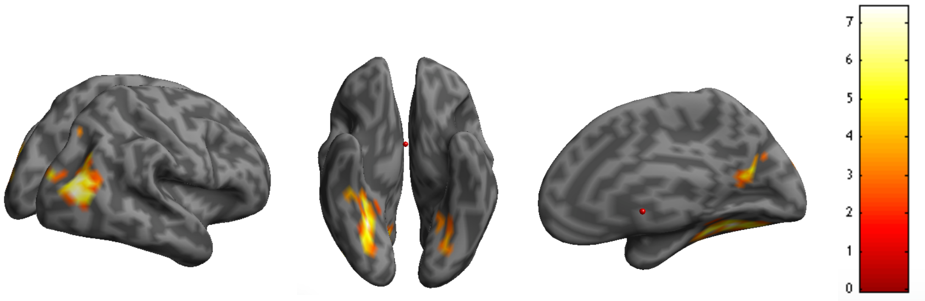


Figure 4.3. Example of the scene localiser showing all three regions of interest; PPA, OPA, RSC, for one participant. A threshold of p<.001$_{\text{uncorrected}}$ was used in order to localise these regions. The colour bar represents t-values.

4.4.2. Time-course analysis

The first part of the analysis was to determine whether there were any hemispheric differences [2 Hemisphere x 2 ROI (OPA/PPA) x 3 Condition (Location; Arrangement; Identity)]. The RM-ANOVA revealed no differences between hemispheres [*F*(1,14)=1.61, *p*=.226] and no interactions between hemisphere and other factors (*p*>.05), therefore data were collapsed across hemispheres for the remainder of the analysis.

A [2 (ROI: OPA; PPA) x 3 (Condition: Location, Arrangement, Identity) RM-ANOVA was run on the three peak HRF time-points (7, 8 & 9 seconds) to determine whether there was any significant release from adaptation for any of the conditions in any of the functionally localised regions of interest (Figure 4.4). The data analysed was the residual mean difference of each condition from repetition. Greenhouse-Geisser correction was reported where data violated sphericity.

Results showed that there was no main effect of ROI [*F*(1,14)=.54, *p*=.474, η2p = .037] or condition [*F*(1.34, 18.76)=.1.47, *p*=.249, η2p = .095] and no significant interaction between ROI and Condition [*F*(1.37, 19.23)=.47, *p*=.559, η2p = .033]. Results of the RM-ANOVA for the right RSC (3 levels of condition) showed no significant effect between any of the conditions [*F*(2,28)=.89, *p*=.424, η2p = .060] (Figure 4.5).

4.4.3. Effect of RS

In addition to examining object-in-scene representations, the current experiment examined whether a more attention-demanding task would evoke RS in scene-selective structures. This required averaging across non-repetition conditions and comparing them relative to repetition in a paired-samples *t*-test. A significant repetition suppression effect was observed for the RSC [*t*(14)= 2.55, *p*=.023] and OPA [*t*(14)= 2.94, *p*=.011]. Conversely, for the PPA, whilst the repetition condition appeared to show attenuation, this effect did not reach significance [*t*(14)=1.40, *p*=.187].

Figure 4.4. HRF plots collapsed across hemispheres for the PPA, OPA and RSC. RS analysis was conducted on the peak three time-points as highlighted in grey.

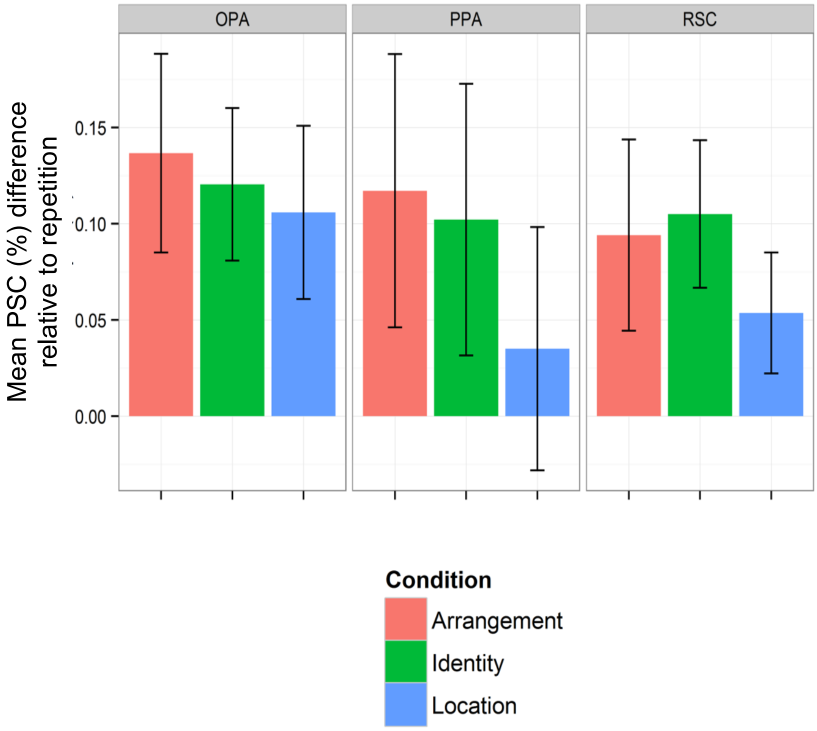


Figure 4.5. Bar graphs showing the mean PSC difference for each experimental condition relative to the repetition condition. Data for each bar is the average PSC extracted from the peak three time points from the time-courses. Bars are representative of standard error (*S.E).*

4.4.4. Exploratory whole-brain analysis

Whole-brain analysis was run to explore whether any regions were sensitive differences in the relationship between objects within scenes. Here we examined the contrasts [Arrangement > Location] and [Location > Arrangement].

Data were warped into normalised MNI space for group-level comparisons. This allowed us to examine if there are any brain regions that show sensitivities to the arrangements of the objects more than the relative location of the objects. Results showed that for the contrast Arrangement > Location at $p<.001_{\text{uncorrected}}$ there were bilateral clusters in the inferior parietal lobule (IPL) that survived FDR cluster correction at $p<.05$ (See Table 1) (Figure *4.6*). This result suggests that the inferior parietal lobe shows sensitivities to local spatial changes of objects in scenes significantly more than global location based changes. No significant clusters were observed for the contrast Location > Arrangement.

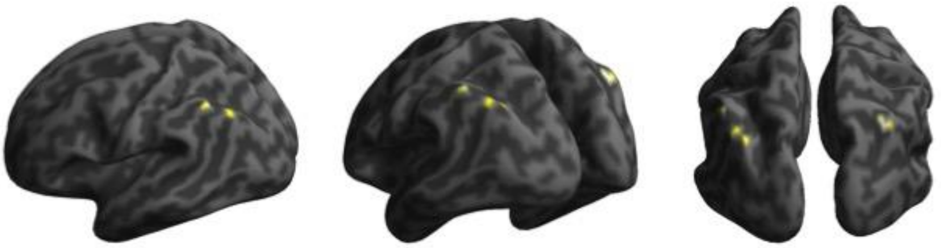


Figure 4.6. Group-level whole-brain analysis for the contrast Arrangement > Location. Results revealed bilateral clusters in the region of the inferior parietal lobe.

Table 1. *Statistics of whole-brain analysis for the contrast Arrangement > Location.*

	FDR-correction	K_E	T-statistic	MNI coordinates
Right IPL	*p* = .025	114	6.56	34, -68, 44
Left IPL	*p* = .025	135	6.29	-48, -62, 50

Note. Results identify bilateral clusters in the IPL in response to object Arrangement compared to Location changes. K_E refers to cluster extent, which is the number of voxels within each cluster.

4.5. Discussion

The present experiment examined differences in the representational content of scene-selective structures for spatial and non-spatial properties of objects in scenes. In addition, we re-examined possible task-related effects that may have explained the absence of RS observed in Chapter 3.

Based on results of Chapter 3, we considered that the inversion task may have distracted participants from attending to the changes between paired scenes. Therefore, in the current study, participants were required to press a button when they detected multiple changes within a scene (Location & Identity or Location & Arrangement). This task required participants to continually focus their attention to the changes between the paired stimuli as opposed to anticipating when an inverted stimulus appeared. We predicted that by re-designing the task to ensure participants were focusing on scene related changes, RS would be observed.

4.5.1. Time-course results

Data included in the time-course analysis reflected the residual differences in PSC between the non-repetition conditions (Location; Arrangement; Identity) and the repetition condition. Results showed that there were no differences in the degree of RS between conditions and this did not differ across ROIs.

Upon visual exploration of the results (PSC of each condition relative to repetition) the OPA numerically showed a larger release in adaptation to Location changes compared to the other two ROI's. This result was not significant and therefore, we cannot further draw conclusions about the processing properties of the OPA. However, given that more recent evidence suggests that the OPA encodes information about scene layout and object locations relative to boundaries within scenes (Julian, Ryan, Hamilton & Epstein, 2016; Kamps, Julian, Kubilius, Kanwisher & Dilks, 2016), it may invite future research to examine how the OPA encodes spatial layout information of objects within natural scenes.

Recent evidence has identified that these regions, specifically the PPA, may encode information relevant for how an individual interacts with an object. Bainbridge and Oliva (2015) identified that the PPA encodes information regarding an objects interaction envelope. Whilst prior research has affiliated the PPA as a region involved in encoding object size (Cate et al., 2011; Troiani, Stigliani, Smith & Epstein, 2014), particularly large objects which can be used for spatial navigation (Janzen & van Turennout, 2004), the interaction envelope identifies that the PPA is sensitive not to the size of the objects per se, but the size of the surrounding spatial envelope that would be determined by an object. This finding suggests that future research examining the complex processing properties of these regions in encoding object-in-scene information is important.

4.5.2. Effect of RS

We predicted that by changing the task to one that focused attention directly to the changes between scenes RS effects would be observed. In order to assess whether any effects of RS occurred within each ROI, data for non-repetition conditions were collapsed (averaged) and directly compared to the data for the repetition condition. In line with our hypothesis, we observed RS in the OPA and RSC but no effect of RS in the PPA. These results support prior claims that suggest direct attention to the stimulus is important for obtaining RS (Henson, Shallice, Gorni-Tempini & Dolan, 2002; Eger, Hendon, Driver & Dolan, 2004; Moore, Yi & Chun, 2013). Together with Chapter 3, these results emphasise the importance of the type of experimental task for evoking BOLD signal during fMRI (Corbetta & Shulman, 2002), and more specifically RS.

Although the OPA and RSC showed releases in adaptation, the ANOVA revealed no dissociable differences among the degree of RS between conditions was observed. Does this indicate that the OPA and RSC encode a wealth of spatial and non-spatial properties of objects in natural scenes? This finding would be consistent with previous evidence implicating these regions in processing multiple spatial properties of objects in scenes (Troiani, Stigliani, Smith & Epstein, 2014). The result of the PPA is somewhat surprising in that it is traditionally the most heavily studied region in scene processing. However, our results may reflect that the PPA is not involved in encoding any of these types of object based information, and might instead process information about the structure of space including more geometric based properties.

In the current study, we observed no effect of RS in the PPA, that is, there was no release in adaptation for any of the conditions relative to repetition. Prior evidence has observed RS in the PPA and evidenced it as an important region for scene encoding (Epstein, Graham & Downing, 2003), thus, our findings for the PPA are inconsistent with those in the literature which have previously observed RS effects in this region. As RS was observed for two ROIs, the lack of RS in PPA is not likely to be task related. Moreover, the absence of RS may reflect a role of the PPA in encoding global spatial information of scenes (Epstein, Graham & Downing, 2003), rather than object based information (MacEvoy & Epstein, 2011).

While we are unable to further infer the precise functional representations held in these regions for object-in-scene information, we suggest that the type of task used is critical for obtaining RS effects (distinguishing differences between repetition and non-repetition conditions). Concluding the RS findings of both fMRI experiments in this thesis we are able to infer that the results of Chapter 3 were likely driven by a lack of direct attention to the changes between paired scenes. This was supported by the results of the current chapter where two of the scene-selective regions showed a significant RS effect (OPA and RSC). Thus, the present experiment provides critical insight for researchers considering using RS in a paired-stimulus design where the importance of direct attention to the stimuli should be acknowledged and implemented within the task design.

4.5.3. Whole-brain analysis

Whole-brain analyses were conducted to examine whether regions beyond our hypothesised ROIs encode information about object Arrangements compared to Locations, and vice versa. This was examined as prior evidence suggests that there is cross-talk between dorsal and ventral streams which *may* encode spatial properties of objects in scenes (Zachariou, Klatzky & Behrmann, 2014). Results highlighted bilateral clusters in the region of the inferior parietal lobe (IPL) where BOLD activation was more significant for changes in object Arrangements compared to object Locations within scenes. The opposing contrast (Location > Arrangement changes) revealed no significant clusters.

Unlike the core scene-selective structures, regions such as the IPL are considered to be involved in multiple complex cognitive processes and behaviours (Fedorenko, Duncan & Kanwisher, 2013). For example, the IPL is implicated in navigation (Maguire et al., 1998), local versus global information processing (Weissman & Woldorff, 2005), spatial orientation (Kravitz, Saleem, Baker & Mishkin, 2011) and allocentric driven information (Committeri et al., 2004; Zaehle et al., 2007).

Spatial reference frames have been applied to the study of both object and scene processing. Allocentric reference frames are thought to represent information concerning the location of objects and their relationship to one another, whereas, egocentric reference frames reflect the relationship between the observer and other objects (Filimon, 2015; Klatzky, 1998) (Figure 4.1). Prior evidence has suggested possible involvement of the parietal lobe in encoding allocentric based information (Committeri et al., 2004; Zaehle et al., 2007). Therefore, it could be suggested that this region may encode information about the Arrangement of objects (more allocentric object-centred) compared to the Location of objects (more egocentric viewer-centred) within scenes. However, due to the design of our study, our whole-brain analyses were built upon feature *changes*. Thus, these findings and interpretations must be considered with caution, as no *a priori* hypotheses were determined for this region.

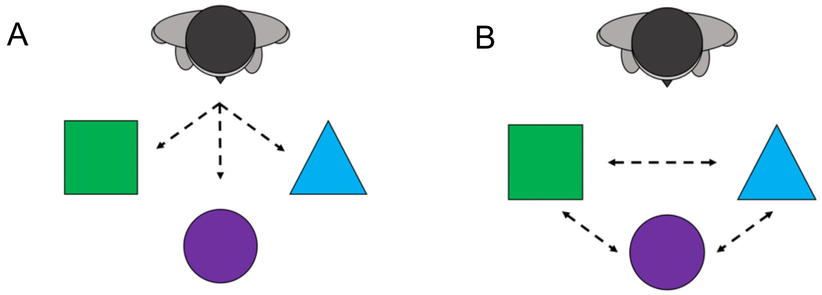


Figure 4.7. Example of two types of references frames. Image (A) is an example of an ego centric reference frame between the individual and the objects. Image (B) is an example of allocentric reference frame for the relationship between the objects independently of the user.

Prior evidence has also linked the IPL with object individuation (Bettencourt & Xu, 2013). Object individuation is the process by which an object is considered indexed or independent of other objects. If we consider the two conditions in whole brain analysis (Arrangement > Location), it could be suggested that Arrangement-related changes reflect stronger object individuation, compared to object Location changes. Arrangement changes visibly indicate that the objects are three individual units, located in the same area of the scene. However, for Location changes, the objects maintain their spatial relationship to one another, and only move in their location within the scene. Therefore, it could be suggested that for Location changes, the three objects might be encoded as one single unit, whereas for Arrangement changes, the objects violate that relationship by changing configuration and their relationship between one another.

Other evidence has linked the IPL with spatial processing, where the IPL plays a role in spatial attention with input from the ventral visual regions (He, Shulman, Snyder & Corbetta, 2007). Lesion studies examining patients with damage to the IPL have identified deficits in spatial attention and spatial working memory (Husain et al., 2001; Wojciulik, Husain, Clarke & Driver, 2001). Given that the IPL has been identified as a region, which plays a role in visual attention (Petersen & Posner, 2012; Posner & Petersen, 1990), results of the present study could reflect an increase in attentional demand toward identifying Arrangement changes compared to Location changes. If this were the case then this result would not reflect spatial processing object objects *per se*, but could be indicative of the increase in attention demand required to encode changes in the Arrangement of objects compared to their Location. Therefore, we interpret these results with caution, and suggest that future experiments examine these possibilities further.

4.5.4. Limitations and considerations for future research

The current study examined spatial and non-spatial object representations across the PPA, RSC and OPA. Whilst no differences in RS were observed between conditions across any of the ROI's, future research should further explore effects of attention and types of task on RS.

The present experiment re-assessed the type of task used in Chapter 3, however, the sample of participants used in the current experiment was different. This presents difficulties in directly comparing the results across the two chapters, as this type of comparison merits a repeated-measures design where each participant takes part in both types of task. Thus, while we are able to conclude that the change in task most likely resulted in changes to RS, future research should directly re-examine task effects on RS in the same sample of participants, using two different forms of task. This type of comparison would provide quantitative evidence of the importance of the task and attention in observing RS effects, specifically for a paired-stimulus design.

Future research should aim to explore these regions as part of wider brain networks. A psychophysiological interaction analysis (PPI) could be used to assay how these regions are functionally coupled during different task states. Within the present study, it was not feasible to use PPI due to the fMRI experimental design (the short trial times would have limited the power). The event-related experimental design in our experiments was used specifically for our questions about object-in-scene representations using RS, whereas PPI is better suited to block-design fMRI studies. Understanding the functional architecture of the scene-selective network would provide a more in-depth understanding by which object and scene information is encoded within the human brain.

Recent functional connectivity research has examined whether these scene-selective regions (PPA, OPA, RSC) are functionally coupled with other cortical structures during episodic memory for scenes and scene perception (Baldassano, Beck & Fei-Fei, 2013; Baldassano, Esteva, Fei-Fei & Beck, 2016). Baldassano et al show that the OPA and PPA form part of a smaller network, involved in processing scene features, whereas the RSC, IPL and PPA connections were driven by episodic memory representations relevant for navigation. Therefore, a more detailed framework is needed to determine how these regions are involved in both perceptual and memory processes relevant for navigation and how they are connected at a wider network level.

Contrary to Baldassano et al., (2016), recent research argues that the OPA is a critical structure for spatial navigation. The OPA has been implicated in processing boundary related information from scenes (Julian, Ryan, Hamilton and Epstein, 2016), as well as information about the egocentric distance of a scene (Persichetti & Dilks, 2016); critically linking it to navigation based processes. Therefore, future research should also consider a navigation-based task that would allow researchers to infer more about the representations held by the OPA, and whether it can be considered a node in a wider scene-based navigation network.

Whilst fMRI allows us to infer differences in neural responses in these regions in relation to spatial and non-spatial properties of objects, other non-neuroimaging methods have contributed significantly to our understanding of object in scene representations. One area yet to be studied in further detail is how variations in scene content influence encoding and memory for space. For example, emotional content is thought to influence how the surrounding scene-space is remembered (Mathews & Mackintosh, 2004). Likewise, salient stimuli such as bodies and faces are also though to affect how participants fixate and attend to natural scenes (Bindemann, Scheepers, Ferguson & Burton, 2010; Mayer, Vuong & Thornton, 2015; Peelen & Kastner, 2014), yet little work has directly examined how memory for space changes as a function of the type of scene content. Therefore, future research should examine not only the contribution of inanimate objects to scene encoding but also animate (people). This would provide an understanding of social influences on spatial cognition.

4.5.5. Concluding remarks

This chapter highlights two important points. First, when measuring repetition suppression effects in paired-stimulus designs, consideration for a task that directly draws the participant’s attention to changes between the paired stimuli is of fundamental importance in evoking significant attenuation effects. This provides insight into the mechanisms that support RS in paired-stimulus designs, which are not as commonly used in the repetition suppression literature. Second, we are able to infer that regions beyond the scene-selective structure are involved in processing information regarding object Arrangement compared to object Locations. However, this needs to be further examined with consideration for the experimental design. In conclusion, while this body of work highlights important methodological considerations for RS, future research should develop these themes examining how variations in scene content influence spatial memory for the scene.

Chapter 5

Boundary extension is distorted by the presence of people in natural scenes

Cardiff University, School of Psychology, Wales, UK.
Bangor University, School of Psychology, Wales, UK.

