n

Revina, Yulia (2017) *Influence of scene surround on cortical feedback to non-stimulated primary visual cortex.* PhD thesis.

http://theses.gla.ac.uk/8016/

Influence of scene surround on cortical feedback to non-stimulated primary visual cortex

Yulia Revina, BSc

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

Institute of Psychology & Neuroscience
College of Medical, Veterinary & Life Sciences
University of Glasgow

March 2017

Contains studies performed in:
Centre for Cognitive Neuroimaging
Department of Psychology
University of Glasgow
Glasgow
G12 8QB
United Kingdom

Abstract

Most of the time we are not passively viewing scenes but want to extract behaviourally relevant information. In addition, objects do not often occur in isolation outside the visual scientist's laboratory but are embedded in complex visual scenes. If the brain is to be adaptive, it needs to process visual information with regards to its context. Thus perception is not purely determined by the specific input to the retina but depends on the surrounding scene, objects, attention, memory, prior knowledge, expectations and predictions.

Traditionally, the visual system in the human brain has been viewed as having a hierarchical organisation with signals travelling in one direction: input from the eyes arrives at "lower" order areas, which then transmit their computations to "higher" order areas. As one moves up the hierarchy, visual areas code more complex and more abstract information, and after the final processing stage, the system gives an output. However, in reality things are not so simple. In fact, in the primary visual cortex (V1), which is one of the first visual processing stages in the brain, external stimuli constitute less than 10% of the total input. The rest of the input originates from internal connections, either within V1 itself or via signals arriving from "higher" areas, back down to V1. In this way, "higher" areas can tell "lower" ones about the bigger picture and the neighbouring elements. This internal processing in the brain is the mechanism which provides context and enriches the information reaching us from the external world.

The signals arriving to V1 from the retina are referred to as *feedforward*, while the signals going in the opposite direction, from higher areas back to V1, are called *feedback*. Each neuron responds to its preferred stimulus in a specific region of the visual field, called the receptive field. Feedforward signals act on the central region of a neuron's receptive field, while feedback signals act on a larger surround region and thus are able to inform the centre about the surrounding context. However, it is not well established which aspects of the surrounding scene define these contextual interactions. This thesis investigated the influence of the scene surround on feedback to V1. We aimed to establish how the scene surround contributes to informative feedback signals.

An introduction about what is already known regarding the function of feedback and the information it transmits is provided in **Chapter 1**. I give an overview of the previous studies which highlight the various contextual roles of feedback, such as perceptual

grouping, contour and object completion, expectation, attention and prediction, as well as being the mechanism allowing visual imagery.

Chapter 2 aimed to address whether feedback provides coarse or fine-grained information about the surrounding scene. Since during normal viewing both feedback and feedforward signals are present, we investigated feedback signals in isolation by using a partial occlusion paradigm to remove meaningful feedforward input in a specific region of the scene. We filtered the scene surrounding the occluded region into a fine-grained and a coarse version. We also varied how much information was shared between the fine-grained and coarse version of the same scene. This was done to investigate whether the information feedback carried was tightly tuned to the spatial scale of the surrounding scene, or whether the information it contained was similar across the two types of the scene surround. We found that the feedback contained signals about both coarse and fine-grained surrounds, but there was also some overlap between these feedback signals. In addition, we found that the feedback information did not correspond to a direct “filling-in” of the missing feedforward input, suggesting that feedback and feedforward signals represent the scene in different ways.

In **Chapter 3** we took a closer look at the amount of meaningful scene surround that is necessary to elicit informative feedback signals. The results showed that increasing the amount of scene information in the surround resulted in more meaningful feedback signals. We confirmed our earlier finding that the feedback information in the occluded region is dissimilar to the corresponding feedforward input when the feedforward region is isolated from the scene surround. Adding the scene surround to the feedforward stimulus increased this feedback/feedforward similarity. Overall, these findings point to the notion that feedback signals combine with feedforward input under normal visual processing. Isolated feedforward input in the absence of the surround provides V1 neurons with impoverished information.

Neighbouring elements of the scene or its overall global structure can be sources of context. In **Chapter 4** we explored which regions of the scene surround contribute the most to the contextual feedback signals arriving at V1 – is this limited to only local neighbouring regions or does the feedback directly contain information about the overall global image structure, taking into account distant retinotopic regions as well? In the first experiment, we used simple global structures made up of four Gabor elements and showed that such simplistic shapes failed to induce contextual feedback into the occluded region.

However, in the presence of feedforward information, we saw that feedback from the local surround combined with identical feedforward input to give rise to different activity patterns in that feedforward region. This suggests that feedback may be recruited differentially depending on whether feedforward stimulation is present or absent. In the second experiment, we used natural scenes and tested whether contextual feedback can originate from a distant retinotopic region in the situation when the local scene surround was not informative. We manipulated scene information in a distant retinotopic region (in the opposite hemisphere) while keeping the local neighbouring surround information the same. The results showed a lack of meaningful feedback in the occluded region, and that feedback from the distant surround had a negligible effect on the identical feedforward information, in contrast to the finding obtained previously with the local surround. These findings suggest that feedback preferentially originates from nearby regions and provides context to disambiguate local feedforward elements. Therefore context about the global scene structure may arise from a series of local surround interactions.

Chapter 5 summarises these findings and discusses the overarching themes regarding the content of feedback and its role in full visual processing. At the end, I propose some future research directions.

Table of Contents

List of Tables

Chapter 2

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix A

Appendix B

List of Publications

Journal articles

Revina Y., Petro, L.S. & Muckli, L. (in preparation). Feedback to nonstimulated V1 contains both high and low spatial frequency information.

Revina Y., Petro, L.S. & Muckli, L. (in preparation). Increased scene information in the surround enhances contextual feedback.

Revina Y., Petro, L.S. & Muckli, L. (in preparation). Feedback signals from the local surround are combined with feedforward information in human V1.

Oral Presentations

Revina Y., Petro L.S., Denk-Florea C., Blum S., Kriegeskorte N. & Muckli L. (2016). Feedback signals from the local surround are combined with feedforward information in human V1. *SINAPSE Annual Scientific Meeting 2016,* Stirling, UK.

Revina Y., Petro L.S., Denk-Florea C., Blum S., Kriegeskorte N. & Muckli L. (2016). Meaningful feedback to occluded V1 is improved by increasing local information in the surround, *Journal of Vision, 16(12),* 572. *Visual Sciences Society Meeting 2016,* St Pete's Beach, FL, USA.

Revina Y., Petro L.S., Denk-Florea C. & Muckli L. (2015). Increased stimulation of the non-classical receptive field region results in more information in occluded V1, *Perception, 44(S1),* 370. *European Conference on Visual Perception,* Liverpool, UK.

Revina Y., Petro L.S., Denk-Florea C. & Muckli L. (2015). Increased stimulation of the non-classical receptive field region results in more information in occluded V1. *Psychology Postgraduate Affairs Group 30th Annual Conference,* Glasgow, UK.

Posters

Revina Y., Petro L.S., Blum S., Kriegeskorte N. & Muckli L. (2016). Feedback signals from the local surround are combined with feedforward information in human V1.

Perception, 45(S2), 47. *European Conference on Visual Perception,* Barcelona, Spain.

Revina Y., Petro L.S., Denk-Florea C., Blum S., Kriegeskorte N. & Muckli L. (2016). Meaningful feedback to occluded V1 is improved by increasing local information in the surround. *Glasgow Neuroscience Day 2016*, Glasgow, UK.

Revina Y., Petro L.S. & Muckli L. (2015). Large amount of surrounding context needs to be visible to induce scene feedback in V1. *Glasgow Neuroscience Day 2015,* Glasgow, UK

Revina Y., Petro L.S. & Muckli L. (2014). Cortical feedback: spatial frequency content and generalisation to feedforward signals. *11th Scottish Neuroscience Meeting*, Glasgow, UK.

Revina Y., Petro L.S., Rao I., Smith F.W. & Muckli L. (2014). Cortical feedback: spatial frequency content and generalisation to feedforward signals. *Human Brain Mapping: 20th Annual Meeting*, Hamburg, Germany.

Revina Y., Petro L.S., Smith F.W. & Muckli L. (2013). Investigating the spatial frequency content of cortical feedback using fMRI. *IMPRS NeuroCom Summer School*, Leipzig, Germany.

Revina Y., Petro L.S., Smith F.W. & Muckli L. (2013). Investigating the spatial frequency content of cortical feedback using fMRI. *Human Brain Mapping: 19th Annual Meeting*, Seattle, USA.

Revina Y., Petro, L.S. & Muckli L. (2013). Investigating the spatial frequency content of cortical feedback using fMRI. *Glasgow Neuroscience Day 2013*, Glasgow, UK.

Acknowledgements

First of all, I would like to thank my supervisor Lars Muckli for providing me the opportunity to work on such an exciting topic. I have always felt inspired by our meetings and discussions about the visual system and I enjoyed puzzling over the data together.

A word of gratitude is due for the BBSRC who funded this PhD. Without their support this project would not have occurred. Thank you also to Frances Crabbe for teaching me how to operate the MRI scanner and running the scans with me.

I would like to thank Lucy for always being available for a chat, to answer a question, look at my graphs and for always reassuring me I was doing fine. Lucy, your tips on writing have also been extremely helpful. I feel I can now express myself in a reasonably coherent way! ☺

Everyone else in the Muckli lab, past and present! Angus, Fiona, Grace, Hanna, Johanna, Luca, Matt, Tyler – thank you for being a fantastic team! Thank you for all your help and expertise, fun conference times, movie nights and just being such a friendly bunch of people!

Thank you also to all the undergraduate and masters students who have worked with me – Cristina, Gavin, Isa and Sebastian. Your help with the data collection and processing was invaluable. And thank you for asking me difficult questions and challenging me.

Everyone at theGIST, thank you for having me as part of the team. I learned a great deal about science communication over the four years and have improved the more formal aspects of my science writing as well.

Steph and Alex – thanks for all the fun times, hill walks and chats. It has made my time in Glasgow all the more fun.

Thank you to my parents for supporting me for all the years I've been at University (feels like forever). Your reassuring words always made me feel better when everything felt like it was going wrong. Thank you also for feeding me when I got too busy to cook! ☺

Finally, a huge thanks goes to my dearest Sam. Thank you for always being supportive no matter what crazy venture I decide to go into. Also thank you for pointing out I am just too stubborn to ever give up, and hence I will be just fine ☺. Your encouragement has stopped me from collapsing under the weight of the imposter syndrome! Lastly, thank you for lending your proof-reading eyes and spotting all [the] missing "the's". Now that I have learned all about grammar, I look forward to proof-reading *your* thesis.

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 other degree at the University of Glasgow or any other institution.

[redacted]

Yulia Revina

Abbreviations

BOLD	Blood oxygen level dependent
CI	Confidence Interval
Cpd	Cycles per degree (of visual angle)
FB	Feedback
FF	Feedforward
FFA	Fusiform face area
fMRI	Functional magnetic resonance imaging
GLM	General Linear Model
HSF	High spatial frequency
L1-6	Cortical layer 1-6
LFP	Local field potential
LGN	Lateral Geniculate Nucleus
LOC	Lateral occipital complex
LSF	Low spatial frequency
MRI	Magnetic resonance imaging
MT	Medial temporal area
MVPA	Multivariate pattern analysis
pRF	Population receptive field mapping
RF	Receptive field
ROI	Region of interest
RSC	Retrosplenial cortex
SD	Standard deviation
SF	Spatial frequency
SVM	Support Vector Machine
V1	Primary visual cortex
V2	An early visual area
V5	Motion sensitive region in extrastriate visual cortex, also known as MT

1 General introduction

Most of the time we are not just passively viewing scenes but want to extract behaviourally relevant information. In addition, objects do not often occur in isolation outside the visual scientist's laboratory but are embedded in complex visual scenes. If the brain is to be adaptive, it needs to process visual information with regards to its context. For example, in **Figure 1.1A**, is the middle object the letter "B" or the number "13"? In **Figure 1.1B**, the squares labelled A and B are actually the same colour, but appear different because the visual system takes the shadow into consideration. The reader is invited to check this for themselves by covering up the other squares and then comparing A and B once more. Thus perception is not purely determined by the specific input to the retina but depends on the surrounding scene and objects, attention, memory, prior knowledge, expectations and predictions. Investigating how the brain achieves this is a question which has kept psychologists busy for many decades (and still is!).

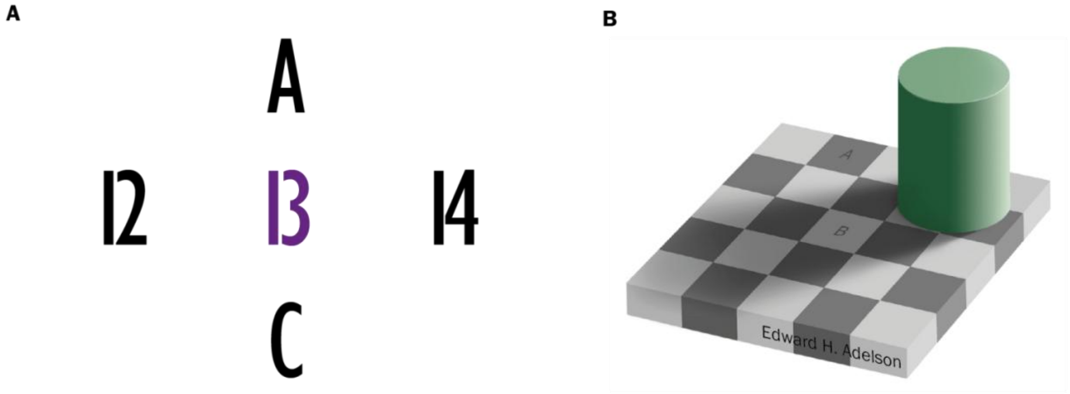


Figure 1.1 | Context is often needed to interpret what we see. A) Is the object in the middle the letter "B" or the number "13"? B) The squares labelled A and B are actually the same colour, but appear different because the visual system takes the shadow into consideration.

Traditionally, the visual system in the human brain has been viewed as having a hierarchical organisation with signals travelling in one direction: input from the eyes arrives at "lower" order areas, which then transmit their computations to "higher" order areas. As one moves up the hierarchy, visual areas code more complex and more abstract information, and after the final processing stage, the system gives an output. However, in reality things are not so simple. In fact, in the primary visual cortex (V1), which is one of the first visual processing stages in the brain, external stimuli constitute less than 10% of the total input (Budd, 1998; Douglas & Martin, 2007). The rest of the input originates from internal connections, either within V1 itself or via signals arriving from “higher” areas, back down to V1. In this way, “higher” areas can tell “lower” ones about the bigger picture and the neighbouring elements. This internal processing in the brain is the mechanism which provides context and enriches the information reaching us from the external world.

