An investigation of factual and counterfactual feedback information in early visual cortex

Angus T. Paton, B.Sc., MA, MRes.

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy.

October 2018.

Containing studies performed in the Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow G12 8QB.

Abstract summary

Primary visual cortex receives approximately 90% of the input to the retina, however this only accounts for around 5% of the input to V1 (Muckli, 2010). The majority of the input to V1 is in fact from other cortical and sub-cortical parts of the brain that arrive there via lateral and feedback pathways. It is therefore critical to our knowledge of visual perception to understand how these feedback responses influence visual processing.

The aim of this thesis is to investigate different sources of non-visual feedback to early visual cortex. To do this we use a combination of an occlusion paradigm, derived from F. W. Smith and Muckli (2010), and functional magnetic resonance imagining. Occlusion offers us a method to inhibit the feedforward flow of information to the retina from a specific part of the visual field. By inhibiting the feedforward information we exploit the highly precise retinotopic organisation of visual cortex by rendering a corresponding patch of cortex free of feedforward input. From this isolated patch of cortex we can ask questions about the information content of purely feedback information.

In **Chapter 3** we investigated whether or not information about valance was present in non-stimulated early visual cortex. We constructed a 900 image set that contained an equal number of images for neutral, positive and negative valance across animal, food and plant categories. We used an m-sequence design to allow us to present image set within a standard period of time for fMRI. We were concerned about low-level image properties being a potential confound, so a large image set would allow us to average out these low-level properties. We occluded the lower-right quadrant of each image and presented each image only once to our subjects. The image set was rated for valance and arousal after fMRI so that individual subjectivity could be accounted for. We used multivariate pattern analysis (MVPA) to decode pairs of neutral, positive and negative valance. We found that in both stimulated and non-stimulated V1, V2 and V3, and the amygdala and pulvinar only information about negative valance could be decoded. In a second analysis we again used MVPA to cross-decode between pairs of valance

and category. By training the classifier on pairs of valance that each contained two categories, we could ask the question of whether the classifier generalises to the left out category for the same pair of valance. We found that valance does generalise across category in both stimulated and non-stimulated cortex, and in the amygdala and pulvinar. These results demonstrate that information about valance, particularly negative valance, is represented in low level visual areas and is generalisable across animal, food and plant categories.

In **Chapter 4** we explored the retinotopic organisation of object and scene sound responses in non-stimulated early visual cortex. We embedded a repeating object sound (axe chopping or motor starting) in to a scene sound (blizzard wind or forest) and used MVPA to read out object or scene information from non-stimulated early visual cortex. We found that object sounds were decodable in the fovea and scene sounds were decodable in the periphery. This finding demonstrates that auditory feedback to visual cortex has an eccentricity bias corresponding to the functional role involved. We suggest that object information feeds back to the fovea for fine-scaled discrimination whereas abstract information feeds back to the periphery to provide a modulatory contextual template for vision.

In a second experiment in **Chapter 4** we further explored the similarity between categorical representations between sound and video stimuli in non-stimulated early visual cortex. We use video stimuli and separate the audio and visual parts in to unimodal stimuli. We occlude the bottom right quadrant of the videos and use MVPA to cross-decode between sounds and videos (and vice-versa) from responses in occluded cortex. We find that a classifier trained on one modality can decode the other in occluded cortex. This finding tells us that there is an overlap in the neural representation of aural and visual stimuli in early visual cortex.

In **Chapter 5** we probe the internal thought processes of subjects after occluding a short video sequence. We use a priming sequence to generate predictions as subjects are asked to imagine how events from a video unfold during occlusion. We then probe these predictions with a series of test frames corresponding to points in time, either close in time to the offset of the video, just before the video would be expected to reappear, the matching frame from when the video would be expected to reappear or a frame from the very distant future. In an adaption

paradigm we find that predictions best match the test frames around the point in time that subjects expect the video to reappear. The test frame from a point close in time to the offset of the video was rarely a match. This tells us that the predictions that subjects make are not related to the offset of the priming sequence but represent a future state of the world that they have not seen. In a second control experiment we show that these predictions are absent when the priming sequence is randomised, and that predictions take between 600ms and 1200ms to fully develop. These findings demonstrate the dynamic flexibility of internal models, that information about these predictions can be read out in early visual cortex and that stronger representations form if given additional time.

In **Chapter 6** we again probe at internal dynamic predictions by using virtual navigation paradigm. We use virtual reality to train subjects in a new environment where they can build strong representations of four categorical rooms (kitchen, bedroom, office and game room). Later in fMRI we provide subjects with a direction cue and a starting room and ask them to predict the upcoming room by combining the information. The starting room is shown as a short video clip with the bottom right quadrant occluded. During the video sequence of the starting room, we find that we can read out information about the future room from non-stimulated early visual cortex. In a second control experiment, when we remove the direction cue information about the future room can no longer be decoded. This finding demonstrates that dynamic predictions about the immediate future are present in early visual cortex during simultaneous visual stimulation and that we can read out these predictions with 3T fMRI.

These findings increase our knowledge about the types of non-visual information available to early visual cortical areas and provide insight in to the influence they have on vision. These results lend support to the idea that early visual areas may act as a blackboard for read and write operations for communication around the brain (Muckli et al., 2015; Mumford, 1991; Murray et al., 2016; Roelfsema & de Lange, 2016; Williams et al., 2008). Current models of predictive coding will need to be updated to account for the brains ability to switch between two different processing streams, one that is factual and related to an external stimulus and one that is stimulus independent and internal

Contents

List of Figures

List of tables

List of publications

Publications

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Phil. Trans. R. Soc. B*, *372*(1714), 20160104.

Paton, A. T., Petro, L. S., & Muckli, L. (in preparation). Auditory feedback shows eccentricity biases and overlap in neural representation with well-matched congruent visual exemplars.

Paton, A. T., Lazarova, Y., Bennet, M., Petro, L. S., & Muckli, L. (in preparation). Reading out the future from feedback responses in extrastriate cortex during visual navigation with fMRI.

Paton, A. T., Cha, J., Petro, L. S., & Muckli, L. (in preparation). Information about non-social negative valance objects is present in non-stimulated early visual cortex.

Petro, L. S., Paton, A. T., Lazarova, Y. & Muckli, L. (in preparation). Future, unseen representation of the world in extrastriate cortex during occlusion.

Oral presentations

Paton, A. T., Petro, L. S., & Muckli, L. (2016). *Investigating Auditory Feedback in Early Visual Areas*. Cortical Feedback Spring School, Jena, Germany.

Poster presentations

Paton, A. T., Petro, L. S., & Muckli, L. (2015). *Investigating Sound Content in Early Visual Cortex*. Poster session presented at European Conference on Visual Perception, Liverpool, England.

Paton, A. T., Petro, L. S., & Muckli, L. (2016). *Investigating Sound Content in Early Visual Areas*. Poster session presented at Vision Science Society, Florida, USA.

Paton, A. T., Petro, L. S., & Muckli, L. (2016). *An Investigation of Sound Content in Early Visual Areas*. Poster session presented at Organisation for Human Brain Mapping, Geneva, Switzerland.

Paton, A. T., Petro, L. S., & Muckli, L. (2017). *An Investigation of Sound Content in Early Visual Cortex*. Poster session presented at Alpine Brain Imaging Meeting, Champéry, Switzerland.

Paton, A. T., Petro, L. S., & Muckli, L. (2017). *An Investigation of Sound Content in Early Visual Areas – Is auditory feedback retinotopic*? Poster session presented at Organisation for Human Brain Mapping, Vancouver, Canada.

Paton, A. T., Petro, L. S., & Muckli, L. (2018). *Decoding visual feedback of pleasant and unpleasant images in V1 with fMRI.* Poster session presented at The Consortium of European Research on Emotion (CERE), Glasgow, Scotland.

Definitions and abbreviations

aMPFC   Anterior medial prefrontal cortex

BOLD   Blood oxygen level dependent

CI   Confidence interval

DNN   Deep neural network

DTI   Diffusion tensor imaging

FB   Feedback

FF   Feedforward

fMRI   Functional magnetic resonance imaging

GLM   General linear model

HF   Hippocampal formation

LGN   Lateral geniculate nucleus

LOC   Lateral occipital cortex

MRI   Magnetic resonance imaging

MTL   Medial temporal lobe

MVPA   Multivariate pattern analysis

PCC   Posterior cingulate cortex

pIPL   Posterior inferior parietal lobule

PHC   Parahippocampal cortex

PRC   Perirhinal cortex

pRF   Population receptive field

RF   Receptive field

ROI	Region of Interest
RSP	Retrosplenial cortex
SD	Standard deviation
SVM	Support vector machine
TMS	Transcranial magnetic stimulation
V1	Primary visual cortex
V2	Secondary visual cortex
V3	Tertiary visual cortex
vMPFC	Ventral medial prefrontal cortex

Acknowledgements

I would like to thank my loving wife Gemma Wild for supporting me with my "science nonsense." Thank you for being tolerant all of those times when I spoke <u>at</u> you at length while you pretended to listen. Also for listening to numerous presentations, giving feedback, and being supportive while I was on my travels presenting my work at various conferences around the world.

I would also like to thank my parents Angus & Denise Paton who without their support I wouldn't have made it this far.

Thank you to my supervisor Lars Muckli for giving me this opportunity. He doesn't know this but when I applied for the position, I was so relaxed during the interview process because I was convinced I wasn't of the right calibre so had no hope and nothing to lose. I am so grateful to you for putting your trust in me and I hope the work in this thesis goes some way to paying that back.

Finally I would like to thank the lab past and present – Fiona, Gemma, Johanna, Lucy, Matt, Michele, Tyler, Yulia R and Yulia L – for your support, friendship and insightful conversation that has helped me to grow over the past four years.

Welcome to the world, Jackson.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any degree at the University of Glasgow or any other institution.

Angus T. Paton

Chapter 1

1 General introduction

How humans perceive the world has long been a topic of debate amongst scientists and philosophers for centuries. At the beginning of the modern study of perception, it was thought that vision began with a stimulus which forms an image on the retina, and through a cascade of forward brain processes a visual percept was formed of that stimulus (Fechner, 1966). The difficulty for vision scientists of the time was that visual information enters a "black box" where the output is a self-generated report of the resulting visual percept. It was only through subsequent post-mortem analysis that inferences about the underlying mechanisms of the brain and their relationship with behaviour could be made. In rare circumstances an individual would survive catastrophic injury to the brain and inferences could be drawn from pre- and post-injury reports, as shown in the famous case of Phineas Gage (Harlow, 1993). Today, modern brain imaging techniques such as functional magnetic resonance imaging (fMRI) allow us to look inside the "black box" and ask questions not only about how humans perceive the world but how vision is achieved at different stages of processing as it is actually happening.

How the brain forms the visual percept is now framed as its ability to solve the inverse problem of vision (Pizlo, 2001). The inverse problem is related to the large amount of information about the relationships of three dimensional objects that is not present in the two dimensional retinal images. To solve this inverse problem the brain must make inferences about these relationships. It does so by using information it has learned and retained about the world. This information is used in a predictive manner, which we call feedback, for hypothesis testing information as it enters and moves through the visual system. Discovering what information the brain stores and how it uses this information to achieve vision is critical to our understanding of the brain.

In this chapter we will present, in brief, how the visual system is organised in the brain and how this organisation gives us clues as to the mechanisms underlying

1

perception. We will then discuss the theoretical frameworks that have been constructed around these observations. We follow this with an overview of the many roles *feedback* plays in visual perception, including discussions on feedback from both visual areas and non-visual areas. Finally, we will introduce the conceptual idea of counterfactual processing before outlining the rationale of this thesis.

1.1   Organisation of the visual system

The visual system of the brain is organised in a hierarchy beginning with the encoding of light information at the retina. The retina is an extension of the brain that contains layers of neurons that convert photons in to nerve impulses through a cascade of chemical processes. Information propagates along the optic nerve before reaching the optic chiasm where the left retinal image is diverted to the left hemisphere and the right retinal image is diverted to the right hemisphere (**Figure 1.1**). In both hemispheres, retinal information reaches the respective left and right superior colliculus and lateral geniculate nucleus where some initial processing begins.
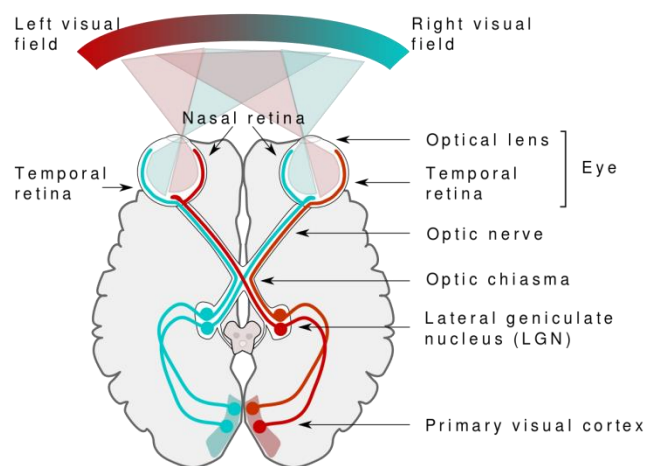
Figure 1.1: A schematic of the left and right retinal images being diverted to the left and right hemispheres of the brain.

The first step in cortical visual processing is primary visual cortex. Primary visual cortex, or striate cortex (V1), gains its name from the visible line of Gennari that represents the myelinated axonal input from the LGN. Neurons in V1 have very small receptive fields estimated at a single degree of visual angle (Levitt & Lund, 2002). V1 neurons are tuned to low-level properties such as orientation, spatial frequency, contrast and ocular dominance (Hubel & Wiesel, 1959). Neurons with similar tuning properties are typically found together forming cortical columns (Hubel & Wiesel, 1968). As a result of these neuronal properties V1 contains a highly precise spatial mapping of the visual field. The lower part of the visual field is mapped inverted to the upper bank of the calcarine sulcus while the upper part of the visual field is mapped inverted to the lower bank of the calcarine sulcus. Due to the retina containing a fovea much of V1 represents the centre of the visual field with less of V1 representing the periphery.

After information is processed at the level of V1, information is next processed at secondary visual cortex, or prestriate cortex (V2, typically referred to as part of extrastriate cortex). V2 is where visual processing begins to diverge in to the dorsal stream for visual-motor processing and the ventral stream for object and form representation (Goodale & Milner, 1992). The receptive field size of neurons in V2 also increases relative to V1. Extrastriate cortex is the next stage in the visual hierarchy which is comprised of V3, V4 and V5. Again, receptive field size increases along the hierarchy with neurons receiving input from more of the visual field. At the level of V5 specific functions begin to emerge such as visual motion and direction selective cells (Zeki, 2015).

Each level of the hierarchy is building on the one below it until the final percept is achieved. The flow of information so far described is a traditional view of visual processing within the context of the classical receptive field model. It is referred to here as being *feedforward* and is typically described as being "driving" such that information is processed at one level of the hierarchy before being processed at the next level. An example of how driving feedforward information moves along the visual hierarchy can be seen in **Figure 1.2** (adapted from Larkum (2013)).
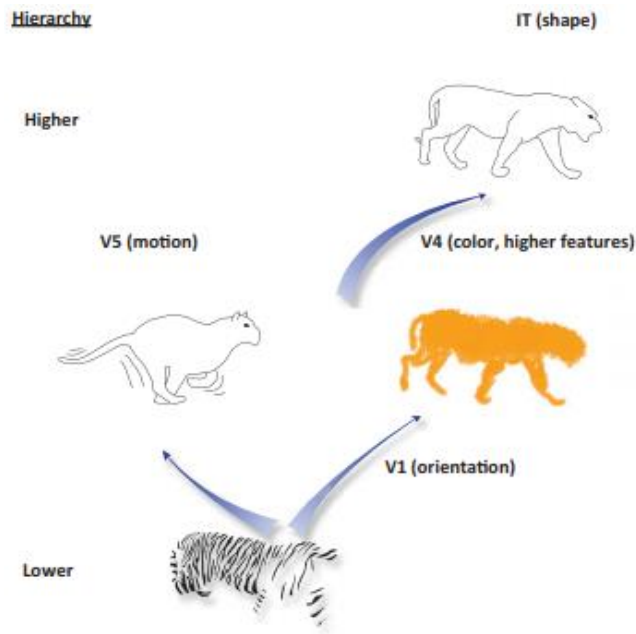
Figure 1.2: A schematic of the feedforward processing steps at different levels of the visual hierarchy.

In this feedforward model of visual processing, low-level features are processed first at the lowest level of the hierarchy. At each subsequent level, more and more complex features are combined until the full image is constructed at the highest level and the visual percept finally reaches consciousness. Feedforward computational models are a good approximation of visual processing as demonstrated by a feedforward deep neural network (DNN) (Güçlü & van Gerven, 2015). However, the problem with feedforward computational models is that they do not reflect the connectivity and dynamics of the human brain.

It has been observed that the tuning properties of V1 neurons can change over time to respond to more global image properties (Lamme & Roelfsema, 2000). This is due to the influence of top-down and lateral processes. The classic receptive field is bounded by a near and far receptive field. Angelucci and Bressloff (2006) suggest that the classic receptive field is driven by input from the LGN with lateral connections extending in to the near surround, whereas the far surround of the receptive field is influenced by top-down processes. This integrated model is referred to as the non-classical receptive field model and demonstrates how top-down information, which we call feedback, can modulate the response to stimuli in a recurrent network.

4

1.2   The recurrent network and prediction

Each of our eyes produces a retinal image. These two images are then combined in to a single image within the brain. However, there are an infinite number of solutions to explain the relationships between the objects contained within the two retinal images. This is an example of an inverse problem. The brain must find a solution to the information collected by the eyes by making inferences from what it already knows about the world. The brain's knowledge of the world is learned from experience and familiarity. To solve the inverse problem the brain makes use of feedback pathways that allow high-level information to guide low-level sensory processing. If we take V1 as an example, V1 has ten times as many feedback connections from V2 as it has feedforward connections from the LGN (Muckli & Petro, 2013). Brain areas further along in the hierarchy send information down the visual hierarchy in effect informing low-level areas with high-level information. As neurons in higher cortical areas have larger receptive fields, low-level receptive fields are being modulated by information from much larger parts of the visual field, as described above in the non-classical receptive field model.

The leading theory to explain how the brain solves the inverse problem of perception is known as *predictive coding* (Rao & Ballard, 1999). One of the first recurrent models to be proposed combining both feedforward and feedback processes to explain vision was put forward by Rao and Ballard in 1999, which was building on previous work by David Mumford. Mumford had proposed the idea that the brain learns statistical regularities that it uses to make sense of sensory input. The brain generates a number of hypotheses which it sends to the thalamus, which in turn chooses the most likely hypothesis as being correct. This information is then broadcast throughout the cortex (Mumford, 1991). Mumford also introduced the idea of reciprocal cortical loops of higher areas communicating abstract information with lower areas, the latter of which deals with more concrete information (Mumford, 1992). Rao and Ballard (1999) built on these concepts introduced by Mumford to explain vision. Their model included the concept of high level predictions informing low level areas. The main difference with this new

5

model was that hypothesis testing was done at each level of the hierarchy and the output of this was a residual error signal. If the error signal was small the prediction was good. If the error signal was large the prediction could be then be updated. This new model was attractive as it also explained certain visual response properties such as 'end-stopping' (Hubel & Wiesel, 1965) and other non-classical receptive field effects (Angelucci & Bressloff, 2006).

One of the limitations of the Rao and Ballard (1999) model however was that it was a linear model which makes it restrictive in its application. A revised model for visual processing was then proposed by T. S. Lee and Mumford (2003) that introduced hierarchical Bayesian inference to the predictive model. Conceptually this revision to the model introduces top-down priors or beliefs that "explain-away" the low-level input, in effect biasing or reinforcing inferences towards these beliefs. This model is similar to the previous model but instead of the sensory input being converted to an error signal, there are alternate layers of error detection and prediction. Another difference between the two models is that the parameters in the model represent the probability distribution of the sample rather than represent the sample itself. The alternating layers of error detection and prediction act to create an inference process that estimates and updates the mean and variance of the probability distribution of the sample, which itself is assumed to be normally distributed.

Hierarchical Bayesian models of predictive coding have been shown to be in good correspondence with the anatomical organisation of the neocortex (Friston, 2003, 2008). However these models were criticised for not being a neuronal model. Hierarchical Bayesian models demonstrate how the computations may work but not how the brain implements them. George and Hawkins (2009) approached this criticism by introducing the Hierarchical Temporal Model. This model approximates how the neocortex generates inferences on to anatomy by way of nodes in a tree-like structure. Using this structure the neocortex builds a model of the world using a spatio-temporal hierarchy. Spatial patterns are stored in the higher nodes, termed internal models (Clark, 2013), which represent larger amounts of space and longer periods of time. George and Hawkins (2009) speculate that this organisation is optimal for learning and generalisation as it reflects the spatio-temporal structure of causes in the external world. But why would the neocortex be

6

organised in this way? This remains an open question, however Friston (2010) suggests that the unifying principle is that the role of the brain is to minimise its free energy; to optimise its computations between alternating layers of error detection and prediction.

An example of how a predictive coding model might work can be seen in **Figure 1.3**. Predictive coding proposes that the brain contains internal models (Clark, 2013) about the world. In the example, the feedforward input is tested against the top-down prediction. The accuracy of the prediction (error signal) propagates through the system.
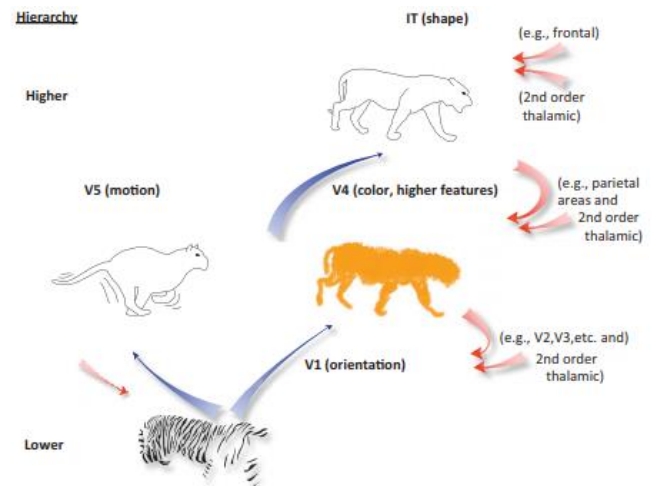
Figure 1.3: Feedback (red) carries predictions. These are compared with sensory input. An error signal (blue) carries information forward about the accuracy of the comparison between prediction and sensory input.

1.3 How feedback influences visual processing at the neuronal level

The visual system is part of the neocortex which contains both excitatory and inhibitory neurons and is characterised by six layers. The primary excitatory neuron in visual cortex is the pyramidal neuron. Pyramidal neurons have been found in all but the first cortical layer. The layer IV pyramidal neuron (**Figure 1.4**, adapted from Spruston (2008)) is important in the context of vision as it receives its synaptic connections, or bottom-up information, primarily from the thalamus. Top-down, or feedback information, arrives at the deep (V and VI) and superficial (I and II) layers.
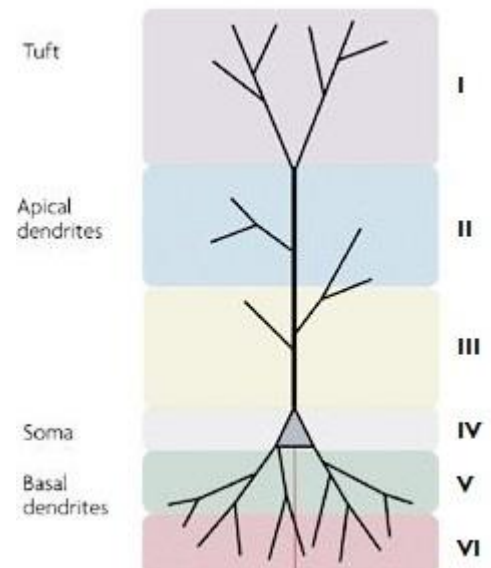
Figure 1.4: Schematic of a layer 4 pyramidal neuron.

7

How and when the neuron fires is dependent on how this information is integrated across the cell. Some cells may be tuned to fire based on coincident input to the tuft and peri-somatic areas (Larkum, 2013). This can be seen in detail in **Figure 1.5** in which Larkum (2013) proposes the backpropagation activated calcium (BAC) firing hypothesis. Pyramidal neurons change their firing properties when they receive both feedforward and feedback input simultaneously (**Figure 1.5 a**). Areas higher in the visual hierarchy can send predictions down the hierarchy to superficial tuft layers of cortex. This information is integrated with the feedforward somatic input at layer 4 and the cell output is modified. In this manner, feedback from higher areas can influence the output of lower-level areas (**Figure 1.5 b**).
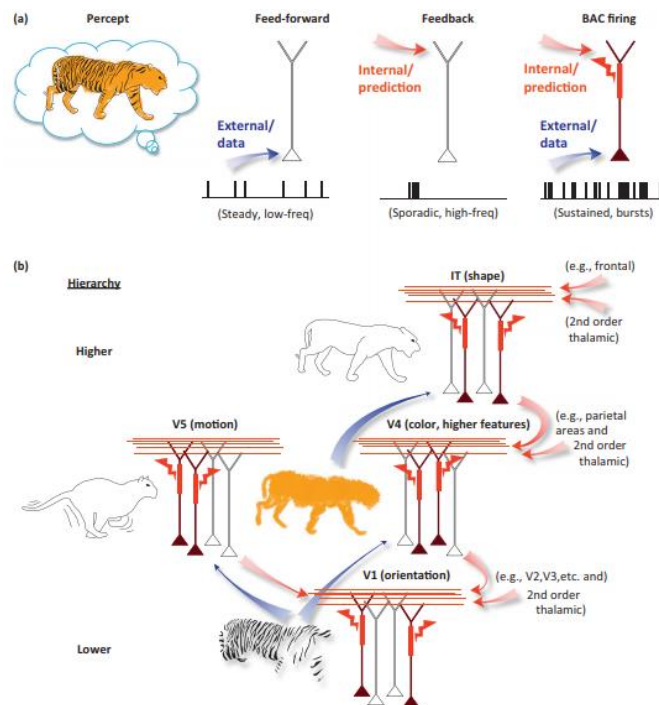
Figure 1.5: (a) Changes in neuronal firing when feedforward (blue) and feedback (red) are coincident. (b) Feedforward input (blue) arrives at the somatic layer and feedback input (red) arrives at the superficial layer.

Recent work in rodent models has revealed that within the set of level 5 pyramidal neurons in primary somatosensory cortex, a subset of these neurons is responsible for perceptual detection (Takahashi, Oertner, Hegemann, & Larkum, 2016). The authors of this paper found that the threshold for perception could be manipulated by increasing and decreasing the Ca^{2+} activity in the apical dendrites. This finding is suggestive of a neuronal mechanism that would allow feedback to modulate whether or not perception is detected.

1.4 The many influences of feedback on vision

The human brain relies on internal models to make inferences about the world. This raises important research questions. What is the information content of feedback responses? What role do they play at different levels of visual processing? One source of information that feedback provides to lower visual areas is that of context. This theory proposes that one role of feedback responses is to provide context for extracting behaviourally relevant information. This is important because the visual world can be ambiguous or noisy. As an example, consider the diagram in **Figure 1.6** (top). Here the centre Figure can be interpreted either as the number 13 if it is read from left to right, or the letter B if it is read from top to bottom. The interpretation of ambiguous objects can change based on the context we find them in.
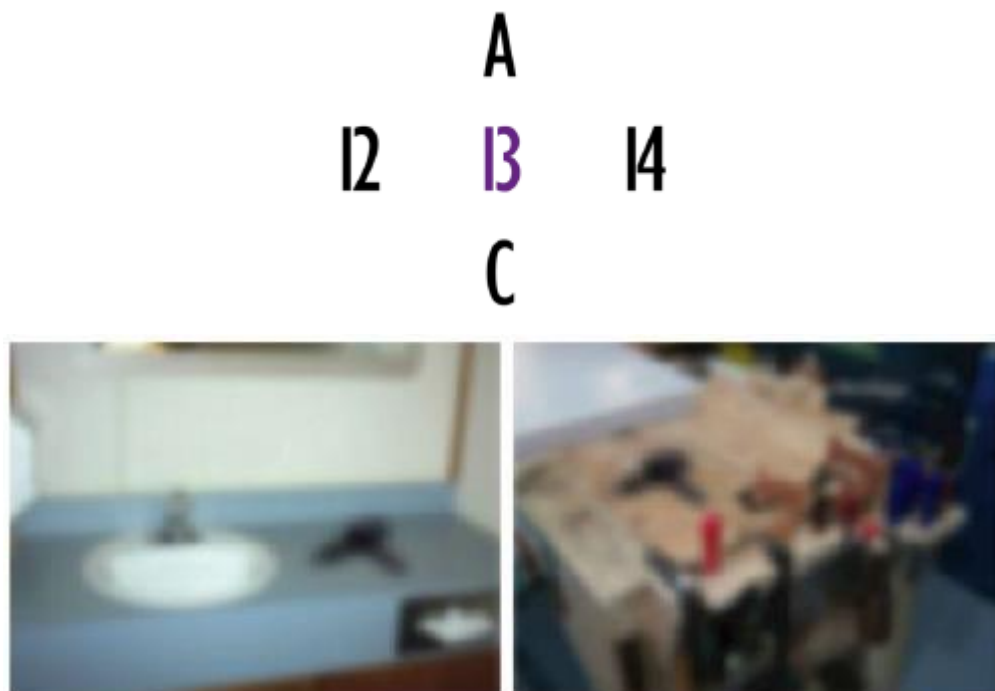
Figure 1.6: (Top) The importance of context can be seen in the interpretation of I3, whether it is read from left to right or from top to bottom. (Bottom) The same object is placed in two different scenes. Depending on the context, the interpretation of the object differs.

In an example by Bar (2004) the same object is placed in two different degraded visual scenes (**Figure 1.6, bottom**). In the left scene, the object is interpreted as a hairdryer as the context of the scene is a bathroom. However, in the right scene the context is now a tool bench and the interpretation of the object is that it is a power drill. The object itself never changes; the context of the surrounding scene shapes our interpretation of what it is.

Context is a useful tool not only for resolving ambiguities of degraded stimuli, but also useful in the case of occlusion. We encounter occlusion frequently in everyday life as objects pass behind other objects as we move around. Neurons in V1 are known to be active even when their receptive fields are occluded due to the influence of feedback (T. S. Lee & Nguyen, 2001). The content of feedback information can be investigated in the laboratory by using techniques such as blindfolding or by occluding part of the visual field. These methods either fully or partially block the feedforward sensory input. Due to the highly precise retinotopic mapping of the visual field in early visual cortex, occlusion results in an area of visual cortex that is cut-off from feedforward input. Therefore, any responses from this "occluded" patch of cortex are the result of top-down processes.