5.1. Abstract

Boundary extension (BE) describes the phenomenon by which an individual falsely remembers a larger expanse of space around a scene than was originally presented. Whilst previous research has examined how variations in scene content modulate BE, few papers have used BE as a tool to directly examine whether highly salient entities, such as people in scenes, would affect how the BE error is observed. Thus, the current experiment follows two lines of investigation. First, does memory for the extent of space alter as a result of the type of entity? Second, does this depend on the degree of saliency evoked by the presence of a person within a scene? In this experiment, participants (*n*=24) viewed pairs of scenes in either close-wide or wide-close trial orders, and were required to make a spatial judgment regarding whether the second scene presentation appeared the 'same', 'closer up' or 'farther away' than the initial presentation. Scenes either contained an individual object, multiple objects, an individual person, multiple people, fixed people, dynamic people or empty landscape scenes. As expected, the BE effect was observed for objects within scenes, however, the reverse asymmetry was observed for people in scenes. These BE effects held across scenes with either individual or multiple entities. In addition, a comparison of BE between dynamic people and fixed people in scenes, revealed that these effects may reflect high levels of saliency attributed to people, for predicting their actions and drawing on social information. Together, these findings suggest that the type of scene content influences memory for the extent of a space, which may reflect an increase in saliency associated with people compared to objects in scenes.

5.2 Introduction

Organisms effortlessly encode multiple pieces of information about the global spatial properties of a scene and its local components (Oliva & Torralba, 2001). This requires recognition of multiple types of visual cues, including spatial cues, landmark cues, object cues and people. Prior research has predominantly studied how inanimate objects influence our memory for the spatial extent of a scene (Intraub & Richardson, 1989). Yet, our understanding regarding how the presence of a person (animate, social and dynamic) influences memory for the spatial extent of a scene is limited. One method that has been used to examine how space is encoded as a function of the scene content is boundary extension (BE) (Bertamini, Jones, Spooner & Hecht, 2005).

5.2.1 Boundary Extension (BE)

BE is a memory phenomenon where participants falsely report a scene as having a larger spatial extent than was previously shown (Intraub & Bodamer, 1993; Intraub & Richardson, 1989). It is considered that BE reflects the process of scene reconstruction, by supporting the continuation of multiple memory representations through amodal completion of scene surfaces and objects (Intraub, 2010; Mullally, Intraub & Maguire, 2012).

BE has traditionally been examined via two distinct approaches. First, researchers have examined the mechanisms that drive the phenomenon (Bertamini, Jones, Spooner & Hecht, 2005; Hale, Brown, McDunn & Siddiqui, 2015; Intraub & Berkowits, 1996; Intraub & Bodamer, 1993; McDunn, Siddiqui & Brown, 2014; Munger, Owens & Conway, 2005). Second, it has been used it as a tool to determine how different scene properties influence memory for the extent of space (Bertamini et al., 2005; Gallagher, Balas, Matheny & Sinha, 2005).

5.2.2. Factors that affect the magnitude of the BE effect

Bertamini, Jones, Spooner and Hecht (2005) studied the influence of different perceptual properties on BE, such as magnification, field of view, object size and scene context. The authors suggest that BE is not modulated by field of view, or scene magnification, but is affected by object size; the bigger the object; the more BE occurred. The authors demonstrated that when observing minimalistic rooms with single focal objects participants report the scene as much harder to judge. This highlights the importance the role of objects and background scene information in evoking BE. This result is supported by the findings of Gottesman and Intraub (2003) who compared BE for both objects with and without a background. BE did not occur for images of objects without a context or background. Moreover, Gallagher, Balas, Matheny and Sinha (2005) revealed that BE was enhanced by the presence of a focal central object, which might reflect its reliance of object-context relations in order build a coherent expansive representation of space through amodal perception (Intraub, Gottesman & Bills, 1998). Does this effect depend on the type of object presented?

5.2.3. BE and scene content influences

Limited research has used BE as a tool to understand whether memory for spatial extent of a scene is dependent upon the *type* of entity. However, there have been some developments in understanding how variations in scene content may influence the magnitude and direction of the effect. Candel, Merckelbach and Zandbergen (2003) observed that BE occurs for both neutral and emotional images. Here, participants took part in a BE drawing task, where they observed an array of emotional/neutral scene photographs, and were required to re-draw the scenes from memory. Their results suggest that participants did show BE for both emotional and neutral scenes.

Conversely, further evidence suggests that memory for the spatial extent of a scene is reduced when a scene contains highly emotional stimuli; referred to as tunnel memory (Safer, Christianson, Autry & Osterlund, 1998). In line with this, Mathews and Mackintosh (2004) identified that individuals with anxiety-like traits show a reduction in BE for scenes that are considered highly emotional (both positive and negative). It was hypothesised that this occurs as a result of participants restricting their awareness/attention from the surrounding space and instead directly attending to the emotional scene content.

Taken together, these experiments indicate that BE is not only influenced by spatial properties of the scene *per se*, but is also influenced in some cases by the presence of highly salient elements. Thus, could BE also be influenced by other stimuli which draw attention such as people?

5.2.4. The current experiment

People in the environment provide different types of information compared to objects, important for social interaction (Adolphs, 2006; Mayer, Vuong & Thornton, 2015). This includes planning future actions, predicting others behaviours, (Fini, Costantini & Committeri, 2014; Simhi & Yovel, 2016), and examining others expressions through the face and body in order to extract information about their identity and their emotions (Birmingham, Bischof & Kingstone, 2008; Fletcher-Watson, Findlay, Leekam & Benson, 2008). Whereas objects can be thought of as fixed in nature, providing rich cues about our location and heading direction within the world (Epstein, 2005; Troiani, Stigliani, Smith & Epstein, 2014). As such, perceiving people in scenes draws upon different social attention mechanisms in comparison to objects (Bindemann, Scheepers, Ferguson & Burton, 2010; Mayer, Vuong & Thornton, 2015).

To date, little research has directly examined whether the type of entity (object or person) within a scene changes the how the spatial extent of a scene is reconstructed in memory and whether memory for the spatial extent of a scene is influenced by these highly salient, social and often occurring factors. Moreover, the majority of research to date has focused on scene processing in relation to inanimate objects, and very little attention has been paid to understand how scene encoding is altered as a function of animate objects such as the presence of people and whether any differences are explained by the number of objects within a scene.

Indeed, there has been a significant drive in the literature recently to understand how we can link core components of cognition to real world examples (Malcolm & Shomstein, 2015; Risko, Richardson & Kingstone, 2016). Thus, the present experiment used a BE paradigm to investigate the following four themes:

1. Is the BE error influenced by the type of scene content (person/object)? We predicted a high degree of BE for objects in scenes, consistent with previous literature (Intraub, Gottesman & Bills, 1998). Conversely, we predict a reduction in BE for scenes containing people, due to their highly salient and social nature compared to objects.
2. Is BE is influenced by the *load* or number of entities? Does this differ between people and objects? This question is motivated by prior research which observed that clutter does affect BE (Gallagher et al., 2005).
3. Is the BE effect reduced more for *dynamic* people (static images of

people in action) compared to *fixed* people (static images of people sat or laying down)? We predicted that if the need to predict others actions does change how the extent of space is remembered then there should be less of a BE effect for dynamic people compared to fixed people.

4. Does BE occur for scenes with no focal object? Prior research has found an increase in BE for scenes containing a focal object and no BE for objects without a background. Therefore, we ask whether BE occurs for landscape type scene with no key focal foreground object. If BE is an error observed as a result of amodal completion and object-context binding to build a coherent representation of space, then BE may only occur for scenes containing objects, as objects provide a grounding and reference of space relevant for navigation.

The present experiment used a BE recognition/rating paradigm where scenes were shown in either CW (close-up view followed by a wide-angle view of the same scene) or WC (wide-angle view followed by a close-up view of the same scene) orders. Here, BE is quantified by an increase in the proportion of same responses for CW trials (diagnostic of the BE error) compared to WC trials. Thus, the main focus was to directly examine the direction of the BE error, by examining the proportion of 'same' responses between CW and WC for each varying scene condition.

5.3. Methods

5.3.1. Participants

Twenty-four participants (8 males, *mean age*: 23.04 years) were recruited from both Cardiff University Psychology department and the external community to take part in the study. Sample sizes for the current experiment were based upon previous BE studies examining scene content variations with BE (Bertamini, Jones, Spooner & Hecht, 2005; Park, Intraub, Yi, Widders & Chun, 2007). All participants had normal or corrected-to-normal vision and could fall within any category of handedness. Full informed consent was provided. Participation was either voluntary, or if they were undergraduates, they were compensated with course credits. Both Bangor University Psychology ethics board, and Cardiff University Psychology ethics panel granted full ethical approval for this study.

5.3.2. Stimuli

The experiment was made up of 14 experimental conditions (individual person; individual object; multiple objects; multiple people; individual fixed person; individual dynamic person and landscapes) presented in both CW and WC orders (Figure 5.1). Stimuli in this experiment are trial unique where each stimulus was only shown in one trial order, either CW or WC, not both. To clarify for the reader, the individual full body person condition is a separate condition and independently analysed from the individual dynamic person condition. Each individual full body person is stood and does not appear to depict any form of action. Conversely, the fixed and dynamic person conditions were either depicting very obvious action or were sat/laying down. Stimuli consisting of people always included the individuals face and the full body. Similarly, object stimuli always contained a full view of each object.

Twenty-three natural full colour photographs (480 x 640 pixel size) were presented randomly in each of the 14 conditions. People and objects in scenes were controlled for size where the proportion of the scene taken up by each did not significantly differ ($p>.05$). Average degree of visual angle was calculated for all conditions where an object or person was present (Individual people, 3.08°; Individual objects 3.14°; Multiple people, 4.52°; multiple objects, 4.19°; fixed people, 4.08°; dynamic people, 3.81°). Average group size for people was 3.02 and for objects was 3.06. The photographs were from mixed sources including digital photographs by the experimenter and some images sampled from Google images. Each image was cropped by 15% on each edge of the photo and re-sized to the same number of pixels, using Photoshop CS3, resulting in a “wide” and a “close” version of each scene (Figure 5.2).

Photoshop was used to create visual distraction masks to limit further processing of scene images after they were removed from the screen. Masks were created in Photoshop using a mosaic filter to distort each component of the scene. The experiment was run on a 15” Apple Macintosh Air and was run using Microsoft PowerPoint.

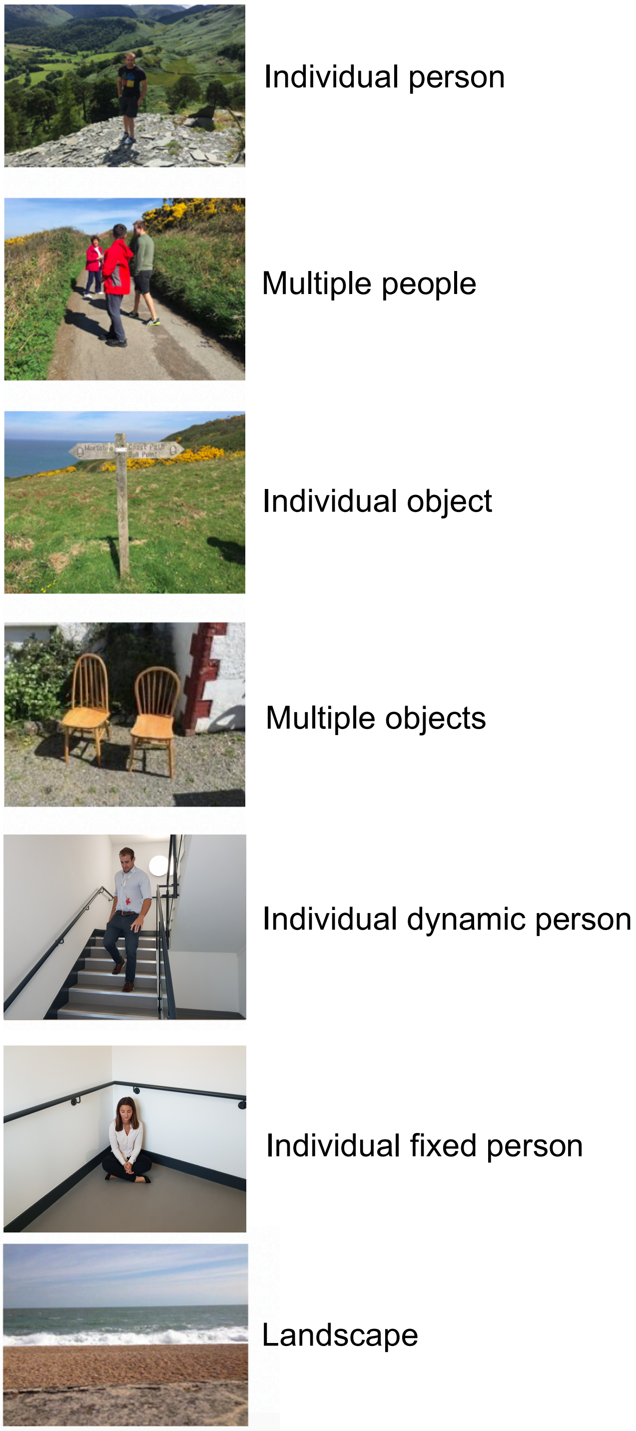


Figure 5.1. An example of the type of stimuli used for each condition. Stimuli containing multiple people and objects could range from 2 – 5 entities. Each condition was shown in Close-Wide and Wide-Close orders resulting in a total of 14 conditions.

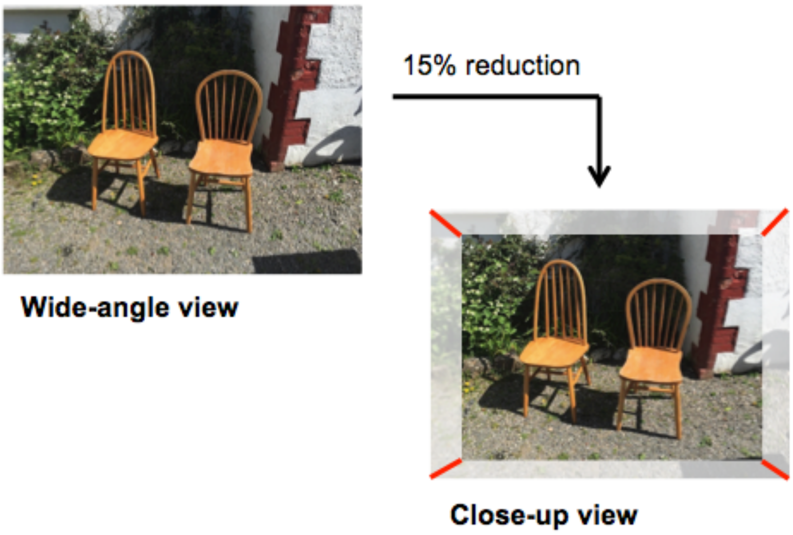


Figure 5.2. Example of wide and close versions of a multiple object condition. All close view images were cropped versions of the full-sized image and resized to match the number of pixels in the original wide-angle image. Red lines are for illustration purposes, not in experimental paradigm.

5.3.3. Task and procedure

The current experiment used a recognition/rating BE paradigm where participants were presented with two scenes sequentially, and verbally responded whether the second scene appeared to be the same, closer up or farther away from the vantage point of the camera than the initial scene. Each participant was exposed to a 4-trial practice run prior to starting the experiment. The experiment was split into three runs, where each participant took part in all runs, equating to a total of 322 trials. Each trial comprised of two scenes (500ms) separated by a visual distractor mask (750ms) and a fixation cross (1000ms). This was then followed by a response screen, which was displayed until participants had responded and were ready to continue to the next trial (Figure 5.3). Participants were required to give a verbal response. The timings of the mask and the fixation were piloted, where it was decided that increasing them (compared to the other typically used timings in the literature) would reduce any potential motion effects between the scenes, especially important for this experiment examining dynamic people in scenes. The experiment was conducted in experimental conditions in small cognitive testing rooms at Cardiff University.

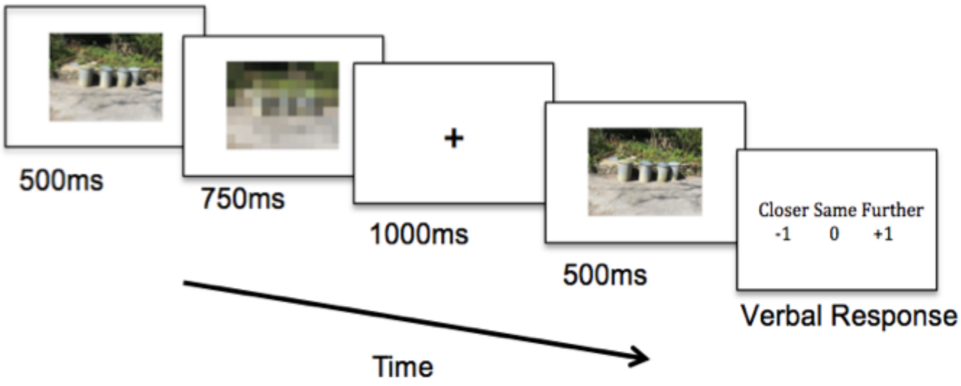


Figure 5.3. A schematic representation of a trial in Experiment 1. Participants made a verbal response to assess whether the second scene was 'closer up', 'same', or 'farther away' than the first scene presentation.

5.3.4. Analysis of the data

5.3.4.1 Mean BE score analysis

Initial analysis included a mean BE score analysis, obtained by averaging across all responses for each condition (closer -1, same 0, farther +1). To standardise the WC responses, all WC means were multiplied by -1 to allow for direct comparison with CW means. In this analysis, paired-t-tests were conducted where BE was observed when the mean for CW was significantly different from WC (an asymmetry in responses), and numerically closer to 0, a standard analysis used in BE experiments (Czigler, Intraub & Stefanics, 2013; Intraub, Gottesman, Willey & Zuk, 1996). This analysis was only conducted on individual objects/people so that results could be equated with those in the literature, examining BE effects for objects within scenes. From here, the main proportion analysis was conducted to look at the direction of the error for people compared to objects.

5.3.4.2 Proportion analysis

The main analysis measured BE by examining the differences in the proportion of 'same' responses between both CW and WC trial types for each person/object condition. In this instance, BE was measured as a higher proportion of 'same' responses to CW compared to WC trials. This analysis provides a sensitive method of quantifying BE by examining the direction of the error itself. Effect sizes are reported as partial eta squared ($\eta^2 p$) for each ANOVA.

The experiment was broken down into three different analyses for each of the different questions. The first one assessed whether there were differences in the BE effect for people/objects in scenes and whether this was modulated by the load. A repeated-measures analysis of variance (RM-ANOVA) was run on the proportion of 'same' responses for each condition [Trial Type (CW/WC) x Entity (objects/people) x Load (individual/multiple)]. A significant three-way interaction was broken down to look at the effect of Load on BE for objects and people respectively.

The second analysis examined whether there was a difference in the direction of the BE error for scenes containing dynamic people compared to fixed people. To test this, a RM-ANOVA [2 trial type (CW/WC) x 2 (fixed/dynamic)] was conducted. Paired *t*-tests were subsequently run to explore the differences between CW and WC for dynamic and fixed people in scenes respectively.

Finally, we examined whether the presence of a focal object influenced memory for the spatial extent of a scene. This analysis was a paired *t*-test between trial type (CW/WC) for empty landscape scenes.

5.4. Results

5.4.1. Mean BE scores – People versus objects

Results from the mean BE scores showed that there was a significant effect of BE for individual objects in scenes ($p<.05$). This was characterised by a significant difference between CW and WC, where CW results were closer to 0, signifying that participants responded 'same' on average more for CW than they did for WC. Results showed that for individual people in scenes, the reverse effect was observed ($p<.05$). That is, a significant difference was observed between CW and WC, but in the opposite direction – scores were closer to 0 for WC compared to CW.

Table 2. *Descriptive statistics for mean BE scores for individual people and objects.*

		Close-wide	Wide-close
People	*M*	0.73	0.43
	S.E	0.03	0.03
Objects	*M*	0.46	0.69
	S.E	0.03	0.03

Note. WC scores were multiplied by -1 in order to allow for a comparative measure from 0 as conducted by Czigler, Intraub & Stefanics (2013). Standard errors (*S.E*) are reported below each mean.

5.4.2. Proportion analysis

5.4.2.1. People versus objects

The initial part of this analysis examined whether variations in the type of scene entity influenced the way in which the BE error occurred. In addition, we ask whether this is modulated by the number of entities within a scene (load). A 2x2x2 [Trial Type (CW/WC) x Entity (objects/people) x Load (individual/multiple)] RM-ANOVA revealed a significant main effect of Entity (people/objects) [$F(1,23)=8.20$, $p=.009$, $\eta^2p = .26$], as well as a significant main effect of Load (individual/multiple) [$F(1,23)=9.24$, $p=.006$, $\eta^2p = .29$]. There was no main effect of Trial Type (CW/WC) [$F(1,24)=.14$, $p=.709$, $\eta^2p = .01$].