This chapter will review some of the things we already know about the primary visual cortex and how internal processing helps us perceive the world by taking surrounding context, prior knowledge and behavioural goals into account. Feedback signals have a role in modulating feedforward input, as well as enabling internal computations in the absence of feedforward stimulation. Feedback has a variety of tasks, such as: the combination of local elements into global structures, generating internal models about the world for providing predictions and expectations, and supporting imagery and working memory. I will start by giving a short overview of the anatomy of the visual system and will move onto explaining in more detail some of the roles the internal signals have in visual processing.

1.1 Primary visual cortex

Vision is important – the human brain dedicates around a quarter of its neocortex to visual processing (Van Essen, 2004; Wandell, Dumoulin, & Brewer, 2007). The visual system has traditionally been thought of as having a hierarchical organisation with “lower” order areas transmitting information to “higher” areas, and as one moves up the hierarchy, visual areas code more complex and more abstract information about the visual input. However, the true relationship is slightly more complex since lower order areas have direct connections to several higher levels of the hierarchy, as well as higher areas having connections back to lower areas (Felleman & Van Essen, 1991). The primary visual cortex is the first stage of the hierarchy. Perhaps the most studied visual area, it is an important region since nearly all of the visual information reaching the cortex passes through here. V1 has also been implicated in visual awareness (Holmes, 1918; Tong, 2003) – an important region indeed! Damage to parts of V1 leads to blindness in the corresponding areas of the visual field. However, feedforward activity in V1 is not in itself sufficient for awareness (Rees et al., 2000). Visual signals from the retina are transmitted to V1 via the lateral geniculate nucleus (LGN) in the thalamus. V1 and the rest of the early visual cortex is organised retinotopically; that is neighbouring parts of the visual field are projected to adjacent parts of the cortical surface (Wandell et al., 2007). In this way, the cortex forms a map of the visual world. However, due to cortical magnification, the central region of the visual field, the fovea, is processed by a larger part of the cortex.

Signals passing from lower order to higher order areas are referred to as *feedforward*. In V1, these inputs from the LGN constitute less than 10% of the total excitatory input (Budd, 1998; Douglas & Martin, 2007). Carandini et al. (2005) estimate that the best current models, mostly based on feedforward processing, can only explain around 40% of the response variance in V1. If most input to V1 does not originate from external stimulation, then internal intra- and inter-areal connections must explain the rest of the response variance. V1 receives many inputs from various cortical areas, both visual and non-visual (Muckli & Petro, 2013). For example, V1 receives ten times as many axons from V2 than it does from the LGN (Budd, 1998). These connections and signals from higher to lower cortical areas are referred to as *feedback*. **Figure 1.2** shows the many brain regions which send feedback to V1. In addition to feedforward and feedback signals, there are *lateral* connections within small regions of each cortical area.

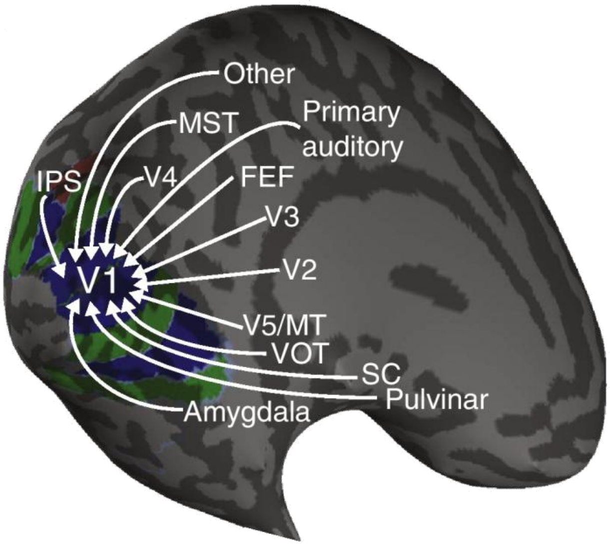


Fig 1.2 | V1 receives feedback from many other cortical areas, both visual and non-visual. For example, visual areas like V2-5, the primary auditory cortex and some subcortical structures such as the amygdala send signals back to V1. Used with permission from Muckli and Petro (2013), *Current Opinion in Neurobiology*.

1.2 Receptive fields

Each V1 neuron responds to a preferred stimulus, such as a line of a particular orientation, in a specific region of the visual field – this is the so called classical or central *receptive field* (RF, Hubel & Wiesel, 1959). However, the RF does not just consist of this central part, but also includes a near and a far surround region. Stimuli presented in the near surround RF, together with the central RF, usually facilitate the response to a visual stimulus. Presenting stimuli of the same orientation in the far surround instead suppresses the response to optimal stimuli in the centre. The LGN input acts on the central RF, with lateral connections extending along the near surround, while the far surround receives feedback from other cortical regions (Angelucci & Bressloff, 2006). The surround RF (comprising of the near and far areas) is also referred to as the non-classical RF, since the presence of stimuli here does not invoke spikes, but can still modulate the response to stimuli in the classical RF. Receptive field size increases along the processing hierarchy (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013), so feedback from higher areas, contributing to the far surround RF, carries information about a larger region of the visual field, compared to the classical RF. In V1 the central RF is around 1° of visual angle (Levitt & Lund, 2002). Surround RFs are on average 4.6 times larger (Angelucci & Bressloff, 2006). Angelucci and Bressloff (2006) found that at eccentricities between 2° and 8° of visual angle, V1 surround sizes were on average 5.1° of visual angle and ranged up to 13°. Ichida, Schwabe, Bressloff and Angelucci (2005) also reported larger surrounds of up to 28° in macaques. Feedback connections convey information to V1 from regions of visual space that are much greater than that conveyed by lateral connections. The size of this feedback region also increases with cortical distance from V1, being larger in the medial temporal area (MT) than in V2 (an area lower in the hierarchy compared to MT), for instance (Angelucci & Bressloff, 2006).

1.3 Why does the brain need feedback?

Traditionally, the visual system has been viewed in a bottom-up, hierarchical manner, with each stage processing more and more complex aspects of the visual stimuli. However, neurons do not act as simple linear feature detectors of bottom-up stimulus input but are also modulated by top-down internal processing (Kayser, Körding, & König, 2004). On the level of single neurons, feedback from the surround RF has been shown to be essential for some properties of neurons. For example, Rao and Ballard (1999) found that removing feedback in their model disrupted the function of endstopped neurons (which normally reduce their response when an optimally oriented line extends beyond the classical RF). Murphy and Sillito (1987) have also shown that removal of feedback from area 17 and 18 in cats (homologous to V1) reduced end-inhibition of LGN cells. On a larger scale, Gilbert and Li (2013) propose that "each neuron is a microcosm of the brain as a whole, with synapses carrying information originating from far flung brain regions." Influences from the surround have effects on neuronal response, and so we must look beyond the classical RF if we are to understand visual processing (Angelucci & Bullier, 2003). Thus in recent years, there has been a shift to acknowledging that top-down processing, via feedback connections from higher areas to lower areas, also plays an important role in visual processing.

The visual system can provide a variety of responses to the same visual input. Perception is not purely determined by what is detected by the retina, but also by cognitive factors, such as attention, memory, expectations and predictions about the upcoming visual stimulation and the task at hand. We are not passively viewing scenes most of the time, but often want to extract behaviourally relevant information. As for recognising specific objects, they often do not occur in isolation outside the vision scientist's laboratory, but are embedded in complex visual scenes. Feedback signals are thought to be the mechanism producing this visual and behavioural context.

Processing visual input in context is helpful because local features may be insufficient to identify a specific object, since the object may be, for example, too small, occluded, camouflaged or ambiguous (Oliva & Torralba, 2007). Moreover, the same feedforward input may arise from several different objects, and thus the overall context of the scene can help to choose one inference over another. For example, in conditions of poor visibility such as blur, the same shape can correspond to either a hairdryer or a drill. In this case, the context of the room, either a bathroom or a garage, would help the perceiver to recognise the true object correctly (**Figure 1.3,** Bar, 2004). Context also affects the efficiency of search and recognition of objects. Objects appearing on a consistent and familiar background are detected more accurately than objects which appear in an inconsistent scene (e.g. Palmer, 1975). In addition, the recognition of objects that are highly associated with a certain context helps the identification of more ambiguous objects which share the same context (Bar & Ullman, 1996). This phenomenon demonstrates that predictable properties of the environment facilitate perception and object recognition. Such prior knowledge can constrain the interpretations of an object we are viewing. Bar (2004) proposed that these context-driven predictions can in fact additionally allow us to choose not to attend to the object at all if none of the possible identities suggested by the context are of immediate behavioural interest.

Nature Reviews | Neuroscience

Figure 1.3 | An identical object can appear as a hairdryer or a drill depending on scene context. Used with permission from Bar (2004). *Nature Reviews Neuroscience*.

Context can come from a variety of sources and can be spatial or temporal. For example, context can be activated by global scene information (Bar, 2004) and we can quickly extract scene category from brief presentations of around 100 ms. Natural scenes have predictable configurations that we are familiar with from previous experience. For example, we know what a generic city, a generic forest, and so on, looks like. Therefore, it is easy to quickly extract context from natural scenes. Haslinger et al. (2012) showed that when viewing natural scenes, the surround, spike history, and local field potentials could explain the firing rate almost as much as the input to the classical RF, highlighting the importance of feedback and context on neuronal firing. Contextual feedback about scene information specifically, may originate from the parahippocampal place area (PPA) and the retrosplenial cortex (RSC), as these cortical regions have been shown to code for spatial as well as non-spatial context (Bar & Aminoff, 2003).

Context can also be temporal, such as learning a particular predictable sequence of events. For example, seeing the letters A, B and C in this order may provide context for expecting the letter D to appear next. Gavornik and Bear (2014) found that after mice were trained on a specific sequence, V1 showed activity for the full sequence, even when certain elements were omitted. The authors interpreted this as the brain making intelligent guesses to form visual percepts from limited information.

The following sections will elaborate some of the contextual influences that feedback underlies – from identifying objects and segmenting them from background, to allowing expectations and prior knowledge to modulate perception.

1.3.1 Global stucture from local elements

In a cluttered environment, the visual system has to identify how local features fit into the global picture. Objects are usually embedded in complex visual scenes, and are often not fully visible, for example, being occluded by other objects in front of them, being in a shadow, in limited viewing angles and so on. When viewing a scene, one role of the visual system is to segment it and form boundaries around various objects. This allows the identification of surfaces, grouping of local elements into single coherent objects and segregation of figures from background. For example, in order to identify which edges belong to the same object on either side of an occluder, the integration of contours and surfaces is needed. Lee and Nguyen (2001) found that V1 responds to illusory contours in Kanizsa figures. Since V1 neurons can receive feedback from a much larger retinotopic area than the classical RF, feedback to V1 can signal about the global image structure and modulate the neurons' response to local features. On a single neuron level, these global to local interactions correspond to the centre-surround interactions of the neurons' receptive fields.

Many authors have proposed that feedback and lateral connections play an important role in perceptual grouping (Angelucci & Bullier, 2003; Ciaramelli, Leo, Del Viva, Burr, & Ladavas, 2007; Coen-Cagli, Kohn, & Schwartz, 2015; Fahrenfort, Scholte, & Lamme, 2007; Gilad, Meirovithz, & Slovin, 2013; Gilad, Pesoa, Ayzenshtat, & Slovin, 2014; Hess & Field, 1999; Scholte, Jolij, Fahrenfort, & Lamme, 2008; Volberg, Wutz, & Greenlee, 2013). For example, Scholte et al. (2008) showed that while boundary detection occurs in a feedforward fashion, surface segregation requires feedback signals towards the early visual areas. Disrupting this feedback activity interferes with figure-ground segregation (Fahrenfort et al., 2007). By using thermal deactivation of a higher area Schmidt, Lomber, Payne and Galuske (2011) demonstrated that feedback connections are necessary for integration of local cues into a global motion percept. In addition, when local elements are perceptually combined into a global shape, activity in V1 is reduced, presumably as a result of grouping processes which are fedback from higher areas (Murray, Kersten, Olshausen, Schrater, & Woods, 2002; de-Wit, Kubilius, Wagemans, & Op de Beeck, 2012).

One way to code for a difference between figure and background or discontinuous elements, could be a differential recruitment of surround suppression, which feedback contributes to (Nassi, Lomber, & Born, 2013). Homogeneous regions of an image are likely to belong to the same surface, such as sand on a beach, while heterogeneities are likely to signal different surfaces and objects. Surround can be recruited to code for this difference. If the image is homogeneous, the output of the RF is suppressed, else the surround suppression mechanism is disengaged (Coen-Cagli et al., 2015) and changes in the local structure are detected, for example, a change in the surface or an edge. In this way, identical feedforward sensory stimuli can be interpreted differently depending on surrounding information.

Feedback is also involved in contour completion. V1 even responds to illusory borders (e.g. Grosof, Shapley, & Hawken, 1993; Kok & de Lange, 2014; Lee & Nguyen, 2001), and can be modulated by surrounding context (Rennig, Karnath, & Huberle, 2013), suggesting top-down effects. Detection of contours requires grouping of related elements into a single coherent object. This cannot be achieved via a simple element by element feedforward coding. For example, Gilad et al. (2013) used voltage sensitive dyes in monkey V1 during contour detection. They found an early response to individual contour and background elements, and a later increased activity in the contour elements, together with the suppression of the background. This late activity went beyond the representation of individual background and contour elements, and was correlated with performance. The authors suggested that V1 is involved in the transformation processes from discrete elements in the early stages to a later representation of a coherent object.

Ability to form global percepts out of local information enables the visual system to be good at recognising objects from partial information and be able to complete the missing image elements. Lerner, Hendler and Malach (2002) investigated where these effects might occur in the visual system. Subjects were presented with three types of images – whole line drawings of objects, grid images where these shapes were occluded by evenly spaced parallel strips and scrambled images, similar to the grid images but where the visible strips were scrambled. The image within each visible strip was still intact. However, the global coherence was broken. Using functional magnetic resonance imaging (fMRI) the authors found higher activation in the lateral occipital complex (LOC) to the grid images compared to the scrambled ones, highlighting a potential locus for “completion effects” in the brain. However, a similar effect was not found in the early retinotopic areas (such as V1), leading the authors to conclude that the early visual areas are more concerned with local feature representation, while completion effects occur later on in the hierarchy. Other studies, however, have shown the role of V1 in image completion, and this effect may depend on top-down feedback.