This method of partial occlusion has proven useful for investigating the role and information content of feedback responses. F. W. Smith and Muckli (2010) developed an occlusion paradigm by masking the bottom right quadrant of complex natural images with a white occluder. Using pattern analysis techniques, they reported that feedback responses were dissimilar for different categories of images in V1 and that these top-down representations of the stimuli could be reliably decoded. They further report that the feedback responses are similar to the feedforward responses when the occluder is removed. This study demonstrates that contextual feedback responses are influencing the feedforward sensory input by providing context that is similar to the missing part of the image. Follow-up studies have shown that contextual feedback information in V1 is found within the superficial layers of the cortical sheet and that the feedback patterns in these layers are tolerant to shifts of the surrounding scene by up to 2° of visual angle (Muckli et al., 2015). Feedback responses in V1 and V2 have also been shown to be generalisable across different categorical exemplars (Morgan, Petro, & Muckli, 2016). Taken together these studies provide evidence that feedback

10

information is categorical and abstract to some degree, allowing it to tolerate shifts in visual angle and to generalise between categorically matching exemplars.

1.4.1 Object tracking through occlusion by top-down processes

Despite objects being partially or even fully occluded we still retain information about those objects. This is demonstrated in occlusion studies with children as young as 4 months old who can learn to predict the re-emergence of a square in motion as it passes behind an occluder (Johnson, Amso, & Slemmer, 2003). This is due to top down processes tracking objects as they move through our visual field. This has been demonstrated in studies using the illusion of apparent motion (Edwards, Vetter, McGruer, Petro, & Muckli, 2017; Muckli, Kohler, Kriegeskorte, & Singer, 2005). In simple apparent motion studies, two white squares blink in an alternating pattern producing the illusion that a single square is moving across the screen. Simple apparent motion paradigms like this can be considered occlusion paradigms as one of the possible hypotheses to explain the illusion is that there is a black occluder hidden in the black background which the stimulus is moving behind. The apparent motion trace across the visual field can be detected in V1 despite the white squares never entering the receptive fields of neurons along the apparent motion path (Muckli et al., 2005). Even when attention is diverted away from the stimuli the apparent motion trace persists. This supports a strong top-down influence of feedback "filling-in" the expected motion. This is further evidenced by Edwards et al. (2017) who show that predictions made pre-saccade are updated to a new retinotopic location post-saccade demonstrating that predictions are highly dynamic.

In a second apparent motion paradigm, Chong, Familiar, and Shim (2016) combine apparent motion with apparent rotation. Two black and white gratings with orientations of 45°±10° and 135°±10° are flashed alternately. Subjects are cued prior that the direction of rotation is either clockwise or anticlockwise. This produces two apparent rotations along the apparent motion trace, either horizontal (90°) or vertical (0°). They find in V1 that the apparent rotation is reconstructed along the apparent motion trace. Critical to this study is that the rotated grating is

11

never seen, not even as an illusion. This is further support for highly dynamic top-down predictions that are tracked across the visual field. However the "filling-in" theory may be best explained by a mental simulation of the rotation facilitated by the flexibility of the internal model.

1.4.2 Figure ground segregation

An important role of vision is the ability to decipher from a cluttered environment what is an object and what is background. The brain achieves this by defining boundaries around objects to segregate them from background information. Neurons in V1 have very small classical receptive fields which only contain information about local features in the feedforward flow of information. To know anything about the global scene the brain must rely on information from feedback processes. Electrophysiology study of awake non-human primate V1 has shown that feedforward processes are sufficient for the purpose of boundary detection (Lamme, Rodriguez-Rodriguez, & Spekreijse, 1999). Boundary detection happens around 80ms post stimulus and is followed by a top-down process filling-in between the edges of the boundary (Lamme et al., 1999; Lamme & Roelfsema, 2000). This finding has been replicated in the human brain. In an electrophysiology and fMRI study by Scholte, Jolij, Fahrenfort, and Lamme (2008), the authors report that boundary detection occurs in early visual cortex around 92ms. This process then spreads towards parietal and temporal lobes. Surface segregation, or filling-in, then occurs around 112ms in temporal areas before spreading to parietal and then occipital areas. Work by Self, van Kerkoerle, Super, and Roelfsema (2013) has revealed how these feedforward and feedback processes occur in the different layers of the cortical sheet. In behaving monkeys, researchers recorded from all layers in V1 while the monkeys performed a Figure-ground segregation task. Recordings revealed that boundary detection occurs first in layer 4 followed by labelling of the Figure region in superficial and deep layers. Finally, region filling-in occurs in layer 1, layer 2 and layer 5 characteristic of feedback processing.

12

1.4.3 Expectation, Attention & Memory

Visual perception can be modulated by expectation. For example objects placed in congruent scenes are detected faster than objects placed in incongruent scenes (Bar, 2004). This is an example of visual expectation. Visual expectation facilitates perception by biasing inferences towards what has been seen before (Summerfield & Egner, 2009) and can shape how we perceive noisy and ambiguous stimuli (**Fig 1.6**). The strength of expectation on perception can be seen in a study by M. L. Smith, Gosselin, and Schyns (2012). In this study, white noise stimuli were presented to subjects with the knowledge that half of the stimuli would contain a low-resolution, noisy, ambiguous face. In fact none of the stimuli contained a face. By subtracting all the non-face stimuli from the self-reported face stimuli, what was revealed was a ghost-like face representation that subjects had projected on to the noise where no face existed. This revealed each subjects internal representation of what an expected face should look like. Visual expectations therefore facilitate perception by sharpening visual representations, as has also been demonstrated experimentally by Kok, Jehee, and De Lange (2012).

Temporal expectation can also influence visual processing prior to the onset of a stimulus. At the neuronal level, neural responses can demonstrate an expectation or preparatory signal before a stimulus is presented. Work by Sirotin and Das (2009) have shown that there are two components to the haemodynamic signal in V1 of non-human primates during a simple fixation task. There is one component that is predictable from neuronal responses to the stimulus and a second component that is entrained to the predicted trial onsets. The expectation of the appearance of a stimulus directs arterial blood to primary visual cortex in anticipation of stimulus onset. In a number of similar experiments, Sharma et al. (2014) show that expectation and attention are tightly interwoven as the temporal expectation signal was absent in the task that did not require explicit attention. Such findings has led some to propose that there is significant overlap of attention and expectation mechanisms (Summerfield & Egner, 2009).

Expectation and attention are also tightly interwoven with memory (Gazzaley & Nobre, 2012). Memory is the mechanism by which brain states are reinstated long after the visual input that created them has ended (Muckli & Petro, 2017). In a recent study by Van Kerkoerle, Self, and Roelfsema (2017), researchers demonstrated the influence of memory on V1 neurons in a behaving monkey. They used current-source density measures to investigate V1 neuronal responses to both attention and memory tasks. In a curve drawing task where the stimulus is shown briefly then hidden, they found increased activity in superficial and deep layers of the cortical sheet after the stimulus was hidden. Interestingly, when a mask was used to disrupt visual working memory the V1 memory activity temporarily disappeared before being reinstated by memory processes.

1.4.4 Reconstructing representations

Feedback to early visual cortex can also be useful for the purpose of reconstructing stimulus representations. This is demonstrated in a fMRI study by Williams et al. (2008) in which cubed, smoothed and spikey pairs of objects were presented in the far periphery of the visual field. Subjects were tasked with either identifying object differences or object category. Researchers found that information about the objects was fed back to the fovea. This finding could be attributed to the influence of attention, as attention is known to spread outwards in the visual field towards areas which do not contain a stimulus (Serences & Boynton, 2007). However in a third experiment investigating the spatial specificity of the information patterns across eccentricity, the authors show that feedback information only occurs in the fovea and not in any other retinotopic locations, such as unstimulated periphery, making their finding unlikely to be attributed to attention and more likely due to reinstating low-level features.

In this same study, behavioural performance steadily increased over time which was correlated with the emergence of category information in the fovea. The researchers suggest that the fovea may act as a blackboard to compute task-relevant visual information (Williams et al., 2008). This idea was also coined by

Mumford (1991) and has been growing in its acceptance not only for vision (Muckli et al., 2015; Roelfsema & de Lange, 2016) but also for other types of sensory input (Murray et al., 2016).

1.5   Non-visual feedback

Top-down input to V1 can originate from both visual and non-visual areas, such as auditory cortices. Within a predictive coding framework, aural input to auditory cortex can trigger rich internal models that feed back to visual cortex (L. S. Petro, Paton, & Muckli, 2017). At the level of the single neuron, V1 is modulated by auditory input. For example, in areas 18 and 19 of the cat visual cortex, 40% of neurons respond to sounds (Morrell, 1972). In mice, tones and bursts of noise invoke response in both superficial and deep layers of primary visual cortex (Iurilli et al., 2012). Interestingly, approximately 10% of level 2 and level 3 neurons in primary visual cortex of awake mice respond to sounds alone (Meijer, Montijn, Pennartz, & Lansink, 2017).

In humans, auditory input influences visual processing at the earliest level of the visual hierarchy. Bursts of noise activate primary visual cortices with hemodynamic response profiles that demonstrate facilitation for multisensory input when compared with unisensory input (Martuzzi et al., 2006). Such simple auditory stimuli have been shown to directly modulate perception. For example, when a single flash is paired with two 'beep' sounds this is perceived as two flashes (Shams, Kamitani, & Shimojo, 2000). Additionally, looming tones facilitate faster detection of perceived motion than either modality does in isolation (Cappe, Thut, Romei, & Murray, 2009). These examples demonstrate a direct modulation of perception at the earliest level of visual processing by simple auditory sounds. However, early visual cortex is also modulated by more complex auditory stimuli. In such cases, it is less clear what the information content actually is, and how this information might facilitate perception.

Early levels of visual processing can be influenced by much more complex sounds. In a series of experiments by Vetter, Smith, and Muckli (2014) it has been

shown that information about categorical sounds can be read out from feedback responses in early visual areas V1, V2 and V3 in healthy subjects who are blindfolded. These categorical response patterns were also shown to be similar amongst congruent categorical exemplars. This evidence provides further support for the theory that feedback responses triggered by complex sensory input are contextual in the earliest levels of visual processing. In this same study the authors also report that these feedback responses can be triggered by just imagining the sounds (Vetter et al., 2014). In a follow up study it was also shown that contextual feedback responses can be read out in non-blindfolded subjects (L. S. Petro et al., 2017). This demonstrates that uninformative feedforward input doesn't "wash-out" the feedback input. Taken together this would suggest that auditory triggered feedback plays a modulatory role on visual perception. This does raise an interesting question about whether the auditory influence is direct from auditory cortex or from high-order internal models. If it is the latter then this would suggest that feedback responses should be similar whether they are triggered by auditory or visual sensory input. However recent evidence does not support this idea (Jung, Larsen, & Walther, 2018).

It is unclear whether auditory information arrives at the encoding space of early visual cortex directly from auditory cortices, or indirectly from higher-order cortical areas involved in audiovisual integration (Koelewijn, Bronkhorst, & Theeuwes, 2010). Evidence for direct auditory cortico-cortical connections to early visual cortex comes from anatomical research. In the macaque brain, monosynaptic afferents from auditory association areas have been reported to terminate in macaque V1 and V2. Auditory afferents that terminate in striate and extrastriate of the macaque do so in layers 1 and 6 of the cortical sheet (Rockland & Ojima, 2003). These connections are also described as being most dense in extrastriate cortex compared to striate cortex, and similarly most dense in peripheral areas compared with fovea areas. Another interesting finding within the macaque model is that auditory afferents terminate close to V2 neurons that have direct feedback connections to V1. Taking both findings together shows a cortical circuit that allows for both a direct and an indirect influence of feedback on neurons in primary visual cortex (Laramée, Kurotani, Rockland, Bronchti, & Boire, 2011). If these connections exist in the human brain, they may act as an anatomical constraint

16

that explains why sounds are consistently found to be decodable in the periphery and to a lesser extent in the fovea.

While it is unclear how anatomical constraints affect feedback to visual cortex, there is support linking auditory cortex with visual cortex functionally in both sighted (Eckert et al., 2008) and blind (Klinge, Eippert, Röder, & Büchel, 2010; Qin, Xuan, Liu, Jiang, & Yu, 2014) individuals. These functional studies also report a bias towards the periphery (Eckert et al., 2008). Functional MRI studies also reveal visual cortex activation for sounds in sighted (Specht & Reul, 2003; Vetter et al., 2014) and blind (Ricciardi, Tozzi, Leo, & Pietrini, 2014). Transcranial magnetic stimulation (TMS) at subthreshold intensity induce the perception of phosphenes when the TMS stimulation is coupled with an auditory stimulus (Bolognini, Senna, Maravita, Pascual-Leone, & Merabet, 2010). The optimal site for stimulation was found to be in peripheral areas of cortex. Applying stimulation to the fovea did not induce perception of phosphenes (Bolognini et al., 2010). A second TMS study replicated this finding and related visual cortex excitability to auditory dimensions of pitch and bandwidth (Spierer, Manuel, Bueti, & Murray, 2013); although they don't explicitly state the effect is stronger in the periphery.

1.5.1 Counterfactual feedback

The brain makes use of extensive feedback connections to make sense of the feedforward sensory input it receives. This is happening all the time while we are engaged with our environment. In this thesis we make the distinction of referring to this type of feedback as being *factual feedback*. By using the term factual feedback, we refer to this type of feedback as being stimulus dependent. This
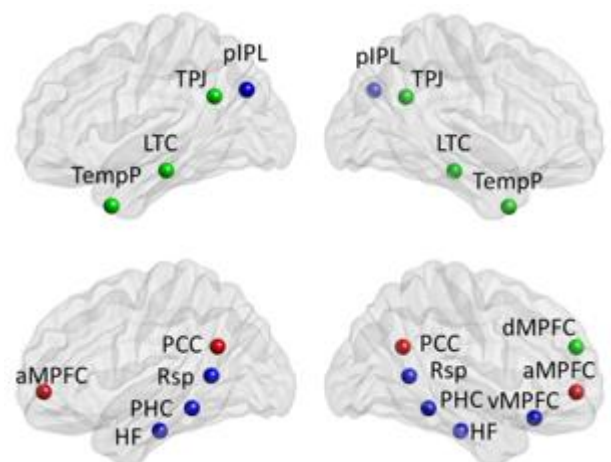
Figure 1.7: An overview of the default mode network.

17

distinction allows us to separate factual feedback from another kind of feedback we call *counterfactual feedback*. Counterfactual feedback relates to feedback that is stimulus independent.

When the brain is not actively engaged in an external task that requires attention, the mind is free to wander and engage in counterfactual thought. For example, when you are at the dentist for a check-up, while you sit in the dentist's chair you can plan what you are going to do after you leave your appointment; which route you are going to take home, whether you are going to stop at the store to pick up something for dinner, what you are going to cook, where in the store those ingredients are, the quickest path to collect them, etc. Counterfactual thinking encompasses a wide range of cognitive processes including mind-wandering (with- or without attention) (Schooler et al., 2011), past & future thought (Schacter, Benoit, & Szpunar, 2017), imagery (Pearson & Kosslyn, 2015) and mental simulation (Hamrick, Battaglia, Griffiths, & Tenenbaum, 2016).

A network of brain regions known as the default mode network is engaged during counterfactual thinking (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010) and this network is anti-correlated with the executive network at both rest (Fox et al., 2005) and during task (Hellyer et al., 2014). The default mode network consists of a midline core (**Figure 1.7**; red) that includes posterior cingulate cortex (PCC) and anterior medial prefrontal cortex (aMPFC). A medial temporal lobe (MTL) subsystem (**Figure 1.7;** blue) includes ventral medial prefrontal cortex (vMPFC), posterior inferior parietal lobule (pIPL), retrosplenial cortex (RSP), parahippocampal cortex (PHC) and hippocampal formation (HF) (Andrews-Hanna et al., 2010; Zhu, Zhu, Shen, Liao, & Yuan, 2017). The midline core is related to self-relevant affective decisions. The MTL subsystem is related to internally constructing mental scenes based on memories. However during future thought, both systems are engaged likely due to constructing mental images from episodic memory (Andrews-Hanna et al., 2010).

Engaging in counterfactual thinking is linked with a phenomenon known as perceptual decoupling (Schooler et al., 2011). Perceptual decoupling has been described to occur when a person engages in thought that is unrelated to the here and now, or when attention to the outside world waxes and wanes over time.

There are many physiological indicators that show when someone is engaging in counterfactual thought, or when a person is perceptually decoupled. For example, longer fixation durations during mindless reading (Franklin, Smallwood, & Schooler, 2011), erratic eye-movements (Reichle, Reineberg, & Schooler, 2010) or increased blinking before people catch themselves mind wandering (Smilek, Carriere, & Cheyne, 2010) and cessation of phasic pupil diameter changes normally observed during tasks to external stimuli (Smallwood et al., 2011). Electrophysiology studies report a reduction in several markers of attentiveness to external stimuli during perceptual decoupling and mind wandering. The P300 event-related potential (ERP) wave is driven by decision making and is found to be attenuated during perceptual decoupling (Smallwood, Beach, Schooler, & Handy, 2008). Similarly visual markers such as the P1 and N1 ERP amplitudes have been shown to also be attenuated (Kam et al., 2011) suggesting a reduction in processing of visual input. During a meditation task it was found that there is also a reduction in the faster alpha and beta frequency oscillations and an increase in slower theta and delta oscillations before people had caught themselves going off-task and engaging in mind wandering (Braboszcz & Delorme, 2011).

When we engage in counterfactual thinking, although our attention to the outside world decreases, the outside world does not go away – we don't suddenly become blind. This is an important concept because current predictive coding models do not account for the alternation between perceiving the external world and perceiving our own internal world. This leads to a number of important research questions. How does counterfactual feedback "take-over" from perception? As a very basic description, at the microscopic level, visual input arrives at the mid-layers of the cortical sheet. Factual feedback arrives at the deep and superficial layers. Where does counterfactual feedback arrive? How do counterfactual feedback and factual feedback co-exist? Incorporating counterfactual feedback in to current prediction models will be an important step in understanding how perception, and ultimately the brain, works.

1.6   Thesis rationale

Visual cortex, particularly primary visual cortex V1, is influenced by many different sources of non-visual input that originate from both cortical and sub-cortical sources (Muckli & Petro, 2013). It is important to understand what influences these non-visual types of feedback have on perception for informing our overall understanding of how vision works. Similarly, it is important to understand how and why low level visual areas help to shape our internal perceptions when the only building blocks the brain has access to are independent of any external stimulus, as is the case with counterfactual feedback.

To this end, the aim of this thesis was to study the influence of non-visual factual and counterfactual feedback to early areas of visual processing. To do this we use a combination of functional magnetic resonance imagining and multivariate pattern analysis (MVPA) (Haxby, 2012) with occlusion paradigms. FMRI is a non-invasive brain imagining technique that measures a blood-oxygen-level dependent (BOLD) signal in cubes (voxels) of tissue. MVPA is a method of analysis that compares patterns of activity across multiple voxels. This method of analysis is particularly useful with occlusion paradigms as we are blocking the feedforward input to neurons within each occluded voxel therefore these voxels contain non-spiking neurons. Changes between voxels are subtle and MVPA is sensitive to these changes. Occlusion paradigms allow us to either partially or fully occlude visual stimuli. Fully occluding an image is the same as not presenting it at all, however we can consider this as 'occlusion by omission', for example, by removing a sequence of frames from a video clip. By occluding part of the visual field we render a corresponding patch of cortex as being resting state. Therefore, the only influence on neurons within the set of occluded voxels is from feedback, or top-down processes.

In **Chapter 3** we were interested in whether or not non-social affective information is fed back to early visual areas. We used a path-guided de-bruijn m-sequence in a fast event-related design to present to subjects a 900 affective image data set were each image had the bottom-right quadrant occluded by an occluder matching the background screen colour. The images were from three categories (animals,

foods, plants) with 100 images for each category for neutral, positive and negative levels of valance. We chose this method as we wanted to know if we could decode affect information, specifically valance, independent of image category and low-level image statistics. After adjusting the images for individual subjectivity based on ratings of valance, we found a consistent story of being able to decode from both feedforward and feedback responses information about negative stimuli. We could not decode positive versus neutral in any region of interest. However we could consistently decode positive versus negative and neutral versus negative images. This tells us that when the classifier has information about positive (or neutral) valance and negative valance, it is the negative valance that is important for classification.

We then went a step further and performed cross-classification. Here we are including category information. We train the classifier to learn to distinguish between two levels of valance for two of the three categories. We then test the classifier on the left out category. We do this leaving one category out as the test category each time, before averaging. We find a similar pattern as before. Only from V2 feedback responses and in the amygdala could positive versus neutral be decoded.

In **Chapter 4** we investigated the peripheral bias associated with auditory feedback in visual cortex. We wanted to know whether different types of sound may feedback to visual cortex at different eccentricities. We hypothesised that the peripheral bias commonly observed with regards to auditory feedback to visual cortex could be related to abstract forms of information such as category, but that more concrete information, for example about objects, may feedback to foveal areas as is the case with visual objects presented in the periphery (Williams et al., 2008). To do this we presented two different scene sounds to subjects. During a scene sound, a brief, repeating object sound could also be played. This created four conditions which could be analysed to decode the scene information keeping the object information the same, or to decode object information keeping the scene information the same. We found that scene information was only decodable in the periphery whereas object information was only decodable in the fovea.

We were also interested in understanding what the information content was of feedback responses in early visual areas to highly-matching categorical sounds and videos. We hypothesised that if both modalities are triggering the same internal models then within a predictive coding framework the information content of feedback responses should also be similar. So far, this has not proven to be the case (Jung et al., 2018). To do this we chose three categorical videos containing dynamic objects and split the auditory and visual components in to unimodal stimuli. We presented audio and visual stimuli independently and alternately, and also occluded the bottom-right quadrant of each video with a white occluder. We found that we could decode the category of sound stimuli from feedback responses in visual cortex as well as category of videos from feedback responses in occluded visual cortex, replicating previous work (L. S. Petro et al., 2017; F. W. Smith & Muckli, 2010; Vetter et al., 2014). This finding adds evidence to the theory that complex sensory stimulation triggers a top-down modulatory influence on low-level processing. Our new finding was that a classifier trained on feedback responses to discriminate between sounds could decode feedback responses about videos (and vice versa). This finding tells us that there is some information in the feedback responses in early visual cortex that is common to both modalities. This finding fits well with predictive coding theories.

In **Chapter 5** we turn our attention to counterfactual feedback. In these experiments we are interested in how internal models can be triggered and then manipulated to make predictions about the immediate future. For example, before crossing the road you may ask yourself, "Can I cross to the other side before the car hits me?" To answer this question the brain needs to have knowledge and mechanisms that allow it to predict the future. To test this, we chose three categorical video clips, decomposed them in to individual frames and presented a sequence of them at 100ms per frame. We would show subjects a sequence of six frames followed by a 800ms blank interval before showing them a final test frame. The test frame could either be the matching frame that would be expected after 800ms, or a mismatching frame from 300ms, 600ms and 4000ms. This design was a fast event-related design and allowed us to test whether or not predictions generated by the image sequence were extrapolated into the immediate future during the blank interval. We could test whether these predictions were more

similar to the video immediately after onset (300ms condition), or similar to the frame that best matches (800ms condition), or something between (600ms condition). Using deconvolution general linear modelling we found that the frame closer in time, the 300ms condition, was never a good match to the prediction. Both the 600ms and 800ms test frames were better matches to the prediction of the brain's extrapolated future. This demonstrates that the brain is capable of extrapolating sensory input in to the immediate future to make predictions. These predictions are not accurate as the 600ms condition proved to be the most predictable test frame suggesting either a "best approximation" made by the brain, or temporal compression where predictions unfold faster in time in the brain than they do in the external world.

In a follow up control study we wanted to test whether the predictions we find in the first experiment were indeed predictions. We modified the paradigm to include a condition were the initial sequence of images was shuffled. This change would make the future unpredictable allowing us to compare predictable and unpredictable conditions. Due to time constraints we only used the 4000ms test frame. We hypothesised that regardless of whether or not the future is matched, we should still find a prediction in the sequential frames condition versus no prediction in the non-sequential frames condition. We also wanted to test whether predictions need more time to build-up, therefore we included two additional conditions were the timing was increased to 200ms per frame. We found that the expected versus unexpected conditions with 100ms timing were indistinguishable. However there was a significant difference in the expected versus unexpected conditions with 200ms timing. This control study demonstrates that what we are testing are predictions about the future based on previously seen information, and that these predictions require somewhere between 100ms and 200ms to 'build-up'.

In **Chapter 6** we were interested in exploring counterfactual processing that occurs concurrently with factual processing. In **Chapter 5** we tested our hypothesis about predictions of the future that are extrapolating over a period of time while subjects are viewing a black screen. There is no informative feedforward input as the prediction is developing. In this new study we wanted to induce counterfactual processing during periods of meaningful feedforward processing. This is an important distinction because it allows us to test whether we

23

can read out factual and counterfactual feedback that is happening at the same time. We also wanted to have control over which internal models were being used to make predictions about the future. In **Chapter 4** subjects had never seen the stimuli before so we could not be sure what information subjects were using to make predictions. To control for this we turned to virtual reality. We spent a year developing a virtual house that subjects could freely explore. We designed the house around a middle cube-shaped room. This design meant there would be four adjoining rooms with each room accessible from the middle room. The rooms were categorical and contained common objects that would normally be found within the context of each specific room (kitchen, bedroom, office, game room). Virtual reality allowed us to train our subjects to become familiar with our virtual house and we could then use videos from the virtual house as stimuli with fMRI.

In a fMRI paradigm we designed trials beginning with a direction cue. This could either be left or right. We followed this with a video sequence of a first person perspective camera moving in a linear direction out of one of two rooms that were designated as starting rooms and opposite one another. A final video sequence was shown of the same camera perspective moving in a linear direction in to one of the two remaining rooms which were designated as end rooms and opposite each other. Subjects in this experiment were asked to predict the upcoming end room during the first video of the start room. This was possible by combining the direction cue with the identity of the first room. Both video sequences had a white occluder in the bottom right quadrant of the video so that we could analyse voxels which only received feedback input. The design of the experiment was orthogonal, allowing us to balance direction cue and destination to ask questions about the factual experience, or direction cue and start room to ask questions about the future. We found that in extrastriate areas we could read out information about the two future rooms from feedback responses.

In **Chapter 7** we discuss these findings in the context of current literature and propose ideas for expanding on the paradigms presented in this thesis to answer the remaining unanswered questions.

2 General methodology

2.1 Functional magnetic resonance imaging

Throughout this thesis we use functional magnetic resonance imagining (fMRI) as our primary method for imaging the brain. FMRI is a non-invasive brain imaging technique that uses blood-oxygen-level dependent (BOLD) contrast. This technique exploits the fact that neurons receive energy directly from the blood. Neurons that are performing computations receive more oxygen from the blood than those neurons that are under less computational load. Differences in the relative levels of oxygenated and deoxygenated blood lead to differences in contrast in the magnetic resonance image. Research has subsequently confirmed that the observed BOLD signal is related to the underlying neurological processing (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001).

2.2 Data analysis

Functional and anatomical data were pre-processed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, Netherlands). The first two volumes of each functional run were discarded to avoid magnetic saturation effects. Functional data underwent slice time correction, 3D motion correction and high-pass temporal filtering using a Fourier basis set (Kay, Rokem, Winawer, Dougherty, & Wandell, 2013). No spatial smoothing was performed. The data was then spatially normalised by transformation into Talairach space. Functional data were aligned with the anatomical data in the form of an inflated surface, and alignment was checked using custom MATLAB scripts. The alignment of all functional runs was checked with respect to their corresponding anatomical data. For each subject alignment was highly correlated (r > 90).

2.2   Occlusion paradigm
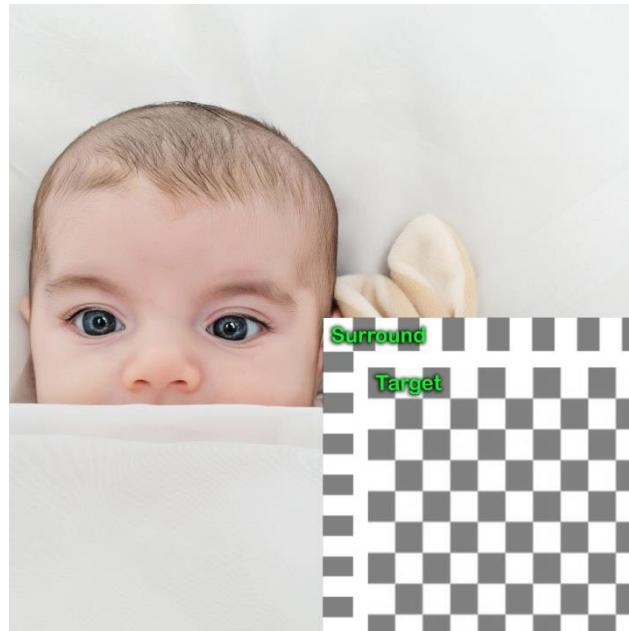
Figure 2.1: An example of an occluded image. The bottom right quadrant is occluded by a white mask. As part of the process to isolate voxels that have receptive fields in the occluded quadrant, we add independent "Surround" and "Target" conditions to our experiments. These conditions flicker inverting chequered boards in the part of the visual field occupied by the occluder.

In Chapters 1, 2 and 4 we use an occlusion paradigm. This paradigm is illustrated in figure 2.1. The lower right quadrant of an image or video stimulus is occluded by a white mask. Due to the retinotopic organisation of visual cortex, this blocks any feedforward information about the stimulus to a patch of cortex that corresponds to the occluded part of the visual field. Any changes to the BOLD signal in voxels that are occluded are therefore attributed to feedback processes.

In the occlusion paradigm we ask subjects to fixate in the centre of the screen. Subjects are monitored by camera to ensure that they are capable of fixating properly.

2.3　Retinotopic mapping and contrast mapping

In each chapter, we collected retinotopic mapping data. We used a rotating flickering chequered board wedge stimulus to measure voxel responses to different parts of the visual field. We also used a rotating flickering chequered board annulus stimulus to measure voxel responses to different eccentricities in the visual field. A correlation analysis of the retinotopic mapping data was projected on to the cortical reconstruction to localise foveal and peripheral V1, V2 and V3 of early visual cortex (EVC).

The t-map from a general linear model (GLM) from experimental runs was projected onto a reconstruction of the cortical surface to localise occluded regions from the contrast between mapping conditions (target > surround). Only significant activity in response to the target mapping contrast in the occluded region bounded by V1, V2, V3, was considered as a ROI for the occluded region.

We further ensured that voxels were occluded by estimating a receptive field for each voxel and selected only those with receptive fields within the occluded area. We further restricted our search for occluded voxels by only selecting those with a t-value greater than 1 for the difference between the "Target" minus "Surround" conditions, as shown in figure 2.1.