Results showed a significant interaction between Entity and Trial Type (CW/WC) [$F(1,23)=101.00$, $p<.001$, $\eta^2p = .82$] suggesting that the type of entity modulates the differences between same responses for CW and WC trials. However no significant interactions were observed between Trial Type and Load [$F(1,23)=.29$, $p=.595$, $\eta^2p = .01$], or the Load with Entity [$F(1,23)=.41$, $p=.524$, $\eta^2p = .02$]. There was a significant three way interaction between all three factors [$F(1,23)=22.46$, $p<.001$, $\eta^2p = .49$] (Figure 5.4). To further explore the significant three-way interaction, 2 x 2 RM-ANOVAs were run for each entity (n=2; object / person) independently.

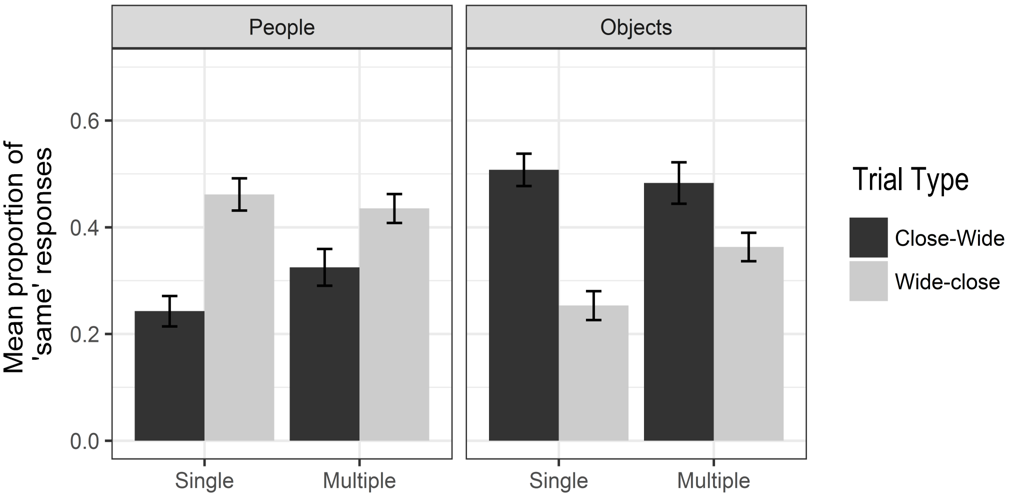


Figure 5.4. Proportion of 'same' responses for both multiple and individual body and object modalities for both CW and WC conditions. Error bars represent the mean standard error (*S.E*).

For people, there was no significant main effect of Load [*F*(1,24)=3.29, *p*=.083, $\eta^2 p$ = .13]. However, there was a significant main effect of Trial Type [*F*(1,24)=18.81, *p*<.001 $\eta^2 p$ = .45], as well as a significant two-way interaction between the Load and Trial Type [*F*(1,24)=8.25, *p*=.009, $\eta^2 p$ = .26]. Post-hoc t-tests showed that there was a significant difference between CW and WC for both individual [*t*(23) = -5.97, *p*<.001] and multiple people [*t*(23) = -2.33, *p*=.029] in the reverse BE direction (higher proportion of same responses for WC compared to CW scenes).

For objects, we observed a significant main effect of Load [*F*(1,24)=6.23, *p*=.020, $\eta^2 p$ = .21] and Trial Type [*F*(1,24)=37.01, *p*<.001, $\eta^2 p$ = .62]. There was also a significant two-way interaction between Load and Trial Type [*F*(1,24)=17.09, *p*=<.001, $\eta^2 p$ = .43]. Paired sample *t*-tests examining the BE effect for objects showed significant differences between CW and WC for both individual objects [*t*(23) = 8.49, *p*<.001] and multiple objects [*t*(23) = 3.07, *p*=.005]. Both these contrasts were representative of BE; a higher proportion of 'same' results for CW compared to WC.

5.4.2.2. Fixed versus dynamic people in scenes

The third part of this experiment explored the effect of fixed versus dynamic body positions of people within scenes. A RM-ANOVA [Trial Type (CW/WC) x Condition (Fixed/Dynamic)] revealed a significant main effect of condition (fixed/dynamic) [*F*(1,23)=21.74, *p*<.001, $\eta^2 p$ = .49], but no significant main effect of Trial Type [*F*(1,23)=2.04, *p*=.166, $\eta^2 p$ =.08]. A significant interaction was observed between Trial Type and Condition (fixed/dynamic) [*F*(1,23)=36.30, *p*<.001, $\eta^2 p$ =.61]. Paired sample *t*-tests revealed a significant difference between Trial Type for dynamic people in scenes [*t*(23)= -4.38, *p*<.001], but no difference for fixed people in scenes [*t*(23)=1.45, *p*=.160] (Figure 5.5).

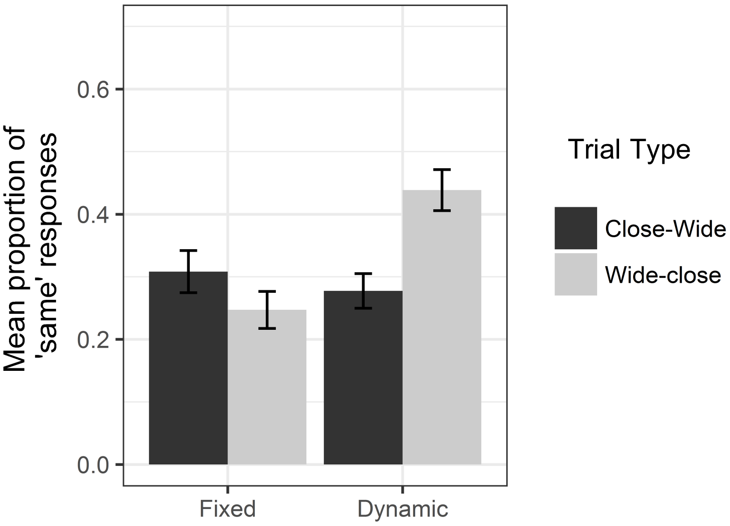


Figure 5.5. Bar plot depicting the proportion of 'same' responses to both fixed and dynamic people in scenes across CW and WC conditions.

5.4.2.3. Empty landscape scenes analysis

Finally, a paired-samples t-test between the trial types (CW/WC) for empty landscape scenes showed no significant differences [*t*(23) = -1.84, *p*= .079].

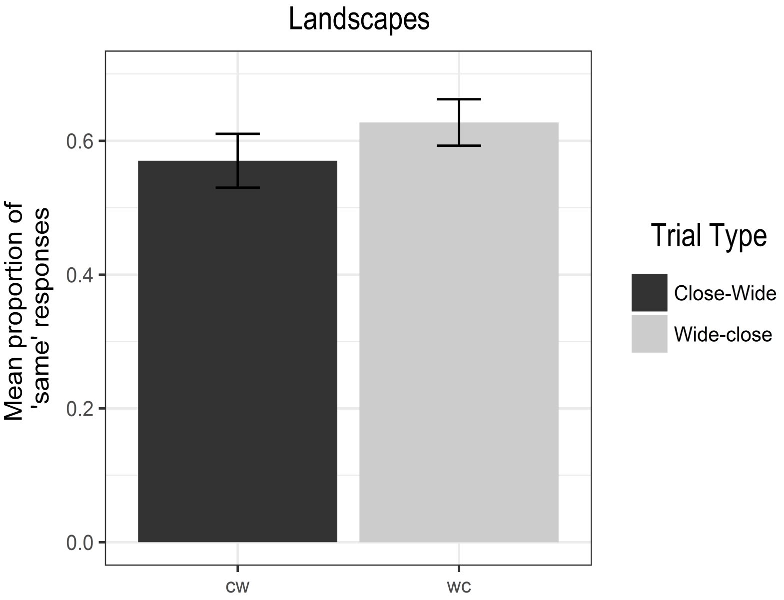


Figure 5.6. Mean proportion of same responses for CW and WC for empty landscape scenes. Error bars represent standard error (*S.E*).

5.5. Discussion

The current experiment examined whether memory for the spatial extent of a scene is modulated by the type of scene content (person or object). Whilst previous evidence observed BE for objects (Intraub, Bender & Mangels, 1992; Intraub & Richardson, 1989), limited research has directly examined BE in response to people in scenes. Across 3 independent analyses we explore 4 themes discussed below.

5.5.1 BE for people versus objects in scenes

Results for the mean BE score analysis showed that there was a significant effect of BE for individual objects within scenes consistent with previous literature (Intraub, Bender & Mangels, 1992b; Intraub, Gottesman, Willey & Zuk, 1996). This was characterised by a significant difference between Trial Type (CW/WC) where the mean score for CW was closer to 0 (Czigler, Intraub & Stefanics, 2013). However, this effect was significantly reversed for individual people in scenes. That is, the mean BE score was significantly closer to 0 for WC compared to CW. These results suggest that BE was observed for objects, but a reverse effect was observed for people, described as boundary contraction.

Results of the proportion analysis revealed a significant BE effect for objects both in individual and multiple conditions, reflected by an increase in the proportion of 'same' responses for objects shown in a CW trial order compared to WC (Intraub & Richardson, 1989; Intraub, Gottesman & Bills, 1998). For scenes that contain people, the reverse effect was observed across both individual and multiple person conditions. This effect was characterised by an increase in the proportion of 'same' responses for WC compared to CW trials. To the best of our knowledge, this effect has not been observed in previous BE research. It provides insight into the mechanisms that drive memory for the spatial extent of a scene, suggesting that it is strongly dependent upon the type of entity within a scene. What explanations are there for the boundary contraction effect observed for people but not objects?

First, these results may be explained by the increase in attention linked with people compared to objects (Downing, Bray, Rogers & Childs, 2004; Kaiser, Strnad, Seidl, Kastner & Peelen, 2014). If one would consider a typical WC BE trial for a scene containing a person, in order to extract socially relevant information participants may minimise their attentional scope to focus on that person (Gallagher, Balas, Matheny & Sinha, 2005). This results in a closer-up representation to be held in memory. When the participant is then presented with the second close-up scene from the trial, they compare this, to their stored close-up representation. The difference between their close-up memory of the scene and the actual close-up scene is minimal and thus, they respond 'same' more often than they respond 'closer'.

Conversely, during the CW condition, participants once again perceptually expand their memory for the person in the initial close-up scene. However, this time they directly compare this representation to the second image, which is the wide-angle view, resulting in a large degree of difference between the two representations. This results in participants saying 'farther' proportionally more than they say 'same'. This account would explain why participants respond 'same' more often for WC trials compared to CW. However, to further explore this reverse effect, replication and further study is necessary.

5.5.2. Load (number of entities)

The second question examined whether the number of entities, or *load,* influenced memory for the spatial extent of a scene. Results showed that load is not *independently* modulated by either the type of entity or trial type (CW/WC). However, a significant three-way interaction suggested that trial type (CW/WC) is modulated by the Load and the type of Entity but in opposite directions for people compared to objects. This is further confirmed by the breakdown of the ANOVA for people and object respectively. Identifying significant interaction between load and trial type for each entity. The results revealed that for people in scenes, Trial Type is influenced by the load but in the opposite direction to objects. That is, the difference between CW and WC for multiple entities is proportionally smaller for both people and objects (but still significant), but this is observed in the opposite direction (more same responses for WC for people and more same responses for CW for objects).

5.5.3. BE for dynamic versus fixed people in scenes

The third question examined the effects of dynamic and fixed individuals in scenes, and whether this would influence the magnitude/direction of BE. It was predicted that there would be less BE for dynamic people in scenes compared to fixed people as when they are sat down, there may be less of a need to predict their actions (Ramnani & Miall, 2004).

Results yielded a strong reverse BE effect (boundary contraction) for dynamic people in scenes, however, for 'fixed' people in scenes there was no difference in the proportion of same responses for CW and WC. These results suggest that dynamic people in scenes (portraying some form of action), compared to fixed people in scenes, change the way in which the expanse of space is encoded and remembered. Does this reflect increased saliency toward predicting others actions?

The threat-signal hypothesis as proposed by Cole, Balcetis and Dunning (2013) suggests that the more threatening a person looks, the closer up they are perceived. The authors suggest that this result supports the need to extract socially relevant information to predict others actions, which conforms with results observed in the current experiment. Indeed, viewing a wide-angle scene as the 'same' for dynamic people may suggest that participants contract their attentional scope to the person, and do not attend to the wider spatial extent of the scene. Conversely, for fixed people who do not appear dynamic, there is less need to predict their future action, and therefore, attention may be less directed towards the person.

Courtney and Hubbard (2004) studied boundary extension in relation to photographs of moving objects using the drawing method of BE. Here, the space was extended in the direction of motion. Whilst this research did not directly control for the direction of movement in people, this would be in important factor to consider in future research that examines people in scenes.

Similar theories and ties can be made with the literature on representational momentum which is defined as a false memory for the stopping point of moving stimuli or objects (Brouwer, Franz & Thornton, 2004). It could be suggested that during tasks containing images of people in action, individuals similarly misremember the spatial frame to which they previously saw the person. While some research has established differences in the mechanisms that drive BE and RM (Munger, Owens & Conway, 2005), our results show a contraction effect for dynamic people, *not* an extension in the direction of the persons facing position.

One question which future research should examine is whether BE occurs for objects which appear dynamic. Although the majority of these objects rely on human interaction (such as driving a car, motorcycle, bicycle and a football), there should be a concerted effort to establish how much these results relate to dynamic people specifically or just motion across all object categories. This links to research in the domain of representational momentum where memory for space is extended in the direction of a moving object (Brouwer, Franz & Thornton, 2004). If boundary contraction is only observed for moving people, then one would speculate that this reflects mechanisms of social attention in order to extract information relating to body language and action perception. Future research should therefore directly compare BE for dynamic objects and people to assess whether these findings are directly related to people and predicting others actions.

Further research will not only provide more information about how we process different entities in a scene context, but also how different social mechanisms change the way in which we process and remember spatial information.

5.5.4. BE for scenes with no focal object

The final manipulation examined whether BE was contingent upon the presence of a focal entity. Given that previous research suggests that objects without a scene context (background) show no BE effect (Intraub, Gottesman & Bills, 1998), we anticipated that scenes without a focal object would also show a reduction in BE. We found little to no evidence of a difference between CW and WC trials for images of landscape scenes with no focal entity. These results suggest that BE was not observed for scenes without objects.

This raises some interesting questions. If BE is a spatial memory error, why is a focal object critical in obtaining the effect? Perhaps BE measures the direct relationship between the object and its context? One could posit that this reflects a mechanism where objects provide a spatial reference to determine our location in the world. If there is no entity present within the scene, this may present difficulties in how individuals determine the shape and dimensions of the space, leading to inability to judge any form of spatial border information. This would need to be further examined in the instance of BE, but provides insight into the possible mechanisms that drive the BE error. This may also support the notion that object-context binding is crucial for amodal perception of space.

On closer inspection of the data, it is evident that participants on average responded 'same' more often than was observed for the other conditions. If our above discussion is correct, this may reflect an individual's difficulty in judging spatial structure without the guidance of an object. This may suggest that objects are critical as units for guiding spatial processing. However, this would need further testing in future research.

5.5.5. How do these results contribute to our knowledge of scene encoding?

In the present chapter, we establish that memory for the spatial extent of a scene is influenced by the type of entity. Specifically, we observe that scenes containing people show a reverse BE effect (boundary contraction) compared to scenes containing objects. This is characterised as a higher proportion of same responses for WC trials compared to CW. Similarly, if people appear dynamic, this reverse effect holds. However, for people who appear fixed, this reverse effect reduces (no difference between the proportion of same responses for CW and WC). To the best of our knowledge, these findings have not previously been documented in the literature.

Birmingham, Bischof and Kingstone (2009a) demonstrated that during eye tracking paradigms, participants are incredibly fast to fixate to areas of the scene that appear 'social' or where people are present. The authors suggest that this is not an effect of saliency, but is related to an individual's need to fixate and extract social information derived from people. Compared to earlier work in the field, more recent experiments have used stimuli where both the face and body are present. This is because prior research studying eye movement patterns across social scenes was limited to faces, where eye movements towards other people face were most prominent (Itti & Koch, 2000). However, evidence suggests that observers extract meaningful social information across both the face and body (Birmingham, Bischof and Kingstone (2009b; Perrett et al., 1992). This guide of eye movements and attention to social components of the scene is in some ways supported by our current results. While we do not study eye movements, we do see a difference in the way that people remember the scene in VSTM when a person is present compared to an object. Moreover, this is particularly strong for people who depict action compared to people who do not. Therefore, our findings from the current study appear to be in line with evidence suggesting that social attention mechanisms may alter how natural complex scenes are remembered.

On a broader level, these results support the notion that scene content variations determine the magnitude of BE (Bertamini, Jones, Spooner & Hecht, 2005) and that scenes containing stimuli considered highly salient can also influence the BE effect (Mathews & Mackintos, 2004). Future research should therefore examine these questions further while implementing a task to assess whether increases in social saliency factors drive differences in BE between people and objects in scenes.

5.5.6. Limitations and future research directions

An important consideration for the present study is whether factors such as field of view (FOV) are important contributing factors to the results. FOV typically refers to the degree of spatial scope that is observed from the camera. Whilst FOV was considered in the collection of stimuli, it could be suggested that general differences in the FOV associated with objects and people in scenes are different, whereby FOV is larger for people compared to objects. Bertamini, Jones, Spooner and Hecht (2005) found that FOV did not have an effect on BE, but their study did not directly compare scenes containing different entities. Future research should consider this by using stimuli whereby each person and object is shown within the same background respectively. This will allow us to determine whether FOV disparities between people and objects influences BE differences.

In addition, it is important to examine the mechanisms underlying the differences in BE between people and objects observed here. It could be suggested that exploring differences in saliency between people and objects would provide further explanation for our findings. Is it the body language of a person that changes the way in which the space is remembered? Is it specific to action? Or could it be that the reverse BE effect is driven by other factors that make people more salient stimuli compared to objects, such as extracting their identities and emotions? Evidence from social psychology literature suggests that we focus attention to both the face and the body of a person more than other parts of the scene (Itti & Koch, 2000). This increase in attention as measured with eye tracking technology suggests that people can be considered more salient within the scene. The mechanisms that underpin such an effect are thought to stem from the need to extract social information from people including their body language and facial expressions (Birmingham, Bischof and Kingstone (2009a). One method to study this would be to examine whether inverting the scenes reduces participants' ability to encode highly salient information from the scenes (Kelley, Chun & Chua, 2003; Shore & Klein, 2000). If this were the case, there should be a more traditional BE effect for inverted scenes containing people compared to upright scenes. In addition, identification of inverted objects should remain unchanged and should be similar for upright and inverted scenes containing objects. This would allow us to assay to what extent spatial encoding of scenes is influenced by mechanisms underlying social interaction.

5.5.7. Conclusions

Together, the findings demonstrate that the type of scene content influences an individual's memory for the spatial extent of a scene. We observed BE for objects within scenes, but a boundary contraction effect for people in scenes. While future research needs to be conducted to further determine why this difference in BE occurs for people, this finding has significant implications for the wider spatial processing literature which do not account for social entities in scenes.

Chapter 6

Investigating the mechanisms that underlie differences in boundary extension for people and objects in scenes

Cardiff University, School of Psychology, Wales, UK.
Bangor University, School of Psychology, Wales, UK.

6.1. Abstract

In the previous chapter, we provide evidence suggesting that memory for the extent of space is modulated by variations in scene content (object/person). To further explore the mechanisms that underpin these differences, the present chapter assesses BE in a similar paradigm, to examine whether these effects can be explained by possible differences in saliency attributed to people compared to objects. Here, participants observed scenes containing either people or objects that were presented either upright or inverted orientations. We predicted that by inverting scenes, participants would be unable to encode highly salient information derived from people, resulting in a boundary extension effect as opposed to boundary contraction. For objects, it was predicted that there would be no change in BE for upright/inverted scenes. Results revealed no difference in the proportion of same responses between CW and WC for upright people in scenes or inverted people in scenes. Conversely, BE was observed for objects in scenes in both upright and inverted orientations. Therefore, inverting the scenes did not significantly change how BE occurred for either people or objects. Moreover, our results reveal that there is still a clear difference in the way that humans reconstruct space from memory for scenes containing people compared to objects. We conclude that memory for the extent of space is influenced by variations in scene content and that based on findings from Chapter 5, this may reflect mechanisms of social action perception.