For example, Ban et al. (2013) investigated the representation of an occluded part of an object in V1 and found that the occluded portion of an object was represented. This was modulated by prior knowledge since the activity was different to that in a “divided” condition, even though the stimulus looked the same when going behind the occluder. This effect of prior knowledge was stronger in V1 than in V2, suggesting these areas may be playing different roles in solving the completion problem. Another study on occlusion found that macaque V1 responds to an occluded bar when the occluder that covers the bar has a perceived depth such that it looks to be in front of it (Sugita, 1999). Since the bar was not visible to the recorded central receptive field, this finding points to the contextual influence of the surround. There was no response to the occluding patch when it was perceived to be at the same depth (similar to Ban et al.’s “divided” condition). The response latency for the bar behind the patch was not different from that for an unoccluded bar. Sugita suggested that it was therefore likely that the feedback signals originated from the areas very close to V1. In a study on neuronal response to whole and partial objects, Tang et al. (2014) recorded intracranial field potentials in epilepsy patients. The features of the object revealed in the partial condition varied from trial to trial. The researchers showed that even with few features present (on average 18%), neural responses in the ventral stream nevertheless retained object selectivity. Crucially, there was a response delay of around 100 ms for the partial objects, suggesting that perhaps recurrent computations based on prior knowledge about the object were needed for the object completion effects to occur. Processing delays were especially pronounced in higher areas along the visual hierarchy. Responses in occluded regions have also been seen with natural scenes. Smith and Muckli (2010) showed that activity patterns in an occluded region of a natural scene contain contextual information about the surrounding scene. Using more dynamic stimuli, such as apparent motion illusions, where two stimuli flashed in an alternating way to produce an illusion of back and forth motion, it has been shown that the apparent motion path is represented in V1, despite no stimulus ever being presented there in a feedforward manner (Muckli, Kohler, Kriegeskorte, & Singer, 2005). A similar result was also demonstrated in ferret area 17 (homologous to primate V1), where spikes were recorded in neurons situated between the stimulus representations (Ahmed et al., 2008).

1.3.2 Facilitating perception

Visual input is interpreted with prior knowledge about the world. The visual system is somewhat “lazy” and does not process everything that is out there to the same degree. Processing every single “pixel” of the visual scene would use up unnecessary energy so the brain takes shortcuts. Information in the world is largely redundant (Attneave, 1954; Barlow, 1961) and a lot of the time is predictable from one instance to the next. For example, as we are walking along a familiar route, the colours of the houses do not suddenly change as we walk past them and we have some expectations of what we will see around the corner. Therefore, it is more efficient to code for departures from the expected rather than analysing all of the visual information. A similar idea is used in digital image processing techniques, using compressed formats such as JPEG. For example, it would be redundant to store and process every pixel in the uniform white background of this page, and therefore only the deviations from white, such as the black letters, would need to be coded. In a similar fashion, the brain can be more efficient, saving bandwidth and energy by only looking for unexpected variations and transmitting these deviations.

Theories of *predictive coding* build on this idea and postulate that the brain aims to predict upcoming perceptual experiences based on internal models constructed from previous sensory experience and then encode the unpredicted input for further processing. The brain is somewhat like a black box: it has no direct access to the outside world and only knows about its internal states and spiking of neurons. Therefore, it has to infer what the stimulus was from the observed effect. A single effect may be consistent with multiple causes and therefore the brain must generate probability distributions associated with each cause (Clark, 2013). The brain is essentially an inference engine, trying to optimise the probabilistic representations of what caused the sensory input. The brain forms a model of the world which it tries to optimise using sensory inputs, with the end goal of reducing *prediction errors* (Friston, 2010). According to Friston, the brain’s task is based on the free-energy principle, aiming to minimise free-energy or surprise. There are two ways to achieve this, by either altering the sensory input (changing actions) or varying the internal state (changing perception). In this way, the brain is constantly optimising its internal models by predicting the stimulus and testing its hypotheses using new sensory input to explain away the prediction errors (Friston, 2010; Lee & Mumford, 2003; Mumford, 1991; Rao & Ballard, 1999). Rao and Ballard (1999) proposed a model in which feedback from higher areas carries predictions, while feedforward signals carry the discrepancies between the predicted and the actual sensory input – the prediction errors. Each hierarchical level attempts to predict responses at each lower level. According to Mumford (1991) higher areas deal with more abstract information, whereas lower areas deal with more concrete data. The brain learns the various patterns that occur in sensory stimuli and generates multiple hypotheses as to what input can cause them. The thalamus then votes on these hypotheses. Descending predictions arising from deep pyramidal cells are compared to incoming sensory signals, and the computed mismatch (prediction error) is transferred in the feedforward stream of the superficial pyramidal cells up to the next cortical higher level to update the internal models (Clark 2013).

Predictive coding theories hypothesise that a predictable stimulus would elicit a lower response because there is no prediction error and no new information needs to be encoded. In line with this, several studies have shown a decrease in activation when the stimulus is predictable. For example, Coen-Cagli et al. (2015) found that in a homogeneous image, the surround suppressed responses in the classical RF, which was not the case when the homogenous surround was removed from around the same feedforward patch of the image. This means that if the centre can be predicted from the surround, the information in the centre is redundant and hence neuronal firing is suppressed. Essentially, the neuron is coding for error or deviation from uniformity rather than a particular piece of content (Rao & Sejnowski, 2002). Another example comes from an fMRI study from Alink, Schwiedrzik, Kohler, Singer and Muckli (2010), who showed that when the onset or motion direction of stimuli can be predicted from the surrounding illusory motion, a smaller blood oxygenation level dependent (BOLD) response in V1 is observed. This effect in V1 is retinotopically specific to the stimulus representation on the apparent motion path. No such effect was observed in V5 (motion processing visual area).

Top-down predictions and expectations can facilitate or bias perception. Bastos et al., (2012) suggested that higher cortical regions create top-down expectations about the stimulus, which are sent to lower visual areas to aid perception. By using multivariate pattern analysis techniques Kok, Jehee and de Lange (2012) demonstrated that these perceptual expectations decreased V1 activity, while at the same time sharpening stimuli representation. This representational sharpening also correlates with behavioural response. Kok, Failing and de Lange (2014) showed that prior expectations evoked feature-specific activity in V1 that was in fact similar to the activity elicited by the corresponding actual stimulus. Top-down expectations are in fact powerful enough to create percepts out of noise. For example, observers can detect patterns in random noise which match their internal representation of a face (Smith, Gosselin & Schyns 2012).

Stimuli which are predictable in the context of apparent motion are detected more often than unpredictable stimuli (Schwiedrzik, Alink, Kohler, Singer, & Muckli, 2007). Predictions help keep perception coherent as demonstrated by studies of the retinal blind spot. In this patch of the retina, no photoreceptors are present; however, we do not see a “gap” in the visual field. Predictions are useful for filling in the missing pieces. For example, Maus and Nijhawan (2008) found that when observers were asked to judge the final position of a bar moving into the blind spot, they perceived the bar to disappear in positions well inside the blind spot. This kind of extrapolation must be due to predictions based on the movement of the bar, and does not occur when the bar stimulus is abruptly “switched off” and becomes invisible outside the blind spot, presumably because feedforward information takes over in coding the veridical offset of the stimulus. Panichello, Cheung and Bar (2013) argue that predictive mechanisms may shape the contents of visual awareness during sensory ambiguity, and when input is less ambiguous, predictions serve more of a facilitatory role. Such predictive modulation is advantageous as it is more useful to generate a meaningful representation of the world, rather than to faithfully represent a noisy signal.

Binocular rivalry – when separate stimuli are shown to each eye and the combined percept alternates between one and the other – is a useful paradigm to study predictions under ambiguity. For example, Chopin and Mamassian (2012) showed that a stimulus will tend to dominate rivalry if it has been presented more frequently in the past. The authors proposed a mechanism for this – predictive adaptation. This follows the simple rule that the distribution of events in recent history should match the one observed in the reference time frame of the remote past. The expectation of the next percepts should match this so that the rule is not broken. The phenomenon of binocular rivalry itself could perhaps be explained with top-down predictions. Hohwy, Roepstorff and Friston (2008) argue that binocular rivalry is a reasonable (knowledge-oriented) response to an ecologically unusual stimulus. One stimulus is chosen to dominate perception, but a large amount of prediction error remains and thus perception shifts to include the unexplained variance. No stable solution is found and hence perception alternates. Presumably seeing a blend of the two stimuli is also very artificial because the visual system’s prior expectations would predict a correlation between the inputs to the two eyes.

When visual input is noisy or degraded, predictions can help perception. In a study where images of objects were gradually revealed at each trial, Eger, Henson, Driver and Dolan (2007) found that images primed with a congruent word were recognised earlier than those primed with an incongruent word. Because the prime was a word rather than a picture, the authors concluded that the observed effect reflected top-down processing. In fact, neuronal response may be highly driven by expectation rather than a specific stimulus itself. For example, Egner, Monti and Summerfield (2010) demonstrated that the fusiform face area (FFA, traditionally associated with face perception) responded strongly also to houses as well as faces under high-face-expectation conditions. Predictive processing with face expectation and face surprise neuronal units might explain the observed effect. The authors argued that visual cortex is better explained as a sum of feature expectation and surprise responses than by bottom-up feature detection.

One source of contextual information is global scene representation (Bar, 2004; Oliva & Torralba, 2007), which feedback may signal about. Global scene structure can be used as a top-down prediction to help bias the interpretation and categorization of the scene. Global scene representations might rely on coarse information (Bar et al., 2006; Oliva & Torralba, 2006) or fine-grained information (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011). Coarse and fine-grained visual information is represented by different *spatial frequency* bands. Spatial frequency (SF) is one of the first features of the visual scene processed by the early visual cortex (e.g. Wilson & Bergen, 1979). The visual system deconstructs the input into different spatial frequency bands. Low spatial frequencies define the coarse fluctuations in light and dark patches and reflect the global information about the image. A blurry picture has low spatial frequency (LSF); high spatial frequencies (HSF), on the other hand, convey more fine-grained detail about abrupt spatial changes in the image, such as edges. Cells can be described as broadly or narrowly tuned depending on their bandpass characteristics – the SF bands they preferentially respond to. LGN cells are broadly tuned, while cortical cells are narrowly tuned. Most cells have bandpass characteristics. Cortical cells tend to have a bandwidth of 1-1.5 octaves, with HSF preference cells more narrowly tuned than those with a LSF preference (De Valois, Albrecht, & Thorell, 1982). The visual system relies on midband more, as not all SFs can be used with equal efficiency (Gold, Bennett, & Sekuler, 1999). Several studies have indicated V1 preference for stimuli in the 0.68-2 cycles per degree of visual angle (cpd) range (Haynes & Rees, 2005; Henriksson, Nurminen, Hyvärinen, & Vanni, 2008; Tong, Harrison, Dewey, & Kamitani, 2012). However, it depends somewhat on the stimulus, in terms of what SF is most optimal. For example, in tasks requiring configural information, LSF is processed faster, while in featural conditions, HSF has an advantage (Goffaux, Hault, Michel, Vuong, & Rossion, 2005). Awasthi, Sowman, Friedman and Williams (2013) also found the importance of LSF for face stimuli, and additionally that HSF supports the processing of place stimuli.

Rajimehr, Devaney, Bilenko, Young and Tootell (2011) lent support to this result by finding that the parahippocampal place area is preferentially activated by HSF information.

The *gist* of a scene refers to the meaningful information that an observer can identify from just a glimpse of a scene, for example, they can extract the category it belongs to. The gist is likely to originate from global scene features, which mostly come from LSF information (Oliva & Torralba, 2006). If the role of feedback is to provide the context of the scene, it may transmit information about the gist – its global structure. LSF information may be well suited, therefore, to provide the context of the scene via top-down mechanisms in order to bias the visual system to a subset of interpretations, with HSFs then filling in the fine-grained details (Bar et al., 2006; Oliva & Torralba, 2006).

Some support for this hypothesis comes from studies showing that LSF is processed faster (Bar et al., 2006; Bar, 2003; Breitmeyer, 2014; Kveraga, Boshyan, & Bar, 2007). Prefrontal cortex, specifically the orbitofrontal cortex (OFC), may have a critical role to play in providing predictions using LSF information. A coarse version of the input is rapidly projected from early visual areas to the OFC. This LSF image is sufficient to activate an initial guess about the input. The predictions originating from the LSF information can then activate the corresponding visual representations in object processing regions in the ventral temporal cortex, to facilitate object recognition by biasing bottom-up processes to concentrate on a small set of the most likely object representations. Kveraga et al. (2007) proposed that coarse information about the scene is carried rapidly to the OFC using magnocellular channels (which are activated by coarse luminance information). They found that stimuli biased toward magnocellular processing differentially activated the OFC compared to parvocellular biased stimuli (more fine-grained). In addition, OFC activity predicted the performance advantage for the magnocellular but not the parvocellular stimuli. Bar et al. (2006) also showed that LSF stimuli elicit functional coupling between early visual areas and the OFC, while HSF ones do not.

LSF may not always be processed first. Schyns and Oliva (1994) demonstrated that exposure to duration changed the interpretation of hybrid stimuli. Short duration produced more accurate matching of LSF hybrids, while long duration – more accurate matching of HSF hybrids. On the other hand, priming studies revealed that after a 30 ms exposure, both LSF and HSF information from a hybrid stimulus (one image in HSF superimposed over another image in LSF) was perceived (Oliva & Schyns, 1997; Parker, Lishman, & Hughes, 1996). Oliva and Torralba (2006) argued that scene gist proceeds in a global manner but does not necessarily rely on LSF.

Schyns and colleagues proposed the flexible usage hypothesis (Schyns & Oliva, 1997) and highlighted the importance of task demands on which SF bands are preferentially processed. If the scene is unknown and needs quick categorisation, then LSF might be more useful to give an initial quick guess. But if the scene category is already established, HSF scales may become more useful in pinning down the particular example of the category (Schyns & Oliva, 1994). For example, if we already know we are viewing images of cities, higher scales would be more useful in differentiating between the Big Ben tower and the Empire State Building. One empirical example for flexible usage comes from Schyns and Oliva (1999), who used a face categorisation task based on gender or emotion, and found that LSF is used for gender categorisation, while HSF is important for emotion. This shows that different SF bands are available to the visual system early on and task demands bias the visual system to preferentially process a particular one. Schyns and Oliva (1997) suggested that scale usage could be flexibly determined by the diagnosticity of scale-specific cues for different categorisations of scenes and faces. Other studies have also shown top-down modulation of scale usage (Sowden, Özgen, Schyns, & Daoutis, 2003; Özgen, Sowden, Schyns, & Daoutis, 2005). For example, Sowden et al. (2003) investigated modulation of SF processing by top-down attention. They used cross-modal cueing, that is, they used a sound cue to indicate the SF to pay attention to in the trial, to make sure the effect was top-down rather than perceptual. The cueing encouraged participants to attend to the SF channel they expected to carry the relevant information. They found that detection of gratings was worse when it appeared in an unexpected SF. The detection of gratings is also worse when the SF varies from trial to trial (Davis & Graham, 1981; Hübner, 1996). This is consistent with the idea that selective monitoring of relevant SF channels is taking place.