2.4　Population receptive field mapping

The retinotopic mapping data was also used to estimate receptive fields for each voxel in early visual cortex (Dumoulin & Wandell, 2008). For each individual voxel, we fit multiple 2D isotropic Gaussian functions to the response profile to both polar and eccentricity mapping data. We accepted the best-fit model as the receptive field for a particular voxel. We eliminated voxels from our analysis that had a low-quality estimate for its receptive field.

2.5   General linear modelling & multivariate pattern analysis

For multivariate pattern analysis we used general linear modelling to estimate a beta (β) value for each voxel independently. This is done for each run independently. The voxel activations are z-scored prior to modelling and each experimental condition is modelled by a single regressor in the model. This produced β values for each condition for each voxel for each run.

The β values were then used to train a linear support vector machine classifier (SVM; using LIBSVM (Chang & Lin, 2011)). Linear SVMs were chosen as research suggests that they generally perform well compared with other classifiers (Ku, Gretton, Macke, & Logothetis, 2008; Misaki, Kim, Bandettini, & Kriegeskorte, 2010; Mitchell et al., 2004) and suffer less from over-fitting than non-linear SVMs (Cox & Savoy, 2003; LaConte, Strother, Cherkassky, Anderson, & Hu, 2005). The linear SVM classifier was trained on n-1 runs of data, where n is the total number of runs in the experiment. The classifier was then tested on the independent nth run of β values. This procedure was repeated for each run for the purpose of cross-validation.

2.7 Statistical testing

Following the cross-validation procedure above, each classifier produced N classifier accuracies. The mean of these was taken as the classifier accuracy for that subject for that region of interest for that condition. The group mean classifier accuracies for each condition and region of interested were then bootstrapped to attain a distribution of classifier accuracies whose 95% confidence interval would contain the true value. We report the median and 95% confidence interval for each condition in each ROI whose 95% confidence interval does not contain the chance classifier accuracy (50%) and is therefore deemed significantly above chance classification.

We chose the approach of using a second level analysis with a t-test as it has been widely used within the field of research (Haxby et al., 2001; Haynes et al.,

28

2007; Spiridon & Kanwisher, 2002). However, a recent paper by Allefeld and colleagues has brought to light that second level analysis using a t-test does not allow one to infer that the observed effect generalises to the population, only that there is an effect within the group of subjects tested (Allefeld, Görgen, & Haynes, 2016). This conclusion about second level t-test analysis should be kept in mind when considering the results within this thesis.

3 Information about non-social negative valance objects is present in non-stimulated early visual cortex

3.1 Abstract

Regions of early visual cortex that do not receive meaningful feedforward stimulation still however contain information about the surrounding visual scene (F. W. Smith & Muckli, 2010). This surrounding contextual information is transferred to early visual cortex by cortical feedback (Muckli et al., 2015). Investigating the contribution of contextual feedback to low-level visual areas is central to understanding how internal brain signals interact with visual processing. Using human functional brain imaging (fMRI) and multivoxel pattern classification, we investigated if information patterns fed back to V1, V2 and V3 included emotional content. We used 900 partially-occluded positive, negative and neutral valance colour images from three distinct categories (animals, foods, plants). The images were presented to subjects during fMRI in a unique path-guided de-Bruijn m-sequence with third order counterbalancing to allow for the modelling of carry over effects of one stimulus on a subsequent stimulus. The unoccluded images were then rated afterwards to account for individual subjectivity. Adopting a classification approach, we can discriminate between valance pairs that include negative valance from stimulated and non-stimulated voxels, independent of category information, in V1, V2 and V3. We also find this pattern in the amygdala and pulvinar. Additionally, using a cross-classification approach, we show that a classifier trained on valance pairs for two categories can generalise to the same valance pair on the remaining test category using stimulated voxels. As before, this is only possible when negative valance information is available to the classifier in each of V1, V2 and V3. For non-stimulated voxels neutral versus negative is generalisable in V1 and neutral versus positive is generalisable in V2. Neutral versus positive is also generalisable in the amygdala. Our data are in accord with previous work that indicates that information about negative valance images are privileged in the set of valance images. We show that negative valance information

is important even at the earliest stage of cortical visual processing and that this information is generalisable across visual categories.

3.2   Introduction

Primary visual cortex (V1) receives the majority of the retinal input, however the retinal input only accounts for approximately 5% of the excitatory information to V1 (Muckli, 2010). Most of the information to V1 is non-geniculate, arriving via lateral and top-down feedback pathways from many different cortical and subcortical sources (Muckli & Petro, 2013). Understanding the interaction between non-geniculate and geniculate brain signals in low-level visual areas is critical to understanding visual processing.

One category of non-geniculate input to the early levels of visual processing is that of emotion information. Emotion information is a unique category of information due to its ability to bias perception. As a result of the brain's overall limited processing capacity, stimuli will compete for processing resources (Petersen & Posner, 2012). The emotional content of a stimulus can influence processing towards a stimulus that contains higher emotional content versus one that contains little or no emotional content (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Pourtois, Schettino, & Vuilleumier, 2013). The emotional content of a stimulus can further bias perception depending on whether or not it is presented centrally or peripherally (Almeida, Soares, & Castelo-Branco, 2015; D'Hondt, Szaffarczyk, Sequeira, & Boucart, 2016) and it can also bias perception based on its valance (D'Hondt et al., 2016).

The influence on brain responses due to the emotional content of a stimulus can vary along several distinct dimensions including valance (negative versus positive), arousal (low versus high) and sociality (social versus non-social) (Britton et al., 2006). The distinction between social and non-social categories is important as they recruit overlapping yet different networks (Britton et al., 2006). Further differences between social and non-social processing can be observed in the distinct temporal windows in which emotional cue extraction occurs for both types

of stimuli. Face cue extraction occurs approximately 170ms after stimulus onset whereas cue extraction for complex emotional scenes occurs approximately 220-230ms after stimulus onset (Bekhtereva, Craddock, & Müller, 2015). Therefore, although both social and non-social stimuli can vary along dimensions of valance and arousal, they should be treated differently.

With respect to social stimuli, it has been shown that brain responses from both stimulated (eye and mouth regions) and non-stimulated retinotopic locations of V1 can be used to discriminate between emotional and non-emotional faces using multivariate pattern analysis (Lucy S Petro, Smith, Schyns, & Muckli, 2013). This finding suggests a top-down influence on low level visual areas by high level social emotion information. It is currently unknown whether information about non-social stimuli can be read out from brain responses in non-stimulated retinotopic visual locations, as has been demonstrated for object category information (Williams et al., 2008). Whether non-geniculate information arrives at low-level visual areas from higher cortical areas or from sub-cortical areas, such as the amygdala or pulvinar, remains an open question.

To test whether information about non-social stimuli can be read out from non-stimulated visual cortex we composed a 900 non-social image data set. The image set contained 300 images from a non-human animal category, 300 images from a food category and 300 images from a plant category. Each category contained 100 images for neutral, positive and negative levels of valance. The image set was used in a fMRI fast event-related design employing a path-guided de Bruijn m-sequence paradigm with third level counterbalancing (Aguirre, Mattar, & Magis-Weinberg, 2011). A unique m-sequence was generated for each subject who viewed each image from the data set only once across ten fMRI runs. Each image was partially occluded by masking the bottom right quadrant of the image with an occluder that matched the RGB colour of the background. Due to the high spatial resolution and retinotopic organisation of visual cortex, occlusion restricts geniculate input to a corresponding patch of visual cortex (F. W. Smith & Muckli, 2010). As V1 is not always involved in processing emotion information, either by TSM interference (Yaple, Vakhrushev, & Jolij, 2016) or due to cortical blindness by bilateral V1 damage (Ajina & Bridge, 2018), we collected brain responses from V2 and V3 as well as the amygdala, the pulvinar and lateral occipital cortex (LOC) to

32

test in our multivariate analysis. After MRI, we asked subjects to rate the unoccluded image set for both valance and arousal on a 9-point Likert scale identical with that used with the International Affective Picture System (IAPS; Lang, Bradley, and Cuthbert (1997)). We did this to obtain individual ratings to control for domain-specific and sex-specific biases (Fernández-Martín, Gutiérrez-García, Capafons, & Calvo, 2017).

In a multivariate classification analysis, we find that we can reliably decode from both stimulated and non-stimulated retinotopic locations in early visual cortex information about negative valance independent of category. This was also true in both the amygdala and pulvinar. In LOC we can read out information about each level of valance independent of category. In a second multivariate cross-classification analysis, we find that information about valance generalises across category. These findings demonstrate that information about non-social valance exists in non-stimulated early visual cortical areas and is in line with previous studies that demonstrate that negative valance is unique in terms of emotion processing within the brain (Kuniecki, Wołoszyn, Domagalik, & Pilarczyk, 2017; Vida & Behrmann, 2017).

3.3 Results

3.3.1 Negative valance information is important in early visual processing.

It has been shown that information about emotional faces can be read out from non-stimulated early visual cortex (Lucy S Petro et al., 2013). It is less clear whether or not information about affective non-social objects can be read out of non-stimulated early visual cortex.

To test whether valance information is present in early visual cortex, we used general linear modelling to first filter out information about category, information about changes in lightness during the sequence of image presentation and information from carry-over effects. We then used general linear modelling for a

valance model on the residual data from the first analysis. This produced multivoxel patterns of beta weights for valance for each run independent of the filtered information. Next, we trained a linear support vector machine classifier (SVM) to distinguish between two levels of valance on N-1 runs of data and then tested the classifier on the Nth run of data. This provided classifier accuracies for 10-folds which were averaged across to produce the final classifier performance. The group classifier accuracies for each pair of valance for reach ROI were bootstrapped to create a distribution from which the median and 95% confidence interval could be found. This was done for V1, V2 and V3 and for the amygdala and pulvinar.

In early visual cortex found that from feedforward responses we could decode neutral versus negative in V1 (median: 67%, CI: 56%:78%, p = 0.01), V2 (median: 68.5%, CI: 59%:79%, p = 0.01) and V3 (median: 69%, CI: 59%:79%, p = 0.01). We could also decode positive versus negative in V1 (median: 74.5%, CI: 67%:82.5%, p = 0.01), V2 (median: 70.5%, CI: 61.5%:78.5%, p = 0.01) and V3 (median: 78%, CI: 70%:86%, p = 0.01).

From feedback responses we could decode neutral versus negative in V2 (median: 60%, CI: 49%:70.5%, p = 0.05). We could also decode positive versus negative in V1 (median: 66.5%, CI: 58%:74%, p = 0.01), V2 (median: 66.5, CI: 56%:78%, p = 0.01) and V3 (median: 67.5%, CI: 56.5%:78%, p = 0.01).

In the amygdala region of interest, we could decode neutral versus negative (median: 65.5%, CI: 56%:74.5%, p = 0.01) and positive versus negative (median: 66.5%, CI: 57.5%:74%, p = 0.01). In the pulvinar region of interest we could only decode positive versus negative (median: 61.5%, CI: 55.5%:67.5%, p = 0.01). In LOC however, we could decode neutral versus positive (median: 60%, CI: 50.5%:66.5%, p = 0.05), neutral versus negative (median: 74%, CI: 64%:84%, p = 0.01) and positive versus negative (median: 79%, CI: 67%:89.5%, p = 0.01).

Our inability to reliably decode positive versus neutral valance across V1, V2, V3, the amygdala and pulvinar tells us that the classifier was using the negative information as the mechanism for performing successful classification. In these regions of interest positive and neutral valance were indistinguishable. However, in LOC the classifier could reliably decode each level of valance.

34

Next, we performed a second analysis without filtering category information. This analysis was a cross-classification analysis. Here we modelled pairs of category and valance individually along with carry-over effects and changes in lightness. This produced multivoxel beta patterns for animal neutral, food neutral, plant neutral, animal positive, food positive, plant positive, animal negative, food negative and plant negative. In a similar cross-validation procedure described above, a classifier was trained to distinguish between two levels of valance for two categories (for example, animal neutral and food neutral versus animal positive and food positive) and tested on the remaining category (for example, plant neutral versus plant positive). Here we aim to test whether valance information overlaps amongst category.

We found that in early visual cortex we could cross-decode from feedforward responses neutral versus negative in V1 (median: 56.34%, CI: 53.84%:58.67%, p = 0.01), V2 (median: 57.34%, CI: 54.34%:60.5%, p = 0.01) and V3 (median: 55%, CI: 51%:60%, p = 0.01). We could also cross-decode positive versus negative in V1 (median: 53.83%, CI: 50.34%:57.33%, p = 0.05), V2 (median: 55.33%, CI: 52.17%:58.67%, p =0.01) and V3 (median: 58.95, CI: 55.15%:63.16%, p = 0.01).

From feedback responses we could cross-decode neutral versus negative only in V1 (median: 55.17%, CI: 50.18%:59.32%, p = 0.05).

Surprisingly we can cross-decode neutral versus positive in V2 (median: 52.49%, CI: 49.99%:54.99%, p = 0.05). We can also only cross-decode neutral versus positive in the amygdala (median: 54.01%, CI: 50.17%:58.67%, p = 0.05).

In the pulvinar we can only cross-decode positive versus neutral (median: 53.5%, CI: 52%:55%, p = 0.01).

As before, negative valance is important for the classifier's ability to distinguish between conditions. However, for the first time we can distinguish between neutral and positive valance from feedback responses in V2 and in the amygdala. In both ROI no other information was cross-decodable. This tells us that valance information is present in non-stimulated early visual cortex and that valance information is generalisable across category.
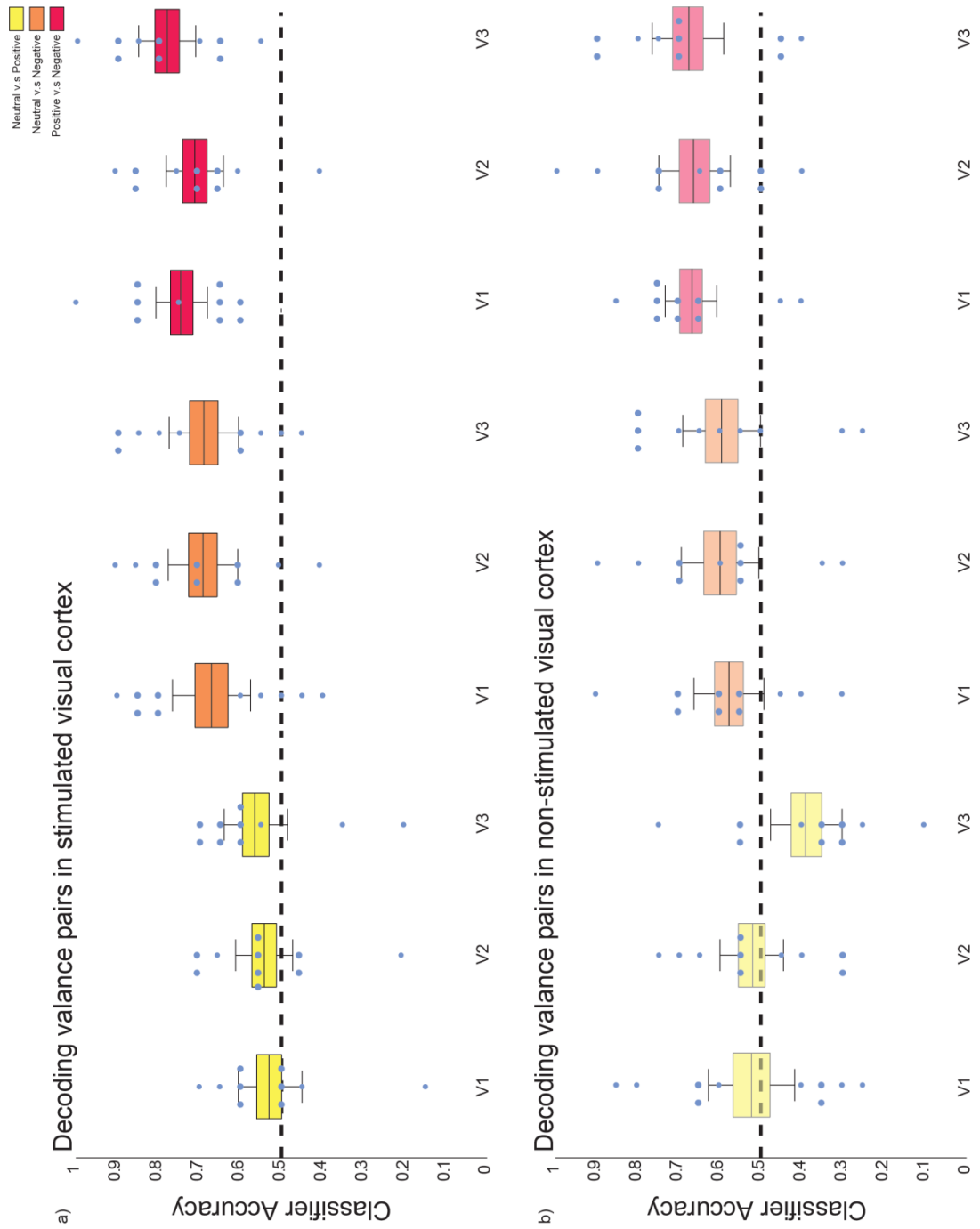
Figure 3.1: Boxplots showing decoding performance for pairs of valance in a) stimulated early visual cortex (feedforward) and b) non-stimulated occluded early visual cortex (feedback). The centre line of a box denotes the median, the whiskers the 95% confidence interval and dots are individual subject data. Chance was 50%.

3.4 Discussion

3.4.1 Negative valance is uniquely represented in visual processing.

In the literature of visual processing of emotion stimuli, negative valance images are a privileged set of stimuli. For example, fixating on an object is faster for negative images versus neutral images (Kuniecki et al., 2017). This is particularly true for a subset of negative valance images that includes snakes (Gomes, Silva, Silva, & Soares, 2017), spiders (LoBue, 2010), guns (Subra, Muller, Fourgassie, Chauvin, & Alexopoulos, 2018) and syringes (Blanchette, 2006). However, the modern negative non-social valance images within this subset do not always demonstrate that they belong to this privileged set, which has led to some researchers suggesting that negative images are only privileged when they represent an evolutionary threat (Vida & Behrmann, 2017). Despite this, studies continue to contradict the evolutionary framework by demonstrating that modern threating images are privileged versus evolutionary threatening images (Subra et al., 2018). This recent finding would suggest that the underlying mechanism that categorises negative valance images in to a privileged set is based on the relevance of the stimulus to the individual. Within this framework, attention is captured for negative images that biases perception to specific stimulus features that are relevant to the individual from moment to moment (Pessoa & Adolphs, 2010).

In our study of valance information processing in early visual cortex, we find that we can read out from both feedforward and feedback responses in V1, V2 and V3, information about negative valance images only. This finding aligns well with the current literature regarding negative valance processing with respect to relevance. In our set of images only 22 out of 300 negative valance images are of snakes or spiders. As these 300 images are averaged across 10 runs in a fast event-related fMRI design, and each snake and spider image are only viewed once, it is unlikely that this is the driving factor for our finding. It is more likely that the relevance of negative stimuli to the individual was the driving factor for this finding in our

37

experiment. This is even more likely considering the images were rated by subjects and individual subjectivity taken in to account.

Additionally in the literature there has been debate about cortical and sub-cortical differences with regards to emotion stimuli representation (Vida & Behrmann, 2017). In our experiment we find that we can also read out from responses in the amygdala and pulvinar information about negative valance images only. This would suggest that in this experiment negative valance information is privileged in both cortical and sub-cortical structures of interest. Surprisingly in LOC, an object-category specific region, we can read out information about all levels of valance independent of categorical information. This difference may be due to the fast rapid-event related design of our study. We present each image for a single second. As we have discussed, the brain responds faster to negative valance images (Kuniecki et al., 2017). By responding faster to negative images, these images will undergo faster processing which may lead to more well-defined representations in low-level visual areas and small, deep sub-cortical structures compared with positive and neutral images. This would explain why we can read out information about positive valance in LOC, a higher cortical area with more refined representations with respect to low level visual areas.

3.4.2 Is valance feedback information in early visual cortex from a cortical or sub-cortical source?

Emotion processing in the human brain is thought to involve both a cortical and sub-cortical pathway (Pessoa & Adolphs, 2010). The sub-cortical pathway includes both the pulvinar and the amygdala. The pulvinar is situated at the posterior of the thalamus and has been described as a hub that the cortex is shaped around. Its unique position hints at its connectivity. The pulvinar has bidirectional connectivity with all visual areas as well as cross-modal and attention networks (Pessoa & Adolphs, 2010). Similarly, white matter connectivity analysis using diffusor tensor imaging (DTI) has also revealed a feedback architecture between early visual areas and the amygdala (Gschwind, Pourtois, Schwartz, Van De Ville, & Vuilleumier, 2011).

Due to the inter-connectivity between the pulvinar and the rest of the cortex, it is thought that the pulvinar is responsible for rapid detection of environmental threats. This can be demonstrated in cases of cortical blindness, where information about fearful faces still enter awareness even when presented in the blind field (Bertini, Cecere, & Làdavas, 2013) likely due to recurrent connectivity between the amygdala and pulvinar and the pulvinar and the cortex (Pessoa & Adolphs, 2010). When the pulvinar is damaged there is a slower response to visual threat (Ward, Danziger, & Bamford, 2005) and overall impaired emotion recognition (Ward, Calder, Parker, & Arend, 2007). Conversely, when the amygdala is damaged there remains an intact ability to detect fearful faces when awareness is directed to specific stimulus features (Adolphs et al., 2005), again likely due to the role the pulvinar plays in relaying information between different parts of the brain.

Although we cannot conclude that there are any differences between sub-cortical and cortical representation of non-social affective images in our study. There are a several limiting factors that should be mentioned. One such limitation is differences in emotion processing in dorsal and ventral visual streams. For example, it has been show that the brain can make rapid distinctions between fearful body expressions versus neutral body expressions as early as 80ms in the dorsal visual stream (Meeren, Hadjikhani, Ahlfors, Hämäläinen, & De Gelder, 2016). As we occluded the bottom-right quadrant of our images, we were restricted to only analysing non-stimulated cortex in the dorsal stream. Therefore, our feedback decoding and cross-decoding analysis are only applicable to dorsal early visual cortex and we cannot make any inferences about the ventral visual stream. There are also differences in how the brain processes affective images with respect to right and left hemispheric specialisation (Gamond, Vecchi, Ferrari, Merabet, & Cattaneo, 2017; Shobe, 2014). As we had no a priori hypothesis about differences in right and left hemispheres, this was not included in our model that the m-sequence was designed from. This omission could also have an impact on our results as the m-sequence design boosts the detection power for the signal expected from the model derived from the hypothesis.

An additional limitation of our study was the issue of arousal. It has been shown that the complexity of an image is correlated with levels of arousal (Madan, Bayer,

39

Gamer, Lonsdorf, & Sommer, 2018), and arousal is related to subjective feelings and enhanced attention (Lindquist, Satpute, Wager, Weber, & Barrett, 2015). In our analysis, arousal is averaged across 300 images for positive, neutral and negative valance but our analysis is not independent of arousal. Arousal therefore remains a confound in this study.

3.5 Methods

3.5.1 Subjects.

Ten subjects (age range 18-33; 5 female) participated in this experiment. Each subject was recruited via the subject data pool maintained by the School of Psychology at the University of Glasgow. There were no exclusion criteria for recruitment. Ethical approval was granted for this study from the College of Science and Engineering at the University of Glasgow.
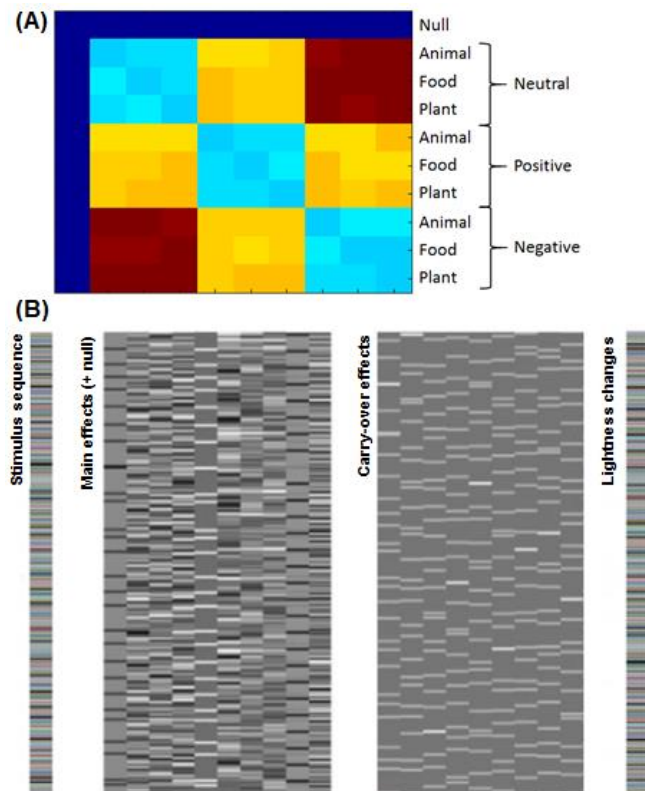
Figure 3.2: (A) Hypothesis model for generating the path-guided m-sequence (B) Modelling the main effects, carry-over effects and changes in lightness.

3.5.2 Stimuli and Procedures.

For this study we collected 900 colour non-social images by way of Google image search. Each image was collected at a high resolution before being resized to 800 pixels by 800 pixels. Each image was classified as belonging to one of nine categories by the experimenter. These were; 'neutral animal', 'positive animal', 'negative animal', 'neutral food', 'positive food', 'negative food', 'neutral plant', 'positive plant', 'negative plant'. During fMRI these images were projected on to a rear projection screen using a projector system (1024 x 768 resolution, 60Hz refresh rate). Images spanned 22 by 15 degrees of visual angle. Subjects were asked to maintain fixation on the centre of the screen and tasked with detecting a one-back repetition of category. The bottom right quadrant of each image was occluded with a 128, 128, 128 RGB occluder. This was also the same colour used as the default background.

A unique path-guided de Bruijn sequence (Aguirre et al., 2011) was generated for each subject. Sequences were generated with third-level counterbalancing resulting in sequences with 1000 elements. This included 100 repetitions of each of the nine conditions plus 100 null trials. A sequence was split in to 10 shorter runs. So that the runs could be stitched back together in to a single run, the last 15 elements of run N were added to the beginning of run N+1. The last 15 elements of run 10 were added to the beginning of run 1 such that the sequence cycles around from start to finish. These overlapping elements were then removed when the runs were stitched back together.

Images were presented for 1 second in a fast event-related design with an inter stimulus interval of 2 seconds. The onset of image presentation was jittered between 0 and 1500 milliseconds and a null trial was 6 seconds long. Each of the 900 images was presented to the subject only once throughout the experiment. At the end of each run a contrast-inverting chequered board was presented for 12 seconds either along the inside of the occluder or deep within the occluded region. These were included for contrast mapping, to identify the occluded region of visual cortex. The contrast mapping conditions were discarded when the 10 runs were stitched into a single run.

41

The experiment was performed using Presentation® software (Version 0.70, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

Figure 3.3: Examples of the affective stimuli used in the experiment.

3.5.3 Image Rating.

After fMRI, subjects were then instructed to rate the 900 unoccluded images for both valance and arousal on a 9-point Likert scale following the same procedure used for the International Affective Picture System database (Lang et al., 1997). These ratings were used to sort the image data set and fMRI data set trials in to positive, neutral and negative for each subject.

Due to individual differences, images that were rated as neutral, positive or negative by the experimenter were not always rated so by subjects. To account for this, we made a small number of corrections to subject ratings. Neutral was designated '5' on the Likert scale. If a subject rated a neutral image either '4' or '6' this was corrected back to '5' as the change from neutral to either negative or positive was deemed negligible (a deviation of one on the scale). However, if a neutral image rated either negative or positive by a deviation of more than one on the scale then this new rating was accepted (see supplementary).

3.5.4 fMRI acquisition.

Data were collected at the Centre for Cognitive Neuroimaging, University of Glasgow. Functional and anatomical MRI data was acquired using a 3 Tesla Tim Trio MRI scanner (Siemens, Erlangen, Germany) with a 32-channel head coil and integrated parallel imaging techniques (IPAT factor: 2). Functional multi-band EPI sequences were used to acquire partial brain volumes of resolution 3mm isotropic, inter-slice thickness of 0.3mm, TR-1s, TE-30, 30 slices, FOV of 220mm and a flip angle of 62°. Anatomical scans (3D MPRAGE) were acquired with a resolution of 1mm isotropic, TR-2.3s and 192 volumes.

3.5.5 Data analysis.

Data were stitched together by removing the 15-trial overlap at the beginning of each run. For classification, we created 3 regressors to model the three categories, 3 regressors to model the three levels of valance, a single regressor to model the null condition, 9 regressors to model the carry-over effects for first, second and third level counterbalancing and a single regressor to model the change in image lightness from the background default lightness. A general linear model analysis was performed that included all regressors minus the 3 regressors for each level of valance. Then on the residuals from this first analysis we performed a second general linear model using only the 3 regressors for valance. The beta estimates from this second GLM were used as patterns for MVPA.

For cross-classification, we created 9 regressors to model the pairwise combinations of valance and category. We included a single regressor to model the null condition, 9 regressors to model the carry-over effects for first, second and third level counterbalancing and a single regressor to model the change in image lightness from the background default lightness. A general linear model analysis was performed that included all regressors. Beta estimates were then used as patterns for MVPA.