6.2. Introduction

When viewing the world, our visual system is overwhelmed with varying types of entities, including spatial elements, objects and people. While a plethora of research has directly examined how inanimate objects contribute to scene encoding and memory, limited research has examined how animate entities such as people influence spatial memory. Here, we extend the findings from Chapter 5 to examine whether previously observed differences in memory for space, are modulated by an increase in saliency attributed for people compared to objects, a difference that has already been documented in the literature (Downing, Bray, Rogers & Childs, 2004; Peelen & Kastner, 2014). Is this influenced by an imperative to extract socially relevant cues from people such as identity, emotion, and predicting others actions to facilitate social interaction (Adolphs, 2003b; Bigelow et al., 2006; Ramnani & Miall, 2004)?

6.2.1 Encoding social information from people

Previous research has revealed that people in scenes are encoded differently to objects. Mayer, Vuong & Thornton (2015) used eye-tracking technology and identified that people in scenes are generally the first entity to be fixated compared to objects. This supports previous research which showed that when presented with images of people, participants attend to others' facial features, such as their eyes (Henderson et al., 2005; Birmingham et al., 2008), primarily because they provide a source for detecting others identities and emotions (Baron-Cohen et al., 1997). In real-world situations, despite the increase in surrounding contextual features, individuals predominantly attend to the eyes of other people as a rich source to gather social information (Birmingham, Bischof & Kingstone, 2008; Cerf, Harel, Huth, Einhäuser & Koch, 2008; Emery, 2000; Smilek, Birmingham, Cameron, Bischof & Kingstone, 2006). This provides evidence to suggest that people salient stimuli, providing social cues to guide social interaction and action prediction (Downing, Bray, Rogers & Childs, 2004).

This is also observed in our findings from Chapter 5, examining BE for dynamic people and fixed people in scenes. Our results identified differences in BE despite faces being regularly shown in both conditions. Therefore, memory for the extent of space cannot be solely linked to processing facial features, but rather via social attention mechanisms linked to both facial expression and body language (Birmingham, Bischof & Kingstone, 2008; Nummenmaa & Calder, 2009). For example, the presence of a person within a scene may draw more attention in order for the observer to extract socially relevant cues, which may change the way attention is allocated towards the rest of the scene.

6.2.2. Effect of highly salient content on scene processing

While a plethora of research has investigated person perception, these studies typically include people isolated from a context (Downing, Bray, Rogers & Childs, 2004; Downing, Jiang, Shuman & Kanwisher, 2001), or explore person perception in visual search/target detection paradigms (Bindemann, Scheepers, Ferguson & Burton, 2010; Kaiser, Oosterhof & Peelen, 2016). However, given our observations in Chapter 5 it is critical to determine why spatial reconstruction differs as a function of the type of entity.

Various scene content manipulations can influence the BE effect and saliency may be a key factor. Evidence suggests that focal visual attention is fundamental for individuals to encode factors that are considered highly salient (Elazary & Itti, 2008). Mathews and Mackintosh (2004) identified that individuals classified as having high anxiety traits show reductions in BE for scenes containing salient emotional stimuli. The authors conclude that this may reflect a narrowing of focal attention to the salient feature within the scene. Here, we apply the same notion independently of emotion. We suggest that compared to objects, people appear more salient within scenes reflected by a narrowing of attentional scope toward the people in order to extract socially relevant information. This would reduce one's ability to encode spatial information beyond the entity within the scene (Mack & Rock, 1998). To test this hypothesis, one could restrict a participant's ability to encode highly social information within a scene.

6.2.3 Current research direction

Here we examine whether the boundary contraction effect revealed for people in scenes (Chapter 5) can be explained by differences in the degree of saliency between people and objects (Downing, Bray, Rogers & Childs, 2004) and whether this inverse BE effect reflects the necessity to extract social cues about a persons' identity, emotions and predict their future actions (Adolphs, 2003b;

Bigelow et al., 2006; Ramnani & Miall, 2004). This would be supported by our findings from Chapter 5, which showed a significant boundary contraction effect for dynamic people, compared to fixed people.

As such, we used an identical style BE paradigm as Chapter 5 whilst including an additional inversion condition to reduce saliency differences between people and objects (Kelley, Chun & Chua, 2003; Shore & Klein, 2000). Therefore, scenes were shown in both upright and inverted orientations as well as CW and WC. Based on prior evidence, we predicted that BE would be observed for both upright and inverted scenes (Beighley & Intraub, 2016). However, if the boundary contraction effect observed for people in scenes does reflect an increase in social attentional demand, a more 'traditional' object-like BE effect should be observed for inverted scenes containing people where saliency effects are thought to be reduced.

A further aim of the current experiment was to control for field of view (FOV). In Chapter 5, the FOV for images of people tended to appear larger compared to images of objects. We therefore controlled for this potential confound by presenting the same background scene with either a person or an object respectively.

6.3. Methods

6.3.1. Participants

Thirty-two (9 males, *mean age*: 21.72 years) Cardiff University undergraduate students in Psychology took part in this experiment for course credits. All participants had normal or corrected-to-normal vision and could fall within any handedness category. Full informed consent was provided prior to the experiment and participants were debriefed following the experiment. Full Ethical approval was provided by Cardiff University to conduct the experiment.

6.3.2. Stimuli

Stimuli consisted of full colour natural scenes containing people and objects independently in both upright and inverted conditions (Figure, 6.1). While Chapter 5 had trial unique scenes for each CW and WC condition, the current experiment used the same stimuli for both trial types (CW/WC). This allowed us to ensure that the effects in Chapter 5 were not driven by differences in the stimuli chosen for CW and WC trials. In addition to the original CW and WC trials this experiment included identical trials where scenes were mixed ‘close-close’ or ‘wide-wide’ scenes. These identical trials were not included in any further analyses but were included to ensure participants observed some authentic *same* trials.

In the present study, both people and objects were matched for size so that the proportion of the scene was taken up by on average the same number of pixels (p>.05). The average degree of visual angle for a person in a scene was 2.72° and for an object it was 2.84°. To control for FOV differences, each scene background was shown with either a person or an object, respectively. Scenes were a mix of the experimenter’s own images and others were sourced from various image search engines. Each person or object in the scene was situated in the same position but this varied between scenes.

Stimuli were split into two sets to ensure that participants did not see all versions of each stimulus in the same background (counterbalanced). For example, each participant took part in one of two versions of the experiment, each of which showed only upright *or* inverted versions of each scene for both people and objects. Participants therefore saw each scene background with an object or person but never in both upright and inverted. This was modelled to remove possible effects of memory for the space and reduced the number of repetitions of the scene background. These two versions were also run in the reverse order (starting with the last trial). Each version contained 216 trials. Masks comprised of coloured mosaic images generated in Adobe Photoshop CC, 2015, and were not derived from the images present in that particular trial. This was to ensure participants were not tracking the location of specific features between the two scenes.

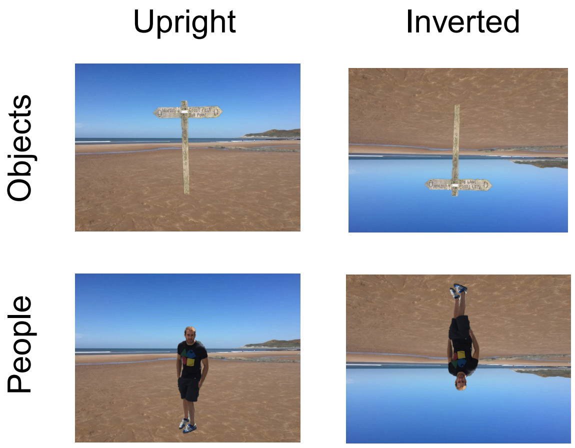


Figure 6.1. Example stimuli: Objects and people in both upright and inverted scenes were shown in sequential pairs, in which he second image relative to the first image could be CW, WC or the same.

6.3.3. *Task and Procedure*

All participants were given detailed instructions about the task prior to participation. Each trial lasted approximately 8 seconds including subject response time. As in Chapter 5, participants responded verbally. For each trial, an initial scene was presented for 500ms, followed by a distractor mask (1000ms), a fixation cross (2000ms) and the second scene (500ms) (Figure 6.2). To further ensure that participants were not focusing on specific features in the scene between presentations, the duration of the mask and the fixation were increased. As in Chapter 5, Participants were required to judge whether the second scene was 'same', 'closer up' or 'farther away' than the first scene of the trial. The experiment was conducted in psychological laboratories at Cardiff University and run on an Apple Macintosh Air Laptop.

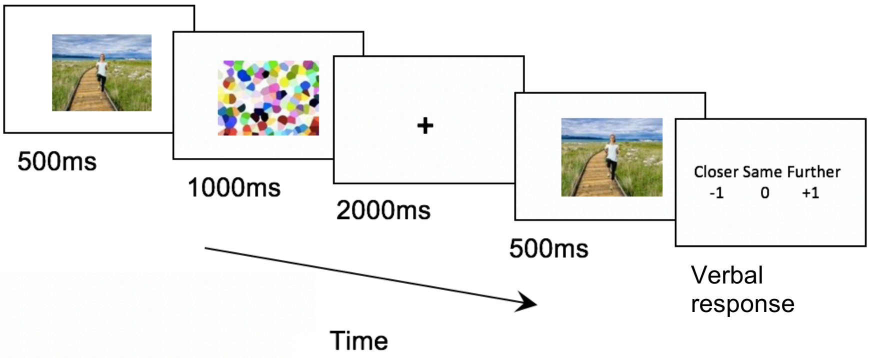


Figure 6.2. A schematic representation of a typical trial order in the second BE experiment.

6.3.4. *Analysis*

Analysis was conducted using the same procedures outlined in Chapter 5. First, BE was measured by calculating mean scores for people and objects for both CW and WC. These scores were then compared with paired *t*-tests for an asymmetry between CW and WC for each entity.

Second, a proportion analysis was conducted to examine the direction of the BE error between the type of entity for CW and WC and whether this was modulated by the orientation. Here, the proportion of 'same' responses for each individual for each condition was input into a RM ANOVA; exploring interaction effects between [Trial Type (CW/WC) x Entity (objects/people) x Orientation [upright/inverted)]. Post-hoc *t*-tests were run to explore the direction of any significant interactions in the ANOVA.

6.4 Results

6.4.1. *Mean BE scores*

Results showed a significant effect of BE for objects in scenes [*t*(31) = -4.00, *p*<.001], where the mean response for CW objects in scenes was significantly closer to 0 than for WC objects. This effect was not observed for people in scenes [*t*(31), -1.21, *p*=.236]. That is, for people in scenes, there was no difference between the mean BE score for CW and WC, suggesting no effect of BE or boundary contraction.

Table3. *Descriptive statistics for mean BE scores.*

		Close-wide	**Wide-close**
People	*M*	0.60	0.54
	S.E	0.04	0.04
Objects	*M*	0.39	0.57
	S.E	0.04	0.03

Note. WC scores are multiplied by -1 to give a positive number by which the CW scores can be directly compared.

6.4.2. Proportion analysis

A 2x2x2 RM-ANOVA [Trial Type (CW/WC) x Entity (object/person) x Orientation [upright/inverted)] revealed a significant main effect of the type of Entity [$F(1,31)=8.01$, $p=.008$, $\eta^2 p$ =.21], but no main effect of Orientation [$F(1,31)=.21$, $p=.654$, $\eta^2 p$ =.01] or Trial Type (CW/WC) [$F(1,31)=3.50$, $p=.071$, $\eta^2 p$ =.10]. Results yielded a significant interaction between Entity and Orientation [$F(1,31)=5.00$, $p=.033$, $\eta^2 p$ =.139]. There was no significant interaction between Orientation and Trial Type (CW/WC) [$F(1,31)=1.38$, $p=.246$, $\eta^2 p$ =.043].

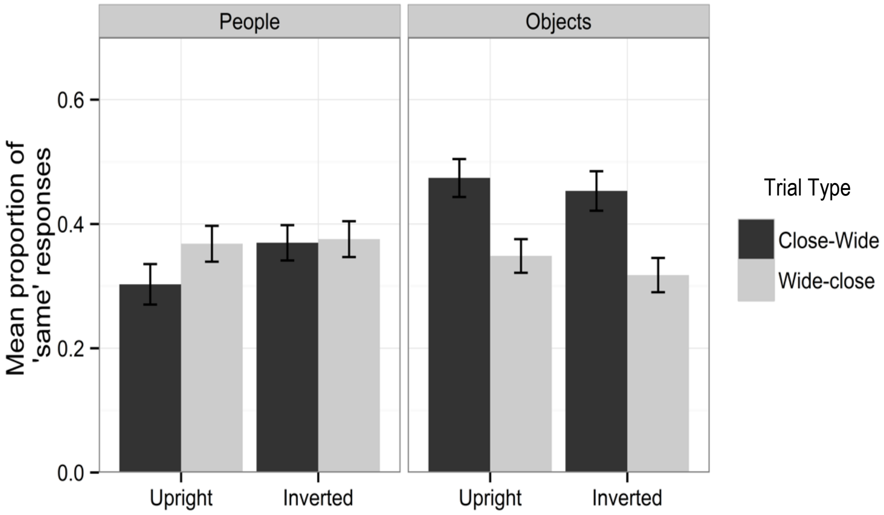


Figure 6.3. Results of Experiment 2 show the proportion of 'same' responses for both people and objects in upright and inverted orientations.

A significant interaction was observed between Entity and Trial Type (*F(*1,31)=16.88, *p*<.001, $\eta^2 p$ =.35), but no significant three-way interaction was observed (*F*(1,31)=.96, *p*=.334, $\eta^2 p$ =.03). Planned paired-sample *t*-tests show a significant difference between CW and WC for both upright objects (*t*(31)=3.37, *p*=.002) and inverted objects (*t*(31)=3.79, *p*=.001) but not for inverted people (*t*(31)= 0.15, *p*=.886), or upright people (*t*(31)=-1.73, *p*=.094) (Figure 6.3).

6.5. Discussion

Our motivation to examine BE for people and objects in scenes was to bridge the gap in understanding for the social-perceptual and attention mechanisms that influence how individuals process their local visual environment which has seldom been explored.

6.5.1. Discussion of findings

This study employed BE to re-examine whether memory for the expanse of space is modulated by the type of entity presented within a scene. It was largely considered that the absence of BE and the presence of boundary contraction for people in scenes (observed in Chapter 5) reflected differences in the degree of saliency between people and objects respectively. Thus, the current study presented participants with scenes in both upright and inverted orientations to examine whether inverting the scenes reduced saliency differences between people and objects (Kelley, Chun & Chua, 2003; Shore & Klein, 2000).

As such, we predicted that for upright people in scenes, there would be a reverse BE effect (boundary contraction), as was observed in Chapter 5. However, by inverting the scenes containing people, the extraction of saliency cues would be reduced, and this would result in increasing the degree of BE error (increase in the proportion of same responses for CW relative to WC). We anticipated that BE error rates for upright and inverted scenes containing objects BE would not differ as identification and recognition of objects is not considered to be affected by inversion (Reed, Stone, Bozova, & Tanaka, 2003).

6.5.1.1. BE for people versus objects in scenes.

Results for the mean BE scores revealed a BE effect for upright objects in scenes, but no effect for upright people in scenes. In addition, our findings from the proportion analysis revealed that memory for the spatial extent of a scene is modulated by the type of entity. This was revealed by a strong interaction effect between the type of entity and trial type (same responses for CW and WC).

In the current study, we observed a significant BE effect for upright scenes containing objects consistent with results from Chapter 5 and previous BE literature (Intraub & Richardson, 1989; Intraub, Bender & Mangels, 1992). BE was also observed for inverted scenes containing objects as predicted based on prior evidence documenting that inverting objects does not reduce the participant's ability to encode its identity (Reed, Stone, Bozova, & Tanaka, 2003). In addition, it has been documented in the BE literature that inverting scenes containing objects does not reduce BE (Beighley & Intraub, 2016). Our finding supports this result by demonstrating that BE still occurs for inverted scenes containing objects.

Results for upright people in scenes showed a tendency for boundary contraction, however this did not reach significance. That is, there was no significant difference between the proportion of "same" responses for CW and WC trials for inverted scenes containing people. Thus, when comparing both these findings from the present experiment and those from Chapter 5, we can conclude that BE is modulated by the type of entity but that the boundary contraction effect requires further exploration in future research.

Together, the findings for the current experiment suggest that the proportion of same responses between CW and WC is modulated by the type of entity within a scene but this interaction is not influenced by the orientation of the scene. This suggests that inverting the scene did not significantly change the way in which the spatial extent of a scene is remembered. From these findings, we are unable to conclude whether the absence of BE error for people in scenes is a direct result of the increase in saliency associated with people compared to objects. In addition, despite a stricter control of FOV, our results still showed that BE error rate was influenced by entity differences within a scene. Therefore, we suggest that these differences are not likely to have influenced results from experiment 1.

6.5.2. What do the results from Chapter 5 and Chapter 6 reveal about scene encoding?

Prior studies examining object-in-scene processing suggest that people are the first entities participants fixate to possibly due to our imperative to extract socially relevant information (Foulsham, Snyder & Carpenter, 1981; Tom Foulsham, Cheng, Tracy, Henrich & Kingstone, 2010; Foulsham & Underwood, 2009; Mayer, Vuong & Thornton, 2015) However, our knowledge regarding how the brain encodes people in scenes is still relatively elusive. Thus, it remains unclear why humans differentially reconstruct the spatial extent of a scene containing a person compared to an object. However, results from the current thesis suggest that these differences might reflect encoding social and highly salient cues about others actions, critical for interaction.

One explanation may be how we interact with each type of cue within the broader world. The interaction envelope as described by Bainbridge and Oliva (2015) suggests that humans form a spatial 'interaction envelope' to account for the space by which we interact with objects. If we apply this hypothesis to object perception, it would explain why participants report CW trials as being the same, but WC being different. This may be because memory for the space is extended to account for 'interaction space' with the objects. However, this would not explain why we see the reverse effect for people in scenes. Therefore, future research should be driven towards better understanding the mechanisms of the interaction envelope hypothesis in relation to BE.

In addition, it is important to explore the current findings further to examine how object recognition directly impacts upon the magnitude of the BE effect in natural scene processing. Hale, Brown and McDunn (2016) showed that increasing the task demand significantly increases the magnitude of the BE effect. Participants were required to make an object recognition judgement followed by a judgement of whether the second scene appears the same, closer or farther than the initial photo. When participants were considered worse at encoding the object identity, BE error increased. This demonstrates that poor initial encoding of the scene leads to a higher BE error. This finding provides insight into the mechanisms that drive BE, showing that it is fundamentally contingent upon attentional resources to an object.

This research may shed light on the results from the current study. Given that people in scenes draw attention it could be suggested that they are encoded with a higher resolution compared to objects. This may be reflected by differences in the way that attention is distributed across a scene. For example, when viewing scenes containing objects, participants may not attend to the scene in so much detail.

However, when there is a person present in the scene, participants contract their attention to focus on the person in order to extract socially relevant cues. As a result, the viewer may be more aware of the borders of the view and therefore do not exhibit a BE error. Whilst this would explain the results from the current study, this would not extend to the boundary contraction finding in Chapter 5.

6.5.3. Considerations for BE as a method of assessing spatial cognition

A crucial element for discussion is the BE approach itself. Whilst the BE method has provided significant insight into understanding the mechanisms by which the brain processes the spatial extent of scenes (Gottesman & Intraub, 2003; Intraub & Richardson, 1989; Park, Intraub, YI, Widders & Chun, 2007); consideration for the type of analyses used is critical. For example, the mean BE score is calculated as the mean score between responses (-1, 0, +1) for CW and the mean responses for WC. However, if a participant responds ‘same’ 2 times, ‘closer’ 4 times and ‘farther’ 4 times during 10 CW trials, this would give them a mean of 0. If they then responded ‘closer’ on 8 WC trials and ‘same’ on 2 trials, the comparison between CW and WC will be asymmetric in the direction of BE. This asymmetry is not indicative of BE error, as it is driven by a symmetry in ‘closer’ and ‘farther’ responses and not by the number of times they make the actual BE error (‘same’). The use of the proportion of ‘same’ responses provides a measure of the direction of the BE error. While results for both the mean BE score and proportion analysis were consistent, our method of analysis is better suited to our question as it provides a measure to detect differences in the direction of the error itself.

In addition, the style of mask used between each scene should be chosen with care. Evidence suggests that BE can be modulated by emotion (Mathews & Mackintosh, 2004; Menetrier, Didierjean & Vieillard, 2013). However, in various BE studies, the distraction mask used is a visual noise mask with central smiley face (Mullally, Intraub & Maguire, 2012; Kim, Dede, Hopkins & Squire, 2016). It could be suggested that this may influences the effect of BE, as it could impact on the way that the following scene is reconstructed. Therefore, future research needs to examine the effect of neutral masks on the magnitude of BE as was considered in for the experiments in the current thesis.