Attention can also modulate responses in the early visual cortex (Gilbert & Sigman, 2007), even in the absence of visual stimulation (Silver, Ress, & Heeger, 2007). Attention effects can also appear later in time, for example, after figure-ground modulation, probably as a result of feedback (Muckli, 2010; Roelfsema, Tolboom, & Khayat, 2007). Jehee, Brady and Tong (2011) showed that attention improved responses to orientation of gratings when orientation was the task-relevant feature, but not when the contrast of the grating had to be attended, suggesting that expectation of a particular task changed visual responses (Petro, Smith, Schyns, & Muckli, 2013). Feature-based attention spreads across the visual field, even to regions of the scene not containing a stimulus (Serences & Boynton, 2007). This may be useful for increasing sensitivity to relevant visual features and facilitating detections of behaviourally relevant stimuli in the immediate future. Attention has also been shown to modulate effects of the contextual surround. A V1 response to a stimulus surrounded by identical context can either be suppressed or enhanced depending on whether the central stimulus or the contextual flanker stimulus is attended (Flevaris & Murray, 2015).

Prediction is not only a useful mechanism for visual perception, but seems to be critical for brain function as a whole. For example, in everyday life, we are making inferences about others' goals, thoughts and personalities. Social interactions require us to reason about hidden causes of behaviour. Predictable stimuli and observed actions such as watching a human make human-like movements, elicit a lower neuronal response, as opposed to a human making robot-like movements (Koster-Hale & Saxe, 2013; Saygin, Chaminade, Ishiguro, Driver, & Frith, 2012). Predictive coding may also be important for motor control – anticipating and explaining away the consequences of one's own motor actions, and models based on prediction are thought to explain the sense of conscious presence in general (Seth, Suzuki, & Critchley, 2012). When these mechanisms go wrong, it may lead to mental disorders, such as schizophrenia, which is marked by delusions and hallucinations. It may be that in schizophrenia, highly-weighted false prediction errors are propagated and revise the brain's internal model of the world. This new (incorrect) model then leads to delusions (Clark, 2013).

1.3.3 Internal computations in the absence of feedforward input

V1 can be active in the absence of bottom-up feedforward input, such as during visual imagery, working memory, sleep or hallucinations. This internally generated activation must therefore be due to feedback. Given the fine-grained nature of V1 in comparison to other brain regions, it is a good candidate area for supporting visual working memory and imagery (Petro, Vizioli, & Muckli, 2014). Harrison and Tong (2009) found that orientations of gratings held in working memory can be decoded from activity patterns in V1. Pratte and Tong (2014) demonstrated that working memory representations in V1 can be spatially specific, especially if the task requires the binding of a feature to a particular location. However, information can also be retained in a non-retinotopic manner (Ester, Serences, & Awh, 2009). Visual working memory mechanisms may be similar to visual imagery more generally (Tong, 2013) and have been shown to have common visual representations (Albers, Kok, Toni, Dijkerman, & de Lange, 2013). Furthermore, Albers et al. found that patterns of activity during mental imagery were similar to those during physical presentation of the stimuli, suggesting actual perception and imagery rely on related neural processes. However, not all authors have found that V1 is activated during internally-generated perceptual experiences. For example, ffytche et al. (1998) found that hallucinations in patients with Charles Bonnet syndrome correlated with activity in the ventral extrastriate cortex, and there was an absence of consistent activity in V1.

V1 does not just receive feedback from higher-level visual areas, but is also involved in processing information that is not strictly under the umbrella of visual perception. For example, V1 is active during sleep, which is thought to be mediated by the hippocampus and required for memory consolidation. Firing sequences evoked by experiences whilst awake are replayed in both early visual cortex and hippocampus during slow-wave sleep in mice (Ji & Wilson, 2007). However, during rapid eye movement (REM) sleep, despite this sleep stage being associated with dreaming, V1 shows an attenuation of activity (Braun et al., 1998). Hindy, Ng and Turk-Browne (2016) highlighted the role of the hippocampus in pattern completion and memory-based expectations in the visual cortex. They showed that sequence information (how much neural evidence is available about the full sequence [cue + action + outcome] from cue + action) in the hippocampus preceded outcome information (how much neural evidence is available from cue + action about the expected outcome) in the visual cortex. This suggests that hippocampus reinstates expected outcomes in the visual cortex.

Information about auditory stimuli can also be decoded in V1. Vetter, Smith and Muckli (2014) showed that category of an auditory stimulus could be read out from the primary visual cortex of blindfolded participants. This information may relate to visual imagery, although participants were not specifically asked to visualize the scene relating to the sound. The authors found a generalisation across different stimuli of the same category, suggesting that this V1 information was abstract and not a fine-grained pictorial representation.

1.3.4 Visual awareness

Lesion studies of V1 have shown that removing V1 disrupts visual awareness (Holmes, 1918; Tong, 2003). Patients with damage to V1 report not seeing any stimuli presented within the region of damage. This could be simply because the feedforward flow of information is halted at the first cortical stage, or alternatively because V1 may serve as a dynamic blackboard for generating perception, using both bottom-up and top-down mechanisms (Tong, 2013). Several studies have highlighted the role of feedback in visual awareness (Orpwood, 2013; Pascual-Leone & Walsh, 2001; Rees, 2007; Ro, Breitmeyer, Burton, Singhal, & Lane, 2003), supporting the latter hypothesis. For example, Pascual-Leone and Walsh (2001) showed that feedback from V5 was needed for increased visual awareness of phosphenes. Feedback may also be important for consciousness in general (Seth et al., 2012).

1.4 How does feedback exert its effects?

Feedback connections are an important source of modulation of the neural signal, rapidly carrying information from more distant retinotopic regions. Changes from differences in the environment cannot be predicted by simple centre-surround interactions, and hence feedback is needed for the full spectrum of possible modulation of neuronal response (Angelucci & Bressloff, 2006; Bullier, 2006). Angelucci et al. (2002) argued that the extent of lateral connections is not sufficient to account for the full effect of modulation from the surround, and in addition, the conduction velocities of lateral axons are too slow for the rapid contextual effects which are often observed (Girard, Hupé, & Bullier, 2001). On the other hand, the conduction velocities of feedback connections are as fast as those of feedforward and ten times faster than those of lateral connections (Girard et al., 2001). Feedforward, lateral and feedback connections all send excitatory inputs to the central RF. Lateral connections additionally connect to local inhibitory neurons, which in turn affect the excitatory neurons. Feedback may have a role in the reorganisation of local connectivity (Bullier, 2006). For example, feedback connections can suppress output by acting on local inhibitory neurons (Angelucci & Bressloff, 2006). In addition, feedback may act on lateral connections which in turn modulate responses in the classical receptive fields (Muckli & Petro, 2013). Feedback signals can enhance V1 activity, and the relative impact of this enhancement can be modulated by subcortical activity. Pafundo, Nicholas, Zhang and Kuhlman (2016) proposed that control of V1 can be rebalanced such that excitatory response properties are more strongly determined by bottom-up inputs or feedback.

The cortex is divided anatomically into six distinct layers, layer 1 being the most superficial, close to the pial surface, and layer 6 being the deepest layer, close to the white matter. Each layer is distinct in some way, for example in terms of the types of cells located there, their density, types of inputs and outputs and so on. Feedforward projections originate in layer 2 and 3 (L2, L3) and terminate in L4. Feedback projections originate in L5 and 6 and terminate in L1 and L5 (Felleman & Van Essen, 1991; Kravitz et al., 2013; Petro & Muckli, 2017). Recent fMRI evidence has shown that information from feedforward signals peaks in the mid-layers, while feedback information peaks in the superficial layers (Muckli et al., 2015).

Traditionally, it has been assumed that neurons simply add up all of the excitatory and inhibitory inputs and if the integrated value exceeds a threshold, then a spike is triggered. However, there is now evidence that pyramidal neurons have two distinct sites of integration, with input to one site modulating the response to input at the other. Pyramidal neurons have an apical tuft located away from the cell body. For cell bodies in L2, 3 and 5, the tuft is in L1. For cell bodies in L6, the tuft is located in L4 (Phillips, 2015). Inputs at the apical tufts of the pyramidal cells (where contextual feedback arrives) amplify response to the basal inputs (where feedforward information enters). This process is called apical amplification (**Figure 1.4**, Phillips, 2015). When apical and basal inputs coincide, the cell's response to its basal inputs is amplified (Larkum, Nevian, Sandler, Polsky, & Schiller, 2009; Larkum, 2013). Thus, via this mechanism feedback can modulate the neuron's response to the feedforward input.

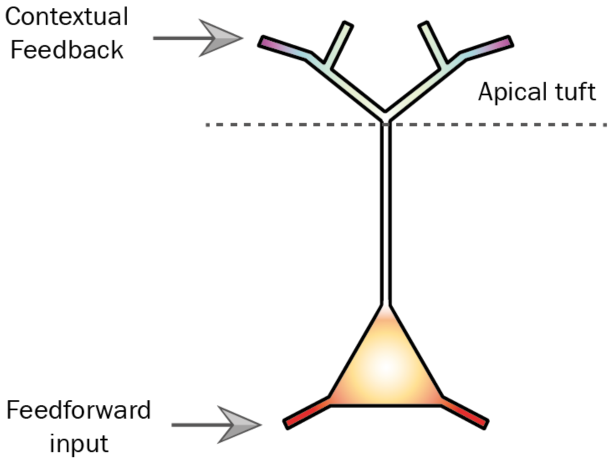


Figure 1.4 | A pyramidal neuron. Contextual feedback acts on the apical tuft and combines with feedforward basal input.

Feedback signals are usually thought of as weak and modulatory, that is, they have slow and diffuse actions and cannot cause spikes of activity, whereas feedforward signals can be thought of as driving. However, there is some evidence which suggests feedback can also have driving effects under some circumstances – strong, rapid effects on the target neurons (Covic & Sherman, 2011; Mignard & Malpeli, 1991; De Pasquale & Sherman, 2011). Hupé et al. (1998) found that, when V1 was active, a small number of V2 cells completely stopped responding to visual stimulation, during inactivation of area MT, i.e. when feedback signals were removed. However, when V1 was inactivated, V2 activity was abolished, even with an intact feedback connection from MT (Girard & Bullier, 1989). This suggests that in the absence of feedforward input from V1, feedback connections from MT cannot drive neurons in V2. Feedback signals may therefore act as a gain control of the neuronal response to the feedforward input (Bullier, 2006).

Another difference between feedback and feedforward signalling is the frequency of oscillatory activity that supports them. Feedforward influences predominate in the gamma band (higher frequencies), while feedback projections are in the alpha-beta band (lower frequencies, Bastos et al., 2012; Fries, 2015; Lee et al., 2013; Michalareas et al., 2016). Buffalo, Fries, Landman, Buschman and Desimone (2011) found that gamma band synchronisation was strongest in the superficial layers (where feedforward projections originate) and alpha-beta frequencies were strongest in the infragranular layers (where feedback projections originate).

1.5 Thesis rationale

Despite contextual modulation from feedback being an important aspect of visual processing, much about the information it transmits and what factors give rise to informative feedback remains unknown. V1 cells receive feedback from the far surround receptive field (Angelucci & Bressloff, 2006) and it has been previously shown that stimulation in the surround provides contextual information about the scene to non-stimulated parts of V1 (Smith & Muckli, 2010). Smith and Muckli showed participants three different natural scenes which had the lower right quadrant of the stimulus occluded. They recorded V1 activity patterns corresponding to this occluded non-stimulated region using fMRI and found that a linear classifier trained on these activity patterns could classify which scene was shown in each trial with above chance accuracy. There was no meaningful feedforward information about the scenes in the occluded region. Therefore these differential activity patterns must originate via feedback from the visible scene surround. However, not much is known regarding the nature of the contextual information from the surround that feedback transmits.

The aim of the thesis was to study the influence of changing information in the surround on V1 responses in non-stimulated retinotopic regions, as well as the same regions of the visual field when receiving bottom-up feedforward input. To investigate this we used an occlusion paradigm, as previously employed by Smith and Muckli (2010). During normal vision, both feedback and feedforward signals are present, and so a useful approach to studying feedback is to isolate it from the feedforward input. To achieve this, we occluded a region of the visual field in the stimulus in order to remove meaningful feedforward input about the image in that retinotopic area. Due to the retinotopic nature of V1, where neighbouring regions of the visual field map onto neighbouring parts of the cortex (Wandell et al., 2007), we were able to look at brain activity relating specifically to the occluded region. To assess brain activity in these regions we used fMRI and multivoxel pattern analysis (MVPA). These techniques are particularly suited to the task at hand because we are looking at regions of the cortex not receiving meaningful feedforward stimulation, and therefore we do not expect an increase in spiking which would lead to an overall increase in activation. Therefore, to compare the brain activity in the occluded region in response to different images, fMRI in conjunction with MVPA is a useful technique, as it is sensitive to non-spiking activity, and related more to computation and oxygen consumption rather than the actual output of the neurons (Logothetis, 2008; Muckli, 2010). Because we are not expecting to see differences in activation, a univariate approach in analysing fMRI data is not appropriate. Therefore, we used a multivariate approach to look at the pattern of brain activation, rather than an overall increase or decrease in activity (for reviews see Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006). Using MVPA we were able to pick up differences in the information pattern between the different conditions.

The specific experiments are described in **Chapters 2-4**. In **Chapter 2**, we were interested in probing the spatial frequency information transmitted by feedback. We presented participants with spatial frequency filtered scenes, a high spatial frequency and a low spatial frequency version. We investigated whether feedback carries coarse information about the scene (more low spatial frequencies) or something more precise and fine-grained (more high spatial frequencies). In addition, we wanted to know whether this information is broadly tuned and is similar across different spatial frequency surrounds, or whether feedback signals are tightly tuned to the spatial frequency of the image. We found that feedback signals were informative for both high and low spatial frequency scenes. This argues against the idea that low spatial frequencies are predominantly involved in generating top-down context (Bar et al., 2006). However, we also showed that while we were able to decode between HSF and LSF feedback, there was also a similarity between the two conditions. A classifier trained on one SF could successfully decode scenes in the other. Finally, we found that the information carried by feedback signals in the occluded region does not represent a direct filling-in of the missing feedforward information, when this feedforward information is presented in isolation (i.e. without the surround). By removing the surround and leaving a small region of the feedforward information, we decreased contextual feedback in that region. We hypothesised that the feedback/feedforward similarity depended on the amount of information available in the surround, since it has previously been shown that feedback signals do show a similar activity pattern to the missing feedforward region when it is part of the full visual scene (Smith & Muckli, 2010).