3.6 Supplementary
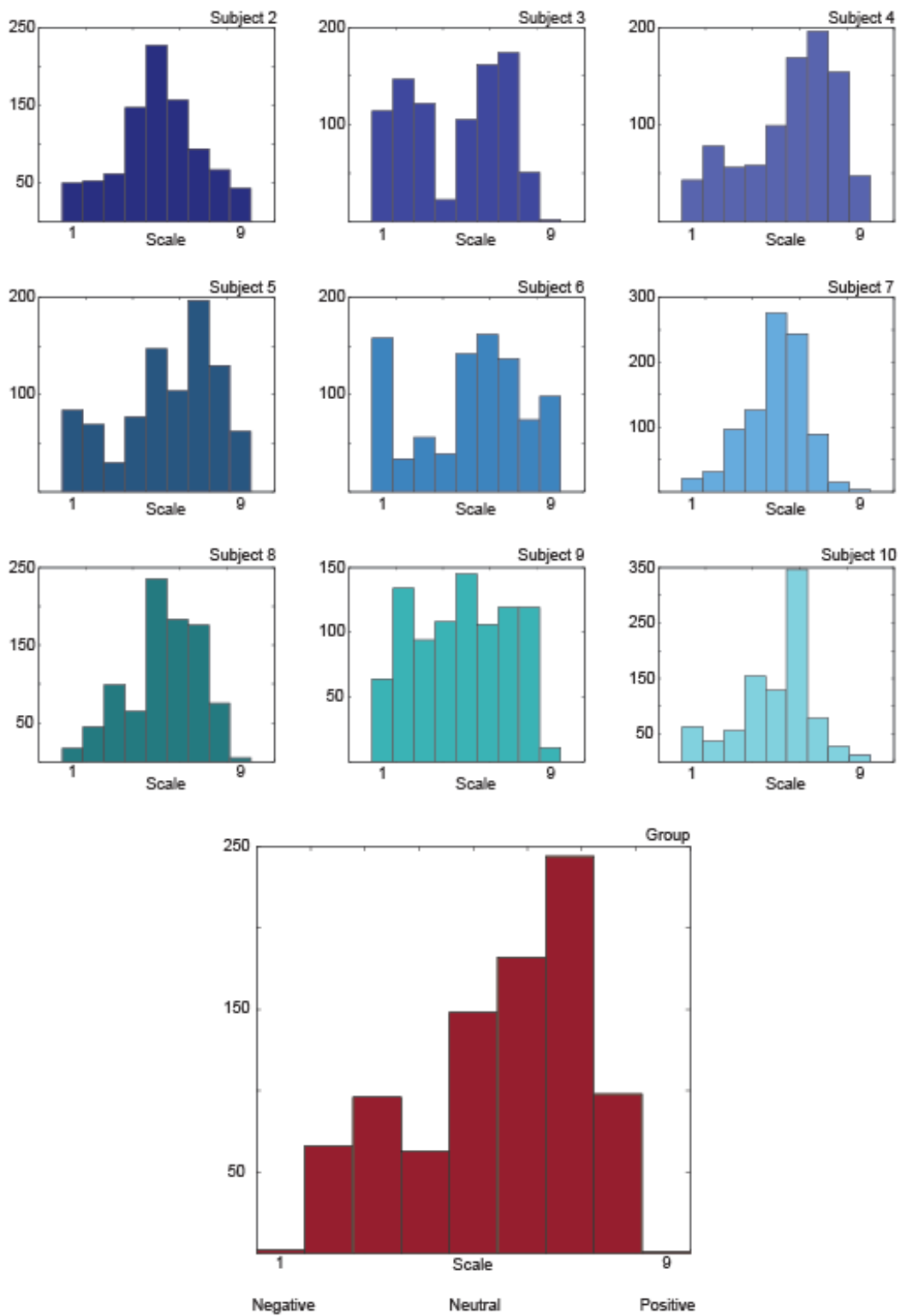
Figure 3.4: Histograms showing the distribution of uncorrected valance ratings for the 900 images for each individual subject (2-10) and the averaged uncorrected valance ratings for the group.
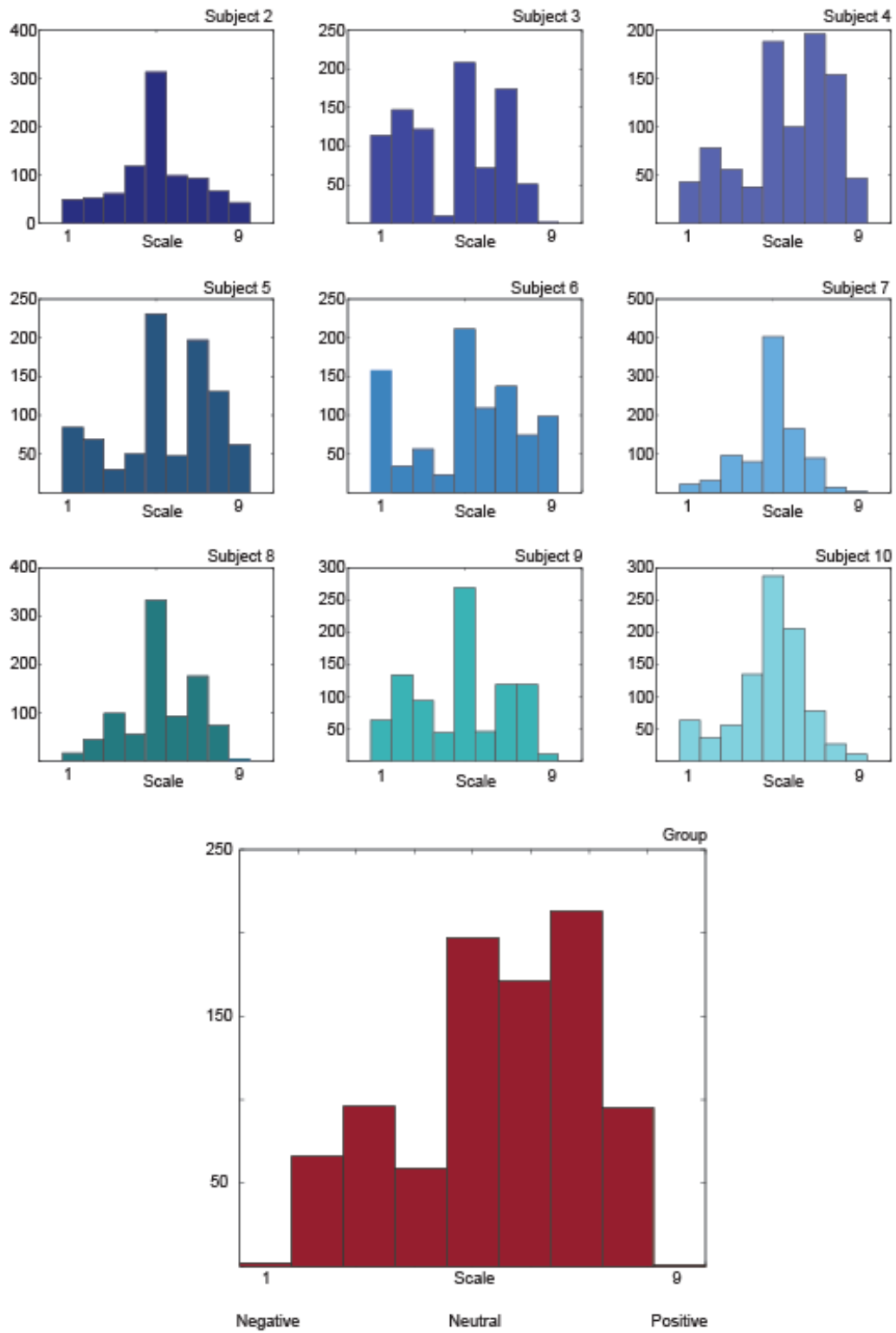
Figure 3.5: Histograms showing the distribution of corrected valance ratings for the 900 images for each individual subject (2-10) and the averaged corrected valance ratings for the group.
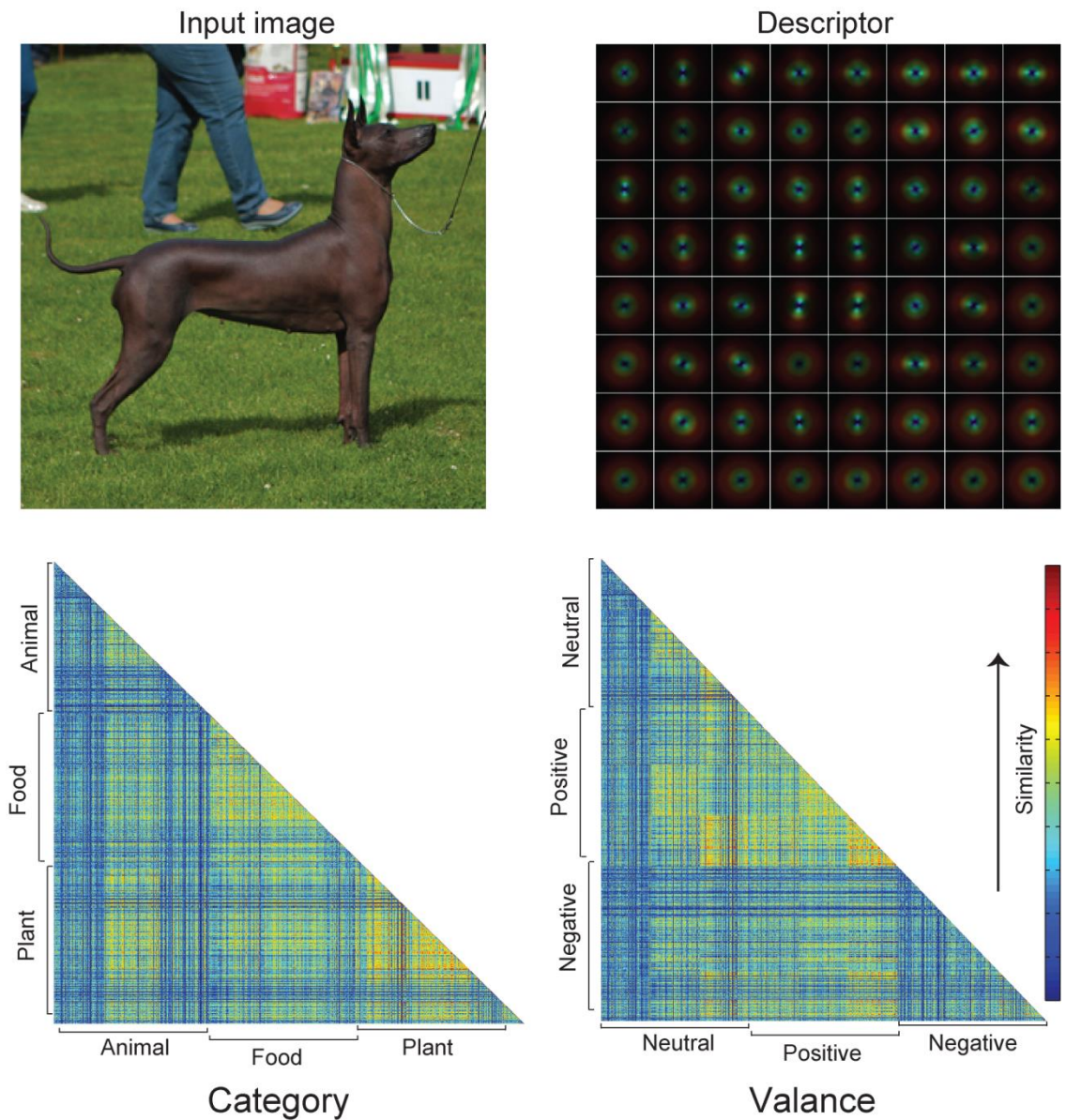
Figure 3.6: A GIST descriptor was computed for each image in the 900 image set. The descriptor represents the image as a lower dimensional representation. These GIST descriptors were then correlated and organised by either category or valance to visually inspect the similarity structure within the image set. Plant and food images are similar within their category whereas animal images were neither similar within category or across category, with the exception of the positive animal image subset which was similar across category. When organised by valance, the positive images were similar within the group and somewhat similar with the neutral and negative images. The negative images were dissimilar within their own group.
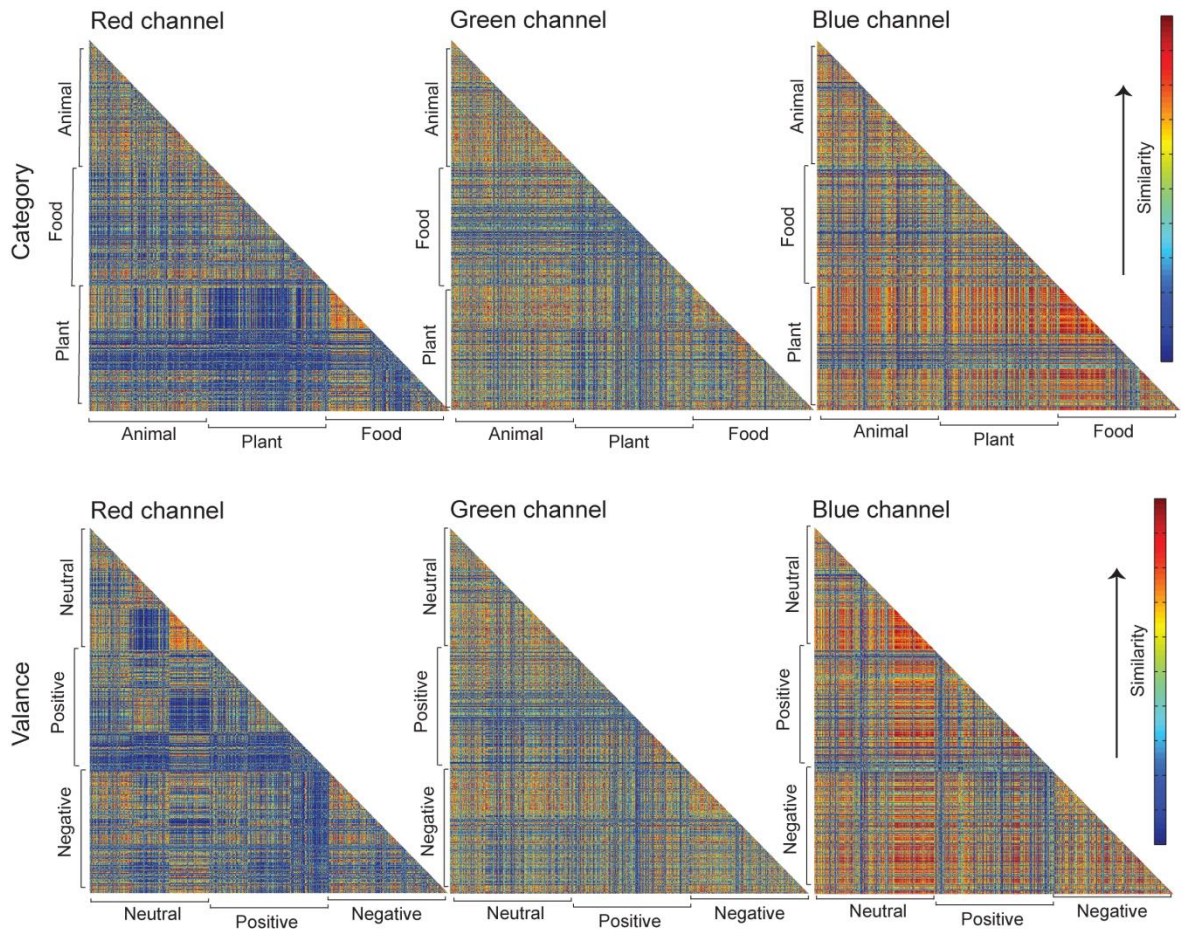
Figure 3.7: Each image was split in to red, green and blue colour channels and colour histograms were computed. For each colour channel, the colour histograms were correlated to visually inspect the similarity structure between images. The resulting correlation matrices were then organised by category or valance.

4 Auditory feedback shows eccentricity biases and overlap in neural representation with well-matched congruent visual exemplars

4.1 Abstract

Early visual cortex receives cortical feedback information about complex auditory stimuli. However the information content and function of complex auditory representations in early visual cortex remains poorly understood. Using functional magnetic resonance imaging, we investigated where in the retinotopic space of early visual cortex we can read out information about auditory scenes and objects. We find that we can read out auditory scene information is in the periphery of early visual cortex and that we can read out auditory object information in the fovea of early visual cortex. We then follow this study by investigating whether complex auditory sounds are cross-decodable with complex silent videos when they are well matched. We replicate previous work showing that auditory information can be read out from early visual cortex, and find that auditory feedback information within an occluded patch of early visual cortex is representationally similar to categorically matching visual feedback information. Our results provide evidence supporting the theory that auditory information in early visual cortex is organised by eccentricity, and that abstract auditory feedback responses in early visual cortex share an overlap in their representation with that of congruent visual feedback responses.

4.2   Introduction

Visual processing is a highly predictive mechanism that utilises top-down expectations to make inferences about simple, complex and noisy visual stimuli. Predictions in visual cortex originate from rich internal mental models that are triggered by all sensory stimuli (Spratling, 2017). A growing body of work is revealing that auditory driven predictions inform perception at almost all stages in the visual hierarchy. For example, auditory scene sounds are decodable in primary and secondary visual cortex (Vetter et al., 2014), and much later in higher areas of visual cortex (Jung et al., 2018).

Complex auditory sounds are composed of many different sound sources. These sound sources can take the form of abstract scene representations or more concrete object representations. No study to date has investigated the retinotopic structure of auditory feedback from different sound sources. There is evidence for a retinotopic structure of feedback responses found in studies of visual object recognition. Objects presented in the periphery are decodable in the fovea, likely due to the need for fine discrimination which the high spatial resolution of the fovea provides (Williams et al., 2008). It is unclear whether an individual sound source in isolation, such as an object, would feedback only to the fovea or be distributed across all eccentricities. On the contrary, the current literature on auditory feedback reveals a bias towards the periphery of early visual areas (Falchier, Clavagnier, Barone, & Kennedy, 2002). There is very little evidence of a functional role of auditory feedback to the fovea.

Auditory feedback in striate and extrastriate cortex are cross-decodable between contextually similar exemplars (Vetter et al., 2014). This demonstrates that the information content between congruent sounds is abstract and contextual at the encoding space of visual cortex. To date, contextually similar visual and auditory exemplars have been shown to be distinct and non-generalisable in early visual cortex (Jung et al., 2018). It is not until much later in the visual hierarchy that contextually matching auditory and visual information converge and become cross-decodable (Jung et al., 2018). This lack of finding is surprising as feedback information about visual stimuli have been shown to be cross-decodable with their

50

feedforward counterparts in early visual cortex (F. W. Smith & Muckli, 2010), across cortical layers in early visual cortex (Muckli et al., 2015), and cross-decodable across stimulus features (Radoslaw Martin Cichy & Teng, 2017), suggesting that visual feedback is generalisable to an extent. If feedback from both modalities is abstract and contextual, it would follow that they both inform perception in a similar manner. It is possible that the lack of cross-decoding between modalities in low-level areas of visual cortex is due to a combination of methodological choices.

To test these ideas we performed two experiments. In the first experiment we combined auditory scenes and objects in a 2x2 block design. Each auditory scene (forest, blizzard) was presented for 12 seconds. One of two auditory objects (axe chopping wood, pull-start motor) was randomly presented in 7 or 8 of 10 possible one second time bins during the scene sound. This made the onset and offset of the object sound unpredictable, requiring attention for the full presentation time. In total there were four types of trial; forest axe, forest motor, blizzard axe, blizzard motor. Linear support vector machine classifiers were trained and tested on scene or object pairs.

In the second experiment we created independent blocks of auditory and visual stimuli from three distinct categorical videos (cars, cows, hammering). To isolate cortical feedback we occluded the lower right quadrant of the video stimuli. By occluding this area of the visual field, we render an area of left dorsal visual cortex as resting state during visual stimulation. To avoid the possibility of spill-over into the occluded region from neighbouring feedforward voxels, inverting chequered boards were used to map out the border and inner target patch of the occluder.

In experiment 1 we find that auditory scene information can be decoded only in the periphery and that auditory object information can be decoded only in the fovea. This demonstrates that auditory feedback arrives at visual cortex from multiple pathways - a peripheral pathway that carries abstract contextual scene information and a foveal pathway that carries object information. In experiment 2 we find that cortical feedback for dynamic auditory and visual scenes are generalisable in occluded visual cortex. Generalising between the two modalities means that there is an overlap in the neural representation of auditory and visual feedback

51

information at the earliest stages of visual processing. Therefore auditory feedback must be informing visual processing in a similar way to visual feedback. We also replicate previous findings that auditory and visual stimuli are decodable in early visual cortex. This new evidence supports the hypothesis that auditory and visual information are processed by anatomically and functionally shared neural populations in early visual cortex.

4.3 Results

4.3.1 Auditory objects feedback to foveal visual cortex.

Although previous studies report that the influence of auditory feedback on perception is biased towards the periphery, it is unclear if this is true for both complex auditory scene and object stimuli. In the first experiment using functional magnetic resonance imaging (fMRI) we tested where in the retinotopic space of early visual cortex auditory objects and scenes can be decoded. To do this we created a 2 by 2 block design (auditory scene [forest/blizzard] x auditory object [axe/motor]). Scene sounds were presented for 12 seconds followed by 12 seconds of baseline. Each scene sound was divided in to an equal number of 1 second time bins in to which a 1 second object sound could be presented (**Fig 4.6a**). A single object sound was randomised in to either 7 or 8 bins, counterbalanced, during a scene sound. This was done to make the object sound unpredictable in its appearance and to make it stand out from the scene.

Subjects were instructed to keep their eyes-open and to fixate on the centre of the screen where a small fixation-cross was placed. They were tasked with responding to a 200ms colour change of the fixation cross during the baseline. We trained a linear support vector machine classifier to discriminate between either the object sounds while the scene sound was the same, or the scene sounds while the object sound was the same. We trained the classifier on n-1 runs of data and decoded object or scenes from the nth run of data. We found that auditory objects could only be decoded in the fovea of V2 (median: 52.26%, CI: 50.47%:54.13%, p=0.05) and V3 (median: 52.03%, CI: 49.89%:54.03%, p=0.05) (**Fig 4.1**) and that

52

auditory scenes could only be decoded in the far periphery of V2 (median: 52.27%, CI: 50.21%:54.88%, p=0.05) (**Fig 4.2**).

This confirms previous work that shows scene sounds are encoded in the periphery of visual cortex. We also show new evidence supporting the idea that the fovea is additionally recruited for auditory object processing, either as a means to segregate objects from background sounds or for finer discrimination.
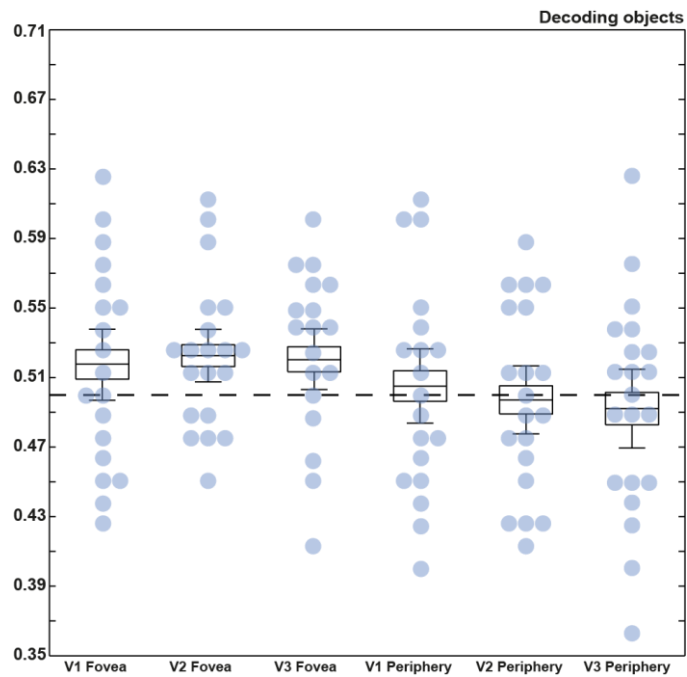
Figure 4.1: Experiment 1, decoding axe versus motor object sounds in early visual cortex.
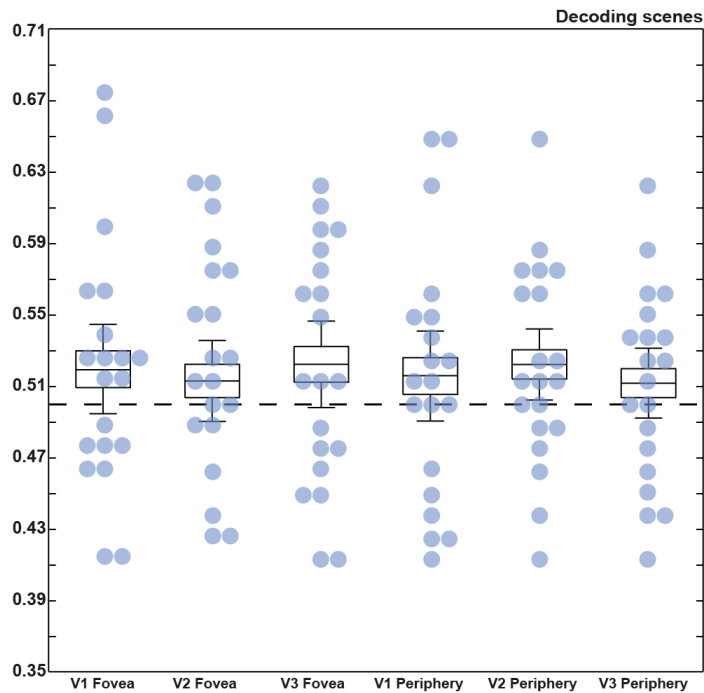
Figure 4.2: Experiment 1, decoding forest versus blizzard scene sounds in early visual cortex.

4.3.1 Auditory and visual feedback is contextually similar.

In the second experiment we investigated whether information about complex auditory scenes could be decoded in retinotopically mapped early visual cortex. We also explored whether image category could be decoded in occluded peripheral visual cortex, and whether a linear support vector machine classifier could generalise between modality. To do this we measured brain activity while eighteen new participants listened to categorical auditory scenes and viewed silent videos in alternating blocks while maintaining central eye-fixation (**Fig 4.6b**). Participants were randomly prompted by fixation colour change to respond by button press with the category of the stimulus. On a single trial basis, pattern classifiers were first trained (for each participant) to discriminate between patterns of beta values for the three auditory scene categories of n-1 runs of data. These pattern classifiers were then applied to the nth run of unseen data to see if they could decode the three auditory scenes. We found that the pattern classifiers could decode auditory category significantly above chance in many different retinotopically mapped regions. These included the fovea of V1 (median: 37.12%

CI: 34.44%:39.90%, p=0.01), V2 (median: 37.13% CI: 34.08%:40.2%, p=0.01), V3 (median: 37.41% CI: 34.06%:40/75%, p=0.01) and the periphery of V1 (median: 35.66% CI: 32.87%:38.33%, p=0.05), V2 (median: 37.96% CI: 35.00%:41.02%, p=0.01) and V3 (median: 37.04 CI: 33.90%:40.10%, p=0.01) (**Fig 4.3**).

Next, we tested whether we could decode visual scene category from occluded voxels in early visual cortex and whether or not a classifier trained on one modality could decode the other. To do this analysis we chose only those voxels that were occluded during image stimulation. An occluded voxel during visual stimulation only receives information from top-down processes involved in cortical feedback. To ensure that occluded voxels did not receive any spill-over from adjacent non-occluded voxels, we applied two thresholding measures. This process results in non-stimulated voxels from dorsal V1, V2, and V3 of the left hemisphere. From these voxels we find that we can decode image category at significantly above chance levels in the occluded patch as a whole (median: 54.26% CI: 49.68%:58.89%, p=0.01), occluded V1 (median: 50.53% CI: 45.58%:55.84%, p=0.01), occluded V2 (median: 44.47% CI: 41.79%:47.32%, p=0.01) and occluded V3 (median: 42.53% CI: 38.32%:46.89%, p=0.01) (**not shown**).

Having demonstrated using a pattern classification approach that we can decode auditory scenes in early visual cortex while participants have their eyes open, and that we can decode visual scenes from occluded voxels, we tested whether feedback information was generalisable between modality. To do this we trained pattern classifiers on n-1 runs of data for one modality and tested the classifier on the nth run of data for the other modality.

We tested whether a pattern classifier trained to discriminate between auditory categories could decode video category. We find that this classifier can decode video category significantly above chance in occluded V3 only (median: 35.21% CI: 33.23%:37.30%, p=0.05) (**Fig 4.4**). We then tested whether a pattern classifier trained to discriminate between video categories could decode auditory category. We find that the visual pattern classifier can decode auditory category at a level significantly above chance in the occluded patch as a whole (median: 35.00% CI: 33.52%:36.57%, p=0.05), occluded V2 (median: 35.10% CI: 33.02%:37.08%, p=0.05) and occluded V3 (median: 36.04% CI: 33.65%:38.33%, p=0.05) (**Fig 4.5**).

55

The success of the cross-classification analysis suggests that there is some overlap in the neuronal representation in early visual cortex of the auditory stimuli that also exists in the visual stimuli. This could be related to the context of the stimulus. We cannot fully discount the role imagery might play in this finding as it is not something we could control in this experiment.

Although our effect sizes are small in these experiments – as is to be expected in an eyes open experiment with auditory stimuli, they are consistent and replicable with our methods across different experiments and different laboratories (van den Hurk, Van Baelen, & de Beeck, 2017).
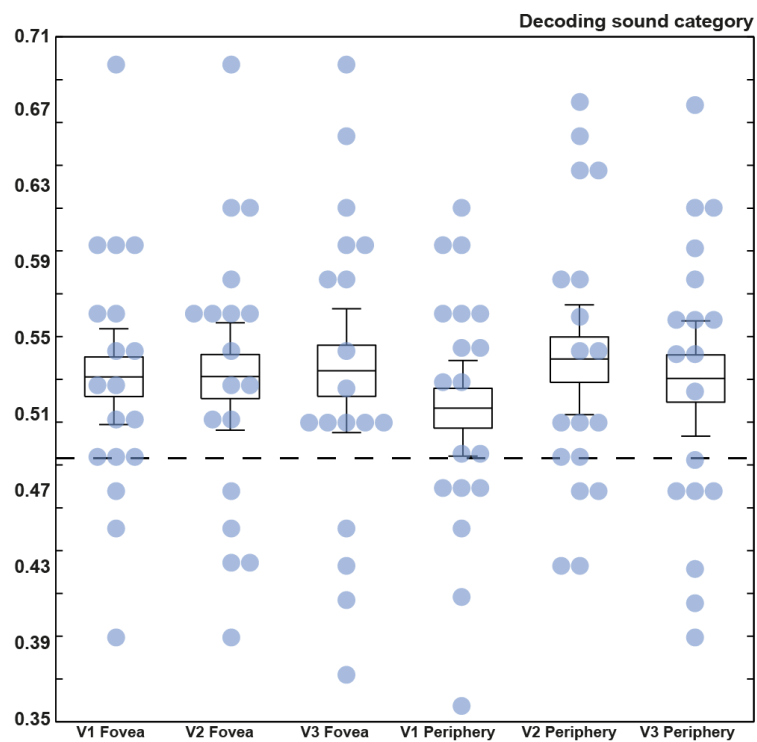
Figure 4.3: Experiment 2, decoding cars, cows and hammer scene object-focused sounds in early visual cortex.
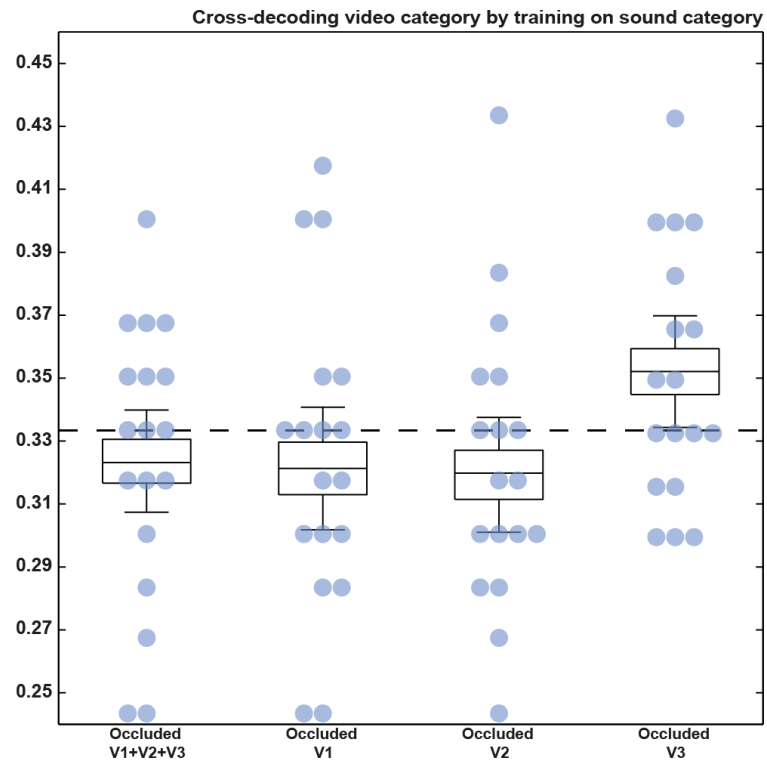
Figure 4.4: Experiment 2, cross-classification in early visual cortex. Training the classifier to discriminate between sound categories then testing it on visual category.
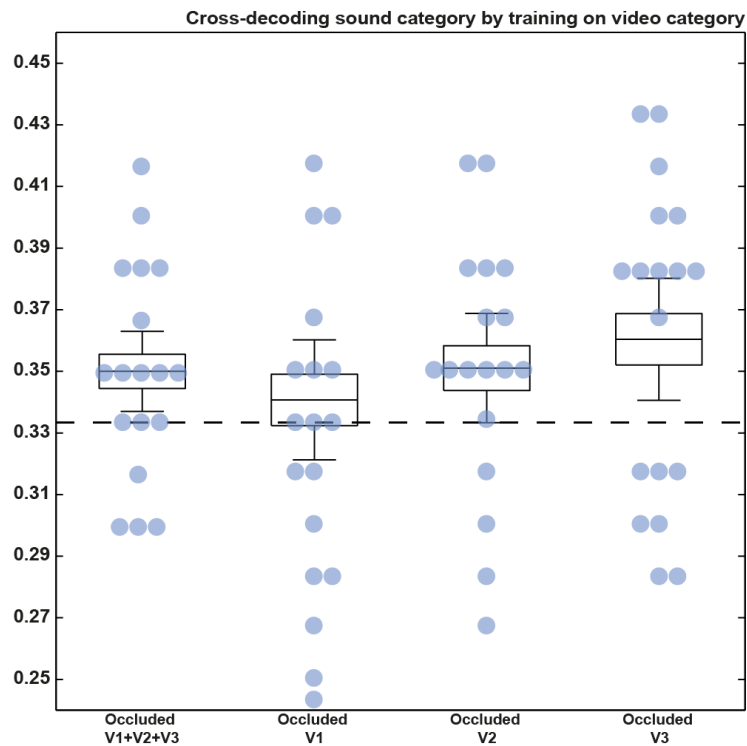
Figure 4.5: Experiment 2, cross-classification in early visual cortex. Training the classifier to discriminate between visual categories then testing it on sound category.