6.5.4. Limitations and future research

It must be considered that while Chapter 6 attempted to control for potential differences in FOV (by showing the person and object in the same scene background respectively); this may have resulted in participants learning the spatial extent of each scene background. Even when participants are instructed to fixate on the scene borders, they still exhibit BE (Gagnier, Dickinson & Intraub, 2013). It could be suggested that repeat exposures to a scene background may not result in participants learning the borders of the scene. However, this needs further investigation in order to guide researchers in future using BE as a method of assessing spatial cognition.

In the present study, we tried to control for differences in saliency attribution between people and objects in natural scenes. It was considered that inverting the scene would reduce participants' ability to extract salient information from people, therefore resulting in a higher proportion of 'same' responses to CW scenes compared to WC, as is observed for objects in scenes. In the present study, this was not the case. It could be suggested that this inversion method did not tap into the right mechanisms that drive the reverse BE for people. Future research should re-examine this observation with other saliency manipulations such as blurring different parts of the person, to investigate if the effects are driven by the need to extract facial expressions or body language to predict future actions.

Moreover, it would be prudent to examine how attention is dispersed across the scene, and whether this is modulated by saliency, specifically, differences between people and objects in scenes in the case of BE. Indeed, eye-tracking research has already identified that people are typically the first entities to be fixated upon in natural scenes (Bindemann, Scheepers, Ferguson & Burton, 2010), however this method should be applied to a BE paradigm to infer how attention to the boundaries of space differs for people in scenes compared to objects. Eye-tracking technology could be used to examine how saccades are distributed across scenes in the presence of people compared to objects. Tracking eye fixations across the scene when there is a person or an object present, would allow one to infer more about the way in which the scene is analysed at the initial scene presentation compared to the second scene presentation.

There were a number of changes to the design of the current experiment in comparison to the first BE study in this thesis. This included not using trial unique stimuli and including the same backgrounds for both people and objects in scenes. This was done to try and control for any stimulus effects that might have been observed in the previous chapter such as FOV (a limitation of trial unique design). While there is currently no evidence to suggest that the number of repetitions of scene background reduces the BE effect, future research should examine this in more depth.

6.5.5. Conclusions

Scenes are complex and require us to processes vast arrays of information and entity types simultaneously. To the best of our knowledge, there is no research that directly examines whether memory for the spatial extent of a scene is influenced by variations in the type of entity (person/object). Finally, this body of work suggests that scene encoding is not only driven by global spatial properties or inanimate objects relevant for navigation (Intraub & Richardson, 1989; Intraub, Bender & Mangels, 1992), but that it is also driven by social constructs and elements, which should be considered in the study of spatial cognition.

Chapter 7

General Discussion

In the present thesis, we examined how variations in scene content are represented in the scene-selective cortex and how these variations influence memory for the spatial extent of a scene. Successful encoding of these properties allows organisms to form a coherent representation of the wider visual world, critical for spatial navigation (Tolman, 1948; (Baumann, Chan & Mattingley, 2012; Chan, Baumann, Bellgrove & Mattingley, 2012); (Marchette, Vass, Ryan & Epstein, 2015).

Various lines of research have attempted to investigate how scenes are more broadly encoded and reconstructed in memory. For example, some propose that scenes are categorised by their global spatial properties such as naturalness/manmade (Greene & Oliva, 2009b; Oliva & Torralba, 2001, 2002; Torralba & Oliva, 2003; Torralba, Oliva, Castelhano & Henderson, 2006) and low-level visual elements of the scene (Rajimehr, Devaney, Bilenko, Young & Tootell, 2011; Watson, Hymers, Hartley & Andrews, 2016). Other evidence suggests that scenes are categorised as a function of object-context relationships (Aminoff, Schacter & Bar, 2008; Bar, 2004; Brockmole & Le-Hoa Vo, 2010; Davenport, 2007; Hollingworth, 2007; Hollingworth & Henderson, 1998; Kalenine, Shapiro, Flumini, Borghi & Buxbaum, 2014) or spatial properties of the objects relative to the user (Bainbridge & Oliva, 2015). Whilst the contributions from these studies have provided us with an understanding regarding how scenes are processed as a function of their local parts and global elements, our knowledge has been limited regarding how the brain encodes spatial and non-spatial properties of objects in scenes and how highly salient entities such as people influence memory for the extent of space.

Thus, the principle aim for the current thesis was to examine how variations in scene content influence scene encoding and memory reconstruction. To address this broader theme, we studied two lines of investigation.

1. How is spatial and non-spatial properties of objects represented in the core scene-selective network?
2. How does scene content (object/person) influence memory for the spatial extent of the scene?

7.1. Discussion of key findings

7.1.1. Chapter 3 - Scene-selective contributions to object-in-scene representations.

The first fMRI experiment employed RS to probe the representational content of core scene-selective structures (PPA; RSC; OPA) in response to changes in Locations, Arrangements and Identities of objects. Our findings showed no statistical differences in the magnitude of the BOLD HRF response between any of the conditions, including repetition. Thus, there was no effect of RS, within or across the three ROIs.

Given the plethora of research that has observed RS effects in these regions, using similar RS designs (Epstein, Graham & Downing, 2003), we considered that these effects might reflect irregularities in the modulation of attention to the task. During scanning, participants were required to press a button on a keypad upon detection of an inverted scene, to maintain their attention throughout the duration of the experiment. In this case, the inversion task may have distracted participants from attending to the relative changes between each pair of scenes, reducing the likelihood of observing RS.

The effect of attention for observing RS is still somewhat unclear, with some evidence suggesting attention is critical for observing RS (Eger, Henson, Driver & Dolan, 2004; Grill-Spector, Henson & Martin, 2006; Moore, Yi & Chun, 2013), and others suggesting that RS is not exclusively dependent upon attention mechanisms (Barron, Garvert & Behrens, 2016; Larsson & Smith, 2012; Larsson, Solomon & Kohn, 2016). Interestingly, it has been suggested that neural adaptation at the single cell level can be observed independently of attention (as suggested by Larsson, Solomon and Kohn (2016). Thus, attention mechanisms in RS during fMRI may be specific to the mechanisms that drive BOLD effects. Moreover, these considerations regarding the modulation of attention during RS was examined in Chapter 4, where participants were required to focus their attention directly to the changes between scenes (with a newly designed task).

7.1.2. Chapter 4 – Object-in-scene representations: Task effects in RS.

In Chapter 4 we adopted the same RS method and stimuli as Chapter 3, while employing a new task that directed participants' attention to the changes between scene pairs. In this experiment, participants were required to respond when they detected multiple changes between paired scenes (for example, both location and identity changes or location and configuration changes). We anticipated that this modified task would draw participants' attention more specifically to the changes between the scenes in each paired trial, and thus, increasing the likelihood of observing attenuation in the repetition condition.

Results showed that RS was observed for both the OPA and RSC. Conversely, for the PPA, there was no observed effect of RS. Results of the PPA are striking in that this region has been most critically linked with scene processing (Epstein & Kanwisher 1998), yet showed no release in adaptation for spatial manipulations within scenes as has been observed previously (Buffalo, Bellgowan & Martin, 2006; Kohler, Crane & Milner, 2002; Pihlajamaki et al., 2004). Prior evidence has suggested that the parahippocampal cortex receives input from dorsal visual regions implicated in visuospatial processing (Baldassano, Beck & Fei-Fei, 2013; Baldassano, Esteva, Fei-Fei & Beck, 2016), thus it is surprising that in this study, the PPA did not elicit a release in adaptation for changes in spatial Locations or Arrangements of objects within scenes. However, it could be suggested that the PPA is not involved in encoding these spatial properties within scenes and is more involved in processing object-context relationships (Aminoff & Tarr, 2015; Bar, 2004) or memory processes relating objects-in-scenes (Aminoff, Gronau & Bar, 2007; Pihlajamaki et al., 2005; Pihlajamaki et al., 2004).

These results provide evidence to suggest that the modified task was sufficient for evoking RS, thus, attention to the changes between scene pairs is critical, supporting evidence suggesting the importance of attention for observing RS (Moore, Yi & Chun, 2009). Whilst there were no significant differences between the experimental conditions, the observation that these two regions showed an RS effect might suggest their potential involvement in representing higher-level information of global scene layout (Oliva & Torralba, 2001) or multiple spatial and non-spatial proprieties of objects in natural scenes (Troiani, Stigliani, Smith & Epstein, 2014).

In addition to the RS time-course analysis, exploratory whole-brain contrasts were run between Arrangement changes (the relationship between the objects) and Location changes (the relationship between the observer and the objects). These contrasts were chosen to examine whether regions beyond our hypothesised ROIs were sensitive to different forms of spatial information about objects in scenes. Results revealed bilateral clusters in the inferior parietal lobe (IPL) for the contrast Arrangement > Location. Speculating this result lead us to consider that the IPL may be more involved in processing allocentric based spatial information about object Arrangements, compared to the egocentric object Locations within scenes. Our whole-brain findings support previous research suggesting that the IPL plays a role in allocentric spatial encoding (Committeri et al., 2004; Zaehle et al., 2007) and object individuation (the degree to which an object appears to be individual and indexed compared to its surroundings) (Bettencourt & Xu, 2013). However, these whole-brain findings were conducted without priory hypotheses and thus, should be interpreted with caution.

Collectively, these fMRI experiments provide insight into the role of attention and task influences in RS experiments (Larsson & Smith, 2012; Larsson, Solomon & Kohn, 2016; Moore, Yi & Chun, 2013; Tootell et al., 1998; Tootell, Hadjikhani, Mendola, Marrett & Dale, 1998). While this will need to be further examined within the same sample of participants, it highlights the importance of the task in observing RS in paired-stimulus designs to examine variations in visual content.

Having initially studied this overarching theme of object-in-scene representations with fMRI RS, we decided to approach the question with a separate complimentary method. One such method that has been investigated with relation to RS is boundary extension (BE). The mechanisms that underpin BE are thought to be similar to those of habituation observed during RS (Park, Intraub, Yi, Widders & Chun, 2007). Therefore, we used BE as a behavioural tool to examine how changes to scene content alter the way in which the spatial extent of the scene is reconstructed in memory. In the second set of experiments we tested whether different types of objects (people vs objects) alter the BE effect for scene construction.

7.1.3. Behavioural Experiments

BE is described as a phenomenon that is characterised by a false inflation in remembered space, beyond the borders of a scene that was previously observed (Intraub & Richardson, 1987). Thus, BE serves as a useful tool in measuring the way in which space around the borders of a scene is remembered.

Prior BE experiments have typically examined scenes containing inanimate entities such as objects (Bertamini, Jones, Spooner & Hecht, 2005; Czigler, Intraub & Stefanics, 2013; Dickinson & LaCombe, 2014; Intraub, Gottesman, Willey & Zuk, 1996). Objects can be thought of as units, which provide cues to infer us of our location and facing direction within our environment (Chan, Baumann, Bellgrove & Mattingley, 2012; Marchette, Vass, Ryan & Epstein, 2014, 2015). However, limited research has directly compared how we encode and remember the extent of space within a scene, in the presence of an object compared to a more salient social stimulus such as a person. For example, these elements are regularly encountered in the world, yet they typically serve two different functions. People are animate entities that we socially interact with, whereas objects are inanimate units that guide our behaviour and serve a function to inform us of our location within the world. Prior to the present thesis, our knowledge of the way in which these two entities differentially influence memory for the extent of space was limited.

In two behavioural experiments, we used BE as a measure to infer whether memory for the extent of space was modulated by the type of scene content (a person or an object).

7.1.4. Chapter 5 – Does the presence of a person influence BE?

The purpose of Chapter 5 was to examine whether memory for the space around the scene boundaries is different in the presence of a social animate entity, such as a person, in comparison to an inanimate object. Given that people may draw more attention in comparison to objects (Downing, Bray, Rogers & Childs, 2004; Fletcher-Watson, Findlay, Leekam & Benson, 2008; Peelen & Kastner, 2014; Seidl, Peelen & Kastner, 2012), and emotional scenes are found to reduce the BE effect (Mathews & Mackintosh, 2004), we predicted a reduction in BE would be observed for people in scenes compared to objects.

Results for both the mean BE score and proportion analyses showed a significant effect of BE for individual objects in scenes, consistent with previous literature (Intraub, Bender & Mangels, 1992; Intraub, Gottesman, Willey & Zuk, 1996). However, this effect was significantly reversed for individual people in scenes. Therefore, our results show a significant BE effect for objects but a boundary contraction effect for people. In addition, the effect of load revealed that when scenes include multiple entities, there is a smaller difference in proportion of same responses between CW and WC for both people and objects in the opposing direction. That is, multiple entities reduce the BE effect for objects in scenes, and reduce the boundary contraction effect for people in scenes, although these differences were still significant. Upon reflection, these differences in spatial memory for people and objects in scenes may reflect higher levels of saliency associated with people compared to objects, driven by social characteristics and a need to predict others actions (Fletcher-Watson, Findlay, Leekham & Benson, 2008; Peelen & Kastner, 2014).

Our second analysis examined whether these differences reflected the dynamic nature of people, driven by an innate requirement to predict their future actions (Ramnani & Miall, 2004). Here, we studied BE in for two further independent conditions, fixed people in scenes and dynamic people in scenes. Results showed that for the dynamic people in scenes, there was a strong boundary contraction effect, whereas for ‘fixed’ people in scenes, there was no significant difference between the proportion of ‘same’ responses for CW and WC scenes. These results provide further evidence to suggest that an increase in saliency (as depicted by dynamic people) changes the way in which the surrounding space along the boundaries of a scene is processed and remembered.

These findings are of particular interest because it highlights that the way in which humans observe the world depends on the type of visual input. In the case of these results, the dynamic nature of people might have required a more contracted focus of attention toward that person to extract relevant social cues, in order to predict their actions. This may have subsequently reduced the scope of their attention to the wider space. However, for people who appeared fixed in scenes, there was less of a need to predict their actions as they were fixed to their location, therefore participants may have distributed their attention toward many aspects of the scene, thus, resulting in no BE effect (no difference between the proportion of ‘same’ responses for CW and WC).

The third analysis examined whether BE was observed for empty scenes (landscape photographs with no focal object). Our observations revealed little to no evidence of a difference between CW and WC landscape scenes with no focal entity. Our findings suggest that scenes without objects do not yield BE, thus, BE may rely upon the integration of both object and context information to form a global understanding of the space, but as the social saliency of the entity increases, the BE effect reduces.

7.1.5. Chapter 6 – Is BE driven by saliency differences between people and object?

In Chapter 6 we aimed to replicate the results from Chapter 5, while better examining potential contributions of saliency between people and objects. Results showed that as previously observed in Chapter 5, there was a significant effect of BE for objects in scenes, but not for people. The reverse direction of the result for people in scenes was similar to that of Chapter 5 however, this did not reach significance (*p*=.09). Therefore, whilst there was no significant boundary contraction effect, these results showed that BE does not occur for people in scenes.

One explanation for the absence of boundary contraction in Chapter 6 might reflect an effect of familiarity for the person in the scene. For example, in this experiment, we better controlled for FOV by using the same scene background for both a person and an object respectively. Therefore, participants were exposed to multiple versions of the scene background, which could have led to a familiarity effect of the spatial scope of the scene. Research has shown that individuals still exhibit a strong BE effect for photographs of familiar environments (Gottesman & Munger, 2010). However, it could be suggested that the effect observed by Gottesman and Munger may be driven by prior knowledge of the space beyond the scene borders, building a more expansive detailed representation of the space, through amodal completion. Conversely, if BE was reduced due to familiarity in this experiment, then this should also have affected the BE score for objects, which was not the case. Given that the effect of stimulus repetition on BE has not been studied previously, it is only possible to speculate this as a plausible explanation.

The second focus for Chapter 6 was to determine whether saliency effects derived from extracting social information, modulated the BE effect observed in Chapter 5. Scenes in Chapter 6 were shown in both upright and inverted orientations (for all conditions). This condition was included as it was thought that inverting scenes would reduce participants' ability to extract salient information (Kelley, Chun & Chua, 2003; Shore & Klein, 2000), associated with people compared to objects. It was predicted that for objects in scenes, there would be no difference in BE between upright and inverted (Beighley & Intraub, 2016; Intraub & Berkowits, 1996). However, for people in scenes, it was thought that inverting the scene would reduce participants' ability to extract socially relevant information (such as identity and predicting their behaviour), and thus, yielding a more similar BE effect to that observed for objects in scenes. Results showed that the BE effect for objects was present regardless of their orientation, supporting the more recent findings of Beighley & Intraub (2016). For the scenes containing people, there was no significant BE or boundary contraction effect for upright or inverted people. Thus, we conclude that inverting the scene did not significantly influence the way in which BE occurred. Without a significant boundary contraction effect for upright people, it is challenging to draw conclusions about the effect of inversion. For example, the results showed no significant difference between CW and WC responses for upright or inverted people. Thus, while we are able to confirm that the type of trial (CW/WC) is modulated by the type of entity, this was not influenced by the additional factor of orientation.

Do these results suggest that the effect of BE differs because of the presence of a person, or might it reflect the absence of an object? Given the plethora of research providing evidence of the importance of objects in scenes in guiding spatial processing (Janzen & Turennout, 2004; Auger, Mullally and Maguire, 2012), it could be suggested that in some instances objects are quite simply used as tools to inform us of the geometric structure of space and guide our behaviour (Marchette, Vass, Ryan & Epstein, 2014). Thus, it could be suggested that scenes containing people are harder to process in terms of spatial extent because they have no spatially relevant cues. That is, without cues to infer an individual of the structure of space the individual finds it harder to make a judgement about the size and depth of the space. This theorem is supported by our finding in Chapter 5 that showed an absence of BE for landscape scenes without a focal object.

Taken together, the BE experiments jointly provide evidence for our initial hypotheses that memory for the spatial extent of a scene differs depending on the type of scene content, more specifically, the type of entity (person/object). We suggest that these findings support one of two possible explanations. First, the presence of a person crucially alters the way in which the spatial extent of a scene is processed and remembered due to an increase in social attentional demand induced by an increase in saliency. Second, it may not be the person *per se* that reduces the effect, but rather the lack of object, which can be used as a cue to the depth and size of space. For example, if we had observed a BE effect of fixed people in scenes then this explanation would not apply and the first explanation would be better fitting.

However, we do not see a BE for any of the conditions containing people. Moreover, these results may be explained partially by both explanations. For example, the contraction effect observed for dynamic people can be explained by the first idea. However, the lack of BE for all conditions containing people would be better explained by the second instance.

7.2. General conclusions and discussion of key accounts of object-in-scene representations

Together the studies in this thesis examined how variations in spatial and non-spatial properties of objects in scenes are encoded within the brain, and how scene content such as the type of entity may impede memory for the expanse of space. Here, I discuss key conclusions drawn from each of the 4 chapters and suggest how they relate to key models and accounts of object-in-scene processing.

Our fMRI chapters highlight the importance of fMRI task design and attention effects in RS, which need to be further explored. The RS effect for the RSC and OPA suggest that these regions may encode a wealth of information about objects within scenes, including both spatial properties and identity properties, similar to that of Troiani, Stigliani, Smith & Epstein (2014). Moreover, our whole-brain findings revealed bilateral activation in the IPL, for object Arrangements more than object Locations, possibly reflecting the cross talk between dorsal and ventral streams for encoding object-in-scene relations (Baldassano, Esteva, Fei-Fei & Beck, 2016; Fang & He, 2005; Santangelo, Di Francesco, Mastroberardino & Macaluso, 2015).

One model which should be considered when reflecting these results is the spatial layout hypothesis (Epstein, 2005). In this model, the PPA is involved in processing information about the shape of space including geometric information and the layout of the scene. While we did not find any significant effects in our fMRI results, we predicted that the PPA would be involved in encoding these types of information more than object based information. However, the literature has revealed many inconsistencies evidence assessing the function of the PPA where some suggest it encodes information about discrete objects and others suggesting it is involved in purely geometric and spatial properties (Epstein, 2005; Epstein, Graham & Downing, 2003; Troiani, Stigliani, Smith & Epstein, 2014). Therefore, future research should continue to assess this model in light of more recent evidence which identifies the PPA as a region that encodes information about object identity.