In **Chapter 3**, we probed this last finding further by assessing how the amount of scene information in the surround affects informative feedback in the occluded region and its similarity to the corresponding feedforward input. We parametrically changed the amount of scene information in the surround and found that first of all, increasing information in the surround increased meaningful feedback signals. Second, we replicated the finding of **Chapter 2** that feedback does not represent a direct filling-in of the isolated feedforward information, and feedback/feedforward similarity increases when we increase information in the surround. Finally, we show that modulating the amount of information in the surround decreases the similarity of activity patterns in regions receiving identical feedforward stimulation. This means that feedback from the surround combines with feedforward signals. Our data highlight the importance of not underestimating feedback signals when aiming to explain full visual processing.

In **Chapter 4**, we investigated which regions of the surround were more important for generating informative feedback signals. Specifically, we asked whether feedback depends more on the information neighbouring the occluded region (which we refer to as the "local" surround) or whether feedback signals transmit information based on the overall structure of the full surrounding region (which we refer to as the "global" surround). We used stimuli composed of orientations which together represented different global structures, and measured whether feedback signals carried information about the neighbouring orientations or the orientation that would be predicted in the occluded region by the overall global shape. We found that these simple stimuli which were less complex than natural scenes did not induce meaningful feedback in the occluded region. However, we did find that information from the local surround combined with feedforward input, a similar finding to that of **Chapter 3**. Finally we asked whether information from a distant region in the surround could influence feedback signals when we kept the local surround constant and uninformative and used more naturalistic images. We found that contextual information from the distant surround did not lead to meaningful feedback signals that differentiated the scenes.

The findings from all three studies are discussed in **Chapter 5** in line with the current literature and I propose ideas for further research for the remaining unanswered questions.

2 Feedback to nonstimulated V1 contains both high and low spatial frequency information

2.1 Abstract

Feedback signals originating from higher visual areas with larger receptive fields modulate the surround receptive field of V1 neurons. In this way, contextual information about other regions of the scene can be transmitted to a neuron's receptive field. Feedforward visual input can be decomposed into different spatial frequency (SF) bands, with low spatial frequencies (LSFs) conveying coarse information, and high spatial frequencies (HSFs) providing fine-grained details. However, it was previously unknown which spatial scales feedback contains. By harnessing the retinotopic nature of V1, the spatial resolution of functional MRI, an established occlusion paradigm (Smith & Muckli, 2010) and pattern analysis methods; we investigated the spatial frequency content of feedback signals, by manipulating the spatial frequency of the scene surrounding a nonstimulated occluded region of V1. Thus we were able to probe feedback signals in the absence of feedforward input. We filtered two natural scenes into a high and low spatial frequency version, and varied the SF overlaps between the HSF and LSF scenes, to manipulate the amount of shared information between them. We found that 1) both high and low spatial frequency surrounds elicited meaningful feedback in the occluded region; 2) there was a similarity between feedback signals for LSF and HSF scenes, even with a lack of SF overlap; 3) but also a difference between HSF and LSF scene feedback; and 4) information in the occluded region is not a direct filling-in of the missing feedforward input. This suggests that feedback transmits both coarse and fine-grained information. However, this information does not have a direct correspondence to the missing feedforward input.

2.2 Introduction

The majority of synaptic inputs to the primary visual cortex (V1) are non-feedforward, instead originating from lateral and feedback connections (Muckli & Petro, 2013). Feedback signals originating from higher visual areas with larger receptive fields modulate the surround receptive field of V1 neurons, also known as the non-classical receptive field (Angelucci & Bressloff, 2006). In this way, contextual information about other regions of the scene can be transmitted to a neuron's receptive field. However, despite the importance of feedback, little is known about the format of information that it carries. Visual input can be decomposed into different spatial frequency (SF) bands, with low spatial frequencies (LSFs) conveying coarse information, and high spatial frequencies (HSFs) providing the fine-grained details. Do feedback signals also carry contextual information about different spatial frequencies, and which spatial frequencies dominate? Feedback could be fine-grained and spatially precise, "filling-in" the context, or spatially diffuse, providing the general expectation of some property, such as a category to which a natural scene belongs.

One source of contextual information is global scene representation (Bar, 2004; Oliva & Torralba, 2007). Therefore, it seems plausible that feedback would also carry information about global scene structure (providing the gist of the scene) and this coarse representation may correspond to low spatial frequency information (Bar, 2004; Oliva & Torralba, 2007). Consistent with the hypothesis that LSF information contributes to top-down signals, several studies have shown that LSF information is processed faster than HSF. Therefore, LSF would have time to influence perception, before fine-grained information is processed (e.g. Bar et al., 2006; Bar, 2003; Breitmeyer, 2014; Kveraga, Boshyan, & Bar, 2007). On the other hand, the gist of a scene may not necessarily rely on low spatial frequencies (Oliva & Torralba, 2006). For example, Walther, Chai, Caddigan, Beck and Fei-Fei (2011) showed that global scene structure can be conveyed by HSFs. Previous studies in our lab have shown that feedback contains information about individual scenes as well as category (Morgan, Petro, & Muckli, 2016), suggesting some fine grained information may remain in feedback.

A useful approach to studying feedback signals is to isolate them from the feedforward input by looking at the effect of surround on regions of a scene which are not receiving feedforward stimulation. Several studies have shown that occluded regions of the visual field contain information about stimulation in the surround (Ban et al., 2013; Shushruth, 2011; Smith & Muckli, 2010; Sugita, 1999). Due to the retinotopic nature of V1 where neighbouring regions of the visual field map onto adjacent parts of the cortex (Wandell, Dumoulin, & Brewer, 2007) it is possible to investigate specific regions of a scene. Using functional MRI (fMRI) and multivoxel pattern analysis (MVPA, see Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006) in human subjects, Smith and

Muckli (2010) showed that occluded regions of the visual field, and thus non-stimulated in a feedforward manner, contain contextual information about the surrounding natural scene.

Using an occlusion paradigm (Smith & Muckli, 2010), fMRI and MVPA we investigated the spatial frequency content of feedback signals, by manipulating the spatial frequency of the scene surround outside the occluded region. We filtered two natural scenes into a high and low spatial frequency version, and additionally varied the SF overlaps between the HSF and LSF scenes. Thus we manipulated the amount of shared information between the two versions of each scene to see how broad feedback signals are in terms of the SF they contain. We found that 1) both high and low spatial frequency surrounds elicited meaningful feedback in the occluded region; 2) there was a similarity between feedback signals for LSF and HSF scenes, even with a lack of SF overlap; 3) but also a difference between HSF and LSF scene feedback; and 4) information in the occluded region is not a direct filling-in of the missing feedforward input. This suggests that feedback transmits both coarse and fine-grained information. However this information does not have a direct correspondence to the missing feedforward input.

2.3 Methods

2.3.1 Subjects

Thirty five subjects from the University of Glasgow participated in the experiment (n = 17 males; mean age: 24.63 years, range: 17 - 42 years). Subjects were paid for their participation. They provided informed written consent and the experiment was conducted in accordance with procedures and protocols approved by the local ethics committee at the University of Glasgow. We excluded two subjects: one due to large motion artefacts and another due to no above threshold activation in the visual cortex on one of the experimental runs. Therefore, we report results from 33 subjects (n = 16 males; mean age: 24.48 years, range: 17 - 42 years).

2.3.2 Stimuli

2.3.2.1 Feedback vs feedforward condition

To test feedback signals in the absence of feedforward stimulation, we used an occlusion paradigm previously employed by Smith and Muckli (2010). For the feedback conditions, the lower right quadrant was occluded by a white rectangle. The white rectangle was placed 0.5° of visual angle diagonally from the centre of the scene, and spanned 11.6° x 9.2°. For some subjects, these parameters were different (see Supplementary Methods, Appendix A). In the feedforward conditions, the corresponding quadrant of the scene was shown.

2.3.2.2 Scenes

We used two natural scene images for each participant. Only two scenes were used for each participant to keep the experiment length reasonable. Twenty three subjects saw a car and a people scene (**Figure 2.1A**), while ten subjects saw a New York and concert scenes (Appendix A). The car and people scenes were 600 x 480 px, which corresponded to 24° x 19.2° visual angle. For five subjects in the Small SF Overlap group and for five subjects in the Large SF Overlap group, New York and concert scenes were used (**Figures S2.1, S2.2**, Appendix A). We used natural scenes as these induce a lot of contextual associations (Bar, 2004). Each scene was filtered to create a high spatial frequency (HSF) and a low spatial frequency (LSF) version (**Figure 2.1A**). In creating these HSF and LSF scenes, we explored a variety of HSF and LSF cut-offs. Therefore, some participants viewed HSF and LSF scenes which shared, to various extents, a subset of SFs, whilst others viewed scenes not sharing any SFs (see **Table 2.1** for SF cut-offs and number of subjects presented with each combination, which we refer to as Overlap groups). We did this to investigate how specific the feedback signals are to the SF band of the surround. If a large amount of SF information needs to be shared between HSF and LSF version of the scene for HSF and LSF feedback to be similar, this would suggest that feedback is tightly tuned to the SF band of the surrounding scene.

Each group of subjects viewed HSF and LSF scenes with different Overlaps in the SF bands between the HSF and the LSF version. Each subject was assigned to one of four Overlap groups (Gap, No Overlap, Small or Large). The HSF and LSF stimuli in the Gap and the No Overlap conditions had no shared information in their SF bands, whereas the Small and the Large Overlap had some SFs which were common between the HSF and the LSF version. There were two Small and two Large Overlap groups with different cut-offs. The Small Overlap stimuli had an overlap of 1 octave, while the Large Overlap stimuli had an overlap of 1.5 octaves. The Gap stimuli had a gap of 1 octave. We chose our particular cut-offs as previous studies have indicated V1 preference for SFs of around 0.68-2 cycles per degree of visual angle (cpd, Haynes & Rees, 2005; Henriksson, Nurminen, Hyvärinen,

& Vanni, 2008; Tong, Harrison, Dewey, & Kamitani, 2012). Each subject only saw one HSF-LSF cut-off pair. **Figure 2.1A** shows stimuli for the 0.81/1.62 cpd Small SF Overlap group. For stimuli in the other conditions, see **Figures S2.1 and S2.2** (Appendix A).

TABLE 2.1 | Spatial frequency (SF) cut-offs used for each pair of stimuli and the SF overlap groups these corresponded to. Values are expressed in cycles per degree (cpd). For a graphical representation, see **Figure S2.3**, Appendix A.

	High pass filter cut-off	**Low pass filter cut-off**	**SF Overlap**
Pair 1 (n = 4)	1.62	0.81	Gap
Pair 2 (n = 6)	0.97	0.97	No Overlap
Pair 3 (n = 2)	0.65	1.30	Small
Pair 4 (n = 10)	0.81	1.62	Small
Pair 5 (n = 2)	0.81	2.03	Large
Pair 6 (n = 9)	0.97	2.43	Large

2.3.2.3 Occluded region mapping

Subjects were presented with three contrast-reversing checkerboards (5 Hz) twice per run. The checkerboards either covered the inner rectangular part of the occluded region (*Target* mapping – approximately 1.5° diagonally from fixation, 10.1° x 7.7° visual angle) or the border between the lower right quadrant and the rest of the stimulus (*Surround* mapping). There were two types of surround stimuli – *Large Surround* (approximately 0.5° diagonally from fixation, 11.6° x 9.2° visual angle) and *Small Surround* (approximately 1.5° diagonally from fixation, 11.6° x 9.2° visual angle) (**Figures 2.1B and 2.1C**).

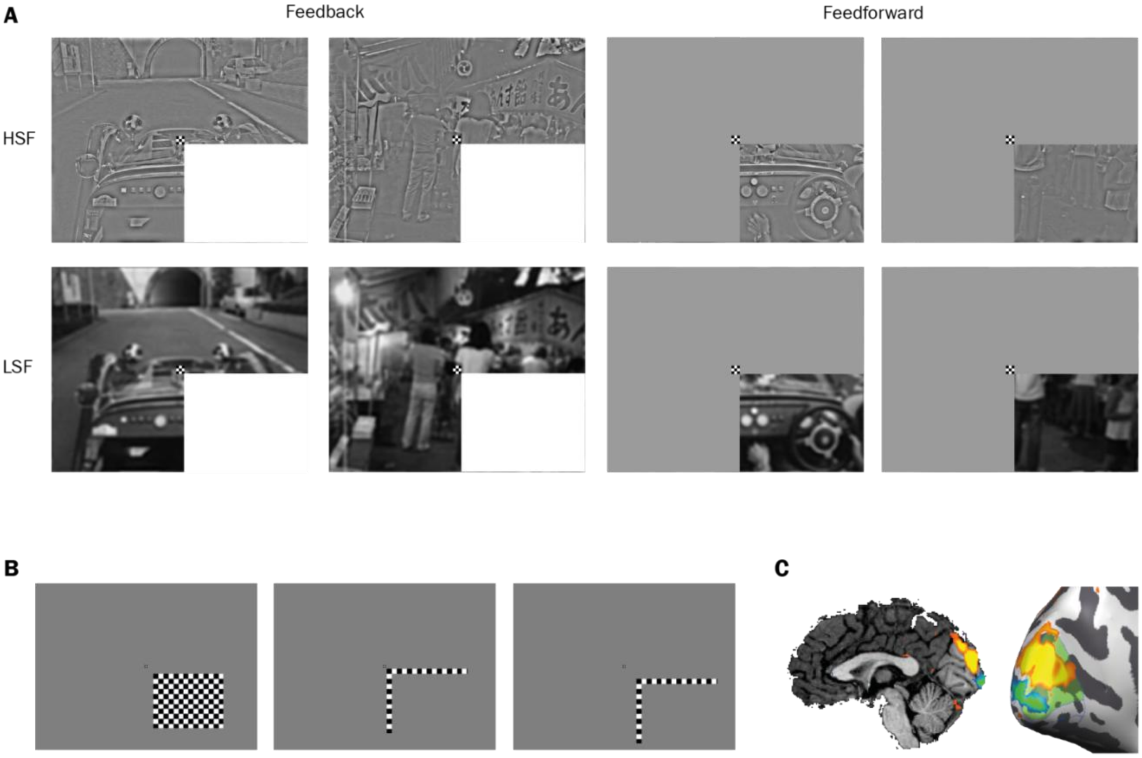


Figure 2.1 | Example stimuli, shown here for the Small Overlap group. In feedback conditions the lower right quadrant was occluded with a white rectangle, while in feedforward conditions the corresponding quadrant was visible. A) Each scene was filtered to produce a high spatial frequency (HSF) or a low spatial frequency (LSF) version. The example images shown here represent the Small Overlap condition, with the HSF scene filtered at 0.81 cpd and LSF scene filtered at 1.62 cpd. B) Checkerboards were used to retinotopically map the occluded region in V1: *Target* (left), *Large Surround* (middle) and *Small Surround* (right). C) The contrast used to map the occluded quadrant (*Target* – *Large Surround*) is shown in the occipital cortex, with V1 shaded in green on the inflated visualization.