4.4   Discussion

4.4.1   Auditory object information is important to the fovea.

Visual feedback information is highly organised in early visual cortex (Marques, Nguyen, Fioreze, & Petreanu, 2018). It has been demonstrated that visual feedback information about objects placed in the periphery can be decoded using multivariate pattern classifiers in non-stimulated fovea, but not in other non-stimulated areas of the periphery (Williams et al., 2008). This information was correlated with perceptual discrimination suggesting that feedback of object information to the fovea is important for discriminating peripheral objects. Consistent with this, in experiment one, we found that we could read out information about auditory objects in non-stimulated foveal V2 and V3 but not in non-stimulated periphery. Further, we show that auditory scene information could be read out only in non-stimulated periphery of V2.

There is evidence suggesting auditory cortex is functionally connected with the periphery of visual cortex (Eckert et al., 2008; Klinge et al., 2010; Qin et al., 2014) but less evidence it is functionally connected with the fovea. Studies using diffusion tensor imaging in humans have shown that Herschel's gyrus is connected to the periphery and para-fovea of the visual cortex (Anton L Beer, Plank, & Greenlee, 2011; Eckert et al., 2008). Transcranial magnetic stimulation (TMS) at subthreshold intensity applied to the occipital visual cortex coupled with an auditory stimulus leads to the perception of phosphenes (Bolognini et al., 2010). In this same study they found that the optimal site of stimulation was in the peripheral visual field. When stimulation was in the central visual field, there was no induction of phosphenes (Bolognini et al., 2010). Further evidence supporting extrastriate cortex as the encoding space of auditory feedback comes from a fMRI study reporting that incongruent video-audio pairings decreases classifier accuracy in V2 and V3, but not in V1, compared with congruent video-audio pairings (de Haas, Schwarzkopf, Urner, & Rees, 2013).

In non-human primate physiology, monosynaptic afferents from auditory association areas have been reported to terminate in striate and extrastriate cortex. These terminations occur in layers 1 and 6 of the cortical laminae following a feedback profile (Anton Ludwig Beer, Plank, Meyer, & Greenlee, 2013). They also follow a density gradient described as decreasing from extrastriate to striate cortex, and decreasing from peripheral areas towards the occipital pole. A similar pattern of connectivity has also been shown in blind mice where primary auditory afferents terminate in both primary and secondary visual cortices. These afferents terminate close to secondary visual cortical neurons with feedback connections to primary visual cortex (Laramée et al., 2011).

Taken together, these results support the hypothesis that auditory feedback influences perception at both the periphery and the fovea. This is further supported by our results from experiment 2. Using sounds that are scenes but contain a focal object sound, decoding context is possible in both the periphery and fovea. Auditory feedback to the fovea may be more specific, containing information about objects and may be driven by higher order areas such as lateral occipital cortex rather than directly from auditory areas.

4.4.2 Auditory and visual feedback information about scenes is contextual and generalisable.

From previous work we have shown that auditory feedback (Vetter et al., 2014) and visual feedback (F. W. Smith & Muckli, 2010) is contextual, which we replicate here. It seems like a straightforward, yet non-trivial step to show that these two types of feedback information contain similar information for contextually matching audio and visual stimuli. In experiment 2, we found that a classifier trained on auditory information in non-stimulated V3 - to discriminate between three sounds - could decode three contextually matching videos from visual feedback information alone. A classifier trained on visual feedback information in non-stimulated V2 or V3 - to discriminate between three videos - could decode three contextually matching sounds from feedback information alone. For the first time we show that auditory and visually induced feedback share an overlap in their neural code at

early levels of visual processing. The overlap in the representation of context between auditory and visual feedback indicates that both types of feedback can guide perception in a similar way.

It is noteworthy that the auditory feedback classifier can only decode visual feedback information in V3, but the visual feedback classifier can decode auditory feedback information in V2, V3 and the occluded area as a whole (V1, V2 and V3). We propose that this is due to auditory feedback being coarser than visual feedback within the domain of vision. A classifier trained on a coarse signal or pattern will result in a poorer model than say a classifier trained on a finer signal or pattern. The quality of the classifier model within this framework would explain the pattern of results we observe. It then remains an open question as to how such a coarse auditory signal aids visual perception. One possibility is that a contextually matching sound facilitates identification of degraded contextually matching visual stimuli (Lu, Zhang, Xu, & Liu, 2018).

It is also interesting to note that a classifier trained on auditory feedback information in early visual cortex cannot decode visual feedforward information. This is also true in the reverse case for the classifier trained on visual feedforward information. The feedforward signal is a fine-grained signal. That we find no overlap in the neural code of these two types of information would indicate that the modulatory effect of feedback is related to higher-order information and not the low level information that early visual cortex extracts from sensory information. This is further supported by those studies that show auditory and feedforward visual information do indeed cross-classify but only in higher order multi-sensory integration areas (Jung et al., 2018).

4.4.3 Theoretical interpretations.

Within a predictive coding framework (Rao & Ballard, 1999), our results can be interpreted as auditory and visual scenes triggering high-level internal models which generate predictions about specific stimulus features. These predictions back-propagate along descending feedback pathways generating error signals

which propagate forward along ascending pathways. When the prediction is well matched the error signal is minimised – the prediction was a good one. If the prediction is less well matched the model is updated and a new prediction is issued. In this manner early sensory areas are adapted with a predictive model for upcoming statistically regular sensory information. This would explain why cross-classification between modality is possible if auditory and visual stimulation are triggering the same internal models.

Our finding that we can decode scene information in the periphery and object information in the fovea indicates that we may have two functionally distinct signals. We observe one signal in the periphery related to content specific scene information, which previous work has shown is generalisable across categorical exemplars (Vetter et al., 2014). Here we show that this signal is also generalisable across modality in extrastriate cortex. One explanation for this peripheral signal could be that it is used as a spatial co-ordinate system for assigning semantic labels to other low-resolution features within this space. An example of this would be keeping track of cars as they enter and exit peripheral vision as you walk down the street. The high spatial resolution of visual cortex compared with auditory cortex would make visual cortex well suited to this task (McDonald, Teder-SaÈlejaÈrvi, & Hillyard, 2000). This could explain why cross-classification is possible in visual areas but not auditory areas. However we do not tonotopically map auditory cortex which could also explain this null finding.

It is intriguing that auditory and visual feedback information is generalisable at the level of extrastriate cortex but not primary visual cortex. It is possible that due to auditory processing beginning in the cochlea and brainstem, that when this processed information arrives in visual cortex, it is coherent only at the level of extrastriate cortex where visual information has started to undergo some level of processing. Auditory feedback to extrastriate cortex does not preclude an auditory influence on primary visual cortex. In the macaque, auditory afferents are shown to terminate next to V2 neurons with feedback projects to V1 (Rockland & Ojima, 2003). Within a predictive coding framework, auditory information at V2 could refine the prediction that will descend to V1. A functional role of auditory feedback to extrastriate cortex could be to refine predictions for informing primary visual cortex.

61

Why auditory and visual feedback is generalisable at all is an intriguing question. Previous work has reported that there is no overlap in the neural representation of auditory and visual stimuli in low level visual cortex (Jung et al., 2018). The difference between our study and the study by Jung et al. (2018) is that we used dynamic videos instead of static images. By controlling how well the regularity of auditory and visual information matched between categorically matching exemplars may explain why we find the overlap in neural representation between modality that Jung et al. (2018) did not. A recent study by Andric, Davis, and Hasson (2017) supports this conclusion. In this study the researchers show that the regularity of auditory and visual stimuli is an important feature for multisensory audiovisual perception, with specific brain systems for detecting regularity mismatch. Regularity matching audio and visual stimuli may explain this finding.

We also observe a second signal in the fovea related to content specific object information. This type of signal could help sensitise the much higher-resolution fovea for the purpose of object foveation prior to orienting behaviour. It has been shown that a burst of pink noise prior to a flash triggers involuntary orienting which in turn aids visual perception (McDonald et al., 2000). Auditory object related feedback to the fovea could contain information about specific stimulus features which guide perception in an anticipatory way. It remains an open question whether signals from object sounds can be generalised across additional congruent exemplars.

Another explanation for our findings is that auditory stimulation triggers visual imagery which is why it is well matched with visual feedback during visual stimulation. One argument against this is a study showing that a visuospatial imagery task does not inhibit decoding of auditory information in early visual cortex (Vetter et al., 2014). This would indicate that visual imagery and auditory driven predictions are separate yet overlapping signals in early visual cortex.

Another possible explanation could be that what we are decoding is attentional modulation in early visual cortex. It has been shown that attentional modulation can occur for both spatial and feature based responses (Peelen & Kastner, 2014) and has its strongest influence towards the periphery (Bressler, Fortenbaugh, Robertson, & Silver, 2013). This could explain why we observe a bias towards the

periphery for both decoding and cross-decoding in our experiments. However all of our sounds show deactivation in early visual cortex with no significant univariate differences. Furthermore at 7T, it has been shown that for sounds BOLD amplitude and information content are independent from one another (Muckli et al., 2015).

In conclusion, we demonstrate that auditory feedback to early visual cortex follows similar eccentricity biases observed in visual processing of visual object and scene representation. Furthermore, we show that auditory and visual feedback share an overlap in their neural representation when the spatial and temporal regularity of the stimuli are well matched. We provide evidence supporting the role of early visual cortex as being an area for multisensory perception.

4.5   Methods

4.5.1   Subjects.

In total, 38 subjects (age range 20-39; 21 female) took part in these experiments. All subjects gave written informed consent and were compensated monetarily for their participation. Subjects were recruited through the University of Glasgow subject data pool and ethical approval was attained from College of Science and Engineering, University of Glasgow.

4.5.2   Stimuli and Procedures

Experiment 1. Here we used two auditory scene sounds and two auditory object sounds. The scene sounds were a blizzard sound and a forest sound, each twelve seconds long. The two objects sounds were an axe sound and a pull-motor sound, each one second long. Scene sounds were presented in a block design in a pseudo-randomised order, with an inter-stimulus interval of twelve seconds. Each of the two object sounds were played either 7 or 8 times (counterbalanced) during

the presentation of a scene sound (**Fig 4.6a**). The presentation of each object sound repetition was randomised so that it stood out as being independent of the scene. This was done to reduce mind wandering or disengagement from the stimulus during the twelve seconds. The design resulted in trials that were forest-axe, forest-motor, blizzard-axe or blizzard-motor. Subjects had their eyes open and were asked to fixate on a chequerboard (9 x 9 pixels or 0.2 x 0.2 degrees in the fovea) in the centre of the screen. Subjects were tasked with responding to a 100ms colour fixation change during the baseline.

Experiment 2. Three scene sounds (Car, Cow, and Hammer) were taken from three videos. The videos were then muted and used as unimodal visual stimuli (**Fig 4.6b**). Videos were displayed in full colour on a rear projection screen using a projector system (1024 x 768 resolution, 60Hz refresh rate). Videos measured 800 x 800 pixels and spanned 22 x 15 degrees of visual angle. A circular mask was placed over the video such that the lower right quadrant was occluded throughout the experiment (occluded region spanned = 10 x 8 degrees). A centralised fixation chequerboard (9 x 9 pixels or 0.2 x 0.2 degrees in the fovea) marked the centre of the screen. A default grey background (128, 128, 128 RGB) was used throughout. Participants were instructed to maintain central fixation at all times and to respond to colour changes of the fixation cross (black/white to red/green) reporting category. Each of four runs consisted of fifteen blocks of independently presented sound and visual stimulation with an inter-stimulus interval of twelve seconds. Each run was followed by two mapping blocks. The mapping blocks involved 12 seconds of stimulation of a contrast inverting chequered board around the inside edge of the occluder followed by 12 seconds of stimulation of a contrast inverting chequered board inside of the occluder (**Fig 4.6c**). This was repeated twice in each run. All three experiments were performed using Presentation® software (Version 0.70, Neurobehavioral Systems, Inc., Berkeley, CA, **www.neurobs.com**).
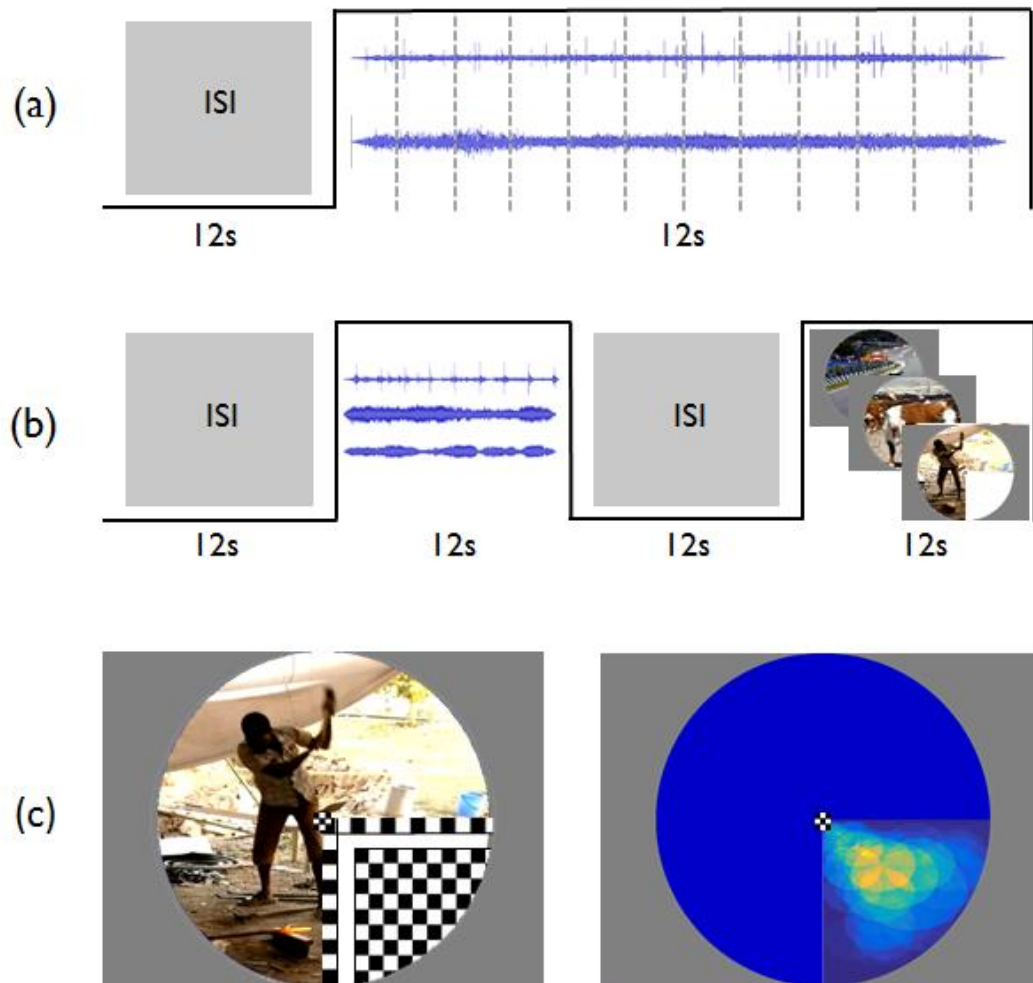
Figure 4.6: (a) Experimental design of experiment 1, (b) Experimental design of experiment 2, (c) Contrast mapping of occluder and occluder surround (left) and pRF estimate 'heat map' within the occluded region (right).

4.5.3 FMRI acquisition.

Data were collected at the Centre for Cognitive Neuroimaging, University of Glasgow. Functional and anatomical MRI data was acquired using a 3 Tesla Tim Trio MRI scanner (Siemens, Erlangen, Germany) with a 32 channel head coil and integrated parallel imaging techniques (IPAT factor: 2). Functional EPI sequences were used to acquire partial brain volumes of resolution 3mm isotropic, inter-slice thickness of 0.3mm, TR-2s, TE-30, 35 slices, FOV of 220mm and a flip angle of 62°. Anatomical scans (3D MPRAGE) were acquired with a resolution of 1mm isotropic, TR-2300ms, TE-2.96ms and 192 volumes.

5 Decoding a future, unseen representation of the world in extrastriate cortex during occlusion

5.1 Abstract

The visual system uses internal representations to generate predictions about upcoming visual information. These predictions are generated in high level cortical areas and are carried to sensory cortices by feedback pathways. Understanding predictive processes is therefore essential for understanding how the brain internally models the external world. During brief movie-viewing, the visual system extrapolates upcoming information based on previously presented frames and information from memory. We probed these forward models in early visual cortex by introducing a 800ms gap after a 600ms movie priming sequence. We used 3T-fMRI to investigate how early visual areas predict images that are expected after a 800ms interval. By probing predictions with test frames that would naturally occur within the movie at 300ms, 600ms, 800ms and 4000ms, we found that forward models best matched the 600ms and 800ms test frames. In a control experiment we found that forward models no longer existed when the priming sequence was shuffled and that accurate predictions take 600ms to 1200ms to form and fully develop. These results indicate that visual cortex extrapolates beyond previous stimulation to generate predictions about future events and that these predictions are temporally imprecise.

5.2   Introduction

An everyday mundane task like crossing the road requires a series of complex cognitive processes. The brain must extract both spatial and temporal information from the local environment and compare this information with what it already knows from memory to determine when it is or is not safe to cross the road. This scenario becomes even more complicated as the flow of traffic increases and multiple targets must be tracked at once. Each vehicle must enter awareness and its trajectory estimated and projected in to the immediate future to determine when to cross. But what are the neural mechanisms by which the brain achieves this task so effortlessly?

Both context and scene analysis are thought to rely on a network involving the medial temporal lobe (MTL), medial parietal cortex (MPC) and the medial prefrontal cortex (MPFC) (Bar & Aminoff, 2003). Sub-regions of the MTL system, including the hippocampus, parahippocampal cortex (PHC) and perirhinal cortex (PRC) code for spatial and context information with both long-term episodic memory (Burgess, Maguire, & O'Keefe, 2002) and short-term working memory (Libby, Hannula, & Ranganath, 2014). The hippocampus also contains allocentric place cells which code for spatial information (O'Keefe, Burgess, Donnett, Jeffery, & Maguire, 1998). Spatial information and memory are thought to be bound by the retrosplenial cortex sub-region of the MPC system (Marchette, Vass, Ryan, & Epstein, 2014; Vann, Aggleton, & Maguire, 2009), likely as a mechanism for navigation using previously experienced events stored in episodic memory (R. A. Epstein, 2008). Retrosplenial cortex is characterised by theta oscillations during rest and has been shown to phase lock with MTL prior to autobiographical memory retrieval (Foster, Kaveh, Dastjerdi, Miller, & Parvizi, 2013).

The networks described share considerable overlap with the default mode network (Andrews-Hanna et al., 2010). The default mode network has been implicated in internalised cognitive processes such as future thinking (Schacter et al., 2017), imagery (Pearson & Kosslyn, 2015) and mental simulations (Hamrick et al., 2016). Extrapolation of internal models could be one mechanism by which the brain achieves the crossing the road task. Such a mechanism would rely heavily on

67

initiating, tracking and updating of both spatial and temporal information of objects in the environment.

Keeping track of temporal events is thought to involve a motor network (Merchant & Yarrow, 2016) that includes the cerebellum (Ivry, Spencer, Zelaznik, & Diedrichsen, 2002). The cerebellum is connected with the cerebrum by a top-down corticopontine-pontocerebellar circuit and a bottom-up cerebellothalamic-thalamocortical circuit (Baumann et al., 2015). When the cerebellum is damaged not only are there sensory and motor impairments (Therrien & Bastian, 2015) but also an inability to adjust temporal models (O'Reilly, Mesulam, & Nobre, 2008) leading to the suggestion that one role of the cerebellum is to update temporal information about internal predictive models (Roth, Synofzik, & Lindner, 2013). The cerebellum is also thought to be linked with the hippocampus which has also been shown to represent temporal information in addition to spatial information (Nielson, Smith, Sreekumar, Dennis, & Sederberg, 2015).

Within a predictive coding framework, we hypothesised that initiated internal predictions would be updated and adjusted over time and that these predictions would filter down to low level visual areas. To test this we developed a paradigm based on signal attenuation as a function of time. In a fast event-related fMRI design, we first presented six frames in sequence from a short video clip. Each frame was presented for 100ms and could be chosen from three different videos. Two videos had an allocentric perspective, one of people moving through a crowded market and the second of a boat passing a dock. The third video had an egocentric perspective of a car driving along a road. Following the initial frames a 800ms period of black screen was shown. Afterwards, a test frame was presented for 100ms. The test frame could be the frame expected after the period of black screen. Alternatively the test frame could be the frame expected after 300ms, 600ms or 4000ms after the offset of the six frame sequence. We surmised that the six sequential frames would initiate a prediction which would be extrapolated across the 800ms of black screen and match with the 800ms test frame. This would result in an attenuated BOLD response to the test frame versus the 4000ms test frame. We included 300ms and 600ms test frames to test the accuracy of predictions. We investigated signal attenuation of BOLD in retinotopic visual areas as well as the hippocampus, retrosplenial cortex and the cerebellum.

68

In a second control experiment we replicated the design of the previous paradigm but included a control condition were the six initial frames are shuffled. We hypothesised that in this control condition there would be no prediction thus no signal attenuation. We compared this condition with the condition where the six frames were in consecutive order. In each condition the test frame was the 4000ms test frame. Again, we hypothesised that regardless of whether the prediction is matched or not the condition that generates a prediction would be attenuated when compared with the shuffled condition. We also included two similar conditions with 200ms frame presentation. These conditions were to test whether or not more time is required to develop more accurate predictions.

5.3   Results

To safely navigate the world around us the brain must extract spatial and temporal information from the environment, parse this information from what it already knows about the world and make predictions about the immediate future. Within a predictive coding framework we hypothesised that these predictions about the immediate future would filter down to low level visual areas for the purpose of hypothesis testing incoming sensory input.

To do this we used a signal attenuation paradigm with fast event-related fMRI (**Figure 5.1**). We presented subjects with six sequential image frames taken from one of three videos. These test frames would initiate a prediction about the state of the world as it unfolds over time. After a 800ms
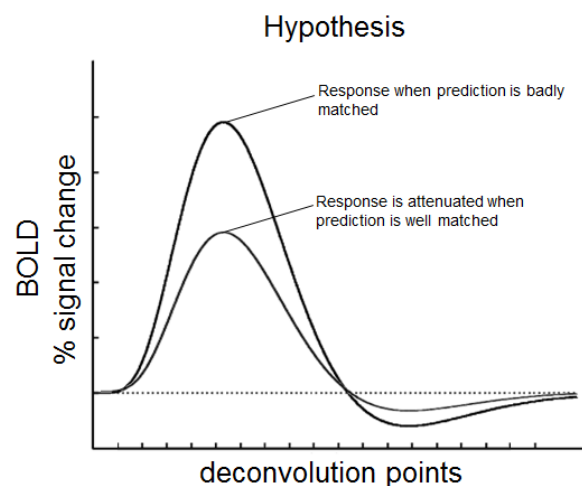
Figure 5.1: The initial sequence of frames generates a prediction about how the environment is unfolding. This prediction can be probed with a test frame. If the test frame matches the prediction then the BOLD response will be attenuated.

blank period we tested these predictions with one of four test frames. The test frames were the frame expected after 300ms, 600ms, 800ms and 4000ms. We tested whether the 800ms test frame matched the prediction as would be expected compared with the 4000ms test frame. We included 300ms and 600ms test frames to test the accuracy of predictions. Our hypothesis was that the BOLD response signal would be attnuated when the prediction matched the test frame (**Fig 5.1**) and that the attenuation would vary depending on prediction accuracy. For analysis we used deconvolution general linear modelling and tested contrasts between the 300ms, 600ms and 800ms test frames with the 4000ms test frame in retinotopically mapped early visual cortex, hippocampus, retrosplenial cortex and cerebellum in experiment one.

In experiment two we tested the hypothesis that BOLD signal would be attenuated for the condition in which the six initial frames are in order versus when they were in a pseudo-random order even when the test frame is a mismatch. To do this we only used the 4000ms test frame. We also tested the hypothesis that predictions need time to develop so we included conditions in which the test frames were presented for both 100ms as in experiment one, and 200ms.

5.3.1 Temporal predictions feedback to extrastriate cortex.

In experiment one, we first analysed differences in signal attenuation between conditions across the peak of the signal. This corresponds to deconvolution points 5, 6 and 7. We found that all test frames were a significant match in left hemisphere ventral V3 (300<4000: $t = 2.23$, $p = 0.03$, 600<4000: $t = 3.7$, $p = 0.0002$, 800<4000: $t = 2.76$, $p = 0.005$) (**table 5.1**).

Next, we analysed differences in signal attenuation between conditions after the peak of the signal. This corresponds to deconvolution points 6, 7 and 8. We found that all test frames significantly matched in left hemisphere ventral V3 (300<4000: $t = 2.48$, $p = 0.01$, 600<4000: $t = 3.67$, $p = 0.0002$, 800<4000: $t = 2.94$, $p = 0.003$).

Test frames 600ms and 800ms were a significant match in left hemisphere V1 (600<4000: t = 2.27, p = 0.02, 800<4000: t = 2.21, p = 0.03) and right hemisphere V1 (600<4000: t = 2.03, p = 0.04, 800<4000: t = 2.32, p = 0.02).

Only the 600ms test frame was a significant match in left hemisphere dorsal V2 (600<4000: t = 2.16, p = 0.03), left hemisphere dorsal V3 (600<4000: t = 2.44, p = 0.01) and right hemisphere dorsal V2 (600<4000: t = 2.12, p = 0.03) (**table 5.2**).

In non-visual regions of interest we find that all test frames match the prediction in the pulvinar (300<4000: t = 2.86, p = 0.004, 600<4000: t = 2.04, p = 0.04, 800<4000: t = 3.147, p = 0.002) and retrosplenial cortex (300<4000: t = 2.44, p = 0.01, 600<4000: t = 2.071, p = 0.038, 800<4000: t = 3.129, p = 0.002). In the cerebellum we find the 800ms a match to the prediction in the anterior cerebellum (800<4000: t = 2.05, p = 0.04) and the anterior vermis (800<4000: t = 2.334, p = 0.02).

Our results from experiment one demonstrate that predictions generated from watching short video clips are extrapolated across short intervals of time. The fact that the 300ms test frame is not a match to the prediction in most cases suggests that the prediction is not just related to the information of the frame sequence at offset but is extrapolated over time. That the prediction matches both the 600ms and 800ms suggests that some temporal constriction happens within the brain. These predictions also appear to be stronger in extrastriate cortex, particularly in the left hemisphere.

5.3.3 Temporal predictions take between 600ms and 1200ms to develop.

In experiment two, we analysed differences in signal attenuation between conditions across the peak of the signal. This corresponds to deconvolution points 5, 6 and 7. We contrasted the sequential frame condition, in which a prediction was expected, with the shuffled frame condition, in which a prediction was not expected. We did this for both 100ms and 200ms frame presentation times. In each case the test frame was the 4000ms frame. When comparing signal attenuation between Expected100ms < NonExpected100ms we only found a significant difference in right hemisphere ventral V2 ($t = 2.29$, $p = 0.02$). When comparing signal attenuation between Expected200ms < NonExpected200ms we found a significant difference in almost every region of interest with the exception of left hemisphere V1, left hemisphere dorsal V2, right hemisphere V1 (**table 5.3**).