Our BE findings provide the most compelling results, where we observe BE for objects within scenes but not people. This was the overall pattern of results across both experiments, however, future research needs to re-examine the possible boundary contraction (reverse BE) for people in scenes which was observed in Chapter 5 but not Chapter 6. Critically, these findings may suggest that the mechanisms associated with encoding objects in scenes differs from those involved with encoding people in scenes. Two possible interpretations were discussed for these results. First, this result could reflect the social nature of people resulting in changes to the way in which attention is distributed across the scene. This idea was supported by our findings showing stronger reverse BE effect for dynamic people in scene compared to fixed people where there was no difference in same responses for CW/WC. Second, our BE results might reflect the importance of objects in encoding the spatial extent of a scene, important for amodal perception. For example, objects are units which we use to inform ourselves of our location in our environment, sometimes relevant for navigation (Chan, Baumann, Bellgrove & Mattingley, 2012; Epstein & Vass, 2014; Marchette, Vass, Ryan & Epstein, 2014). Thus, objects may be critical for grounding our understanding of the space and our location within it. As such, the presence of an object provides spatial reference, which provides context and meaning for the brain to reconstruct space beyond the border through amodal perception. However, future research should test these explanations to infer if BE relies upon object-context relationships and if salient entities prohibit this from occurring, as they typically do not provide strong contextual associations.

Our findings for the two BE studies can be considered in light of the multi-source model of memory (Intraub & Dickinson, 2008; Intraub, 2010). The multi-source model suggests that scene representation occurs as a function of amodal continuation of space, prior knowledge and previously learned object-context relationships. Our BE findings for objects in scenes replicates prior studies (Intraub, Bender & Mangels, 1992; Intraub, Gottesman, Willey & Zuk, 1996) which support this model. However, our results for people in scenes did not show BE. Could it be that this model is only appropriate for considering objects in scenes? Given that we rarely associate people in specific contexts (unless they are linked to specific jobs or are familiar to us) it might be that the brain is unable to draw upon any of these elements of information by which to create a representation of space. Therefore, there should be an extension of the multi-source model of scene memory to account for people in scenes. Future research should explore this further to determine the mechanisms behind why we do not see traditional BE effects for people in scenes.

Collectively, this thesis contributes novel findings suggesting that they type of content can modulate how scenes are reconstructed in memory, and extend prior research highlighting the importance of attention-demanding tasks in observing RS effects in fMRI.

7.3. Limitations and directions for future research

7.3.1. Mechanisms of BOLD and RS

One limitation relates to the method of RS itself. fMRI is an indirect measure of neural activity, derived from the contrast ratio between oxyhaemoglobin and deoxyhaemoglobin in the blood supplied to the brain (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001). The precise mechanisms that underpin RS are still heavily debated (James & Gauthier, 2006; McMahon & Olson, 2007). In humans, there appears to be a relationship between RS effect in fMRI and repetition priming (Dobbins, Schnyer, Verfaellie & Schacter, 2004; Voss, Hauner & Paller, 2009), however this relationship is not always observed for non-human primates (McMahon & Olson, 2007). Thus, to make inferences about the derived signal in RS, one needs to know which of the mechanisms (sharpening; fatigue; facilitation) is driving RS for a particular paradigm (Grill-Spector, Henson & Martin, 2006; Larsson, Solomon & Kohn, 2016). Without this, we are limited in our ability to understand the relationship between the BOLD and the behaviour of neurons in a given population. Thus, the mechanisms that drive fluctuations in RS during fMRI still need to be established.

While we are able to draw some conclusions for our two fMRI studies independently, the putative impact of attention on RS cannot be directly attributed to the novel attention task as the inferences were made in two independent samples. Therefore, future research should directly examine attention mechanisms and task effects influence the degree of RS within the same sample and with an array of tasks which are designed to specifically examine the effects of task on RS.

7.3.2. Toward an understanding of network dynamics in scene processing

Early fMRI studies revealed functionally defined regions in the cortex that showed selectivity to one condition compared to another (contrast method/activation analysis) (Epstein & Kanwisher, 1998). This method of determining brain function significantly advanced our understanding of the segregation of different functional brain structures. Since these early approaches fMRI methods have been developed, including RS, which allow us to probe the representational content of regions within the brain.

In the present thesis, the focus was to examine whether RS effects between conditions differed across ROIs. This method permitted us to assess topological differences across the scene-selective network, for encoding spatial and non-spatial information about objects in scenes. While this approach informs how activation differs between regions, it does not tell us if/how these regions are functionally connected and does not infer the network dynamics. The fMRI methods in the present thesis were specifically designed for RS, and thus, the design of the experiment is not optimal for investigating functional connectivity analyses.

Recent fMRI approaches now explore network dynamics of scene processing to examine how other regions beyond the PPA, RSC and OPA are implicated in spatial encoding and memory for scene information (Baldassano, Esteva, Fei-Fei & Beck, 2016). Baldassano, Esteva, Fei-Fei and Beck (2016) used functional connectivity analysis on data from the Human Connectome Project, to examine how these scene-selective regions are functionally organised in wider networks that process scene information. Their results provide insight into two possible networks that code different forms of information (Figure 7.1). They suggest the first network consists of the OPA and the posterior PPA, critically involved in coding feature information such as scene layout. In a second network, they observe that the anterior portion of the PPA functionally correlates with the RSC; fundamental in processing information relevant for navigation and episodic memory.

With the contributions of Baldassano, Esteva, Fei-Fei and Beck (2016) we are able to understand how different kinds of scene information are processed at the wider network level. With these more recent insights, future research should examine more closely, the functional architecture of object-in-scene representations. In this instance one analysis approach would be to employ a psychophysiological interaction analysis (PPI), which explores condition specific connectivity patterns between a seed region (a previously defined region of interest) and the rest of the brain.

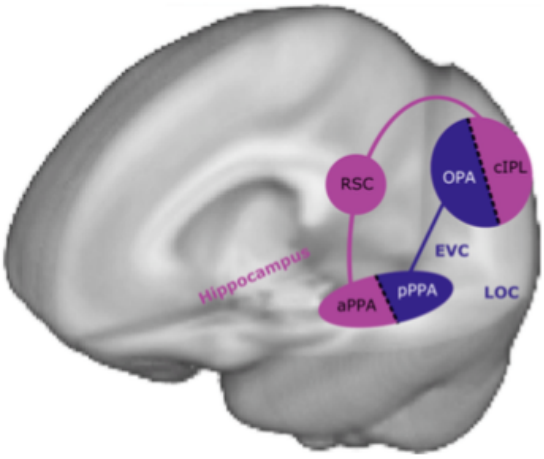


Figure 7.1. A representation of scene-selective organisation in the PPA, RSC, OPA, and IPL. Figure from Baldassano et al., (2016).

7.3.3. The effect of saliency in modulating BE effects

One of the most important contributions in this thesis is our observation that BE does not occur for people in scenes. In our first experiment, we observed a boundary contraction effect (reverse BE effect) for people in scenes. In our second experiment, we inverted the scenes to examine whether our prior results reflect the social/highly salient nature of people in scenes. However, given that upright people in scenes showed no boundary contraction effect, we were unable to draw conclusions regarding the influence of the inversion task on people in scenes. However, overall, our findings showed that the entity by trial type interaction was not influenced by the orientation of the scene. One possible consideration is that the inversion condition was not strong enough to reduce participants' ability to extract socially relevant cues from people in scenes.

In line with evidence studying BE for emotional scenes (Mathews & Mackintosh, 2004), it could be suggested that our results may reflect the social nature of people in scenes, thus determining the mechanisms that underpin these differences in BE for people and objects needs further exploration. Could there be two distinct mechanisms supporting spatial encoding? One dedicated for processing fixed objects and spatial elements of the scene required for aiding in spatial navigation? And a second one that processes space in relation to the dynamic demands driven by many non-fixed entities in our environment; such as people and animals?

Eye-tracking approaches could be used to examine how fixations are dispersed across the scene differently when there is a person compared to an object present. This could be done by comparing the distribution of fixations to the first image presentation in each trial with the second and comparing these potential differences between the two entities. This would allow us to explore whether these differences occur at the initial encoding phase or if they occur at the comparative phase (when the second image is compared to the mental representation formed for the first image). Additionally, given that prior research suggests initial fixations in scenes are made towards people (Bindemann, Scheepers, Ferguson & Burton, 2010; Mayer, Vuong & Thornton, 2015), fixations may be more densely clustered around the person whereas for objects, fixations may be more sparsely distributed across the scene. This approach would allow us to infer how visual attention to the scene differs as a function of the type of entity.

Recent evidence has documented that poor encoding increases BE for objects in scenes (Hale, McDunn & Brown, 2016), however, it might be that the differences observed between people and objects reflect differences in the way that scenes are attended and encoded. For example, it might be that for scenes containing objects, the *gist* can be inferred without high resolution encoding. However, for scenes containing people, a high degree of attention to the person is required to encode social cues, thus participants are more aware of the spatial extent of the scene (resulting in *no* BE). This theory has previously been discussed in the BE literature, where evidence shows that when attention is increased to the central foreground object, BE effect reduces. This is consistent with our findings where no BE was observed for scenes containing people. This idea is also supported by research suggesting that people (both bodies and/or faces) within scenes capture attention (Bindemann, Scheepers, Ferguson & Burton, 2010; Mayer, Vuong & Thornton, 2015). How could this be studied further? One way to examine this would be to include an additional task to the BE one requiring participants to encode other information beyond the person within a scene. If their attention is driven away from the person in the scene, then encoding of the scene might be poorer, increasing the

BE effect. This would allow us to determine if these differences in spatial memory for people and objects in scenes are driven by how scenes are encoded (the ‘resolution’ of initial scene encoding).

Likewise, future research should assess the importance of objects in BE, do they ground our understanding for space? If the object in a scene is critical for observing BE, then this might be measured using fMRI. Park, Intraub, Yi, Widders and Chun (2007) studied neural effects of BE in the core scene-selective regions and revealed that the PPA, showed significant attenuation for CW trials but not for WC trials. This is because the PPA encoded the second wide-angle view of the scene (in a CW trial) as the same as the first closer up presentation. Thus, if BE were reliant upon the presence of an object, one might predict that attenuation would be observed in the PPA for CW scenes containing objects, but not for CW landscape scenes with no focal object. This would allow us identify whether it is the presence of an object in its context that drives the BE.

7.3.4. Towards a broader model of object-in-scene understanding

Aminoff and Tarr (2015) suggest that scene processing may be driven by mechanisms of *associative processing*, a framework to describe processing of scenes through multiple inputs such as object-context relations, memory for prior experiences, amodal representations and predictions/expectations of object relationships. These associations reflect the interaction between processes such as visual recognition and memory for past visual experience (Aminoff, Kveraga & Bar, 2013). Using fMRI, the authors revealed the scene-selective regions are important nodes for encoding learned associations between objects in non-scene arrays. These include novel locations and identities of objects. Their results further implicated the PPA as a region for encoding spatial relations such as locations, whereas the RSC and OPA were more critically implicated in encoding object identity information. Their finding suggests that these regions do not just encode varying visual dimensions of scene information (open/closed, natural/manmade), but instead encode a complex combination of associative cues at a more general level driven by learned associations.

Although their study provides a novel framework for understanding scene processing, the stimuli are not representative of natural scenes. It could be suggested that the use of natural scenes may change the way in which these associations are learned. Future work should therefore re-explore this framework using natural scene contexts to assess whether this is generalizable. Whilst this framework provides interesting insight into the mechanisms of associative processing in the domain of scenes, it may not account for studies that detect activation within these regions for novel scenes and novel object-context relations that are not learned. Thus, future research should extend this framework to examine the network dynamics of associative processing and examine whether these effects can be observed with novel associations.

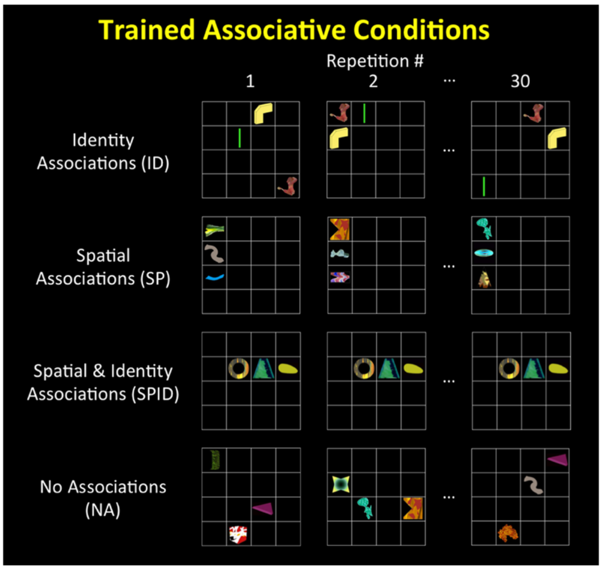


Figure 7.2. Example of trained associations, comprising of both spatial and non-spatial conditions. Image from Aminoff and Tarr (2015).

7.3.5. Limitation of the boundary extension method

A final limitation is that of the BE method. In the case of the current thesis we used BE as a tool to understand how scene content variations influence memory spatial extent of a scene. In the BE literature, there is a lack of standardised methods. For example, many researchers collect data on a 3-point or 5-point scale where participants respond ‘closer’, ‘same’ or ‘farther’. Form here the data is then collapsed to provide an overall BE measure (Intraub & Dickinson, 2008). As discussed in the BE empirical chapters, this method of analysis may not actually measure the effect of BE. It may simply reflect the balance of ‘farther’ and ‘closer’ responses. In the present thesis, we devised a novel way of analysing the data by examining the proportion of ‘same’ responses between CW and WC. Here we reported BE as an increase in the proportion of same responses for CW compared to WC.

While we were able to replicate a BE effect for objects in both of our experiments, further testing of this analysis method is required to ensure that it can be replicated.

7.5. Applications for the current research

When examining the cognitive processes that underlie human perceptual experience, one must also consider the applications that were initially considered at the beginning of the journey.

This thesis aimed to not only provide novel contributions to advance our knowledge of how variations in scene content are encoded and remembered, but with hope to provide insight for researchers examining situations where spatial processing may not be intact.

One example of a disease where spatial processing can become impaired is Alzheimer's disease (AD). Prior evidence examining the role of genetic risk factors in the development of AD have shown that scene encoding is affected in young adults (Salvato, Patai, McCloud & Nobre, 2016; Shine, Hodgetts, Postans, Lawrence & Graham, 2015). These striking findings highlight the importance of spatial cognition research in determining the networks and structures that are involved in these processes. By understanding how these regions encode spatial and non-spatial information, we will be able to extend our models to these young individuals to determine how they encode these types of information and whether deficits at a younger age can predict risk of the development of AD.

A more practical application for spatial cognition research is that of driverless cars. Humans are able to effortless encode a magnitude of visual information from their environment, including both spatial properties and the types of entities. However, when developing technology these cognitive processes including identification and recognition of various elements have to be built in. This level of computation and cognition for machines can be based upon models derived from human cognition. Therefore, it is critical to extend and disseminate the findings from visual attention and visual spatial cognition to these wider applied models of computational learning in machines.

7.6. Final comments

The body of research presented in the current thesis adopted a mix of behavioural and fMRI approaches, to advance our knowledge of object-in-scene representations during encoding and memory. The reason behind investigating this particular line of research was to examine how humans encode spatial and non-spatial properties of objects within scenes, and how variations in the type of scene content influences memory for the extent of space. Our findings provide insight into how scene content variations influence building of coherent spatial representations of the world (cognitive map, Tolman, 1948), and how this may be altered by the presence of social elements within our environment. In addition, we highlight the importance of methodological considerations for RS in a paired-stimulus design.

With careful development of these experiments and consideration for the limitations associated with these methods, we will be able to advance this body of work to infer more about the mechanisms driving scene processing with a focus to determine how organisms encode and remember both spatial and non-spatial information within their local visual environment. The behavioural findings provide novel insight for both spatial cognition and social perception work. By developing these ideas and exploring further how different scene elements influence visual scene processing, we will be better able to understand the mechanisms that support spatial cognition, and how this is implemented in both health and disease.

References

Adolphs, R. (2003a). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4*(3), 165-178. doi:10.1038/nrn1056

Adolphs, R. (2003b). Investigating the cognitive neuroscience of social behavior. *Neuropsychologia, 41*(2), 119-126.

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Res, 1079*(1), 25-35. doi:10.1016/j.brainres.2005.12.127

Aggleton, J. P. (2010). Understanding retrosplenial amnesia: insights from animal studies. *Neuropsychologia, 48*(8), 2328-2338. doi:10.1016/j.neuropsychologia.2009.09.030

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain, 122 (Pt 9)*, 1613-1628.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in cognitive sciences, 15*(3), 122-131. doi:10.1016/j.tics.2011.01.003

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cereb Cortex, 17*(7), 1493-1503. doi:10.1093/cercor/bhl078

Aminoff, E., Schacter, D. L., & Bar, M. (2008). The cortical underpinnings of context-based memory distortion. *J Cogn Neurosci, 20*(12), 2226-2237. doi:10.1162/jocn.2008.20156

Aminoff, E. M., Kveraga, K., & Bar, M. (2013). The role of the parahippocampal cortex in cognition. *Trends Cogn Sci, 17*(8), 379-390. doi:10.1016/j.tics.2013.06.009

Aminoff, E. M., & Tarr, M. J. (2015). Associative Processing Is Inherent in Scene Perception. *PLoS One, 10*(6), e0128840. doi:10.1371/journal.pone.0128840

Andrew Browning, N., Grossberg, S., & Mingolla, E. (2009). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Netw, 22*(10), 1383-1398. doi:10.1016/j.neunet.2009.05.007

Archer, J., Hay, D. C., & Young, A. W. (1994). Movement, face processing and schizophrenia: evidence of a differential deficit in expression analysis. *Br J*

Clin Psychol, 33 (Pt 4), 517-528.

Attwell, D., Buchan, A. M., Charpak, S., Lauritzen, M., Macvicar, B. A., & Newman, E. A. (2010). Glial and neuronal control of brain blood flow. *Nature, 468*(7321), 232-243. doi:10.1038/nature09613

Auger, S. D., & Maguire, E. A. (2013). Assessing the mechanism of response in the retrosplenial cortex of good and poor navigators. *Cortex, 49*(10), 2904-2913. doi:10.1016/j.cortex.2013.08.002

Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *PLoS One, 7*(8), e43620. doi:10.1371/journal.pone.0043620

Azam, M., Potlapalli, H., Janet, J., & Luo, R. C. (1996). Outdoor landmark recognition using segmentation, fractal model and neural network. *Image Understanding Workshop, 1996 Proceedings, Vols I and Ii*, 189-203.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47-89.

Bainbridge, W. A., & Oliva, A. (2015). Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage, 122*, 408-416. doi:10.1016/j.neuroimage.2015.07.066

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *Neuroimage, 75*, 228-237. doi:10.1016/j.neuroimage.2013.02.073

Baldassano, C., Esteva, A., Fei-Fei, L., & Beck, D. M. (2016). Two Distinct Scene-Processing Networks Connecting Vision and Memory. *eNeuro, 3*(5). doi:10.1523/ENEURO.0178-16.2016

Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *J Vis, 16*(2), 9. doi:10.1167/16.2.9

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617-629. doi:10.1038/nrn1476

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38*(2), 347-358.

Barrash, J. (1998). A historical review of topographical disorientation and its neuroanatomical correlates. *J Clin Exp Neuropsychol, 20*(6), 807-827. doi:10.1076/jcen.20.6.807.1114

Barron, H. C., Garvert, M. M., & Behrens, T. E. (2016). Repetition suppression: a means to index neural representations using BOLD? *Philos Trans R Soc Lond B Biol Sci, 371*(1705). doi:10.1098/rstb.2015.0355

Baumann, O., Chan, E., & Mattingley, J. B. (2010). Dissociable neural circuits for encoding and retrieval of object locations during active navigation in humans. *Neuroimage, 49*(3), 2816-2825. doi:10.1016/j.neuroimage.2009.10.021

Baumann, O., Chan, E., & Mattingley, J. B. (2012). Distinct neural networks underlie encoding of categorical versus coordinate spatial relations during active navigation. *Neuroimage, 60*(3), 1630-1637. doi:10.1016/j.neuroimage.2012.01.089

Baumann, O., & Mattingley, J. B. (2016). Functional Organization of the Parahippocampal Cortex: Dissociable Roles for Context Representations and the Perception of Visual Scenes. *J Neurosci, 36*(8), 2536-2542. doi:10.1523/JNEUROSCI.3368-15.2016

Beighley, S., & Intraub, H. (2016). Does inversion affect boundary extension for briefly-presented views? *Visual Cognition, 24*(3), 252-259. doi:10.1080/13506285.2016.1229369

Bertamini, M., Jones, L. A., Spooner, A., & Hecht, H. (2005). Boundary extension: The role of magnification, object size, context, and binocular information. *Journal of Experimental Psychology-Human Perception and Performance, 31*(6), 1288-1307. doi:10.1037/0096-1523.31.6.1288

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. *J Cogn Neurosci, 25*(10), 1711-1722. doi:10.1162/jocn_a_00422

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol Rev, 94*(2), 115-147.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol, 14*(2), 143-177.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *J Exp Psychol, 103*(3), 597-600.