2.3.3 Task & procedure

Scenes were presented on a uniform grey background using MRI compatible goggles (NordicNeuroLab) with 800 x 600 px screen resolution, which corresponded to 32° x 24° visual angle. For each subject there were 8 types of trial (2 scenes, high or low SF, occluded [FB] or non-occluded [FF]). In each 12 s trial the stimulus was flashed on and off (200 ms on/ 200 ms off) 30 times. This maximised the signal to noise ratio (Kay, Naselaris, Prenger, & Gallant, 2008). Each trial type was presented sequentially, with the trial order randomized in each sequence. Each sequence lasted 96 s (8 x 12 s). A 12 s fixation period was included before and after each sequence of trials. Each experimental run lasted 10 min 48 s, consisting of four trial sequences and two mapping sequences (2 x *Target* and two *Surrounds*). There were four experimental runs in total. The subjects' task was to fixate on a central checkerboard and report a fixation colour change with a button press. Subjects pressed a different button, depending on whether the colour change occurred during scene 1 or scene 2. The purposes of the task were to ensure that the subject attended the stimuli and to minimize eye movements. This was important for accurate retinotopic mapping of the occluded region in the lower right quadrant of the visual field.

After the experimental runs, we performed a retinotopic mapping procedure to estimate the borders of the early visual areas V1-3. The mapping procedure was a standard polar-angle protocol consisting of either wedge shaped checkerboards arranged in a “bow-tie” or a single wedge which started in the right horizontal meridian and rotated clockwise (12 rotations per scan, wedge angle: 22.5°, scan time: 13 min 28 s). Subjects were also familiarised with the unfiltered non-occluded images with a short practice run prior to entering the scanner.

2.3.4 MRI acquisition

MRI data were collected using a 3T Siemens Tim Trio System with a 12-channel head coil. Blood oxygen level dependent (BOLD) signals were measured with an echo-planar imaging sequence (echo time: 30 ms, repetition time: 1000 ms, field of view 210 mm, flip angle: 62°, 18 axial slices). The spatial resolution for functional data was 3 x 3 x 3 mm. Each experimental run had 648 volumes. Retinotopic mapping consisted of 424 volumes (bow-tie) or 808 (single wedge). The 18 slices were positioned to maximize coverage of occipital cortex. A high resolution 3D anatomical scan (3D Magnetization Prepared Rapid Gradient Echo, 1 x 1 x 1 mm resolution) was also recorded (192 volumes).

2.3.5 MRI data processing

Functional data for each experimental run and retinotopic mapping were corrected for slice time (cubic spline interpolation) and 3D motion (Trilinear/Sinc interpolation), temporally filtered (high-pass filtered at 6 cycles with GLM-Fourier, and linearly detrended), and spatially normalized into Talairach space with Brain Voyager QX 2.8 (Brain Innovation). Subsequently, the anatomical data were used to create an inflated cortical surface and functional data were overlaid.

2.3.6 Voxel selection and analysis

Excessive subject movement between runs is likely to affect correspondence between voxels from one run to another. This in turn could affect our analysis, as we selected our region of interest (ROI) based on the averaged functional data of all 4 runs. To determine whether there was good alignment between functional data covering the visual areas, we calculated an alignment value for each subject by measuring Pearson's correlation in a ROI in the visual cortex between the four functional runs. The median alignment value across the subjects was 98% and single subject values ranged from 92% to 99%.

The occluded region was mapped using a general linear model (GLM) contrast of the *Target* region against the *Large Surround*, as described previously in Smith and Muckli (2010). The ROI was selected from activation in V1 only. To further minimize spillover activity from neighbouring stimulated areas, voxels from the ROI were then selected for analysis on the basis of the difference between *Target* and *Large Surround* t-value being greater than 1.

2.3.7 Controls for MVPA analysis

2.3.7.1 Analyses with extended safety boundary around the occluded region

To further make sure our findings of scene information in the quadrant were not due to spillover activity from the feedforward surround, we ran a separate analysis with a more stringent method of voxel selection. We selected our region of interest in BrainVoyager as the contrast of *Target* mapping being higher than both the *Large* and the *Small Surround*. In addition, we selected voxels with t-values fitting the criteria of (*Target – Large Surround*) > 1 & (*Target – Small Surround*) > 1. This further minimized any voxels which may have been responding to feedforward stimulation at the border of the occluded region.

2.3.7.2 Analyses with subjects with above chance classification in at least one feedforward condition

In the feedforward conditions the quadrant receives visual stimulation. Hence there should be a difference in activity patterns between the images for these conditions. If no above-chance classification was found for either of the feedforward conditions, it may suggest the subject was not paying attention or was making too many eye movements. Thus we also show additional results for only subjects with at least one feedforward condition classifying above chance. Significance above chance was measured using permutation analysis with 1000 trials. This excluded three subjects in the Small Overlap group and two in the Large Overlap group.

2.3.8 Multivoxel Pattern Classification Analysis

The voxels matching all these criteria were entered into the linear classifier (Support Vector Machine [SVM]). For classification analyses, we trained the classifier to decode the two scenes in each condition. For cross-classification analyses we trained the classifier on one experimental condition and tested on the other. The classifier used single-trial activity patterns (beta values) for training, and was then tested on either single trial (8 trials x 4 sequences) or average activity patterns for each of the 8 trial types. The classifier was trained on 3 of the runs and tested on the remaining run (i.e. one-run-out cross-validation).

We bootstrapped (1000 samples) the classifier performances for individual subjects, in order to estimate the single subject mean. We then bootstrapped (1000 samples) these values to estimate the group mean and associated variance. The confidence intervals (CIs) were defined with an alpha level of 0.05. Classifier performances were deemed to be significantly above chance (50%) if the 95% confidence intervals did not overlap it. We were not able to conduct an ANOVA to check for group differences as the data were not parametric. Differences between group classifier performances were thus assessed via a permutation test (1000 samples) of the differences between the group means (*p* values not corrected for multiple comparisons).

2.4 Results

2.4.1 Both HSF and LSF scene surrounds induce meaningful feedback signals

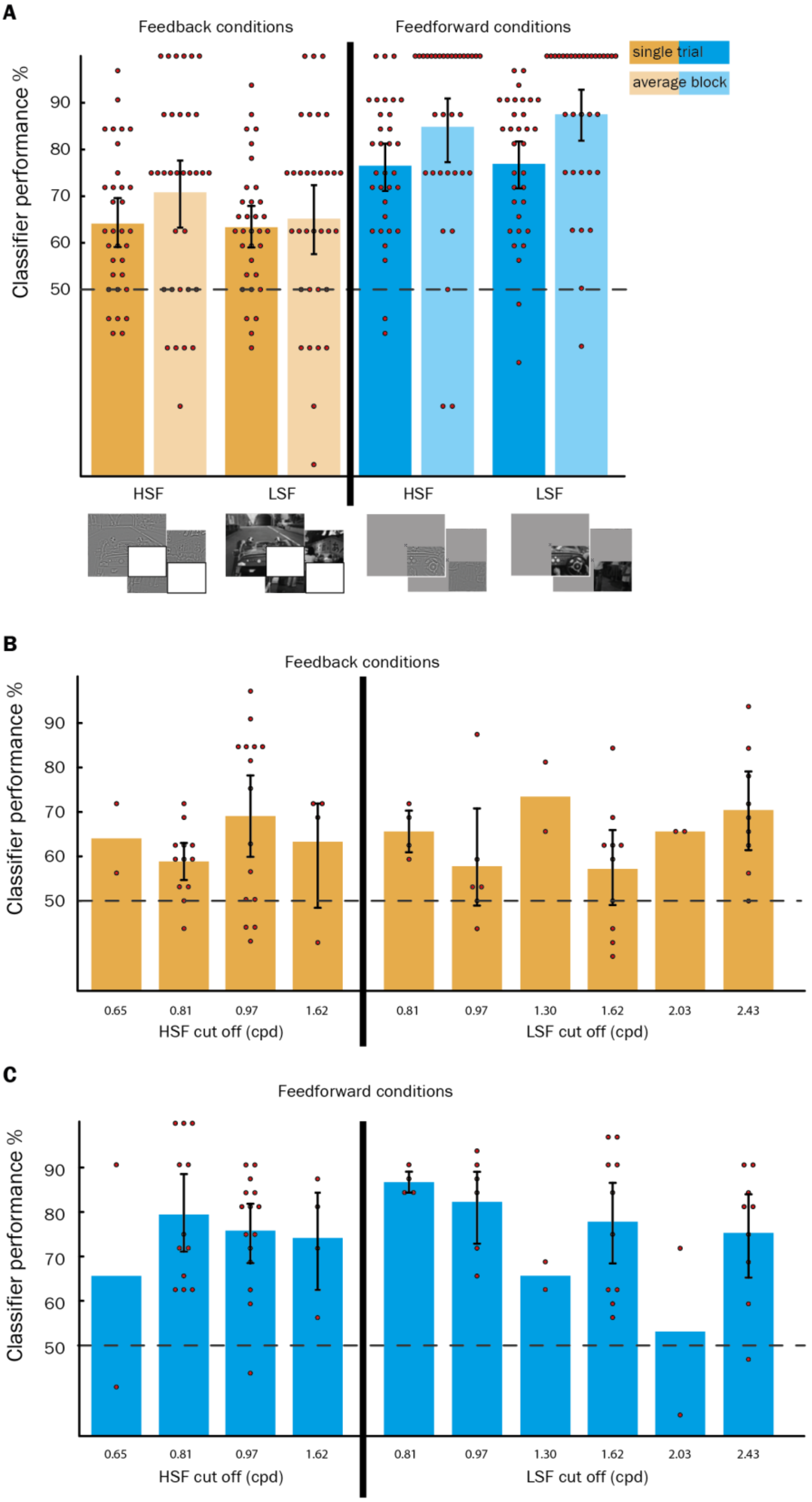
First, we tested whether the spatial frequency filtered surround induces meaningful information in the occluded quadrant. We trained the Support Vector Machine (SVM) classifier to decode between the two scenes using voxel patterns responding to the lower right quadrant. SVM classification performance was used to estimate whether the quadrant contained informative feedback signals about the scene.

Collapsing across the different SF cut-off groups, classifier performance for decoding between the two scenes was above chance for both HSF and LSF scenes, in both feedback and feedforward conditions (**Figure 2.2A**). Decoding during feedforward (FF) conditions was significantly higher than during feedback (FB) conditions for high spatial frequency

(*FF HSF*, single trial classification [ST]: 76.52%, confidence interval [CI] [0.0540 0.0473] vs *FB HSF*, ST: 64.11%, CI [0.0502 0.0549], p = 0.001; *FF HSF*, average block classification [AB]: 84.85%, CI [0.0758 0.0606] vs *FB HSF*, AB: 70.83%, CI [0.0758 0.0682], p = 0.008) and low spatial frequency (*FF LSF*, ST: 76.89%, CI [0.0521 0.0483] vs *FB LSF*, ST: 63.35%, CI [0.0436 0.0455], p < 0.001; *FF LSF*, AB: 87.50%, CI [0.0568 0.0530] vs *FB LSF*, AB: 65.15%, CI [0.0758 0.0720], p < 0.001) conditions. There was no difference in classifier performance between HSF and LSF conditions, for either the feedback or feedforward stimuli.

Figure 2.2B shows the single trial classification performance for decoding between the two scenes in each of the different cut-offs for HSF and LSF stimuli in the feedback conditions. **Figure 2.2C** shows the same for the feedforward conditions.

[Next page] Figure 2.2 | Classification performance for decoding the two scenes in HSF and LSF conditions, for feedback and feedforward stimuli. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. We did not plot confidence intervals for groups with only 2 subjects. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. A) Classifier performance for HSF and LSF conditions, collapsed over different SF cut-offs. N = 33. Images are example stimuli used for a subset of subjects. B) Classifier performance split by different SF cut-offs for feedback conditions. C) Classifier performance split by different SF cut-offs for feedforward conditions.


A
Feedback conditions
Feedforward conditions
single trial
average block
Classifier performance %
90
80
70
60
50
HSF
LSF
HSF
LSF
B
Feedback conditions
Classifier performance %
90
80
70
60
50
0.65
0.81
0.97
1.62
HSF cut off (cpd)
0.81
0.97
1.30
1.62
2.03
2.43
LSF cut off (cpd)
C
Feedforward conditions
Classifier performance %
90
80
70
60
50
0.65
0.81
0.97
1.62
HSF cut off (cpd)
0.81
0.97
1.30
1.62
2.03
2.43
LSF cut off (cpd)

2.4.2 Similarity of feedback across HSF and LSF stimuli

Secondly, we tested whether the classifier can generalise over spatial frequencies, in other words, decode between the scenes even when it was trained on HSF and tested on LSF (and vice versa). We trained the classifier on either the HSF version of the two scenes and tested on LSF, or trained on LSF and tested on HSF. We performed this analysis for the different Overlap groups, to see how the amount of shared spatial frequency information between the HSF and LSF version of the scene would affect this generalisation. We predicted that if feedback is specific to the SF range of the scene surround then the larger the overlap the better the classifier would perform since there would be more shared information between HSF and LSF. Alternatively, if feedback is similar across different SF surrounds, then we predicted we would see similar levels of generalisation across all Overlap groups.

2.4.2.1 Training on HSF and testing on LSF

In feedforward conditions, classifier performance was above chance for all Overlaps groups, for both single trial and average block analysis, and was not significantly different between groups (**Figure 2.3A**, right). For feedback conditions, classifier performance was above chance for Gap (ST only: 58.59%, CI [0.0469, 0.0703]), Small Overlap (ST only: 55.99%, CI [0.0547 0.0573]) and Large Overlap (ST: 64.20%, CI [0.0909 0.0938]; AB: 70.45%, CI [0.1250 0.1364]) groups (**Figure 2.3A**, left). Cross-classifier performance for Large Overlap was higher than for Small Overlap for average block classification, $p = 0.048$.

2.4.2.2 Training on LSF and testing on HSF

For feedforward conditions, classifier performance was above chance for Gap, Small Overlap and Large Overlap, for both single trial and average block analysis, and was not significantly different between groups (**Figure 2.3B**, right). For feedback conditions (**Figure 2.3B**, left), classifier performance was above chance for Gap (ST only: 64.84%, CI [0.0703 0.0703]), No overlap (ST: 67.19%, CI [0.0833 0.0729]; AB: 66.67%, CI [0.1250 0.1042]), Small Overlap (ST: 57.03%, CI [0.0599 0.0625]; AB: 61.46%, CI [0.0938 0.1042]) and Large Overlap (ST: 63.64%, CI [0.0511 0.0511]; AB: 67.05%, CI [0.0795 0.0909]). There was no significant difference between the groups.

The results show that the classifier can generalise over spatial frequencies. We can train on HSF and successfully decode between the LSF scenes, and vice versa. We see that a lot of shared information in terms of spatial frequency is not necessary for cross-classification, as we could cross-classify in the Gap, No Overlap and Small conditions. Large Overlap is the only condition that was above chance for both directions of cross-classification and for both single trial and average block classification, which may suggest that classifier generalises better across SF in the presence of more shared information. However, the fact that cross-classification works for the other conditions, in single trial classification at least, suggests generalisation can still occur without a large amount of shared information.