When predictions are well developed they are found in both hemispheres but again show a bias towards extrastriate cortex.
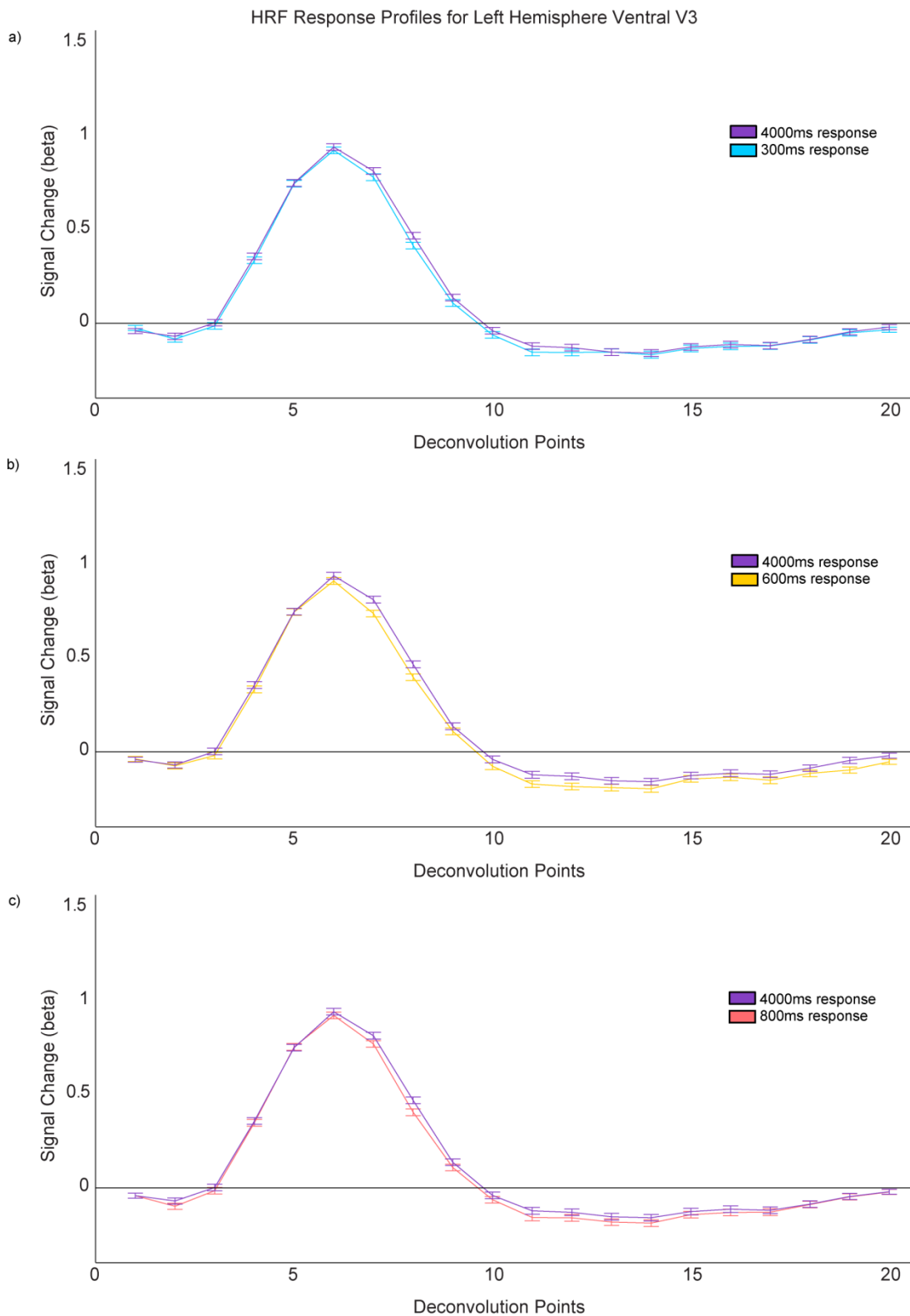
Figure 5.2: a) Response profile of the HRF for the 4000ms condition and the 300ms condition, showing significant difference across deconvolution points 6,7,8 b) Response profile of the HRF for the 4000ms condition and the 600ms condition, showing significant difference across deconvolution points 6,7,8 c) Response profile of the HRF for the 4000ms condition and the 800ms condition, showing significant difference across deconvolution points 6,7,8.

73

Table 5.1: Experiment 1. General linear model contrasts showing their respective t and p values for each region of interest, across deconvolution points 5, 6 & 7.

ROI	300 < 4000		600 < 4000		800 < 4000	
	T-value	P-value	T-value	P-value	T-value	P-value
LH V1	1.21	0.22	1.81	0.07	1.53	0.12
LH V2d	0.88	0.38	1.65	0.09	-0.4	0.68
LH V2v	1.22	0.22	1.68	0.09	1.38	0.17
LH V3d	1.58	0.11	1.92	0.055	0.6	0.55
LH V3v	2.23	0.03*	3.7	0.0002**	2.76	0.005*
RH V1	1.21	0.22	1.64	0.1	1.82	0.07
RH V2d	1.14	0.25	1.57	0.12	0.57	0.57
RH V2v	1.00	0.32	1.19	0.24	1.4	0.16
RH V3d	0.1	0.91	0.93	0.35	0.82	0.41
RH V3v	1.42	0.15	1.29	0.2	1.69	0.09

* Significant after correction **Significant after correction for multiple comparisons

Table 5.2: Experiment 1. General linear model contrasts showing their respective t and p values for each region of interest, across deconvolution points 6, 7 & 8.

ROI	300 < 4000		600 < 4000		800 < 4000	
	T-value	P-value	T	P-value	T	P-value
LH V1	1.41	0.16	2.27	0.02*	2.21	0.03*
LH V2d	1.08	0.28	2.16	0.03*	0.5	0.61
LH V2v	1.22	0.22	1.85	0.07	1.86	0.06
LH V3d	1.71	0.09	2.44	0.01*	1.48	0.14
LH V3v	2.48	0.01*	3.67	0.0002**	2.94	0.003*
RH V1	1.37	0.17	2.03	0.04*	2.32	0.02*
RH V2d	1.48	0.14	2.12	0.03*	1.29	0.2
RH V2v	1.03	0.30	1.18	0.24	1.56	0.12
RH V3d	0.47	0.64	1.33	0.19	1.24	0.22
RH V3v	1.34	0.18	1.26	0.21	1.88	0.06

* Significant after correction **Significant after correction for multiple comparisons

74

Table 5.3: Experiment 2. General linear model contrasts showing their respective t and p values for each region of interest, across deconvolution points 5, 6 & 7.

| ROI | Exp 100 < NonExp 100 | | Exp 200 < NonExp 200 | |
	T-value	P-value	T-value	P-value
LH V1	0.5	0.62	1.31	0.19
LH V2d	0.69	0.49	1.69	0.09
LH V2v	1.67	0.1	2.72	0.007*
LH V3d	-0.097	0.92	2.07	0.04*
LH V3v	1.62	0.1	2.47	0.01*
RH V1	0.92	0.36	1.76	0.08
RH V2d	0.64	0.52	2.83	0.005**
RH V2v	2.29	0.02*	2.8	0.005**
RH V3d	0.89	0.37	2.25	0.02*
RH V3v	1.82	0.07	2.83	0.005**

* Significant after correction **Significant after correction for multiple comparisons

5.4 Discussion

5.4.1 Visual predictions extrapolate over time in the absence of visual input.

Input to the visual system contains both spatial and temporal information. The world we see is not a steady state but a structured set of patterns that change over time. Predictive coding theories tell us that the brain actively predicts the incoming spatial and temporal information to optimise its internal model of the world (T. S. Lee & Mumford, 2003). Accurate prediction relies on detecting patterns and predicting how they will develop over time. One prediction of this framework is that the brain will continue to generate predictions after the sensory stimulation is removed. The question remains, is the prediction related to the spatial information of the stimulus prior to stimulus removal, or is the prediction extrapolated over time becoming a prediction that is unrelated to the stimulus?

We tested this in experiment one. We found that a sequence of images viewed as a short video clip triggers a prediction. When the visual stimulation is removed the prediction continues to unfold. We probed this prediction after a blank interval of 800ms with a series of test frames. We found that the prediction best matched with the 600ms and 800ms test frames. The prediction rarely matched with the 300ms test frame. This tells us that the prediction is unrelated to the 300ms test frame. Therefore it is likely that the prediction generated by the priming sequence continues to be extrapolated over the entire length of the 800ms blank interval which is why it best matches the 600ms and 800ms test frames. After the interval, the prediction matches a future, unseen, version of the world.

The generated prediction containing information about a future, unseen, version of the world is likely due to mental simulation processes (Ullman, Spelke, Battaglia, & Tenenbaum, 2017). Mental simulation suggests that the brain can make predictions about the future based on simulations of internal models. Children as young as 5-months old can make inferences about how things in the world will behave and learn to generalise this information to objects with similar properties (Hespos, Ferry, Anderson, Hollenbeck, & Rips, 2016). Recent computational

76

modelling work suggests that the brain may achieve this in a two-pronged approach. First by making inferences in a bottom-up approach based on visual features, then by using this feature-based inference as a starting point for top-down simulations (Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2018). Our findings fit within this theory. The priming sequence provides the brain with spatial and temporal features that act as a starting point to make top-down predictions about the immediate future.

5.4.2 Future predictions about the world are temporally imprecise.

In our first experiment we use a priming sequence of frames to induce a prediction about the future state of the world. After 800ms of black screen we probe these predictions with a test frame. The test frame can either be the frame that would appear after 800ms from the offset of the priming sequence, a frame close in time to the priming sequence (300ms) or a frame in-between (600ms). We find that the 600ms and 800ms test frames are the best match to the subject's prediction. Subject predictions best match both the expected frame (800ms) and an earlier frame (600ms) suggest that the predictions are temporally imprecise.

Keeping track of time within the brain is thought to involve a motor network (Merchant & Yarrow, 2016) involving the cerebellum (Ivry et al., 2002). Precise timing is important for many different types of behaviour, however some motor actions are known to compress temporal judgements in both adults and children (Droit-Volet, 2017). For example short intervals of time between two briefly flashed perisaccadic visual stimuli are judged to be shorter in time than they actually are (Morrone, Ross, & Burr, 2005). Temporal compression in this example was explained by modelling the space and time transformations that are remapped in the receptive fields of visual neurons. This mechanism is described as being optimal during fixation but it is not updated perisaccadically which accounts for the timing misjudgement (Binda, Cicchini, Burr, & Morrone, 2009). Judgement of the duration of tones has also been shown to be affected by time compression when the tone is preceded by auditory click trains, flickering visual squares, expanding

77

visual circles and even white noise (Wearden, Williams, & Jones, 2017). Although our paradigm used fixation throughout, small eye movements during the blank interval could account could induce remapping of top-down information in the non-classical receptive field. Remapping of top-down predictions to new retinotopic locations has been show in a study involving apparent motion (Edwards et al., 2017).

Temporal compression is not limited to motor actions. A study of perceptual decoupling during mind wandering has shown that judgements of time are also compressed (Terhune, Croucher, Marcusson-Clavertz, & Macdonald, 2017). The authors attribute this to transient fluctuations in attentional states. This theory fits well with the finding from our experiment. During the 800ms blank interval subjects are tasked with extrapolating the sequence of events that they were shown in the priming sequence. Attentional lapses during this baseline period could account for the temporal imprecision we find in subject predictions when probed with the test frames.

Another possibility is that the imprecision of predictions is a useful functional feature for applying current internal models to new situations. Flexible predictions could be used as a starting point for constructing new internal models or dealing with the uncertainty of future events.

5.4.3 Temporal predictions arrive at extra-striate cortex but not striate cortex.

Primary visual cortex is known to receive most of its input from non-geniculate sources (Muckli & Petro, 2013). This has led to V1 being described as a blackboard for performing computations on task-relevant low-level visual information by one cortical area that can then be read by other cortical and sub-cortical areas (Roelfsema & de Lange, 2016). In our second experiment we increase the presentation time of frames to allow accurate predictions to develop. We find that we can read-out these predictions in extra-striate areas. Primary

visual cortex is the only region of interest in both the left and right hemispheres where we are unable to read-out predictions.

One explanation for this finding could be to do with the roles memory and imagery play in our paradigm. Mental imagery is known to be decodable in higher cortical areas (Reddy, Tsuchiya, & Serre, 2010). Whether imagery can be read-out from V1 appears to depend on how memory is engaged by the paradigm and the complexity of the stimuli. When a paradigm engages short-term memory or working memory, imagery is decodable in V1 for stimulus contrast (Xing, Ledgeway, McGraw, & Schluppeck, 2013), orientation (Harrison & Tong, 2009) and category (Dijkstra, Bosch, & van Gerven, 2017). However when paradigms engage long-term memory, imagery is no longer decodable in V1 for high-level representational information such as scenes (Radoslaw M Cichy, Heinzle, & Haynes, 2011), objects (S.-H. Lee, Kravitz, & Baker, 2012) and category (Reddy et al., 2010). The exception is when paradigms engage long-term memory with respect to basic or low-level features such as gratings (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Bergmann, Genç, Kohler, Singer, & Pearson, 2016), Figure contrast (Schlegel et al., 2013) and location (Radoslaw M Cichy et al., 2011). Another exception is from a study that used encoding models that encoded for low-level features. These low-level feature models were used to discern between the identities of famous paintings from imagery responses in V1 (Naselaris, Olman, Stansbury, Ugurbil, & Gallant, 2015). This dichotomy between short and long-term memory and its relatedness to complex imagery is also revealed from BOLD activity in memory related cortical areas. Imagery using short-term memory generally results in larger BOLD amplitude when compared with imagery using long-term memory (Ishai, Haxby, & Ungerleider, 2002). In our paradigm subjects were using a sequence of images to make judgements about the future state of the world. This judgement likely relied upon expectations derived from long-term memory to make complex predictions about the future of the world (Burgess et al., 2002). We also find that predictions were related to a test frame between 600ms and 800ms, but not from a 300ms frame close in time to the priming sequence. Therefore the judgements subjects were making were less likely to be related to low-level features. If they were we would expect to see

predictions matching for all test frames. Making complex future predictions from episodic memory could explain why we do not find information in V1.

5.5 Methods

5.5.1 Subjects.

Twenty three subjects (13 in experiment one, 10 in experiment two, age range 18-32; 12 female) were recruited at the School of Psychology at the University of Glasgow. Subjects were recruited with either normal or corrected vision and hearing. Ethical approval for this study was granted by the College of Science and Engineering at the University of Glasgow.
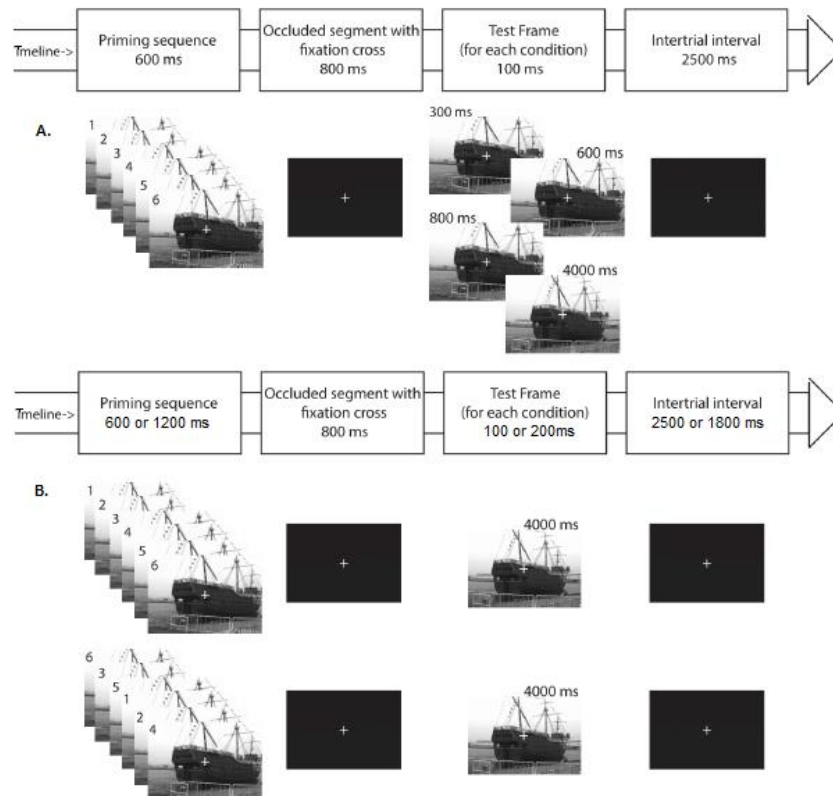
Figure 5.3: Experimental design for experiment one (A) and experiment two (B). A priming sequence in order generates a prediction. The prediction is extrapolated over time during the black fixation and probed with a test frame.

5.5.2 Stimuli and Procedures.

Three categorical videos were decomposed in to 120 individual frames. Two of the videos were from an allocentric perspective of people moving through a market square and a boat passing a dock. The third video was from an egocentric perspective of a car moving along a road. All of the test frames were greyscale.

In experiment one (**Fig 5.2a**), trials were 4000ms long. They comprised of six sequential frames chosen from the pool of 120 for each video and presented for 100ms each. A baseline period followed with a white fixation cross on a black background. Subjects were instructed to fixate throughout. After this a test frame was shown. The test frame would either be the frame expected after the period of baseline (the 800ms test frame) or a mismatching test frame (either the 300ms, 600ms or 4000ms test frame). The test frame was also shown for 100ms. The task for the subject was to respond by button press whether the test frame matched their expectation or was a mismatch to their expectation. The onset of the prime sequence during a trial was jittered as part of the design. By choosing six frames much a larger set allows each trial to have unique set of spatial and temporal features. In total we had four conditions plus a null condition, each repeated 32 times. This resulted in each run being 10 minutes and 40 seconds long. We collected 4 runs.

In experiment two (**Fig 5.2b**) we used the same design and trial structure as in experiment one. In this experiment our conditions included the prime sequence in order or a shuffled prime sequence. In each case the test frame was the 4000ms frame. These conditions either had a 100ms frame presentation time or a 200ms frame presentation time resulting in four conditions in total. In this experiment subjects were tasked with responding by button press whether the priming sequence was in order or out of order.

5.5.3 FMRI acquisition.

Data were collected at the Centre of Cognitive Neuroimaging (CCNi) at the University of Glasgow. Data were collected using a 3T Trim Trio MRI scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. Anatomical scans (3D MPRAGE) were collected with a resolution of 1mm isotropic using a TR of 2300ms, TE of 2.96ms, for 192 volumes. Functional 2-factor multi-band EPI sequences were collected with a resolution of 3mm isotropic, TR of 1000ms, TE of 26ms and coverage of 38 slices.

5.5.4 Data analysis.

We used BrainVoyager QX 2.8 (Brain Innovation, Maastricht, Netherlands) to pre-process the functional and anatomical data. First we removed the first two volumes of each run of functional data to avoid magnetic saturation effects. Next, we performed slice-time scan correction, 3D motion correction and finally high-pass temporal filtering using a 6 cycle (5 cycle for exp. 2) Fourier basis set (Kay et al., 2013). The functional data was then spatially normalised by transformation in to Talairach space and aligned with the anatomical data. The success of the alignment was checked using custom MATLAB scripts. Alignment was acceptable if the correlation between runs was $r > 0.9$. The alignment between the functional runs and the anatomical data are checked by visual inspection.

To localise V1, V2 and V3 we projected polar and eccentricity data on to a reconstruction of the grey-matter white-matter boundary. Ventral and dorsal V1, V2 and V3 were then further localised by general linear modelling all four conditions for all 20 deconvolution points. We chose only those voxels in regions of interest that were significantly active. We localised the hippocampus, retrosplenial cortex and thalamus using a Freesurfer brain atlas. The pulvinar was localised from the thalamus ROI by taking the posterior commissure as the most anterior point of the pulvinar, as described by using a Talairach atlas (Fischer & Whitney, 2012). The cerebellum was anatomically segmented into four sub-

regions; vermis, anterior cerebellum, poster cerebellum above the horizontal fissure and posterior cerebellum below the horizontal fissure.

For each region of interest we performed a 20-point deconvolution general linear model that included all subjects. In experiment one we analysed contrasts between test frames 300ms<4000ms, 600ms<4000ms and 800ms<4000ms. We analysed specifically the three points across the peak response (deconvolution points 5, 6 and 7, see **Table 5.1**) and descending from the peak response (deconvolution points 6, 7 and 8, see **Table 5.2**). In experiment two we analysed contrasts between expected and non-expected conditions for both 100ms and 200ms conditions. We analysed specifically the three points across the peak response (deconvolution points 5, 6 and 7, see **Table 5.3**).

6 Reading out the future from feedback responses in extrastriate cortex during visual navigation with fMRI

6.1 Abstract

The goal of the visual system is to solve the inverse problem of vision through the application and optimisation of prior knowledge. The brain achieves this via recurrent predictive processing from the highest visual areas down to primary visual cortex. However, the brain can also engage in perception-like thought concurrently during complex visual tasks, such as navigating your way to work in the morning. This counterfactual processing is exclusively top-down. It is unclear how the visual system maintains both factual and counterfactual streams of information. We tested this using functional magnetic resonance imaging with video clips taken from a virtual house environment. Subjects were first trained to learn the layout and the different contextual rooms of a virtual house. Then in fMRI they were given a direction cue followed by a partially occluded video of exiting one room, followed by a partially occluded video of entering a second room. The direction cue combined with the identity of the first room revealed the identity of the upcoming room before it was shown to the subject. Using multivariate pattern analysis, we show that we can decode from feedback responses (N=20) in V2 and V3 the identity of the future room during the first video sequence. In a control experiment in which the directional cue is removed (N=12) we show that we can no longer decode information about the future room. This finding demonstrates that feedback information about counterfactual processing is present in extrastriate cortex while subjects perform a navigational task.

6.2   Introduction

The human brain is highly interconnected. Communication within the brain largely takes place from one brain area to another by feedback or lateral communication. Feedback and lateral communication vastly outnumbers the traditional feedforward flow of sensory processing (Muckli & Petro, 2013). The abundance of feedback and lateral communication is contextualised by predictive coding theories (Friston, 2010; George & Hawkins, 2009; T. S. Lee & Mumford, 2003; Rao & Ballard, 1999). When the brain is engaged in a task, the feedforward sensory information is compared with information from rich internal mental models (Clark, 2013) that are learned by the brain from experience and familiarity. Information from internal models descends through all levels of sensory hierarchies to predict and refine the feedforward sensory information at each hierarchical level.

When not engaged in a task the brain is free to wander. During mind-wandering, task-related networks become less engaged and the default mode network becomes more engaged (Mooneyham et al., 2017). The default mode network is a network of structures that allow the brain to engage in counterfactual thought such as mind-wandering (Mooneyham et al., 2017), imagining (Pearson & Kosslyn, 2015), mental simulation (Ullman et al., 2017) and future thought (Schacter et al., 2017). Counterfactual thinking is an exclusively internalised process. The content of counterfactual thought, for example for imagery, has been shown to be decodable at the lowest levels of sensory processing (Albers et al., 2013; Bergmann et al., 2016; Dijkstra et al., 2017; Harrison & Tong, 2009; Xing et al., 2013). Critically, counterfactual thought does not stop the processing and monitoring of feedforward sensory information. This is most evident when we read a book. We read the black words on the white page yet we can simultaneously engage in counterfactual thought when we imagine rich and complex scenes in our mind. This leads us to the question of how does the brain achieve this? One possible solution is that information about the factual sensory experience and the internally generated counterfactual experience co-exist. A second solution is that feedback about the factual sensory environment is suppressed when we engage in counterfactual thinking.

85

To test these ideas we constructed a virtual 4-room "house" that included a kitchen, bedroom, office room and a game room. The rooms were orientated in a plus-sign configuration. The virtual house allowed us to train subjects in a new environment they had not experienced before. Subjects would have to use information from the training experience and from memory to form new internal models of the virtual house that we could then test in fMRI. During fMRI, trials were constructed starting with a 2 second directional cue (left or right) followed by a 10 second inter-stimulus interval. Next a 6 second video of a first-person perspective camera walking out of one of two rooms (bedroom or game room) was shown followed by a 12 second inter-stimulus interval. Finally a second 6 second video of a first-person perspective camera was shown walking in to one of two rooms (kitchen or office room). The directional cue and the identity of the room from the first video reveal the destination that will be shown in the final video. Subjects were tasked with maintaining central gaze fixation and told to imagine or predict the anticipated room. Each video had the bottom right quadrant occluded so that we could identify a group of voxels that were excluded from any feedforward sensory information.

We find that we can decode from brain responses in the right hemisphere (feedforward + feedback) the identity of the two rooms shown in the first video and the identity of the two rooms shown in the second video in V1, V2 and V3 replicating previous work (F. W. Smith & Muckli, 2010). We also find that we can decode from brain responses in occluded cortex (feedback only) the identity of the two future rooms during the presentation of the first video in V2 and V3. It is not possible to decode the identity of the two future rooms in primary visual cortex. We show that internal expectations about the immediate future can be read out of extrastriate cortex using 3-Tesla fMRI and multivariate pattern analysis.

86

6.3 Results

6.3.1 Information about both present and future context exists concurrently.

To date it has been shown that contextual information is made available to early visual cortex for the purposes of hypothesis testing incoming visual information (Shipp, 2016). In this experiment using functional magnetic resonance imaging (fMRI) we tested whether contextual information about the immediate future is also made available to early visual cortex. To do this we trained subjects in a four room virtual house before showing them video clips of the virtual house in a fMRI environment. During fMRI subjects were first cued with a left or right direction that lasted 2 seconds. After a 10 second baseline period, subjects were then shown a 6 second video clip of a first person perspective view of walking out of either the bedroom or games room. After a further 12 second baseline period subjects were again shown a first person perspective view of walking in to either the kitchen or the office room. Subjects were tasked with reporting whether the room they see walking in to is the room they were expecting to see based on the cued direction and the first room. Subjects were further tasked with maintaining central gaze fixation throughout the experiment.

We trained and tested four linear support vector machine (SVM) classifiers to discriminate between multivariate patterns for the following conditions; to decode the identity of the two starting rooms during the presentation of the first video, to decode the identity of the two unseen upcoming rooms during the presentation of the first video, to decode the identity of the two end rooms during the presentation of the second video, to decode the identity of the two past rooms during the presentation of the second video. Each classifier was trained on N-1 runs of data and tested on the Nth run. This produced N classifier accuracies which were averaged before the group classifier accuracies were bootstrapped to create a distribution containing the true classifier accuracy within the 95% confidence interval. This was done for both feedforward regions of interest (right hemisphere V1, V2 and V3) and feedback regions of interest (occluded V1, V2 and V3).

We found that we could reliably decode from feedforward responses the identity of the two rooms shown in both the first and second videos in V1 (first: median: 72.59%, CI: 67.12%:77.67%, p=0.05; second: median: 95.86%, CI: 93.59%:97.58%, p=0.01), V2 (first: median: 70.32%, CI: 66.17%:74.55%, p=0.05; second: median: 96.72%, CI: 95.40%:97.97%, p=0.01), and V3 (first: median: 69.92%, CI: 65.79%:74.06%, p=0.05; second: median: 93.84%, CI: 91.41%:96.01%, p=0.01). We could only reliably decode the identity of the second video from feedback responses in V1 (median: 66.72%, CI: 62.34%:70.64%, p=0.05), V2 (median: 58.36%, CI: 54.22%:62.65%, p=0.05), and V3 (median: 57.67%, CI: 53.77%:61.57%, p=0.05). Finally, we could also decode the identity of the two future rooms during the presentation of the first video only from feedback responses in extrastriate regions V2 (median: 52.90%, CI: 50.72%:55.08%, p=0.05) and V3 (median: 57.67%, CI: 53.77%:61.57%, p=0.05).

In the control study we remove the direction cue and find that we can still decode the factual experience from feedforward responses for both videos from (first video: median: 69.27%, CI: 64.19%:74.49%, p=0.01, second video: median: 98.03%, CI: 97.24%:98.81%, p=0.01) and V3 (first video: median: 62.38%, CI: 57.04%:68.37%, p=0.01, second video: median: 95.32%, CI: 93.24%:97.26%, p=0.01). However we can no longer decode any information about the future rooms. This control study demonstrates that in the primary experiment the information is related to the immediate future derived from both the cue and the context of the first room.

In this study we replicate previous work showing that context can be decoded from both feedforward and feedback responses about the current visual experience (F. W. Smith & Muckli, 2010). Importantly, we show that we can also decode from feedback responses information about the context of upcoming or expected environments before they are experienced.
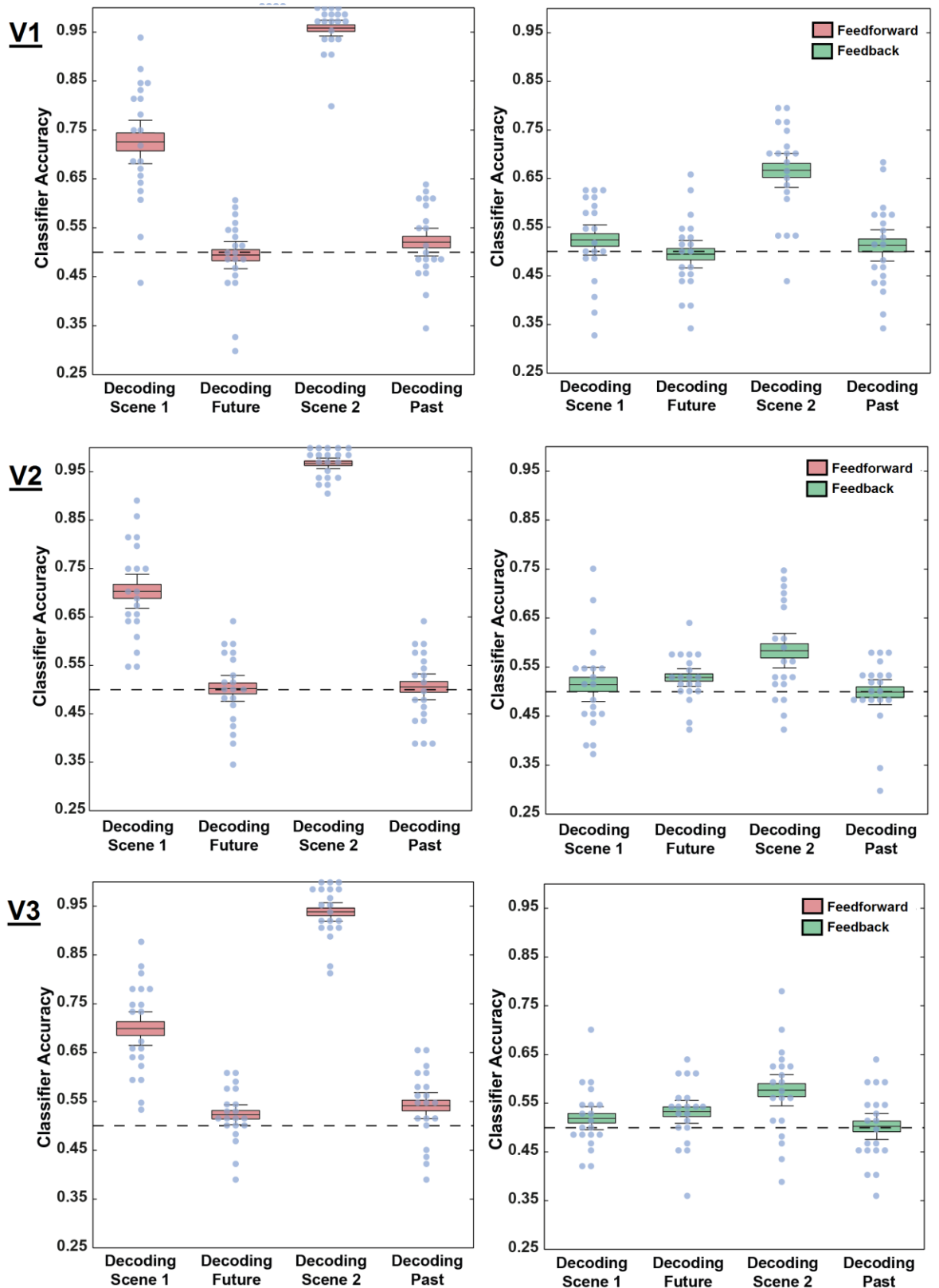
Figure 6.1: Showing decoding performance for feedforward responses in V1 (top left), feedback responses in V1 (top right), feedforward responses in V2 (middle left), feedback responses in V2 (middle right), feedforward responses in V3 (bottom left) & feedback responses in V3 (bottom right).