Bigelow, N. O., Paradiso, S., Adolphs, R., Moser, D. J., Arndt, S., Heberlein, A., . . .

Andreasen, N. C. (2006). Perception of socially relevant stimuli in schizophrenia. *Schizophr Res, 83*(2-3), 257-267. doi:10.1016/j.schres.2005.12.856

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *J Exp Psychol Hum Percept Perform, 36*(6), 1477-1485. doi:10.1037/a0019057

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: the roles of action, competition and social content. *Q J Exp Psychol (Hove), 61*(7), 986-998.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009a). Saliency does not account for fixations to eyes within social scenes. *Vision research*, *49*(24), 2992-3000.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009b). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, *17*(6-7), 904-924.

Borji, A., Sihite, D. N., & Itti, L. (2013). Objects do not predict fixations better than early saliency: a re-analysis of Einhauser et al.'s data. *J Vis, 13*(10), 18. doi:10.1167/13.10.18

Bouchon, C., Nazzi, T., & Gervain, J. (2015). Hemispheric asymmetries in repetition enhancement and suppression effects in the newborn brain. *PloS one*, *10*(10), e0140160.

Braak, H., & Braak, E. (1991a). Morphological changes in the human cerebral cortex in dementia. *J Hirnforsch, 32*(3), 277-282.

Braak, H., & Braak, E. (1991b). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol, 82*(4), 239-259.

Brockmole, J. R., & Le-Hoa Vo, M. (2010). Semantic memory for contextual regularities within and across scene categories: evidence from eye movements. *Atten Percept Psychophys, 72*(7), 1803-1813. doi:10.3758/APP.72.7.1803

Brouwer, A.-M., Franz, V. H., & Thornton, I. M. (2004). Representational momentum in perception and grasping: Translating versus transforming objects.

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

Buffalo, E. A., Bellgowan, P. S., & Martin, A. (2006). Distinct roles for medial temporal lobe structures in memory for objects and their locations. *Learn Mem, 13*(5), 638-643. doi:10.1101/lm.251906

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *J Neurosci, 30*(21), 7269-7280. doi:10.1523/JNEUROSCI.5551-09.2010

Buxton, R. B., Griffeth, V. E., Simon, A. B., Moradi, F., & Shmuel, A. (2014). Variability of the coupling of blood flow and oxygen metabolism responses in the brain: a problem for interpreting BOLD studies but potentially a new window on the underlying neural activity. *Front Neurosci, 8*, 139. doi:10.3389/fnins.2014.00139

Candel, I., Merckelbach, H., & Zandbergen, M. (2003). Boundary distortions for neutral and emotional pictures. *Psychonomic Bulletin & Review, 10*(3), 691-695. doi:Doi 10.3758/Bf03196533

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol Psychol, 54*(1-3), 241-257.

Castelhano, M. S., & Witherspoon, R. L. (2016). How You Use It Matters: Object Function Guides Attention During Visual Search in Scenes. *Psychol Sci, 27*(5), 606-621. doi:10.1177/0956797616629130

Cerf, M., Harel, J., Huth, A., Einhäuser, W., & Koch, C. (2008). *Decoding what people see from where they look: Predicting visual stimuli from scanpaths.* Paper presented at the International Workshop on Attention in Cognitive Systems.

Chan, E., Baumann, O., Bellgrove, M. A., & Mattingley, J. B. (2012). From objects to landmarks: the function of visual location information in spatial navigation. *Front Psychol, 3*, 304. doi:10.3389/fpsyg.2012.00304

Charles Leek, E., & Johnston, S. J. (2006). A polarity effect in misoriented object recognition: The role of polar features in the computation of orientation-invariant shape representations. *Visual Cognition, 13*(5), 573-600.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*(2), 149-178.

Cole, S., Balcetis, E., & Dunning, D. (2013). Affective signals of threat increase perceived proximity. *Psychol Sci, 24*(1), 34-40. doi:10.1177/0956797612446953

Committeri, G., Galati, G., Paradis, A. L., Pizzamiglio, L., Berthoz, A., & LeBihan, D. (2004). Reference frames for spatial cognition: different brain areas are involved in viewer-, object-, and landmark-centered judgments about object location. *J Cogn Neurosci, 16*(9), 1517-1535. doi:10.1162/0898929042568550

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Courtney, J., & Hubbard, T. (2004). *Possible asymmetries and effects of attention in boundary extension.* Paper presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis, MN.

Creem, S. H., & Proffitt, D. R. (2001). Defining the cortical visual systems: "what", "where", and "how". *Acta Psychol (Amst), 107*(1-3), 43-68.

Cukur, T., Huth, A. G., Nishimoto, S., & Gallant, J. L. (2016). Functional Subdomains within Scene-Selective Cortex: Parahippocampal Place Area, Retrosplenial Complex, and Occipital Place Area. *J Neurosci, 36*(40), 10257-10273. doi:10.1523/JNEUROSCI.4033-14.2016

Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Attention, Perception, & Psychophysics*, *62*(4), 673-683.

Czigler, I., Intraub, H., & Stefanics, G. (2013a). Prediction beyond the borders: ERP indices of boundary extension-related error. *PLoS One, 8*(9), e74245. doi:10.1371/journal.pone.0074245
10.1371/journal.pone.0074245

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Mem Cognit, 35*(3), 393-401.

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

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A, 93*(24), 13494-13499.

Di Salle, F., Formisano, E., Linden, D. E., Goebel, R., Bonavita, S., Pepino, A.,Tedeschi, G. (1999). Exploring brain function with magnetic resonance imaging. *Eur J Radiol, 30*(2), 84-94.

Dickinson, C. A., & LaCombe, D. C. (2014). Objects influence the shape of remembered views: Examining global and local aspects of boundary extension. *Perception, 43*(8), 731-753. doi:10.1068/p7631

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci, 31*(31), 11305-11312. doi:10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *J Neurosci, 33*(4), 1331-1336a. doi:10.1523/JNEUROSCI.4081-12.2013

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature, 428*(6980), 316-319. doi:10.1038/nature02400

Dossani, R. H., Missios, S., & Nanda, A. (2015). The Legacy of Henry Molaison (1926-2008) and the Impact of His Bilateral Mesial Temporal Lobe Surgery on the Study of Human Memory. *World Neurosurg, 84*(4), 1127-1135. doi:10.1016/j.wneu.2015.04.031

Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition, 93*(1), B27-38. doi:10.1016/j.cognition.2003.10.010

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470-2473. doi:10.1126/science.1063414

Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2004). BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. *J Neurophysiol, 92*(2), 1241-1247. doi:10.1152/jn.00206.2004

Eklund, A., Nichols, T., & Knutsson, H. (2015). Can parametric statistical methods be trusted for fMRI based group studies?. *arXiv preprint arXiv:1511.01863*.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of vision*, *8*(3), 3-3.

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci Biobehav Rev, 24*(6), 581-604.

Epstein, R., Deyoe, E. A., Press, D. Z., Rosen, A. C., & Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cogn Neuropsychol, 18*(6), 481-508.

doi:10.1080/02643290125929

Epstein, R., Graham, K. S., & Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron, 37*(5), 865-876.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron, 23*(1), 115-125.

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

Epstein, R. A. (2005). The cortical basis of visual scene processing. *Visual Cognition, 12*(6), 954-978. doi:10.1080/13506280444000607

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci, 12*(10), 388-396. doi:10.1016/j.tics.2008.07.004

Epstein, R. A., & Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb Cortex, 17*(7), 1680-1693. doi:10.1093/cercor/bhl079

Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *J Neurophysiol, 97*(5), 3670-3683. doi:10.1152/jn.00003.2007

Epstein, R. A., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience, 17*(1), 73-83. doi:Doi 10.1162/0898929052879987

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J Neurosci, 27*(23), 6141-6149. doi:10.1523/JNEUROSCI.0799-07.2007

Epstein, R. A., Parker, W. E., & Feiler, A. M. (2008). Two kinds of FMRI repetition suppression? Evidence for dissociable neural mechanisms. *J Neurophysiol, 99*(6), 2877-2886. doi:10.1152/jn.90376.2008

Epstein, R. A., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philos Trans R Soc Lond B Biol Sci, 369*(1635), 20120533. doi:10.1098/rstb.2012.0533

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature neuroscience, 8*(10), 1380-1385.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A, 110*(41), 16616-16621. doi:10.1073/pnas.1315235110

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Prog Brain Res, 155*, 3-21. doi:10.1016/S0079-6123(06)55001-0

Feuerriegel, D. (2016). Selecting appropriate designs and comparison conditions in repetition paradigms. *Cortex, 80*, 196-205. doi:10.1016/j.cortex.2015.10.022

Filimon, F. (2015). Are all spatial reference frames egocentric? Reinterpreting evidence for allocentric, object-centered, or world-centered reference frames. *Frontiers in human neuroscience, 9*.

Fini, C., Costantini, M., & Committeri, G. (2014). Sharing Space: The Presence of Other Bodies Extends the Space Judged as Near. *PLoS One, 9*(12), e114719. doi:10.1371/journal.pone.0114719

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*(4), 571-583.

Foulsham, C. K., 2nd, Snyder, G. G., 3rd, & Carpenter, R. J., 3rd. (1981). Papillary cystadenoma lymphomatosum of the larynx. *Otolaryngol Head Neck Surg, 89*(6), 960-964.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117*(3), 319-331.

Foulsham, T., & Underwood, G. (2009). Does conspicuity enhance distraction? Saliency and eye landing position when searching for objects. *Q J Exp Psychol (Hove), 62*(6), 1088-1098. doi:10.1080/17470210802602433

Gagnier, K. M., Dickinson, C. A., & Intraub, H. (2013). Fixating picture boundaries does not eliminate boundary extension: Implications for scene representation. *Quarterly Journal of Experimental Psychology, 66*(11), 2161-2186. doi:10.1080/17470218.2013.775595

Gallagher, K., Balas, B., Matheny, J. and Sinha, P. (2005) The Effects of Scene Cate- gory and Content on Boundary Extension. In B. Bara, L. Barsalou and M. Buc- ciarelli (eds.) Proceedings of the 27th Annual Meeting of the Cognitive Science Society. Stresa, Italy: Cognitive Science Society.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends Cogn Sci, 8*(9), 396-403. doi:10.1016/j.tics.2004.07.002

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. (2013). Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *J Cogn Neurosci, 25*(6), 961-968. doi:10.1162/jocn_a_00372

Ganel, T., & Goodale, M. A. (2003). Visual control of action but not perception requires analytical processing of object shape. *Nature, 426*(6967), 664-667. doi:10.1038/nature02156

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*(4), 870-878.

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., & Denis, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. *Neuroreport, 8*(3), 739-744.

Gottesman, C. V., & Intraub, H. (1999). Wide-angle memories of close-up scenes: A demonstration of boundary extension. *Behavior Research Methods Instruments & Computers, 31*(1), 86-93. doi:Doi 10.3758/Bf03207697

Gottesman C. V. (2011). Mental layout extrapolations prime spatial processing of scenes. *Journal of Experimental Psychology: Human Perception & Performance*, 37, 382–395

Gottesman, C. V., & Intraub, H. (2002). Surface construal and the mental representation of scenes. *J Exp Psychol Hum Percept Perform, 28*(3), 589-599.

Gottesman, C. V., & Intraub, H. (2003). Constraints on spatial extrapolation in the mental representation of scenes: View-boundaries vs. object-boundaries. *Visual Cognition, 10*(7), 875-893. doi:10.1080/13506280344000130

Gottesman, C. V., & Munger, M. P. (2010). Is Boundary Extension Different When You've Been There? Memory for Familiar and Unfamiliar Campus Pictures. *Journal of Vision, 10*(7), 1240-1240.

Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: rapid scene understanding benefits from prior experience. *Atten Percept Psychophys, 77*(4), 1239-1251. doi:10.3758/s13414-015-0859-8

Greene, M. R., & Oliva, A. (2009a). The Briefest of Glances: The Time Course of

Natural Scene Understanding. *Psychological Science, 20*(4), 464-472. doi:10.1111/j.1467-9280.2009.02316.x

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

Greene, M. R., & Oliva, A. (2010). High-Level Aftereffects to Global Scene Properties. *Journal of Experimental Psychology-Human Perception and Performance, 36*(6), 1430-1442. doi:10.1037/a0019058

Grezes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: an fMRI study of implicit processing. *Eur J Neurosci, 17*(12), 2735-2740.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci, 10*(1), 14-23. doi:10.1016/j.tics.2005.11.006

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp, 6*(4), 316-328.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst), 107*(1-3), 293-321.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371-388.

Guo, F., Preston, T. J., Das, K., Giesbrecht, B., & Eckstein, M. P. (2012). Feature-independent neural coding of target detection during search of natural scenes. *Journal of Neuroscience*, *32*(28), 9499-9510.

Habib, M., & Sirigu, A. (1987). Pure topographical disorientation: a definition and anatomical basis. *Cortex, 23*(1), 73-85.

Hale, R. G., Brown, J. M., & McDunn, B. A. (2016). Increasing task demand by obstructing object recognition increases boundary extension. *Psychon Bull Rev, 23*(5), 1497-1503. doi:10.3758/s13423-016-1018-5

Hale, R. G., Brown, J. M., McDunn, B. A., & Siddiqui, A. P. (2015). An influence of extremal edges on boundary extension. *Psychonomic Bulletin & Review, 22*(4), 961-966. doi:10.3758/s13423-014-0751-x

Harel, A., Groen, II, Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The Temporal Dynamics of Scene Processing: A Multifaceted EEG Investigation. *eNeuro, 3*(5). doi:10.1523/ENEURO.0139-16.2016

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: gradients of object and spatial layout information. *Cereb Cortex, 23*(4), 947-957. doi:10.1093/cercor/bhs091

Hart, B. M., Schmidt, H. C., Roth, C., & Einhauser, W. (2013). Fixations on objects in natural scenes: dissociating importance from salience. *Front Psychol, 4*, 455. doi:10.3389/fpsyg.2013.00455

He, B. J., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2007). The role of impaired neuronal communication in neurological disorders. *Current opinion in neurology, 20*(6), 655-660.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annu Rev Psychol, 50*, 243-271. doi:10.1146/annurev.psych.50.1.243

Henderson, J. M., Larson, C. L., & Zhu, D. C. (2007). Cortical activation to indoor versus outdoor scenes: an fMRI study. *Experimental Brain Research*, *179*(1), 75-84.

Henderson, J. M., Zhu, D. C., & Larson, C. L. (2011). Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Visual Cognition, 19*(7), 910-927. doi:10.1080/13506285.2011.596852

Henson, R. N., & Rugg, M. D. (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia, 41*(3), 263-270.

Henson, R. N., Shallice, T., Gorno-Tempini, M. L., & Dolan, R. J. (2002). Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cereb Cortex, 12*(2), 178-186.

Hodgetts, C. J., Shine, J. P., Lawrence, A. D., Downing, P. E., & Graham, K. S. (2016). Evidencing a place for the hippocampus within the core scene processing network. *Hum Brain Mapp, 37*(11), 3779-3794. doi:10.1002/hbm.23275

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *J Exp Psychol Hum Percept Perform, 33*(1), 31-47. doi:10.1037/0096-1523.33.1.31

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *J Exp Psychol Gen, 127*(4), 398-415.

Hollingworth, A., & Henderson, J. M. (1999). Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination. *Acta Psychol (Amst), 102*(2-3), 319-343.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113.

Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: a dissociation between explicit change detection and visual memory. *Percept Psychophys, 66*(5), 800-807.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: visually specific information is retained in memory from previously attended objects in natural scenes. *Psychon Bull Rev, 8*(4), 761-768.

Hsu, Y. F., Hamalainen, J. A., & Waszak, F. (2014). Repetition suppression comprises both attention-independent and attention-dependent processes. *Neuroimage, 98*, 168-175. doi:10.1016/j.neuroimage.2014.04.084

Husain, M., Mannan, S., Hodgson, T., Wojciulik, E., Driver, J., & Kennard, C. (2001). Impaired spatial working memory across saccades contributes to abnormal search in parietal neglect. *Brain, 124*(Pt 5), 941-952.

Iannetti, G. D., & Wise, R. G. (2007). BOLD functional MRI in disease and pharmacological studies: room for improvement? *Magn Reson Imaging, 25*(6), 978-988. doi:10.1016/j.mri.2007.03.018

Intraub, H. (2010). Rethinking Scene Perception: A Multisource Model. *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 52, 52*, 231-264. doi:10.1016/S0079-7421(10)52006-1

Intraub, H. (2012). Rethinking visual scene perception. *Wiley Interdisciplinary Reviews-Cognitive Science, 3*(1), 117-127. doi:10.1002/wcs.149

Intraub, H., Bender, R. S., & Mangels, J. A. (1992). Looking at Pictures but Remembering Scenes. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(1), 180-191. doi:Doi 10.1037/0278-7393.18.1.180

Intraub, H., & Berkowits, D. (1996). Beyond the edges of a picture. *American*

Journal of Psychology, 109(4), 581-598. doi:Doi 10.2307/1423396

Intraub, H., & Bodamer, J. L. (1993). Boundary Extension - Fundamental Aspect of Pictorial Representation or Encoding Artifact. *Journal of Experimental Psychology-Learning Memory and Cognition, 19*(6), 1387-1397. doi:Doi 10.1037//0278-7393.19.6.1387.

Intraub, H., & Dickinson, C. A. (2008). False memory 1/20th of a second later: What the early onset of boundary extension reveals about perception. *Psychological Science*, *19*(10), 1007-1014.

Intraub, H., Gottesman, C. V., & Bills, A. J. (1998). Effects of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology-Learning Memory and Cognition, 24*(1), 186-201. doi:Doi 10.1037/0278-7393.24.1.186

Intraub, H., Gottesman, C. V., Willey, E. V., & Zuk, I. J. (1996). Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language, 35*(2), 118-134. doi:DOI 10.1006/jmla.1996.0007

Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes. *J Exp Psychol Learn Mem Cogn, 15*(2), 179-187.

Intraub H. (1997). The representation of visual scenes. *Trends in the Cognitive Sciences*, 1, 217–221

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203. doi:10.1038/35058500

James, T. W., & Gauthier, I. (2006). Repetition-induced changes in BOLD response reflect accumulation of neural activity. *Hum Brain Mapp, 27*(1), 37-46. doi:10.1002/hbm.20165

Janzen, G., & van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nature neuroscience, 7*(6), 673-677. doi:10.1038/nn1257

Jessen, F., Manka, C., Scheef, L., Granath, D. O., Schild, H. H., & Heun, R. (2002). Novelty detection and repetition suppression in a passive picture viewing task: a possible approach for the evaluation of neuropsychiatric disorders. *Hum Brain Mapp, 17*(4), 230-236. doi:10.1002/hbm.10071

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: fast categorization and object interference. *Vision Res, 47*(26), 3286-3297. doi:10.1016/j.visres.2007.09.013

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The Occipital Place Area Is Causally Involved in Representing Environmental Boundaries during Navigation. *Curr Biol, 26*(8), 1104-1109. doi:10.1016/j.cub.2016.02.066

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *J Neurosci, 36*(41), 10522-10528. doi:10.1523/JNEUROSCI.1385-16.2016

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc Natl Acad Sci U S A, 111*(30), 11217-11222. doi:10.1073/pnas.1400559111

Kaiser, D., Strnad, L., Seidl, K. N., Kastner, S., & Peelen, M. V. (2014). Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J Neurophysiol, 111*(1), 82-90. doi:10.1152/jn.00371.2013

Kalenine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2014). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychon Bull Rev, 21*(3), 645-651. doi:10.3758/s13423-013-0536-7

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *Neuroimage, 132*, 417-424. doi:10.1016/j.neuroimage.2016.02.062

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual cognition*, *17*(6-7), 979-1003.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annu Rev Neurosci, 23*, 315-341. doi:10.1146/annurev.neuro.23.1.315

Kelley, T. A., Chun, M. M., & Chua, K. P. (2003). Effects of scene inversion on change detection of targets matched for visual salience. *J Vis, 3*(1), 1-5. doi:10.1167/3.1.1

Kim, D., & Nevatia, R. . (1994). A method for recognition and localization of generic objects for indoor navigation. . *Proceedings of the Second IEEE Workshop on (pp. 280-288). IEEE.*(IEEE.), 280-288).