2.4.3 Lack of similarity between feedback and corresponding feedforward information

Are feedback signals similar to the corresponding feedforward information? For example, are feedback signals in the HSF condition similar to the activity pattern relating to the HSF feedforward stimulation, or do they carry different information? To test feedback and feedforward similarity, we first trained the classifier to decode between the two scenes on feedback conditions and tested on feedforward conditions (and vice versa), for both HSF and LSF (**Figure 2.4A**). Classifier performance was at chance level for both HSF and LSF scenes, and for both directions of cross-classification. This suggests that the information feedback provides in the occluded region is different to the corresponding feedforward information. Secondly, we trained the classifier to decode HSF vs LSF on feedback conditions and tested its ability to decode in the feedforward conditions (and vice versa, **Figure 2.4B**). Classifier was above chance only for Scene 2 when training on feedback and testing on feedforward (ST: 54.92%, CI [0.0275 0.0256]; AB: 56.82%, CI [0.0379 0.0379]).

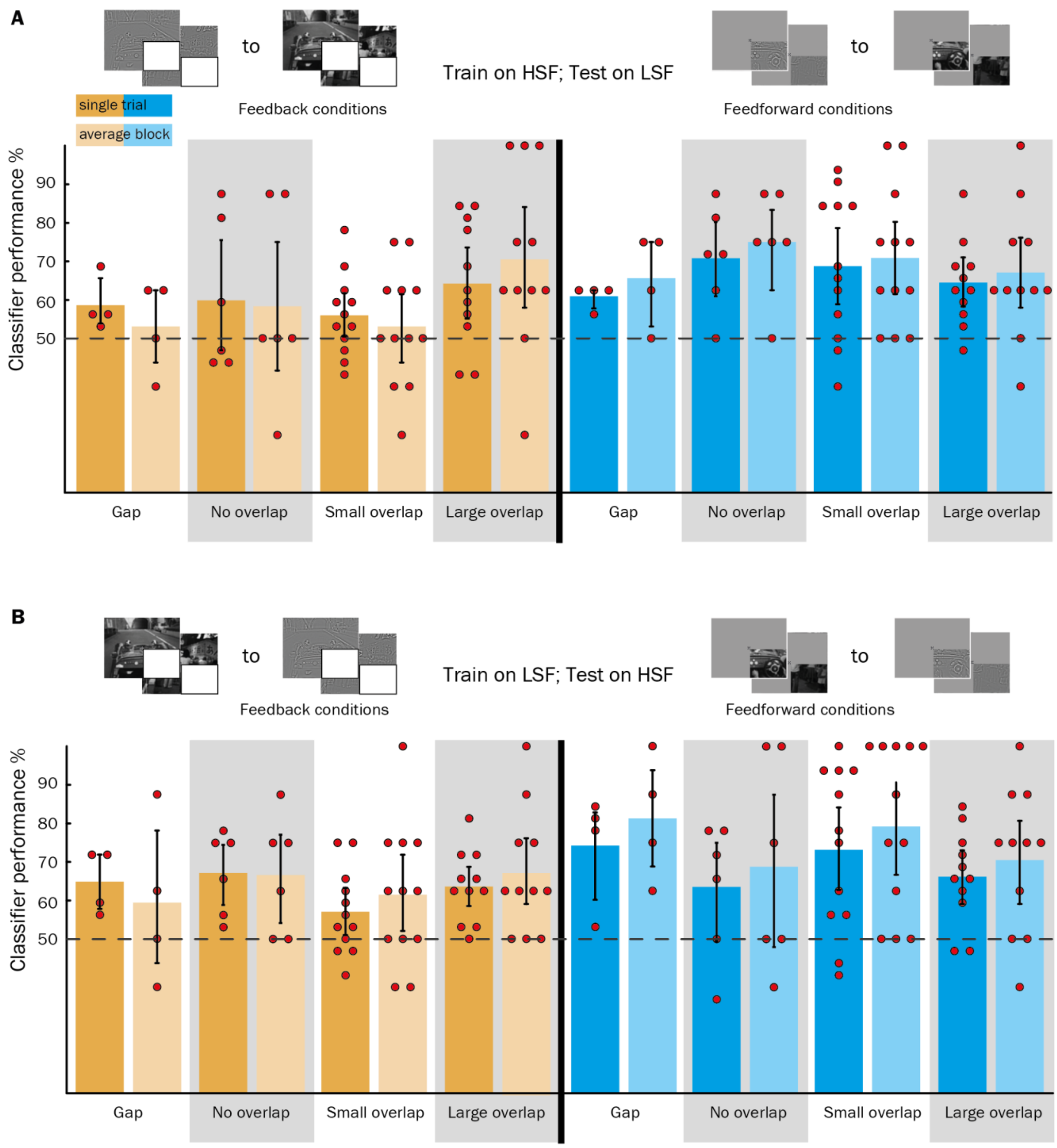


Figure 2.3 | Cross-classification performance for training to decode the two scenes in one SF and testing in the other, for different Overlap groups. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. Gap: n = 4; No Overlap, n = 6; Small Overlap, n = 12; Large Overlap, n = 11. Images are example stimuli used for a subset of subjects. A) Classifier performance for training on HSF and testing on LSF. B) Classifier performance for training on LSF and testing on HSF.

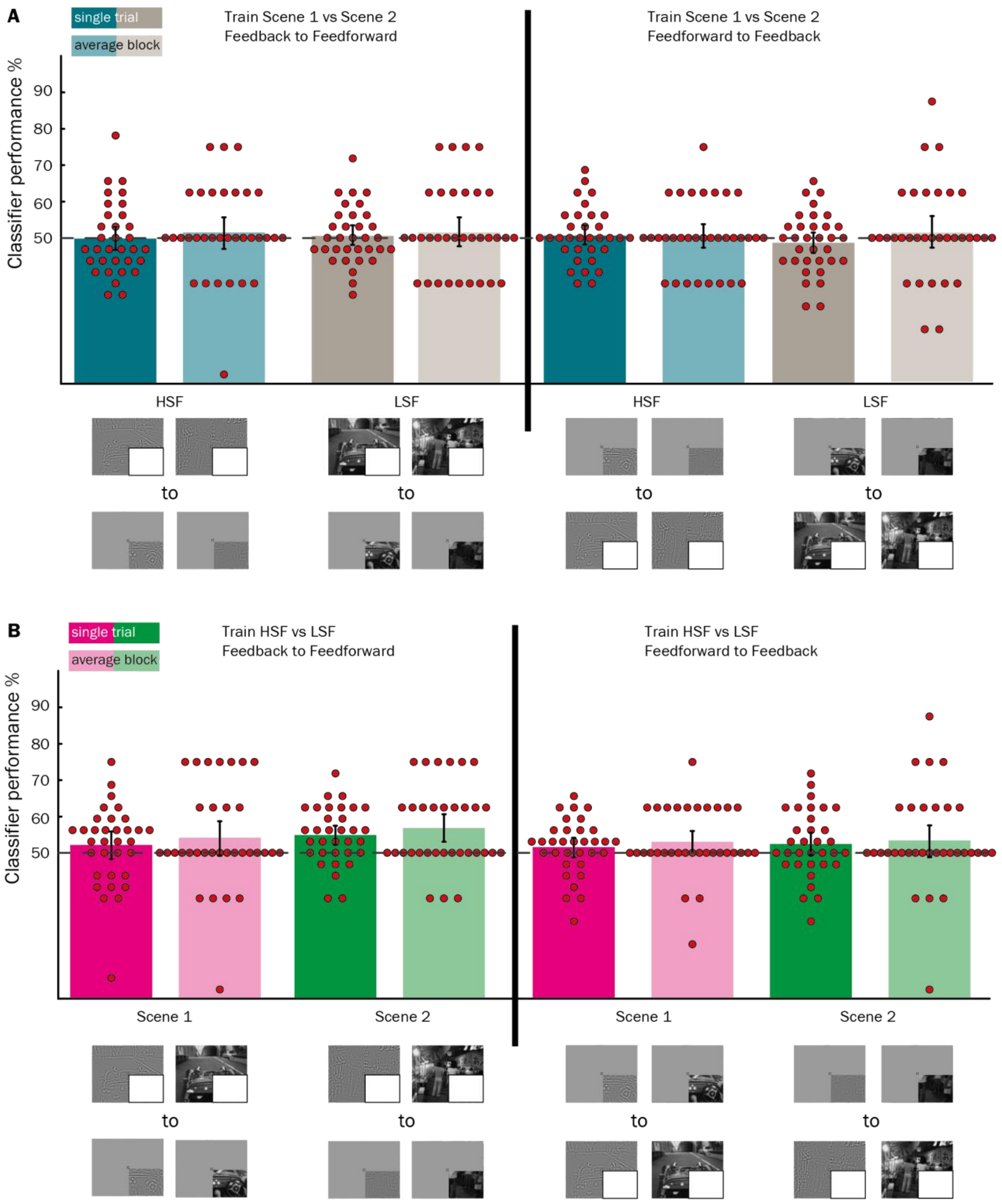


Figure 2.4 | Cross-classification performance for training on feedback and testing on feedforward conditions (and vice versa). Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. N = 33. Images are example stimuli used for a subset of subjects. A) Classifier performance for decoding Scene 1 vs Scene 2 in HSF and LSF conditions. B) Classifier performance for decoding HSF vs LSF in Scene 1 and Scene 2.

2.4.4 Does feedback carry spatial frequency information not related to a specific scene?

We wanted to see whether spatial frequency related feedback in one scene is similar to that in another scene. In other words, does feedback carry any information about the spatial frequency of the surround that is not specific to a particular scene? We trained the classifier to decode between HSF vs LSF on Scene 1 and tested whether it could decode between HSF vs LSF on Scene 2 (and vice versa). Classifier performance was above chance for both feedback and feedforward conditions, and for both directions of cross-classification (**Figure 2.5**). Feedback appears to carry information about high or low spatial frequency that is similar across different scenes. In other words, there is some degree of similarity between the information in, for instance, the occluded region of the HSF version of Scene 1 and the occluded region of the HSF version of Scene 2.

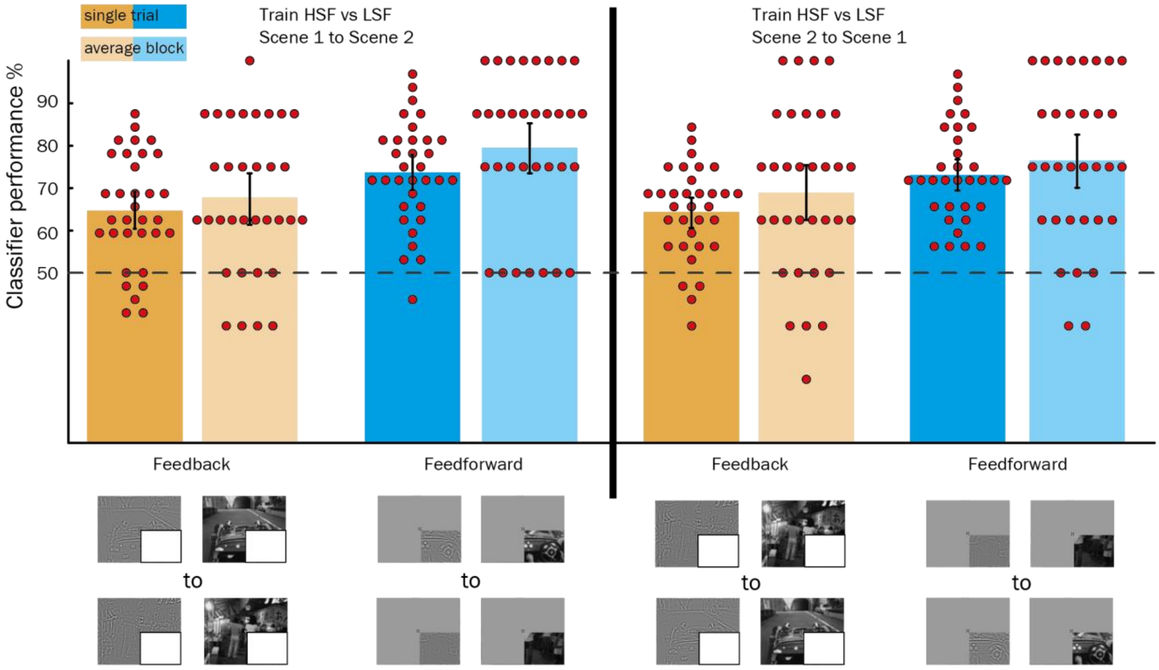


Figure 2.5 | Cross-classification performance for training to decode HSF vs LSF in Scene 1 and testing in Scene 2 (and vice versa), for feedback and feedforward conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. N = 33. Images are example stimuli used for a subset of subjects.

2.4.5 Analyses with more stringent criteria

Analyses using a more conservative method of voxel selection in the occluded region, using both the *Large* and the *Small Surround* mapping (see Methods), led to a similar pattern of results (**Figure S2.5**, Appendix A), suggesting our results are unlikely to be due to “spillover” activation from the surround feedforward stimulation. Analysis using only subjects who had at least one feedforward condition above chance, also showed a similar pattern of results, suggesting that the results observed are unlikely to be due to excessive eye movements and subjects not paying attention during the experiment (**Figure S2.6, S2.7**, Appendix A).

2.5 Discussion

The present study aimed to investigate the spatial frequency information carried by feedback signals to an occluded region of the visual field in V1. First, we have replicated the findings of Smith and Muckli (2010), by showing that non-stimulated V1 receives contextual feedback from the surrounding regions of the scene, and this can occur even with reduced information, in other words, when the surrounding scene only contains information in certain spatial frequencies. Second, we show that feedback carries information about both low and high spatial frequencies, suggesting that it transmits both coarse and fine-grained information, respectively.