6.4   Discussion

6.4.1 Feedback switches between the factual and the counterfactual.

Information processing within the brain is predominantly achieved through feedback and lateral communication. When the brain is engaged in a task, a task-related network triggers feedback to low level sensory areas. When the brain is not engaged in a task but is engaged in counterfactual thought, this also triggers feedback to low level sensory areas. However when attention is turned inwards, the external world does not disappear. Therefore we can infer two conclusions from this. The first is that when engaged in counterfactual thought the factual experience, or the external world, is monitored to some extent. This is consistent with our findings. We find that we can read out from feedforward brain responses in V1, V2 and V3 the identity of the perceived rooms during the presentation of the first video while subjects are engaged in counterfactual thought. Our ability to decode information about concurrent counterfactual thought is evidence that factual information is processed through the visual system while attention is turned inwards and possibly disengaged from the external world. To what extent attention to the external world is disengaged remains unclear. The fact that we can decode both types of information suggests that attention may fluctuate between both external and internal thought.

The second inference is that both sensory processing and counterfactual processing involve feedback. It is unclear whether feedback information exists for both the factual and counterfactual experience, or switches between them both depending on the task. We tested this question with 3-tesla fMRI and find that we can decode the identity of the future upcoming rooms during the presentation of the first video from feedback responses in V2 and V3. We cannot decode the identity of the perceived rooms during the first video from feedback responses. This is also the case in the control study suggesting that this is due to a methodological issue, likely due to the centre of the visual field during the first video sequence containing uninformative information about the context of the room.

90

A second explanation could be related to perceptual decoupling. When the default mode network is engaged and attention is turned inwards for mind wandering, future thinking or mental simulation, perception becomes decoupled. To date there has been no neuroscientific evidence to explain the mechanism involved in perceptual decoupling or specifically the transition between being engaged with the external world towards being engaged in internal counterfactual thought. Perceptual decoupling could explain this null finding in both experiments if subjects are using the first video sequence to orient themselves within the virtual house.

Due to the sluggishness of the BOLD response, it cannot be ruled out from this experiment that the two signals occur independently at different times. The first video could trigger top-down feedback about the context of the room. As this response diminishes, and the subject switches to task, a second response could be initiated about the future room. This process would result in finding two neural signals in extrastriate cortex at the same time. As this question was not the focus of this experiment, we did not control for the timing of events that could trigger predictive signals.

6.4.2 Counterfactual feedback is found in extrastriate cortex.

We replicate previous work by showing that during the presentation of the second video we can decode the identity of the two perceived rooms from feedback responses in V1, V2 and V3 (F. W. Smith & Muckli, 2010). Primary visual cortex is the first significant step in visual processing and receives feedback information from many different sources including vision, audition and motor cortex. Interestingly, during the first video we can decode the identity of the two expected future rooms in V2 and V3 but not from V1. This raises the question of why extrastriate cortex may be important for counterfactual thinking.

One possible explanation is that when we engage in counterfactual processing striate cortex is reserved for monitoring the statistical regularities of the environment. One such type of regularity would be physical regularities. This is

91

consistent with the subjective experience when we drive along the motorway our mind can wander and engage in counterfactual thought, yet we remain in control of the car. Our attention is then drawn back to the task of driving when the physical regularities of the motorway in front of us changes, for example when a bend in the road appears or a car on the horizon appears. There is evidence to support this idea. The tuning properties of V1 receptive fields match the statistics of the natural world (Simoncelli and Olshausen, 2001). V1 is tuned for monitoring statistical regularity. Work by Takahashi et al. (2016) has revealed that sensory perception is modulated by Ca^{2+} activity in the apical dendrites in a subset of layer 5 pyramidal neurons in primary somatosensory cortex in mice. By manipulating the activity of apical dendrites the threshold for sensory perception could be modulated. If a similar mechanism exists in the human brain for a subset of pyramidal neurons in visual cortex, this could provide a mechanism by which perception shifts between the external and internal modulated by the content of feedback.

Another possible explanation is that it is computationally expensive to send predictions to all of early visual cortex for all of the visual field, especially V1 where receptive fields are smallest. Following a free energy principle of predictive coding (Friston, 2010) it would be less computationally expensive to send predictions only to those areas with larger receptive fields such as extrastriate cortex.

6.4.3 Theoretical interpretations.

Predictive coding theories provide a framework to explain why the brain relies heavily on feedback communication (Shipp, 2016). Top-down predictions or expectations are triggered from rich internal mental models that descend through the cortical hierarchy to hypothesis test sensory input. One problem with predictive coding theories is that they don't fully account for our ability to decouple our percept of the external world and engage in counterfactual thought. When a person engages in counterfactual thought, the sensory input is not always necessary or relevant to the counterfactual task. Yet our experience of the external

world does not simply switch-off. Therefore the brain must have a way for information about the factual experience and the counterfactual experience to either co-exist, or switch between the two. This could be achieved by a neural circuit that accommodates two separate non-overlapping processing pathways for both factual and counterfactual feedback. Recent evidence has shown that a subset of layer 5 pyramidal neurons in somatosensory cortex can modulated the threshold for perception by modulating input to the apical dendrites (Takahashi et al., 2016). This type of mechanism would allow for internalised thought while allowing for the monitoring of sensory input. An alternative to this is that there is a shared neural mechanism whereby both types of information co-exist but oscillate at different frequencies.

6.5 Methods

6.5.1 Subjects.

Thirty two subjects (age range 18-33; 16 female, N=12 for control study) were recruited from the subject data pool at the School of Psychology, University of Glasgow. Subjects were recruited without any exclusion criteria. Ethical approval was granted for this study from the College of Science and Engineering at the University of Glasgow.

6.5.2 Stimuli and Procedures.

A 3D virtual house was modelled for realism in Blender 3D and then exported to the Unity3D game engine for the purpose of using virtual reality. Images from the virtual house were rendered from Blender at 30 frames per second to create video stimuli showing a first person perspective camera moving through the house. The videos were created to present 4 different conditions. The conditions were 1) walking out of the bedroom or game room and 2) entering the kitchen or office. The camera placement was controlled in each video such that specific features,

93

such as where the walls, ceiling and floor meet, or door aperture position were the same in each frame across all 4 videos. The only difference between the videos was the contextual objects.

Subjects were first trained in virtual reality to learn the layout of the virtual house. Each subject spent 15 minutes freely exploring followed by 15 minutes of performing simple retrieval tasks instructed by the experimenter. Each room contained a number of different objects that could be interacted with. Objects were categorically matching to their room of origin.
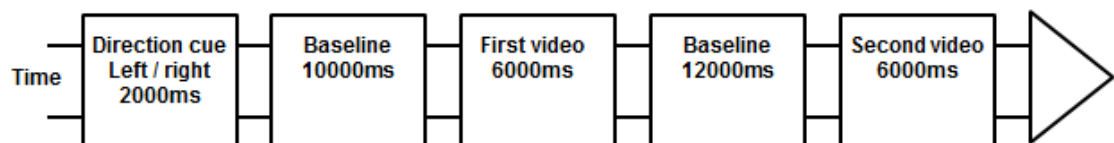
Figure 6.2: Trial sequence design.

After training, subjects were taken to MRI where they performed the experiment. Experimental trials were 40 seconds long. They consisted of a 2 second screen-centred cue for direction, either left or right. This was followed by a 10 second period of baseline. The first video clip followed for a period of 6 seconds. This video showed the camera perspective walking out of a room which could either be the bedroom or the game room. Another period of baseline followed for 12 seconds. Then the second 6 second video was shown of the camera perspective walking in to one of two rooms, either the kitchen or the office. This was followed by another 4 second period of baseline. Subjects were tasked with fixating on the centre of the screen throughout the experiment, and tasked with predicting or expecting the upcoming room during the presentation of the first video. They responded by button press during the second video if their expectation was correct.

In the control study, the direction cue was removed and subjects were tasked with responding during the second video whether the starting room was now on their left or right with respect to the direction they are facing.

6.5.3 FMRI acquisition.

We collected our data at the Centre of Cognitive Neuroimaging at the University of Glasgow. Data were acquired with a 3-Tesla Trim Trio MRI scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. Functional EPI sequences were used to collect partial brain volumes of resolution 2mmx2mmx2mm with an inter-slice thickness of 0.2mm with a TR of 1390ms, TE of 30ms and coverage of 22 slices. Anatomical scans (3D MPRAGE) were collected with a resolution of 1mmx1mmx1mm with a TR of 2300ms, TE of 2.96ms, across 192 volumes.

6.5.6 Pattern classification.
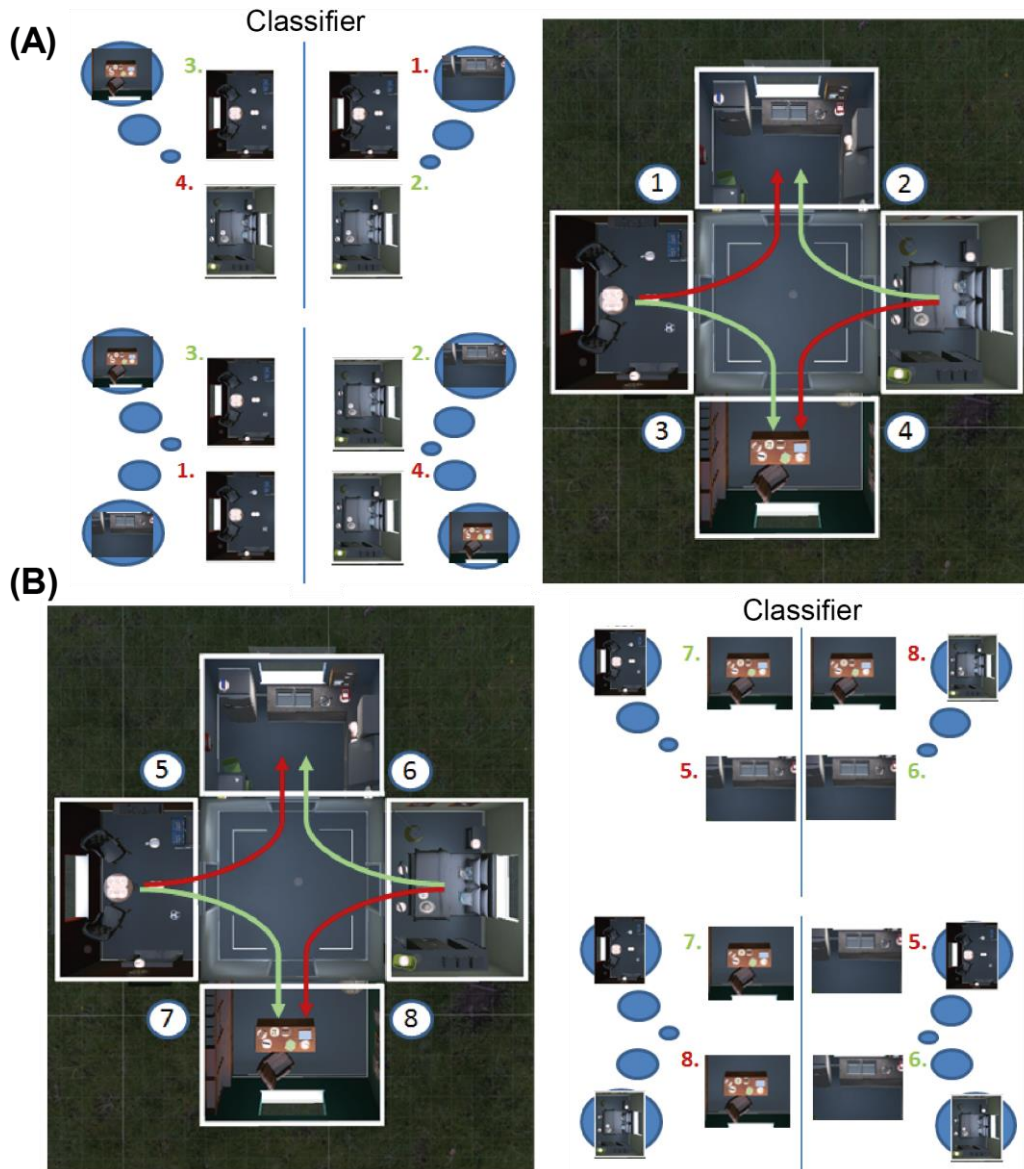
Figure 6.3: Classifiers were constructed as shown for A) (top) decoding the future unseen room (bottom) decoding the current scene in the first videos B) (top) decoding the past (bottom) decoding the current scene in the second videos

For decoding the identity of the two perceived rooms during either the first or second videos we trained the classifier to discriminate between bedroom versus games room trials in the first video and kitchen versus office trials in the second video. This allows us to balance each side of classifier for both directional information and expected information. For decoding the identity of the future

rooms during the first video we train the classifier on bedroom and game room trials where the expectation is the kitchen versus bedroom and game room trials where the expectation is the office. Due to the orthogonal design of the experiment this classifier is also balanced for directional information. To decode the past rooms we train the classifier on kitchen and office trials where the past room was the bedroom versus kitchen and office trials where the past room was the game room. As before this classifier is balanced for directional information. Each classifier was trained on n-1 runs of data and tested on the nth run in a leave-one-run-out cross-validation procedure.

7   General Discussion

7.1   Chapter summaries

This thesis aimed to investigate what role non-visual feedback plays in visual perception by attempting to gain a better understanding of what information arrives at low-level visual areas via feedback pathways. There are two types of feedback that we have defined in this thesis. The first is factual feedback that is dependent on a stimulus. The second is counterfactual feedback that is stimulus independent. This is an important distinction to make within the context of this thesis as the literature on memory would define recall, or the remembering of events that happened in the past, as being factual feedback. However memories are not as salient as vision and can often be false (Brewin & Andrews, 2017; Davies & Granhag, 2017). For simplicity, our definition relates one type of feedback to the external environment and the other to the internal environment independent of what is going on in the outside world.

To study both types of feedback in early visual cortex we exploited the high spatial resolution and retinotopic organisation of early visual cortex by using occlusion. In each of our paradigms, we either occluded part of the visual field or the entirety of the visual field for a brief period of time. The occluder acts as a gatekeeper for feedforward sensory information. By occluding part (or all) of the visual field we can control the flow of feedforward information to a corresponding patch of visual cortex. Inhibiting the flow of feedforward information effectively renders the corresponding patch of visual cortex as non-spiking cortex. This is particularly useful as fMRI combined with MVPA is sensitive to non-spiking activity, which allows us to investigate the modulatory influence of top-down information (Muckli, 2010).

In **Chapter 3** we tested the hypothesis that non-stimulated early visual cortex contains valance information about non-social affective images. We adopted a multivariate pattern analysis approach and found that we could decode pairs of valance conditions independent of category when one of those conditions was

negative. This finding supports our hypothesis and tells us that negative valance was being represented differently to positive and negative valance in non-stimulated early visual areas. We found a similar pattern of results in stimulated early visual cortex as well as in the amygdala and the pulvinar. In lateral occipital cortex we found that we could decode positive, negative and neutral valance. A second finding from **Chapter 3** was that cross-decoding revealed that valance is generalisable across category for negative valance versus positive/neutral valance in stimulated V1, V2 and V3. From feedback brain responses we found that neutral versus negative was generalisable in V1 and neutral versus positive was generalisable in V2 and the amygdala. We conclude that feedback pathways do provide early visual areas with information about valance and this information is represented similarly across the categories we used in this study.

In **Chapter 4** we built on previous work showing that sound categories could be decoded from early visual cortex using multivariate pattern analysis (Vetter et al., 2014). We aimed to replicate this finding and test the hypothesis that visual and auditory triggered categorical feedback is representationally similar. An important methodological change that we made was that in the previous study by Vetter et al. (2014) subjects were blindfolded whereas we adopted an eyes-open fixation paradigm. This meant that during auditory stimulation visual cortex was receiving uninformative feedforward information. We found that we could cross-decode between auditory and visual categories in non-stimulated visual cortex, supporting our hypothesis. We also successfully replicated previous work by Vetter et al. (2014). In **Chapter 4** we also tested a second hypothesis that auditory feedback to visual cortex would have retinotopic differences depending on whether the sound was a scene sound or an object sound. We found that scene sounds could only be decoded from peripheral V2 while object sounds could only be decoded in foveal V2 and V3. These findings tell us that categorically matching auditory and visual stimuli trigger feedback responses that share an overlap in their information content. The triggered feedback also plays an important modulatory functional role in the periphery for abstract scene information or in the fovea for concrete object information that may require foveation or finer discrimination at a high spatial scale.

In **Chapter 5** we made the switch from studying what we call factual feedback to what we call counterfactual feedback. We have so far learned something about the modulatory role feedback can play on low-level visual areas. In this study we were interested in learning how the brain, within a Bayesian predictive coding framework, anticipates the immediate future before it is seen. We used a sequence of image frames to create a short video that would trigger expectations about how the sequence of events might unfold after the video is switched off. In an adaption paradigm we used a test frame to probe how well the test frame matched the sequence as it unfolded in our subjects mind 800ms after the image sequence offset. The manipulation was that the test frame could be an expected frame from 300ms, 600ms, 800ms or 4000ms after the offset of the priming sequence. The remarkable finding was that in early visual cortex the 600ms and 800ms test frames best matched the internally extrapolated sequence of events inside the subjects mind, related to a future state of the world that has never seen by the subject. In a second experiment in **Chapter 5** we demonstrate that predictions about unseen events – events that unfold in a subject's mind - require 600ms to 1200ms to fully develop.

In **Chapter 6** we further explore the brain's ability to make predictions about the immediate future. In **Chapter 5** we had no control over the prior information that subjects used to extrapolate the priming sequence beyond the occluder into the immediate, unseen future. One theory on how the brain achieves this is by manipulating dynamic internal models (Clark, 2013). The priming sequence therefore acts as a jumping off point, but what prior knowledge or internal models subjects are using to make judgements about an unseen future was unknown. In an attempt to control prior knowledge about a stimulus set we turned to virtual reality to create a virtual house for use in a navigation paradigm. The rationale for doing this was so that our subjects could become familiar with a unique environment before being exposed to our stimuli. This would allow memories and internal models to be created within a 3D reference frame. Although the virtual house was filled with common household furniture and objects, any use of existing internal models would be fitted to the virtual house template for making judgements about the future in our paradigm. In the fMRI paradigm subjects combined a direction cue with the current room location to make judgements about

a future room they would eventually navigate to. Using multivariate pattern analysis we show that we can read out from non-stimulated brain responses in extrastriate cortex information about the future room identity. In a control experiment where the direction cue is removed this effect disappears. Together these findings demonstrate that feedback plays a role in anticipating the future in a highly predictive manner.

7.2   What is the information content of feedback responses across our paradigms?

One of the major questions that permeate this thesis is, what is the information that feedback processes are providing to early visual cortex? The primary analysis we utilise through-out this thesis is multivariate pattern analysis. Multivariate pattern analysis can inform us of differences between two types of information but it cannot tell us what that information is (Norman, Polyn, Detre, & Haxby, 2006). However, by using decoding and cross-decoding analysis and careful consideration with regards to the design of the paradigm in question, we can draw inferences about what the information is likely to be and what it is unlikely to be.

In **Chapter 3** we investigate valance information in non-stimulated early visual cortex. We use an m-sequence design to present to subjects 900 images of neutral, positive and negative valance. In our decoding analysis we average across 300 neutral images, 300 positive images and 300 negative images over ten experimental runs to produce multivariate patterns for analysis. We filter out category information, carry-over effects and the effect of changes in lightness from one image to the next. Therefore our analysis is independent of these filtered information categories. We find that we can decode negative valance. We can be confident that this finding is unrelated to object category, low-level visual features or fluctuations in lightness. Information from memory processes related to the familiarity of the content of the image set are also unlikely to be contributing to the decoding of negative valance that we observe as each image is only seen once by a subject. It is possible that each image is triggering object related internal models that descend feedback pathways providing information about the subject matter of

the image. However our steps in averaging across a large number of images and filtering category information should remove this. If this was not true, we would expect all pairwise decoding of valance to work. As we don't find neutral versus positive to reliably decode we can conclude that the successful decoding of neutral versus negative and positive versus negative conditions is due to the information content of the negative valance category. The null finding for neutral versus positive decoding also rules out a role of arousal. If arousal were the driving factor we would expect to see a difference between these two conditions also.

That negative valance information is present in feedback responses is further supported by the cross-decoding analysis. We find negative valance generalises across category in stimulated early visual cortex and in non-stimulated V1. This finding tells us that negative valance information is shared across category and is also represented similarly across category. When we compare our findings with the literature on emotion processing in vision, we find many parallels showing that negative valance is a privileged type of information (Blanchette, 2006; Gomes et al., 2017; Kuniecki et al., 2017; LoBue, 2010; Subra et al., 2018). Therefore it is likely that in this experiment the information content of feedback responses to viewing negative valance images is about negative valance. But what does this mean?

The mostly likely scenario is that negative valance information is privileged in visual processing to facilitate visual processes towards visual features related to danger (Vida & Behrmann, 2017). This theory has been expressed in the literature from both an evolutionary perspective (Gomes et al., 2017) and from the perspective of an objects relevance to an individual (Blanchette, 2006). If true, one hypothesis that we could postulate from what we have learned from the work in **Chapter 4** is that valance information about peripheral threats would feed back to foveal retinotopic locations. Similar to peripheral object information being routed to the fovea, but not other non-stimulated peripheral regions (Williams et al., 2008), we could reasonably expect valance information to be localised to the fovea. There is some evidence to support this idea. Almeida et al. (2015) studied sub-cortical responses to negative snake images presented in both the fovea and periphery and found differences in sub-cortical responses to periphery and foveal

conditions. These sub-cortical differences may follow the same pattern as outlined by Williams et al. (2008) in early visual areas.

In **Chapter 4** we explored the similarity between feedback responses about contextually congruent auditory and visual stimuli. Within a predictive coding framework the incoming sensory information is tested against internally generated predictions (Shipp, 2016). If true, feedback about the internal model should be similar whether it was triggered by an auditory stimulus or a visual stimulus. We tested this hypothesis and found evidence to support it by observing an overlap in the information content of feedback responses for categorically congruent sound and visual stimuli within occluded early visual cortex. Interestingly there has been no evidence to support this hypothesis in previous work (Jung et al., 2018). One obvious difference that may cut to the heart of the discussion is the difference in stimuli choice between our study and the study by Jung et al. (2018). In our study we compared feedback responses between spatially and temporally dynamic sounds and videos. However in the study by Jung et al. (2018) they compared feedback responses between spatially and temporally dynamic sounds and static scene images. If we conclude that there is no overlap in the neural representation between categorically matching dynamic sounds and static images in early visual cortex, but there is overlap in the neural representation between categorically matching dynamic sounds and dynamic videos, then the overlap in information must be to do with the similarity of how the spatial and temporal information is changing over time. This could be to do with the matching onset and offset of particular stimulus features. Therefore the information content of audio and visual feedback responses in early visual cortex in our experiment is not purely categorical, but perhaps related to the regularity of the dynamics of the stimuli. Recent work by Andric et al. (2017) has shown that the regularity of auditory and visual stimuli is an important feature for multisensory integration in visual cortex leading them to conclude visual cortex is a junction for integration of temporally-extended auditory and visual input.

In a second experiment in **Chapter 4** we also explored the peripheral bias that is common throughout the human and non-human primate literature regarding auditory feedback to early visual cortex (Anton L Beer et al., 2011; Anton Ludwig Beer et al., 2013; Eckert et al., 2008; Klinge et al., 2010; Laramée et al., 2011;

103

Murray et al., 2016; Qin et al., 2014; Rockland & Ojima, 2003). It has been shown that for objects presented to the periphery information about those objects is directed to the fovea via feedback pathways but not to other non-stimulated peripheral areas (Williams et al., 2008). This finding is likely due a precursor to foveation, likely sensitising the fovea in a preparatory manner. We hypothesised that within a predictive coding framework auditory object information should also be directed to the fovea for a similar purpose, therefore ignoring the peripheral bias for auditory feedback. We found evidence to support this hypothesis and further showed that auditory objects were only decodable in the fovea and auditory scenes only decodable in the periphery. As the two objects were distinguishable from their feedback responses in the fovea the information content must contain object specific information. What the object specific information is related to is more difficult to infer. There is however some recent functional connectivity work relating object information from feedback responses in early visual cortex to lateral occipital cortex (Bennett, Petro, & Muckli, 2016) suggesting it is indeed object specific.

In **Chapter 5** we switched focus slightly to a study of counterfactual feedback. In this experiment we used a priming sequence of images to trigger predictions and then we test the prediction after a short period of occlusion (800ms) with a test frame. In this attenuation study we find that the prediction best matches the 600ms and 800ms test frames but not the 300ms or 4000ms test frame. In other words, the prediction best matches the test frame(s) that would more-or-less be expected after the period of occlusion. This is a remarkable finding. It means that the brain has generated a prediction about a future unseen state of the world.

In this experiment we asked subjects to predict how the priming sequence might unfold during occlusion. We did not tell subjects how to do this task explicitly, so we can't be sure of the strategies that they used. However we can make some suggestion as to what the information might be. One suggestion is that the prediction is related to a simulation of the future (Ullman et al., 2017). This is derived from the idea that the brain can simulate rigid body physics. An example of this can be taken from the game Jenga. Jenga is a game where players take turns to remove a single block from a 54 block tower. This can only be accomplished if the brain can run simulations on the stability of the tower before a specific block

104

has been removed. There is evidence supporting the brain's ability to run rigid body simulations (Hamrick et al., 2016; Ullman et al., 2018). This would explain why in our experiment the best matching test frame (600ms & 800ms) is a better match than the earliest test frame (300ms), which itself is more closely related to the offset of the priming sequence but not the prediction.

Another suggestion is that distal cues in the priming sequence act as a reference point for subjective judgements of an objects position in space. Work over the past four decades has revealed cells in the hippocampus representing an animal's location in space (O'Keefe, 1979), and cells in entorhinal cortex representing an animal's spatial environment (Hafting, Fyhn, Molden, Moser, & Moser, 2005). It is possible that the priming sequence provides an idea of 'where I am' and more distal cues give an idea of 'where I expect to be.' There is emerging evidence that an animal's current subjective location is represented in both the hippocampus and in V1 (Saleem, Diamanti, Fournier, Harris, & Carandini, 2018). However there is little or no evidence supporting the idea of a predicted future subjective location. With regards to subjective location, one of the three priming sequences used in this experiment was of egocentric motion. The other two priming sequences were of allocentric motion. The effect we observe could be driven by predictive subjective location judgements from the single egocentric condition. Alternatively the effect could be driven from all conditions if the brain also makes predictive future judgements about allocentric motion. There is some evidence supporting this idea from apparent motion studies that show that the apparent motion trace is updated to new retinotopic locations pre-saccade (Edwards et al., 2017). Judgements about the future state of egocentric motion, allocentric motion, or both, best explain the finding we report in this chapter.

In **Chapter 6** we investigated whether we could read out from feedback responses information regarding expectations about the future during visual stimulation. We trained subjects in a virtual environment to develop strong internal models of our virtual world through exposure and familiarity. We find that when a subject has information about where they are in a virtual space and information about which direction to take, we can read out information about the expected destination before the subject physically arrives there. For the second time across two paradigms we find evidence for a neural signal that is unrelated to the current

105

external state of the world. One suggestion for what the information content of this signal could be is that it is categorical room information about the future unseen room. This could be tested by having two virtual house configurations and using cross-classification analysis to test whether the future information is generalisable between the contextually congruent future rooms across the two house environments. In other words, do two pairs of independent rooms trigger similar predictions if the destination is the same?

The information could also be related to future judgements of subjective location in space, as previously discussed. It has been shown in mice that subjective location is represented in both the hippocampus, related to place cells (O'Keefe, 1979), and in V1 (Saleem et al., 2018). However in our experiment we do not find information about the future room in V1, but in extrastriate cortex. This does not mean that V1 does not contain information, but we cannot reject the null hypothesis. Finding information about the future room in extrastriate cortex in humans versus V1 in mice could be related to differences between visual acuity across species. Mice have low visual acuity (Prusky, West, & Douglas, 2000). Representing subjective location at a low spatial scale in mouse V1 may be analogous to representing subjective future location at a lower spatial scale in human V2. The differences could also be related to representing subjective judgements of current location and testing them with incoming sensory input in V1 versus representing future subjective location independent of sensory input. Furthermore, the future room information we detect could be related not only to place cells but also to grid cells, which provide a spatial representation of the virtual environment as a whole. Predictions about whether you would expect to be in one future room versus another future room could explain our finding. However it is unknown whether or not grid cells respond in an anticipatory way for expected or planned movement. They do however exhibit spatial representation during mental simulations (Bellmund, Deuker, Schröder, & Doeller, 2016) and imagined navigation (Horner, Bisby, Zotow, Bush, & Burgess, 2016) which would support this idea.

In conclusion, we set out to study non-visual feedback to early visual cortex and the role feedback may play in visual perception. We learned that non-visual feedback in low-level non-stimulated cortex represents information about both the

external world and the internal world. This is an important finding with respect to current predictive coding models. Feedback plays a role in hypothesis testing incoming sensory information (George & Hawkins, 2009; Rao & Ballard, 1999; Shipp, 2016). These predictive models are yet to explain how the brain can switch from one stream of perception (about the external world) to another (the "inner-eye") while simultaneously monitoring sensory input. Predictive coding models will need to be updated to include this new type of information.

7.3 Feedback responses to early visual cortex are retinotopically organised

In **Chapter 4** we find that we can decode auditory scene information in the periphery of V2 and auditory object information in the fovea of V2 and V3. This finding indicates that we have two distinct signals. One signal in the periphery related to scene information and a second signal in the fovea related to object information. When object and scene information is combined into a single sound we find that we could read this information out from both the fovea and the periphery.