Kim, S., Dede, A. J., Hopkins, R. O., & Squire, L. R. (2015). Memory, scene construction, and the human hippocampus. *Proc Natl Acad Sci U S A, 112*(15), 4767-4772. doi:10.1073/pnas.1503863112

Klatzky, R. L. (1998). *Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections.* Paper presented at the Spatial cognition.

Kohler, S., Crane, J., & Milner, B. (2002). Differential contributions of the parahippocampal place area and the anterior hippocampus to human memory for scenes. *Hippocampus, 12*(6), 718-723. doi:10.1002/hipo.10077

Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron, 39*(4), 681-691.

Kohn, A., & Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nat Neurosci, 7*(7), 764-772. doi:10.1038/nn1267

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506-1509. doi:10.1126/science.1061133

Kovacs, G., & Vogels, R. (2014). When does repetition suppression depend on repetition probability? *Front Hum Neurosci, 8*, 685. doi:10.3389/fnhum.2014.00685

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. *J Neurosci, 31*(20), 7322-7333. doi:10.1523/JNEUROSCI.4588-10.2011

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience, 12*(4), 217-230. doi:10.1038/nrn3008

Lancaster, T. M., Ihssen, N., Brindley, L. M., & Linden, D. E. (2017). Further support for association between GWAS variant for positive emotion and reward systems. *Transl Psychiatry, 7*(1), e1018. doi:10.1038/tp.2016.289

Larsson, J., & Smith, A. T. (2012). fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb Cortex, 22*(3), 567-576. doi:10.1093/cercor/bhr119

Larsson, J., Solomon, S. G., & Kohn, A. (2016). fMRI adaptation revisited. *Cortex,* *80*, 154-160. doi:10.1016/j.cortex.2015.10.026

LeDoux, J. (1998). *The emotional brain: The mysterious underpinnings of emotional life*: Simon and Schuster.

Linsley, D., & MacEvoy, S. P. (2014). Evidence for participation by object-selective visual cortex in scene category judgments. *J Vis, 14*(9). doi:10.1167/14.9.19

Linsley, D., & MacEvoy, S. P. (2015). Encoding-Stage Crosstalk Between Object- and Spatial Property-Based Scene Processing Pathways. *Cereb Cortex, 25*(8), 2267-2281. doi:10.1093/cercor/bhu034

Logothetis, N. K. (2010). Neurovascular Uncoupling: Much Ado about Nothing. *Front Neuroenergetics, 2*. doi:10.3389/fnene.2010.00002

Logothetis, N. K., Guggenberger, H., Peled, S., & Pauls, J. (1999). Functional imaging of the monkey brain. *Nat Neurosci, 2*(6), 555-562. doi:10.1038/9210

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157. doi:10.1038/35084005

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annu. Rev. Physiol., 66*, 735-769.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex, 4*(5), 523-531.

MacEvoy, S. P., & Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *J Neurophysiol, 98*(4), 2089-2098. doi:10.1152/jn.00438.2007

MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci, 14*(10), 1323-1329. doi:10.1038/nn.2903

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. *Visual attention, 8*, 55-76.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand J Psychol, 42*(3), 225-238.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: a human navigation network. *Science, 280*(5365), 921-924.

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *J Exp Psychol Gen, 144*(2), 257-263. doi:10.1037/xge0000060

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci, 17*(11), 1598-1606. doi:10.1038/nn.3834

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside Looking In: Landmark Generalization in the Human Navigational System. *J Neurosci, 35*(44), 14896-14908. doi:10.1523/JNEUROSCI.2270-15.2015

Mathews, A., & Mackintosh, B. (2004). Take a closer look: emotion modifies the boundary extension effect. *Emotion, 4*(1), 36-45. doi:10.1037/1528-3542.4.1.36

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do People "Pop Out"? *PLoS One, 10*(10), e0139618. doi:10.1371/journal.pone.0139618

McDunn, B. A., Siddiqui, A. P., & Brown, J. M. (2014). Seeking the boundary of boundary extension. *Psychonomic Bulletin & Review, 21*(2), 370-375. doi:10.3758/s13423-013-0494-0

McMahon, D. B., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *J Neurophysiol, 97*(5), 3532-3543. doi:10.1152/jn.01042.2006

Mendez, M. F., & Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia, 41*(10), 1387-1395.

Menetrier, E., Didierjean, A., & Vieillard, S. (2013). Is boundary extension emotionally selective? *Quarterly Journal of Experimental Psychology, 66*(4), 635-641. doi:10.1080/17470218.2013.764332

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*(5146), 520-522.

Moore, K. S., Yi, D. J., & Chun, M. (2013). The effect of attention on repetition suppression and multivoxel pattern similarity. *J Cogn Neurosci, 25*(8), 1305-1314. doi:10.1162/jocn_a_00387

Mullally, S. L., Intraub, H., & Maguire, E. A. (2012a). Attenuated Boundary Extension Produces a Paradoxical Memory Advantage in Amnesic Patients. *Current Biology, 22*(4), 261-268. doi:10.1016/j.cub.2012.01.001

Mullally, S. L., & Maguire, E. A. (2011). A new role for the parahippocampal cortex in representing space. *J Neurosci, 31*(20), 7441-7449. doi:10.1523/JNEUROSCI.0267-11.2011

Mullin, C. R., & Steeves, J. K. (2011). TMS to the lateral occipital cortex disrupts object processing but facilitates scene processing. *J Cogn Neurosci, 23*(12), 4174-4184. doi:10.1162/jocn_a_00095

Munger, M. P., Owens, T. R., & Conway, J. E. (2005). Are boundary extension and representational momentum related? *Visual Cognition, 12*(6), 1041-1056. doi:10.1080/13506280444000643

Nakada, T. (1999). [BOLD functional MRI: practical pitfalls]. *Rinsho Shinkeigaku, 39*(1), 39-41.

Nasr, S., Liu, N., Devaney, K. J., Yue, X. M., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H. (2011). Scene-Selective Cortical Regions in Human and Nonhuman Primates. *Journal of Neuroscience, 31*(39), 13771-13785. doi:10.1523/Jneurosci.2792-11.2011

Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. (2011). Scene-selective cortical regions in human and nonhuman primates. *Journal of Neuroscience*, *31*(39), 13771-13785.

Nasr, S., & Tootell, R. B. (2016). Visual field biases for near and far stimuli in disparity selective columns in human visual cortex. *Neuroimage*. doi:10.1016/j.neuroimage.2016.09.012

Nieto-Castanon, A., Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2003). Region of interest based analysis of functional imaging data. *Neuroimage, 19*(4), 1303-1316.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in cognitive sciences, 13*(3), 135-143.

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *J Cogn Neurosci, 12*(6), 1013-1023.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A, 87*(24), 9868-9872.

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. doi:Doi 10.1023/A:1011139631724

Oliva, A., & Torralba, A. (2002). Scene-centered description from spatial envelope properties. *Biologically Motivated Computer Vision, Proceedings, 2525*, 263-272.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Visual Perception, Pt 2: Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception, 155*, 23-36. doi:10.1016/S0079-6123(06)55002-2

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*(4), 1333-1340.

Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage, 47*(4), 1747-1756. doi:10.1016/j.neuroimage.2009.04.058

Park, S., & Chun, M. M. (2014). The Constructive Nature of Scene Perception. *Scene Vision: Making Sense of What We See*, 45.

Park, S., Intraub, H., Yi, D. J., Widders, D., & Chun, M. M. (2007). Beyond the edges of a view: Boundary extension in human scene-selective visual cortex. *Neuron, 54*(2), 335-342. doi:10.1016/j.neuron.2007.04.006

Park, S., Konkle, T., & Oliva, A. (2015). Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. *Cereb Cortex, 25*(7), 1792-1805. doi:10.1093/cercor/bht418

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *J Neurophysiol, 93*(1), 603-608. doi:10.1152/jn.00513.2004

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94-97.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci, 18*(5), 242-250. doi:10.1016/j.tics.2014.02.004

Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex, 77*, 155-163. doi:10.1016/j.cortex.2016.02.006

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu Rev Neurosci, 35*, 73-89. doi:10.1146/annurev-neuro-062111-150525

Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, T., Aronen, H. J., & Soininen,

H. (2005). Distinct and overlapping fMRI activation networks for processing of novel identities and locations of objects. *Eur J Neurosci, 22*(8), 2095-2105. doi:10.1111/j.1460-9568.2005.04380.x

Pihlajamaki, M., Tanila, H., Kononen, M., Hanninen, T., Hamalainen, A., Soininen, H., & Aronen, H. J. (2004). Visual presentation of novel objects and new spatial arrangements of objects differentially activates the medial temporal lobe subareas in humans. *Eur J Neurosci, 19*(7), 1939-1949. doi:10.1111/j.1460-9568.2004.03282.x

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Soc Cogn Affect Neurosci, 2*(1), 67-70. doi:10.1093/scan/nsm006

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci, 13*, 25-42. doi:10.1146/annurev.ne.13.030190.000325

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965-966.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol, 9*(4), e1000608. doi:10.1371/journal.pbio.1000608

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature neuroscience, 7*(1), 85-90.

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Res, 44*(19), 2301-2311. doi:10.1016/j.visres.2004.04.006

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat Neurosci, 2*(11), 1019-1025. doi:10.1038/14819

Risko, E. F., Richardson, D. C., & Kingstone, A. (2016). Breaking the fourth wall of cognitive science: Real-world social attention and the dual function of gaze. *Current Directions in Psychological Science, 25*(1), 70-74.

Rosenbaum, R. S., Gao, F., Richards, B., Black, S. E., & Moscovitch, M. (2005). "Where to?" remote memory for spatial relations and landmark identity in former taxi drivers with Alzheimer's disease and encephalitis. *J Cogn Neurosci, 17*(3), 446-462. doi:10.1162/0898929053279496

Ross, M. G., & Oliva, A. (2010). Estimating perception of scene layout properties from global image features. *Journal of Vision, 10*(1). doi:Artn 210.1167/10.1.2

Rutishauser, U., Walther, D., Koch, C., & Perona, P. (2004). *Is bottom-up attention useful for object recognition?* Paper presented at the Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on.

Safer, M. A., Christianson, S. A., Autry, M. W., & Osterlund, K. (1998). Tunnel memory for traumatic events. *Applied Cognitive Psychology, 12*(2), 99-117. doi:Doi 10.1002/(Sici)1099-0720(199804)12:2<99::Aid-Acp509>3.0.Co;2-7

Salvato, G., Patai, E. Z., McCloud, T., & Nobre, A. C. (2016). Apolipoprotein ε4 breaks the association between declarative long-term memory and memory-based orienting of spatial attention in middle-aged individuals. *Cortex, 82*, 206-216.

Santangelo, V., Di Francesco, S. A., Mastroberardino, S., & Macaluso, E. (2015). Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory. *Hum Brain Mapp, 36*(12), 5003-5017. doi:10.1002/hbm.22984

Sastyin, G., Niimi, R., & Yokosawa, K. (2015). Does object view influence the scene consistency effect? *Atten Percept Psychophys, 77*(3), 856-866. doi:10.3758/s13414-014-0817-x

Sayres, R., & Grill-Spector, K. (2006). Object-selective cortex exhibits performance-independent repetition suppression. *J Neurophysiol, 95*(2), 995-1007. doi:10.1152/jn.00500.2005

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science*, *5*(4), 195-200.

Scoville, W. M., B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry, 20*(1), 11-21.

Seidl, K. N., Peelen, M. V., & Kastner, S. (2012). Neural evidence for distracter suppression during visual search in real-world scenes. *J Neurosci, 32*(34), 11812-11819. doi:10.1523/JNEUROSCI.1693-12.2012

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cereb Cortex, 14*(12), 1346-1357. doi:10.1093/cercor/bhh095

Shine, J. P., Hodgetts, C. J., Postans, M., Lawrence, A. D., & Graham, K. S. (2015). APOE-epsilon4 selectively modulates posteromedial cortex activity during scene perception and short-term memory in young healthy adults. *Sci Rep, 5*, 16322. doi:10.1038/srep16322

Shore, D. I., & Klein, R. M. (2000). The effects of scene inversion on change blindness. *J Gen Psychol, 127*(1), 27-43. doi:10.1080/00221300009598569

Simhi, N., & Yovel, G. (2016). The contribution of the body and motion to whole person recognition. *Vision research, 122*, 12-20.

Singh-Curry, V., & Husain, M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia, 47*(6), 1434-1448. doi:10.1016/j.neuropsychologia.2008.11.033

Smilek, D., Birmingham, E., Cameron, D., Bischof, W., & Kingstone, A. (2006). Cognitive Ethology and exploring attention in real-world scenes. *Brain Res, 1080*(1), 101-119. doi:10.1016/j.brainres.2005.12.090

Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron, 35*(6), 1157-1165.

Steck, S. D., & Mallot, H. A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence: Teleoperators and Virtual Environments, 9*(1), 69-83.

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J Cogn Neurosci, 16*(6), 955-965. doi:10.1162/0898929041502715

Sulpizio, V., Boccia, M., Guariglia, C., & Galati, G. (2016). Functional connectivity between posterior hippocampus and retrosplenial complex predicts individual differences in navigational ability. *Hippocampus, 26*(7), 841-847. doi:10.1002/hipo.22592

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci, 11*(9), 1004-1006. doi:10.1038/nn.2163

Summerfield, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Front Hum Neurosci, 5*, 67. doi:10.3389/fnhum.2011.00067

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1-2), 1-20.

Tolman, E. C. (1948). Cognitive maps in rats and men.

Tommasi L., Chiandetti C., Pecchia T., Sovrano V. A., Vallortigara G. (2012). From natural geometry to spatial cognition. Neurosci. Biobehav. Rev. 36, 799–82410.1016/j.neubiorev.2011.12.007

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron, 21*(6), 1409-1422.

Tootell, R. B., Hadjikhani, N. K., Mendola, J. D., Marrett, S., & Dale, A. M. (1998). From retinotopy to recognition: fMRI in human visual cortex. *Trends Cogn Sci, 2*(5), 174-183.

Torralba, A. (2009). How many pixels make an image? *Vis Neurosci, 26*(1), 123-131. doi:10.1017/S0952523808080930

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14*(3), 391-412.

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. doi:10.1037/0033-295x.113.4.766

Treisman, A. (1998). Feature binding, attention and object perception. *Philos Trans R Soc Lond B Biol Sci, 353*(1373), 1295-1306. doi:10.1098/rstb.1998.0284

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cereb Cortex, 24*(4), 883-897. doi:10.1093/cercor/bhs364

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol, 4*(2), 157-165.

van Koningsbruggen, M. G., Peelen, M. V., & Downing, P. E. (2013). A causal role for the extrastriate body area in detecting people in real-world scenes. *J Neurosci, 33*(16), 7003-7010. doi:10.1523/JNEUROSCI.2853-12.2013

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30*(6), 655-668.

Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *J Neurosci, 33*(14), 6133-6142. doi:10.1523/JNEUROSCI.3873-12.2013

Vass, L. K., & Epstein, R. A. (2016). Common Neural Representations for Visually Guided Reorientation and Spatial Imagery. *Cereb Cortex*. doi:10.1093/cercor/bhv343

Vlcek, K., & Laczo, J. (2014). Neural correlates of spatial navigation changes in mild cognitive impairment and Alzheimer's disease. *Front Behav Neurosci, 8*, 89. doi:10.3389/fnbeh.2014.00089

Vo, M. L., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *J Vis, 10*(3), 14 11-13. doi:10.1167/10.3.14

Võ, M. L. H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742.

Voss, J. L., Hauner, K. K., & Paller, K. A. (2009). Establishing a relationship between activity reduction in human perirhinal cortex and priming. *Hippocampus, 19*(9), 773-778. doi:10.1002/hipo.20608

Vuilleumier, P. (2000). Faces call for attention: evidence from patients with visual extinction. *Neuropsychologia, 38*(5), 693-700.

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. doi:10.1523/Jneurosci.0559-09.2009

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc Natl Acad Sci U S A, 108*(23), 9661-9666. doi:10.1073/pnas.1015666108

Wang, Q., Sporns, O., & Burkhalter, A. (2012). Network analysis of corticocortical connections reveals ventral and dorsal processing streams in mouse visual cortex. *J Neurosci, 32*(13), 4386-4399. doi:10.1523/JNEUROSCI.6063-11.2012

Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage, 124*(Pt A), 107-117. doi:10.1016/j.neuroimage.2015.08.058

Weiner, K. S., Sayres, R., Vinberg, J., & Grill-Spector, K. (2010). fMRI-adaptation and category selectivity in human ventral temporal cortex: regional differences across time scales. *Journal of neurophysiology*, *103*(6), 3349-3365.

Weissman, D. H., & Woldorff, M. G. (2005). Hemispheric asymmetries for different components of global/local attention occur in distinct temporo-parietal loci. *Cereb Cortex, 15*(6), 870-876. doi:10.1093/cercor/bhh187

Wig, G. S., Grafton, S. T., Demos, K. E., & Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nat Neurosci, 8*(9), 1228-1233. doi:10.1038/nn1515

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol, 8*(2), 227-233.

Wilcox, T., & Biondi, M. (2015). Functional Activation in the Ventral Object Processing Pathway during the First Year. *Front Syst Neurosci, 9*, 180. doi:10.3389/fnsys.2015.00180

Wojciulik, E., Husain, M., Clarke, K., & Driver, J. (2001). Spatial working memory deficit in unilateral neglect. *Neuropsychologia, 39*(4), 390-396.

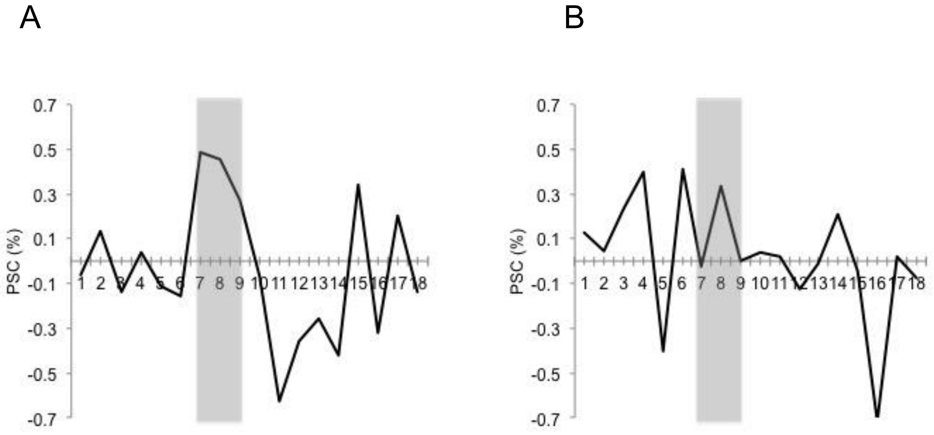
Xu, Y., Liu, J., & Kanwisher, N. (2005). The M170 is selective for faces, not for expertise. *Neuropsychologia, 43*(4), 588-597. doi:10.1016/j.neuropsychologia.2004.07.016

Yovel, G., & Kanwisher, N. (2004). Face perception: domain specific, not process specific. *Neuron, 44*(5), 889-898. doi:10.1016/j.neuron.2004.11.018

Zachariou, V., Klatzky, R., & Behrmann, M. (2014). Ventral and dorsal visual stream contributions to the perception of object shape and object location. *J Cogn Neurosci, 26*(1), 189-209. doi:10.1162/jocn_a_00475

Zaehle, T., Jordan, K., Wustenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Res, 1137*(1), 92-103. doi:10.1016/j.brainres.2006.12.044

Appendix A



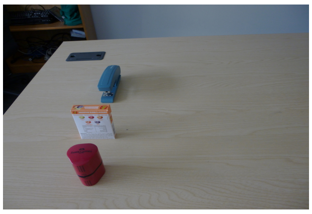
Example of a HRF (A) that did pass quality control, and a HRF (B) that did not. Runs which were deemed not to be of satisfactory quality were removed from the analysis for that particular ROI. Note that the graph includes data collapsed across conditions to ensure that it was not possible to infer which conditions were driving effects observed.

Appendix B

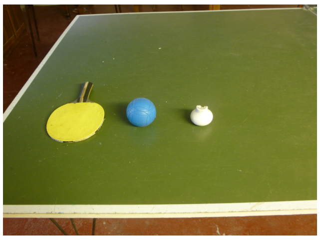















Example of a ‘base’ image for each scene family (n=19) used for both the fMRI experiments.