The simplest explanation for finding an auditory object related signal in the fovea is that auditory objects require high-spatial resolution for the purpose of finer discrimination, a role that the fovea provides and the periphery does not (McDonald et al., 2000). This idea is supported by vision studies that show that peripherally presented objects feedback information to the fovea (Bennett et al., 2016; Williams et al., 2008). It is also consistent with central field biases of object specific cortical areas such as lateral occipital cortex (Grill-Spector, Kourtzi, & Kanwisher, 2001) and eccentricity biases for objects in higher visual cortical areas (Hasson, Levy, Behrmann, Hendler, & Malach, 2002). Interestingly Hasson et al. (2002) discovered that some objects, such as buildings, are biased to peripheral representation only, which would violate this explanation. However they suggest buildings are a special case citing R. Epstein and Kanwisher (1998) who suggest buildings are landmarks representing local space. While we find that scene information - abstract information about local space - could only be decoded in the

107

periphery, replicating previous work (Vetter et al., 2014), this is not the case for the visual counterpart which can be decoded in both periphery and fovea (Bennett et al., 2016). This difference could be explained by fluctuations between scene and object processing over prolonged stimulation periods, suggested by studies that show that viewers prefer to look at objects over background (Nuthmann & Henderson, 2010).

Auditory scene information may be biased to the periphery due to local space biases (R. Epstein & Kanwisher, 1998; Hasson et al., 2002) such as categorical context (Vetter et al., 2014). From an anatomical perspective the periphery is anatomically more proximal to auditory cortex than the fovea ,so it is not surprising that direct anatomical projections from primary auditory cortex to visual cortex terminate in the periphery of non-human primates (Rockland & Ojima, 2003) and also possibly in humans (Anton L Beer et al., 2011). An additional functional role for auditory feedback in the periphery beyond representing local space context could be that it acts as a low (relative to the fovea) spatial co-ordinate system for monitoring low-resolution stimulus features or statistical regularities while leaving the fovea available for fine discrimination tasks. Vetter et al. (2014) demonstrated that natural scene sounds could be read out of the periphery reliably but not the fovea during both a working memory task and a visuo-spatial imagery task which would support this theory.

In **Chapter 5** and **Chapter 6** we investigate counterfactual feedback information during occlusion and during concurrent visual stimulation. Counterfactual feedback is information that is initialised internally and is independent of an external stimulus. A curious finding we observe across both paradigms is that counterfactual feedback is localised to extrastriate retinotopic locations. This could be a coincidence. Feedback in general has been shown in V1 in a number of paradigms (Muckli et al., 2015; Lucy S Petro et al., 2013; F. W. Smith & Muckli, 2010). However, with respect to stimulus independent feedback as seen in **Chapter 6**, it is possible that V1 may be reserved for background functions during counterfactual thinking.

When a person engages in counterfactual thinking, they engage a brain network called the default mode network which is anti-correlated with the executive

network (Hellyer et al., 2014). During the engagement of this network for mind-wandering, for example, a phenomenon known as perceptual decoupling occurs. The neural mechanism of perceptual decoupling remains largely unknown. One study has related perceptual decoupling during auditory stimulation to a decrease in functional connectivity in the auditory pathway (Hove et al., 2015). Perceptual decoupling has been described as a reduced awareness of the external environment. It is possible that in vision perceptual decoupling utilises the spatial fidelity of extrastriate cortex for manipulating dynamic internal models (imagining) but spares V1 for monitoring the external environment outside of conscious perception. This would fit with the read/write blackboard hypothesis of V1, whereby other cortical areas can write information to V1 that can be read by other cortical and sub-cortical areas (Muckli et al., 2015; Mumford, 1991; Murray et al., 2016; Roelfsema & de Lange, 2016; Williams et al., 2008).

We can explore this idea further by considering under what circumstances feedback about counterfactual thought is found in V1 and when it is not found. Imagery is a good candidate paradigm for this as it is the most widely studied example of counterfactual thinking that also uses brain imaging techniques. A pattern emerges across a variety of paradigms invoking imagery. When high-level representations are back projected during imagery to low level visual areas they are absent from V1 (Radoslaw M Cichy et al., 2011; S.-H. Lee et al., 2012; Reddy et al., 2010). However when these back projected representations are manipulated by working memory they are then present in V1 (Dijkstra et al., 2017). By contrast, low-level representations appear to always be available to V1 (Albers et al., 2013; Bergmann et al., 2016; Radoslaw M Cichy et al., 2011; Harrison & Tong, 2009; Schlegel et al., 2013; Xing et al., 2013) likely due to V1 being tuned to low-level features. These studies point to the importance of task as a factor for whether the retinotopic space of V1 is recruited for representing high-level representations. Put in other words, if the task of imagery requires fine, detailed spatial scale then V1 is recruited. If the task of imagery requires only the broad strokes then V2 is of sufficient spatial scale.

7.4   Limitations & future directions

One of the major limitations in fMRI research is time. To circumvent this problem we used a path-guided de Bruijn m-sequence in **Chapter 3** to present to subjects a large image set in a typical amount of time for an fMRI experiment. These sequences are useful as they provide a measure of signal detection power and are particularly useful for the study of categories (Aguirre et al., 2011). However in the fast event-related design we could only reliable decode information about negative valance information, likely due to its privileged nature in cortical/sub-cortical processing of danger (Anderson et al., 2003; Pourtois et al., 2013). Whether positive valance information can be read out from non-stimulated early visual cortex is unclear. This hypothesis could be tested and our finding replicated by reducing the order of counterbalancing (less trials) and increasing the stimulus presentation time. We have briefly mentioned the importance of task. In this study we used a category 1-back task. This could be another factor explaining why we couldn't decode positive information but could decode negative information (privileged) in a fast event-related design. If subjects were attending to category and less to valance it would explain why only the privileged information was found.

In **Chapter 4** we show for the first time neural overlap between auditory and visually triggered feedback in non-stimulated early visual cortex. Attempts have been made in the past to test the hypothesis that there is shared neural representation between feedback for auditory and visual modality without success (Jung et al., 2018). Therefore our finding should be replicated to provide additional evidence. As an additional step more exemplars should be included for the purpose of cross-classification. Such an analysis would shed light on whether or not cross-classification is achieved due to specific stimulus features shared between modality or if it is related to high-level categorical representations.

In this same chapter we explore the retinotopic nature of auditory object and scene feedback to early visual cortex. Due to the realisation that an auditory scene is a composition of object sounds we decided to embed a rogue object sound in to a background scene sound and vary the onset and offset of the object sound to make it stand out from the background. This likely impacted the power of our

analysis as the classifier was searching for object information within scene information from an already coarse signal. This issue could have been alleviated by presenting objects in isolation to form an object sound, or in pairs to form a simple scene sound. We also used an unrelated baseline task in this study, which likely also impacted our statistical power as the sounds were listened to passively. If we had used a task related to objects and scenes we would have increased our overall statistical power.

Chapter 5 and **6** was the beginning of our study on counterfactual thinking. These studies were informative as they demonstrated a different type of feedback that could be read out from extrastriate cortex. The main question raised from these studies is, how can the brain switch between two different streams of perception – one that is factual and one that is counterfactual? To answer this question we plan to use high-field 7-Tesla fMRI (Olman et al., 2012) to explore the feedback profile of both counterfactual and factual feedback in different layers of the cortical sheet. We believe that by showing these two different response profiles in the same, or different, cortical layers will necessitate a change to current predictive coding models to account for the role feedback plays in both the counterfactual percept (not testing against sensory information) versus monitoring the external world for changes (testing against a possibly limited sensory input).

8   Bibliography

Adolphs, R., Gosselin, F., Buchanan, T. W., Tranel, D., Schyns, P., & Damasio, A. R. (2005). A mechanism for impaired fear recognition after amygdala damage. *Nature, 433*(7021), 68.

Aguirre, G. K., Mattar, M. G., & Magis-Weinberg, L. (2011). de Bruijn cycles for neural decoding. *Neuroimage, 56*(3), 1293-1300.

Ajina, S., & Bridge, H. (2018). Subcortical pathways to extrastriate visual cortex underlie residual vision following bilateral damage to V1. *Neuropsychologia*.

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology, 23*(15), 1427-1431.

Allefeld, C., Görgen, K., & Haynes, J.-D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *Neuroimage, 141*, 378-392.

Almeida, I., Soares, S. C., & Castelo-Branco, M. (2015). The distinct role of the amygdala, superior colliculus and pulvinar in processing of central and peripheral snakes. *PloS one, 10*(6), e0129949.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. (2003). Neural correlates of the automatic processing of threat facial signals. *Journal of Neuroscience, 23*(13), 5627-5633.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron, 65*(4), 550-562.

Andric, M., Davis, B., & Hasson, U. (2017). Visual cortex signals a mismatch between regularity of auditory and visual streams. *Neuroimage, 157*, 648-659.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in brain research, 154*, 93-120.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*(8), 617.

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

Baumann, O., Borra, R. J., Bower, J. M., Cullen, K. E., Habas, C., Ivry, R. B., . . . Moulton, E. A. (2015). Consensus paper: the role of the cerebellum in perceptual processes. *The Cerebellum, 14*(2), 197-220.

Beer, A. L., Plank, T., & Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental Brain Research, 213*(2-3), 299.

Beer, A. L., Plank, T., Meyer, G., & Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Frontiers in integrative neuroscience, 7*, 5.

Bekhtereva, V., Craddock, M., & Müller, M. M. (2015). Attentional bias to affective faces and complex IAPS images in early visual cortex follows emotional cue extraction. *Neuroimage, 112*, 254-266.

Bellmund, J. L., Deuker, L., Schröder, T. N., & Doeller, C. F. (2016). Grid-cell representations in mental simulation. *Elife, 5*, e17089.

Bennett, M., Petro, L., & Muckli, L. (2016). Investigating cortical feedback of objects and background scene to foveal and peripheral V1 using fMRI. *Journal of Vision, 16*(12), 568-568.

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2016). Smaller primary visual cortex is associated with stronger, but less precise mental imagery. *Cerebral Cortex, 26*(9), 3838-3850.

Bertini, C., Cecere, R., & Làdavas, E. (2013). I am blind, but I "see" fear. *Cortex, 49*(4), 985-993.

Binda, P., Cicchini, G. M., Burr, D. C., & Morrone, M. C. (2009). Spatiotemporal distortions of visual perception at the time of saccades. *Journal of Neuroscience, 29*(42), 13147-13157.

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *The Quarterly Journal of Experimental Psychology, 59*(8), 1484-1504.

Bolognini, N., Senna, I., Maravita, A., Pascual-Leone, A., & Merabet, L. B. (2010). Auditory enhancement of visual phosphene perception: The effect of temporal and spatial factors and of stimulus intensity. *Neuroscience letters, 477*(3), 109-114.

Braboszcz, C., & Delorme, A. (2011). Lost in thoughts: neural markers of low alertness during mind wandering. *Neuroimage, 54*(4), 3040-3047.

Bressler, D. W., Fortenbaugh, F. C., Robertson, L. C., & Silver, M. A. (2013). Visual spatial attention enhances the amplitude of positive and negative fMRI responses to visual stimulation in an eccentricity-dependent manner. *Vision Research, 85*, 104-112.

Brewin, C. R., & Andrews, B. (2017). Creating memories for false autobiographical events in childhood: A systematic review. *Applied Cognitive Psychology, 31*(1), 2-23.

Britton, J. C., Phan, K. L., Taylor, S. F., Welsh, R. C., Berridge, K. C., & Liberzon, I. (2006). Neural correlates of social and nonsocial emotions: An fMRI study. *Neuroimage, 31*(1), 397-409.

Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron, 35*(4), 625-641.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia, 47*(4), 1045-1052. doi:https://doi.org/10.1016/j.neuropsychologia.2008.11.003

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST), 2*(3), 27.

Chong, E., Familiar, A. M., & Shim, W. M. (2016). Reconstructing representations of dynamic visual objects in early visual cortex. *Proceedings of the National Academy of Sciences, 113*(5), 1453-1458.

Cichy, R. M., Heinzle, J., & Haynes, J.-D. (2011). Imagery and perception share cortical representations of content and location. *Cerebral Cortex, 22*(2), 372-380.

Cichy, R. M., & Teng, S. (2017). Resolving the neural dynamics of visual and auditory scene processing in the human brain: a methodological approach. *Phil. Trans. R. Soc. B, 372*(1714), 20160108.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences, 36*(3), 181-204.

Cox, D. D., & Savoy, R. (2003). fMRI Brain Reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19*(2), 261-270.

D'Hondt, F., Szaffarczyk, S., Sequeira, H., & Boucart, M. (2016). Explicit and implicit emotional processing in peripheral vision: A saccadic choice paradigm. *Biological psychology, 119*, 91-100.

Davies, G., & Granhag, P. A. (2017). Introduction to target article and commentaries: A systematic review of the experimental literature on the creation of false memories of childhood events by adults. *Applied Cognitive Psychology, 31*(1), 1-1.

de Haas, B., Schwarzkopf, D. S., Urner, M., & Rees, G. (2013). Auditory modulation of visual stimulus encoding in human retinotopic cortex. *Neuroimage, 70*, 258-267.

Dijkstra, N., Bosch, S., & van Gerven, M. A. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience*, 3022-3016.

Droit-Volet, S. (2017). Time dilation in children and adults: The idea of a slower internal clock in young children tested with different click frequencies. *Behavioural processes, 138*, 152-159.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*(2), 647-660.

Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., & Menon, V. (2008). A cross-modal system linking primary auditory and visual cortices: Evidence from intrinsic fMRI connectivity analysis. *Human brain mapping, 29*(7), 848-857.

Edwards, G., Vetter, P., McGruer, F., Petro, L. S., & Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *Scientific reports, 7*(1), 16538.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392*(6676), 598.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences, 12*(10), 388-396.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*(13), 5749-5759.

Fechner, G. (1966). Elements of psychophysics. Vol. I.

Fernández-Martín, A., Gutiérrez-García, A., Capafons, J., & Calvo, M. G. (2017). Adaptive attunement of selective covert attention to evolutionary-relevant emotional visual scenes. *Consciousness and cognition, 51*, 223-235.

Fischer, J., & Whitney, D. (2012). Attention gates visual coding in the human pulvinar. *Nature communications, 3*, 1051.

Foster, B. L., Kaveh, A., Dastjerdi, M., Miller, K. J., & Parvizi, J. (2013). Human retrosplenial cortex displays transient theta phase locking with medial temporal cortex prior to activation during autobiographical memory retrieval. *Journal of Neuroscience, 33*(25), 10439-10446.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678.

Franklin, M. S., Smallwood, J., & Schooler, J. W. (2011). Catching the mind in flight: Using behavioral indices to detect mindless reading in real time. *Psychonomic Bulletin & Review, 18*(5), 992-997.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks, 16*(9), 1325-1352.

Friston, K. (2008). Hierarchical models in the brain. *PLoS computational biology, 4*(11), e1000211.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127.

Gamond, L., Vecchi, T., Ferrari, C., Merabet, L. B., & Cattaneo, Z. (2017). Emotion processing in early blind and sighted individuals. *Neuropsychology, 31*(5), 516.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences, 16*(2), 129-135.

George, D., & Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS computational biology, 5*(10), e1000532.

Gomes, N., Silva, S., Silva, C. F., & Soares, S. C. (2017). Beware the serpent: the advantage of ecologically-relevant stimuli in accessing visual awareness. *Evolution and Human Behavior, 38*(2), 227-234.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences, 15*(1), 20-25.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41*(10-11), 1409-1422.

Gschwind, M., Pourtois, G., Schwartz, S., Van De Ville, D., & Vuilleumier, P. (2011). White-matter connectivity between face-responsive regions in the human brain. *Cerebral Cortex, 22*(7), 1564-1576.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience, 35*(27), 10005-10014.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature, 436*(7052), 801.

Hamrick, J. B., Battaglia, P. W., Griffiths, T. L., & Tenenbaum, J. B. (2016). Inferring mass in complex scenes by mental simulation. *Cognition, 157*, 61-76.

Harlow, J. M. (1993). Recovery from the passage of an iron bar through the head. *History of Psychiatry, 4*(14), 274-281.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34*(3), 479-490.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage, 62*(2), 852-855.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology, 17*(4), 323-328.

Hellyer, P. J., Shanahan, M., Scott, G., Wise, R. J., Sharp, D. J., & Leech, R. (2014). The control of global brain dynamics: opposing actions of frontoparietal control and default mode networks on attention. *Journal of Neuroscience, 34*(2), 451-461.

Hespos, S. J., Ferry, A. L., Anderson, E. M., Hollenbeck, E. N., & Rips, L. J. (2016). Five-month-old infants have general knowledge of how nonsolid substances behave and interact. *Psychological Science, 27*(2), 244-256.

Horner, A. J., Bisby, J. A., Zotow, E., Bush, D., & Burgess, N. (2016). Grid-like processing of imagined navigation. *Current Biology, 26*(6), 842-847.

Hove, M. J., Stelzer, J., Nierhaus, T., Thiel, S. D., Gundlach, C., Margulies, D. S., . . . Merker, B. (2015). Brain network reconfiguration and perceptual decoupling during an absorptive state of consciousness. *Cerebral Cortex, 26*(7), 3116-3124.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of neurophysiology, 28*(2), 229-289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology, 195*(1), 215-243.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *Neuroimage, 17*(4), 1729-1741.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., . . . Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. *Neuron, 73*(4), 814-828.

Ivry, R. B., Spencer, R. M., Zelaznik, H. N., & Diedrichsen, J. (2002). The cerebellum and event timing. *Annals of the new York Academy of Sciences, 978*(1), 302-317.

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences, 100*(18), 10568-10573.

Jung, Y., Larsen, B., & Walther, D. B. (2018). Modality-independent coding of scene categories in prefrontal cortex. *Journal of Neuroscience*, 0272-0218.

Kam, J. W., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow fluctuations in attentional control of sensory cortex. *Journal of cognitive neuroscience, 23*(2), 460-470.

Kay, K., Rokem, A., Winawer, J., Dougherty, R., & Wandell, B. (2013). GLMdenoise: a fast, automated technique for denoising task-based fMRI data. *Frontiers in neuroscience, 7*, 247.

Klinge, C., Eippert, F., Röder, B., & Büchel, C. (2010). Corticocortical connections mediate primary visual cortex responses to auditory stimulation in the blind. *Journal of Neuroscience, 30*(38), 12798-12805.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica, 134*(3), 372-384.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Ku, S.-p., Gretton, A., Macke, J., & Logothetis, N. K. (2008). Comparison of pattern recognition methods in classifying high-resolution BOLD signals obtained at high magnetic field in monkeys. *Magnetic resonance imaging, 26*(7), 1007-1014.

Kuniecki, M., Wołoszyn, K. B., Domagalik, A., & Pilarczyk, J. (2017). Effects of Scene Properties and Emotional Valence on Brain Activations: A Fixation-Related fMRI Study. *Frontiers in human neuroscience, 11*, 429.

LaConte, S., Strother, S., Cherkassky, V., Anderson, J., & Hu, X. (2005). Support vector machines for temporal classification of block design fMRI data. *Neuroimage, 26*(2), 317-329.

Lamme, V. A., Rodriguez-Rodriguez, V., & Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex, 9*(4), 406-413.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences, 23*(11), 571-579.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, 39-58.

Laramée, M.-E., Kurotani, T., Rockland, K., Bronchti, G., & Boire, D. (2011). Indirect pathway between the primary auditory and visual cortices through layer V pyramidal neurons in V2L in mouse and the effects of bilateral enucleation. *European Journal of Neuroscience, 34*(1), 65-78.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences, 36*(3), 141-151.

Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage, 59*(4), 4064-4073.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *JOSA A, 20*(7), 1434-1448.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences, 98*(4), 1907-1911.

Levitt, J. B., & Lund, J. S. (2002). The spatial extent over which neurons in macaque striate cortex pool visual signals. *Visual neuroscience, 19*(4), 439-452.

Libby, L. A., Hannula, D. E., & Ranganath, C. (2014). Medial temporal lobe coding of item and spatial information during relational binding in working memory. *Journal of Neuroscience, 34*(43), 14233-14242.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex, 26*(5), 1910-1922.

LoBue, V. (2010). And along came a spider: An attentional bias for the detection of spiders in young children and adults. *Journal of experimental child psychology, 107*(1), 59-66.

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.

Lu, L., Zhang, G., Xu, J., & Liu, B. (2018). Semantically Congruent Sounds Facilitate the Decoding of Degraded Images. *Neuroscience, 377*, 12-25.

Madan, C. R., Bayer, J., Gamer, M., Lonsdorf, T. B., & Sommer, T. (2018). Visual complexity and affect: ratings reflect more than meets the eye. *Frontiers in psychology, 8*, 2368.

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. *Nature neuroscience, 17*(11), 1598.

Marques, T., Nguyen, J., Fioreze, G., & Petreanu, L. (2018). The functional organization of cortical feedback inputs to primary visual cortex. *Nature neuroscience, 21*(5), 757.

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2006). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex, 17*(7), 1672-1679.

McDonald, J. J., Teder-SaÈlejaÈrvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906.

Meeren, H. K., Hadjikhani, N., Ahlfors, S. P., Hämäläinen, M. S., & De Gelder, B. (2016). Early preferential responses to fear stimuli in human right dorsal visual stream-a MEG study. *Scientific reports, 6*, 24831.

Meijer, G. T., Montijn, J. S., Pennartz, C. M., & Lansink, C. S. (2017). Audio-visual modulation in mouse V1 depends on cross-modal stimulus configuration and congruency. *Journal of Neuroscience*, 0468-0417.

Merchant, H., & Yarrow, K. (2016). How the motor system both encodes and influences our sense of time. *Current Opinion in Behavioral Sciences, 8*, 22-27.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage, 53*(1), 103-118.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., & Newman, S. (2004). Learning to decode cognitive states from brain images. *Machine learning, 57*(1-2), 145-175.

Mooneyham, B. W., Mrazek, M. D., Mrazek, A. J., Mrazek, K. L., Phillips, D. T., & Schooler, J. W. (2017). States of mind: Characterizing the neural bases of focus and mind-wandering through dynamic functional connectivity. *Journal of cognitive neuroscience, 29*(3), 495-506.

Morgan, A. T., Petro, L. S., & Muckli, L. (2016). Cortical feedback to V1 and V2 contains unique information about high-level scene structure.

Morrell, F. (1972). Visual system's view of acoustic space. *Nature, 238*(5358), 44.

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature neuroscience, 8*(7), 950.

Muckli, L. (2010). What are we missing here? Brain imaging evidence for higher cognitive functions in primary visual cortex V1. *International Journal of Imaging Systems and Technology, 20*(2), 131-139.

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., . . . Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology, 25*(20), 2690-2695.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS biology, 3*(8), e265.

Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current opinion in neurobiology, 23*(2), 195-201.

Muckli, L., & Petro, L. S. (2017). The significance of memory in sensory cortex. *Trends in neurosciences, 40*(5), 255-256.

Mumford, D. (1991). On the computational architecture of the neocortex. *Biological cybernetics, 65*(2), 135-145.

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics, 66*(3), 241-251.

Murray, M. M., Thelen, A., Thut, G., Romei, V., Martuzzi, R., & Matusz, P. J. (2016). The multisensory function of the human primary visual cortex. *Neuropsychologia, 83*, 161-169.

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage, 105*, 215-228.

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences, 112*(35), 11078-11083.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-430.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8), 20-20.

O'Keefe, J. (1979). A review of the hippocampal place cells. *Progress in neurobiology, 13*(4), 419-439.

O'Keefe, J., Burgess, N., Donnett, J. G., Jeffery, K. J., & Maguire, E. A. (1998). Place cells, navigational accuracy, and the human hippocampus. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353*(1373), 1333-1340.

O'Reilly, J. X., Mesulam, M. M., & Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience, 28*(9), 2252-2260.

Olman, C. A., Harel, N., Feinberg, D. A., He, S., Zhang, P., Ugurbil, K., & Yacoub, E. (2012). Layer-specific fMRI reflects different neuronal computations at different depths in human V1. *PloS one, 7*(3), e32536.

Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: ending the imagery debate. *Proceedings of the National Academy of Sciences, 112*(33), 10089-10092.

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends in Cognitive Sciences, 18*(5), 242-250.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a'low road'to'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35*, 73-89.

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1714). doi:10.1098/rstb.2016.0104

Petro, L. S., Smith, F. W., Schyns, P. G., & Muckli, L. (2013). Decoding face categories in diagnostic subregions of primary visual cortex. *European Journal of Neuroscience, 37*(7), 1130-1139.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research, 41*(24), 3145-3161.

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: what is magic and what is not. *Biological psychology, 92*(3), 492-512.

Prusky, G. T., West, P. W., & Douglas, R. M. (2000). Behavioral assessment of visual acuity in mice and rats. *Vision Research, 40*(16), 2201-2209.

Qin, W., Xuan, Y., Liu, Y., Jiang, T., & Yu, C. (2014). Functional connectivity density in congenitally and late blind subjects. *Cerebral Cortex, 25*(9), 2507-2516.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience, 2*(1), 79.

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *Neuroimage, 50*(2), 818-825.

Reichle, E. D., Reineberg, A. E., & Schooler, J. W. (2010). Eye movements during mindless reading. *Psychological Science, 21*(9), 1300-1310.

Ricciardi, E., Tozzi, L., Leo, A., & Pietrini, P. (2014). Modality dependent cross-modal functional reorganization following congenital visual deprivation within occipital areas: a meta-analysis of tactile and auditory studies. *Multisensory research, 27*(3-4), 247-262.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50*(1-2), 19-26.

Roelfsema, P. R., & de Lange, F. P. (2016). Early visual cortex as a multiscale cognitive blackboard. *Annual review of vision science, 2*, 131-151.

Roth, M. J., Synofzik, M., & Lindner, A. (2013). The cerebellum optimizes perceptual predictions about external sensory events. *Current Biology, 23*(10), 930-935.

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., & Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature*, 1.

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: Mechanisms and functions. *Current Opinion in Behavioral Sciences, 17*, 41-50.

Schlegel, A., Kohler, P. J., Fogelson, S. V., Alexander, P., Konuthula, D., & Tse, P. U. (2013). Network structure and dynamics of the mental workspace. *Proceedings of the National Academy of Sciences, 110*(40), 16277-16282.

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. (2008). Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *Journal of cognitive neuroscience, 20*(11), 2097-2109.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences, 15*(7), 319-326.

Self, M. W., van Kerkoerle, T., Super, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Current Biology, 23*(21), 2121-2129.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2), 301-312.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature, 408*(6814), 788.

Sharma, J., Sugihara, H., Katz, Y., Schummers, J., Tenenbaum, J., & Sur, M. (2014). Spatial attention and temporal expectation under timed uncertainty predictably modulate neuronal responses in monkey V1. *Cerebral Cortex, 25*(9), 2894-2906.

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in psychology, 7*, 1792.

Shobe, E. R. (2014). Independent and collaborative contributions of the cerebral hemispheres to emotional processing. *Frontiers in human neuroscience, 8*, 230.

Sirotin, Y. B., & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature, 457*(7228), 475.

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of cognitive neuroscience, 20*(3), 458-469.

Smallwood, J., Brown, K. S., Tipper, C., Giesbrecht, B., Franklin, M. S., Mrazek, M. D., . . . Schooler, J. W. (2011). Pupillometric evidence for the decoupling of attention from perceptual input during offline thought. *PloS one, 6*(3), e18298.

Smilek, D., Carriere, J. S., & Cheyne, J. A. (2010). Out of mind, out of sight: eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21*(6), 786-789.

Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences, 107*(46), 20099-20103. doi:10.1073/pnas.1000233107

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology, 22*(3), 191-196.

Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: an auditory rapid event-related fMRI-task. *Neuroimage, 20*(4), 1944-1954.

Spierer, L., Manuel, A. L., Bueti, D., & Murray, M. M. (2013). Contributions of pitch and bandwidth to sound-induced enhancement of visual cortex excitability in humans. *Cortex, 49*(10), 2728-2734.

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

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and cognition, 112*, 92-97.

Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience, 9*(3), 206.

Subra, B., Muller, D., Fourgassie, L., Chauvin, A., & Alexopoulos, T. (2018). Of guns and snakes: testing a modern threat superiority effect. *Cognition and emotion, 32*(1), 81-91.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403-409.

Takahashi, N., Oertner, T. G., Hegemann, P., & Larkum, M. E. (2016). Active cortical dendrites modulate perception. *Science, 354*(6319), 1587-1590.

Terhune, D. B., Croucher, M., Marcusson-Clavertz, D., & Macdonald, J. S. (2017). Time contracts and temporal precision declines when the mind wanders. *Journal of Experimental Psychology: Human Perception and Performance, 43*(11), 1864.

Therrien, A. S., & Bastian, A. J. (2015). Cerebellar damage impairs internal predictions for sensory and motor function. *Current opinion in neurobiology, 33*, 127-133.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences, 21*(9), 649-665.

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive psychology, 104*, 57-82.

van den Hurk, J., Van Baelen, M., & de Beeck, H. P. O. (2017). Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proceedings of the National Academy of Sciences*, 201612862.

Van Kerkoerle, T., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature communications, 8*, 13804.

Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience, 10*(11), 792.

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology, 24*(11), 1256-1262.

Vida, M. D., & Behrmann, M. (2017). Subcortical Facilitation of Behavioral Responses to Threat. *Scientific reports, 7*(1), 13087.

Ward, R., Calder, A. J., Parker, M., & Arend, I. (2007). Emotion recognition following human pulvinar damage. *Neuropsychologia, 45*(8), 1973-1978.

Ward, R., Danziger, S., & Bamford, S. (2005). Response to visual threat following damage to the pulvinar. *Current Biology, 15*(6), 571-573.

Wearden, J., Williams, E. A., & Jones, L. A. (2017). What speeds up the internal clock? Effects of clicks and flicker on duration judgements and reaction time. *The Quarterly Journal of Experimental Psychology, 70*(3), 488-503.

Williams, M. A., Baker, C. I., De Beeck, H. P. O., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature neuroscience, 11*(12), 1439.

Xing, Y., Ledgeway, T., McGraw, P. V., & Schluppeck, D. (2013). Decoding working memory of stimulus contrast in early visual cortex. *Journal of Neuroscience, 33*(25), 10301-10311.

Yaple, Z. A., Vakhrushev, R., & Jolij, J. (2016). Investigating emotional top down modulation of ambiguous faces by single pulse TMS on early visual cortices. *Frontiers in neuroscience, 10*, 305.

Zeki, S. (2015). Area V5—a microcosm of the visual brain. *Frontiers in integrative neuroscience, 9*, 21.

Zhu, X., Zhu, Q., Shen, H., Liao, W., & Yuan, F. (2017). Rumination and default mode network subsystems connectivity in first-episode, drug-naive young patients with major depressive disorder. *Scientific reports, 7*, 43105.