2.5.1 Feedback contains both coarse and fine-grained information

We found scene-specific information patterns in the occluded region, with both LSF and HSF surrounds. This suggests that both LSF and HSF surrounding information gave rise to meaningful feedback signals. In addition to theories suggesting LSFs are important for providing contextual information and contribute to top-down expectations (e.g. Bar et al., 2006; Bar, 2003; Breitmeyer, 2014; Kveraga et al., 2007), we show that HSF scene information is also sufficient to trigger contextual feedback. Our results are more in line with the flexible usage hypothesis (Schyns & Oliva, 1997), which proposes that both HSF and LSF information can be processed first and demands of the task can bias the visual system to attend to the most informative scale. Both HSF and LSF information is available at the onset of visual processing (Schyns & Oliva, 1994) and when presented with a hybrid stimulus (HSF and LSF superimposed in one image) people attend to the most relevant SF channel for the task at hand (e.g. Oliva & Schyns, 1997; Schyns & Oliva, 1999; Sowden, Özgen, Schyns, & Daoutis, 2003). Furthermore, Oliva and Torralba (2006) argue that scene gist might be provided by global scene structure, but which might not necessarily use LSF. In addition, Walther et al. (2011) showed that global scene structure is preserved in line drawings and underlies scene category representation. It is plausible that LSF is an important source of contextual information in natural viewing when all spatial scales are available. However, in our filtered scenes, informative context was presented only in one type of spatial scale. Thus, in the HSF stimulus, high spatial frequencies were the only informative spatial scale and therefore this spatial scale was used for top-down context. Walther et al. found a similarity between brain activity in relation to line drawings (HSF) and coloured photographs of the same image (full SF spectrum), suggesting that impoverished HSF information is sufficient for scene identification. Rajimehr, Devaney, Bilenko, Young and Tootell (2011) showed that the parahippocampal place area (PPA), which is known to process scenes and spatial context (Bar & Aminoff, 2003), responds preferentially to HSF information. Since we used scenes in the present experiment, it is plausible that feedback may originate from the PPA (as well as other cortical regions) and therefore transmit contextual information about HSFs back to V1.

While we found that HSF surrounds can elicit feedback, suggesting some fine-grained information may be transmitted, we also tested the breadth of SF tuning for feedback signals. Walther et al. (2011) found that they could cross-classify from line drawings to coloured photographs, and vice versa. However, the coloured photographs had a full SF spectrum and therefore there was shared information between the two types of image. To address how broad feedforward and feedback signals are in terms of spatial frequency tuning, we tested cross-classification from one SF type to another when only some or none of the SF information was shared between the two versions. We found we could cross-classify from one SF to the other, in both feedforward and feedback conditions even when there was no overlap in terms of the spatial frequency spectrum between the HSF and LSF version (for example, in the Gap condition), suggesting brain activity patterns were similar for the HSF and LSF version, commensurate with a broad tuning. Both HSF and LSF surrounds elicited meaningful feedback and since it was similar for the two versions of the scene, one hypothesis could be that the surround elicits feedback containing the same general template (in some SF band) for the particular image, regardless of the SF it is presented in. However, we would argue that feedback for the HSF and LSF scene is not identical as we could also decode between HSF and LSF in the occluded region and generalise this decoding to another scene. This suggests that the tuning of feedback does depend at least somewhat on the surround of the scene – presumably HSF for HSF surround and LSF for LSF surround.

We have demonstrated that on the one hand feedback signals share a similarity between the HSF and LSF scenes even when there is little shared information in terms of the spatial frequency of the scene surround. On the other hand, we found that the HSF and LSF surrounds elicited informative feedback signals equally well, and we could also decode between HSF and LSF versions of the same scene. This suggests that feedback may contain coarse information about the scene, but nevertheless retain some fine-grained information. This is in line with previous work in our lab (Morgan et al., 2016) which showed that feedback contains information about both the category of a scene (such as a forest, corresponding to coarse structure) as well as individual scenes within a category (fine-grained structure of a particular example of a forest scene).

One possible limitation is that we also saw SF generalisation for the feedforward conditions. Even when there was no shared SF information in the HSF and LSF version, we could cross-classify from HSF to LSF, and vice versa. This suggests that our SF cut-offs were not wide enough to lead to completely separate brain activity patterns in the feedforward regions. Since feedback signals are likely to be even coarser than feedforward signals (Muckli et al., 2015), it is unsurprising we could generalise across SF in feedback conditions as well. With a wider Gap condition, we might be able to better probe how broad feedback signals are in comparison to feedforward. Alternatively, this generalisation could be possible because there is a relationship between where the object boundaries are in the different SF bands. A blurry or a sharp edge is still the same edge, and hence the brain activity pattern is similar, if the scene representation in the visual system is related to figure-ground segregation and object identification.

2.5.2 Feedback signals do not correspond to a direct filling-in of the missing feedforward information

We saw that feedback is meaningful for both HSF and LSF scenes. However, how do the activity patterns in the occluded region compare to those in the corresponding feedforward region of the scene? Are feedback signals in the HSF and LSF scenes similar to the corresponding feedforward signals? To answer this question we trained the classifier on feedback conditions and tested whether it can use the same information to decode the stimuli in the feedforward conditions. We did not find a similarity between feedback and feedforward signals, suggesting that feedback signals do not represent a direct filling-in of the feedforward information. This is in contrast to the findings of Smith and Muckli (2010). However, they used a full scene as the feedforward condition, whereas in the present study we used a feedforward quadrant with no surround. In **Chapter 3**, we demonstrate that this feedback and feedforward similarity depends on the amount of surrounding scene information. This is because removing the scene surround outside the feedforward quadrant removes the contextual surrounding feedback, which drives this similarity effect. This finding might be surprising if we consider that feedback has been implicated in transmitting predictions and expectations about the scene (e.g. Clark, 2013; Friston, 2010; Rao & Ballard, 1999; Bastos et al., 2012; Kok, Jehee, & de Lange, 2012), and we might therefore expect feedback to represent the missing feedforward information. It is possible that the missing scene information is still represented, but in a different format. For example, it may be that the information is coarser in terms of its content because of the larger visual fields in higher visual areas or less precise retinotopically (e.g. de-Wit, Kubilius, Wagemans, & Op de Beeck, 2012) or because feedback and feedforward signals project to different cortical layers (Muckli et al., 2015; Rockland & Pandya, 1979). Muckli et al. showed using high resolution fMRI that during normal visual stimulation, feedforward information peaks in the mid-layers, while contextual feedback information peaks in the superficial layers.

2.5.3 Low level properties

One question arises about the level of information that feedback transmits. Does it code for high level information, for example the scene category, or does it have some low level information, such information about the spatial frequency band that is not specific to a particular scene category? We were able to decode HSF vs LSF on Scene 1 and generalise this to decoding HSF vs LSF on Scene 2. This similarity of the SF information between the different scenes suggests that feedback carries general information about the spatial frequency band of the surround that is unrelated to the specific structure of the scene in question. However, this finding may be explained by differences in contrast since the HSF scenes were mostly grey whereas the LSF scenes would be better described as patches of white and black. It would be useful to further test this finding with SF filtered scenes which are better matched for contrast.

2.5.4 Lateral connections

We can be sure that there is no meaningful feedforward stimulation in the occluded region. However, the information that we do record there could originate from both lateral and feedback connections. Although lateral connections might have a role, we would argue that they cannot fully explain the observed effect, as they only span a relatively small region of retinotopic space, and are not sufficient to account for the full range of surround modulatory response (Angelucci et al., 2002). Lateral connections are 2.3 times larger than the classical receptive field (Sceniak, Ringach, Hawken, & Shapley, 1999) and can project to around 2° from the classical RF in V1 at eccentricities of 2-6° (Stettler, Das, Bennett, & Gilbert, 2002).

2.5.5 Conclusion

Using pattern analysis techniques we probed the information content of feedback signals, and show that they contain both high and low spatial frequency information about the surrounding scene. Further to behavioural studies demonstrating that both HSF and LSF information can be used for scene processing depending on which spatial scale is informative, we show on a neural level that both HSF and LSF scene surrounds can contribute to top-down contextual feedback. In addition, we demonstrate that although we can decode between HSF and LSF feedback, there are also similarities in feedback for the two versions of the scene, suggesting its tuning is quite broad. Finally, we find that the information in the occluded region is not a direct filling-in of the missing feedforward input.

3 Increased scene information in the surround enhances contextual feedback

3.1 Abstract

The majority of synaptic inputs to the primary visual cortex (V1) are non-feedforward, instead originating from lateral and feedback connections (Muckli & Petro, 2013). Animal electrophysiology experiments show that feedback signals originating from higher visual areas with larger receptive fields modulate the surround receptive field of V1 neurons (Angelucci & Bressloff, 2006). Surround modulation of feedforward processing has never been parametrically investigated in human V1, despite theories of visual processing specifying that both feedback and feedforward signals are required for healthy visual cognition (Larkum, 2013). In the present study we investigated how much contextual surrounding information is necessary for scene-specific feedback in non-stimulated regions, using an occlusion paradigm with natural scenes (Smith & Muckli, 2010), functional magnetic resonance imaging (fMRI) and pattern analysis methods. We gradually revealed the surrounding scene using Gaussian bubbles (Gosselin & Schyns, 2001) of varying sizes around an occluded or visible lower right quadrant. We show that 1) increasing the visibility of the surround, and thus increasing information in surround receptive fields, increases meaningful feedback in the occluded region; 2) information in the non-stimulated region does not represent a direct filling-in of the missing feedforward input; and 3) full visual processing is a mixture of feedback and feedforward signals.

3.2 Introduction

The majority of synaptic inputs to primary visual cortex (V1) are non-feedforward, instead originating from lateral and feedback connections (Muckli & Petro, 2013). Animal electrophysiology experiments show that feedback signals originating from higher visual areas with larger receptive fields modulate the surround receptive field of V1 neurons, also known as the non-classical receptive field (Angelucci & Bressloff, 2006). Surround modulation of feedforward processing has never been parametrically investigated in human

V1, despite theories of visual processing specifying that both feedback and feedforward signals are required for healthy visual cognition (Larkum, 2013). It is unclear how much stimulation in the surround is necessary for informative feedback signals, and how much information feedback contributes to visual processing during feedforward stimulation.

A useful approach to studying feedback signals is to isolate them from the feedforward input by looking at the effect of surround on regions of a scene which are not receiving feedforward stimulation. Due to the retinotopic nature of V1 where neighbouring regions of the visual field map onto adjacent parts of the cortex (Wandell, Dumoulin, & Brewer, 2007) it is possible to investigate specific regions of a scene. Several studies have shown that occluded regions of the visual field contain information about stimulation in the surround (Ban et al., 2013; Shushruth, 2011; Smith & Muckli, 2010; Sugita, 1999). Smith and Muckli (2010) showed using functional magnetic resonance imaging (fMRI) and multivoxel pattern analysis (MVPA, see Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006) in human subjects that occluded regions of the visual field, and thus non-stimulated in a feedforward manner, contain contextual information about the surrounding natural scene.

Smith and Muckli (2010) found that with three quarters of the scene visible it was possible to record contextual information in the occluded region, while Tang et al. (2014) showed that with only 9-25% of an image revealed, neural responses in the visual ventral steam retained object selectivity, suggesting that scene completion is possible from very limited information. In the present study we investigated how much contextual surrounding information is necessary for scene-specific feedback in non-stimulated regions of V1, using an occlusion paradigm with natural scenes (Smith & Muckli, 2010), fMRI and MVPA. We gradually revealed the surrounding scene using Gaussian bubbles (Gosselin & Schyns, 2001) of varying sizes around an occluded or visible lower right quadrant. We show that 1) increasing the visibility of the surround, and thus increasing information in the surround receptive fields, increases meaningful feedback in the occluded region; 2) information in the non-stimulated region does not represent a direct filling-in of the missing feedforward input; and 3) feedback from the surround is essential for full visual processing.

3.3 Methods

3.3.1 Subjects

Twenty nine subjects from the University of Glasgow participated in the experiment (n = 13 males; mean age: 24.28 years, range: 19-41 years). Subjects were paid for their participation. They provided informed written consent and the experiment was conducted in accordance with procedures and protocols approved by the local ethics committee at the University of Glasgow. We excluded subjects for chance level classification performance for at least one feedforward condition (actual image visible in the quadrant, control condition). Additionally we excluded subjects with poor alignment between the four functional runs (see *Voxel Selection and Analysis*). Thus, eight subjects were excluded from further analysis and in the following we report results from 21 subjects (n = 10 males; mean age: 25.29 years, range 19-41 years).

3.3.2 Stimuli

3.3.2.1 Feedback vs feedforward condition

To test feedback signals in the absence of feedforward stimulation, we used an occlusion paradigm previously employed by Smith and Muckli (2010). For the feedback conditions, the lower right quadrant was occluded by a white rectangle. The white rectangle was placed 0.5° of visual angle diagonally from the centre of the scene, and spanned 11.6° x 9.2°. In the feedforward conditions, the corresponding quadrant of the scene was shown. Thus, in these conditions, we looked at the full visual processing, i.e. both feedforward signals and feedback from the surround.

3.3.2.2 Scenes

We used two natural scene images for each participant. Natural scenes were used as these induce a lot of contextual associations (Bar, 2004). Each scene was 600 x 480 pixels and spanned 24° x 19.2° of visual angle. We did not normalize the images in terms of low-level visual features, such as luminance, contrast or energy at each spatial frequency because we wanted the scenes to look as natural as possible. Smith and Muckli (2010, Experiment 4) previously showed that feedback signals cannot be solely attributed to these low-level visual features.

To investigate the contribution of surrounding contextual information on the brain activity patterns corresponding to the lower right quadrant, we manipulated the visibility of the surrounding ¾ of the scene with Gaussian apertures (referred to as “bubbles” from this point onwards, Gosselin & Schyns, 2001) of various sizes to produce the following types of stimuli: ¼ (no surrounding scene shown), Small Bubbles (standard deviation [SD] = 50 x 32 pixels), Medium Bubbles (SD = 90 x 56 pixels), Large Bubbles (SD = 125 x 100 pixels) and Full (surround fully visible). The study consisted of four experiments, with each subject participating in only one of them (see **Figure 3.1A** for number of subjects in each experiment). In each experiment, stimuli were shown in four different conditions (**Figure 3.1A**). We kept the number of conditions to four in each experiment partly to test different combinations of stimuli conditions and partly to keep the experiment length reasonable. In Experiment 1, we used stimuli in the Full Feedback occluded condition, ¼ feedforward, Small Bubbles feedforward and Medium Bubbles feedforward conditions. In Experiment 2, we replaced Small and Medium Bubbles with Large Bubbles and the Fully Visible scene. In Experiment 3, we added the Fully Visible scene to test whether more contextual feedback would be seen in the Small and Medium Bubbles conditions if participants were more familiar with the full scene. In Experiment 4, we tested the effect of reducing the surrounding information around the occluded region using Small, Medium and Large Bubbles feedback conditions.

3.3.2.3 Occluded region mapping

Subjects were presented with three contrast-reversing checkerboards (5 Hz) twice per run. The checkerboards either covered an inner rectangular part of the occluded region (*Target* – 2.5° diagonally from centre, 10.2° x 7.8° visual angle) or the border between the lower right quadrant and the rest of the stimulus (*Surround*). There were two types of surround checkerboard stimuli (**Figure 3.1B**) – *Large Surround* (0.5° diagonally from fixation, 11.6° x 9.2° visual angle) and *Small Surround* (1.5° diagonally from fixation). The activation in the early visual areas for the (*Target – Large Surround*) contrast is shown in **Figure 3.1C**.

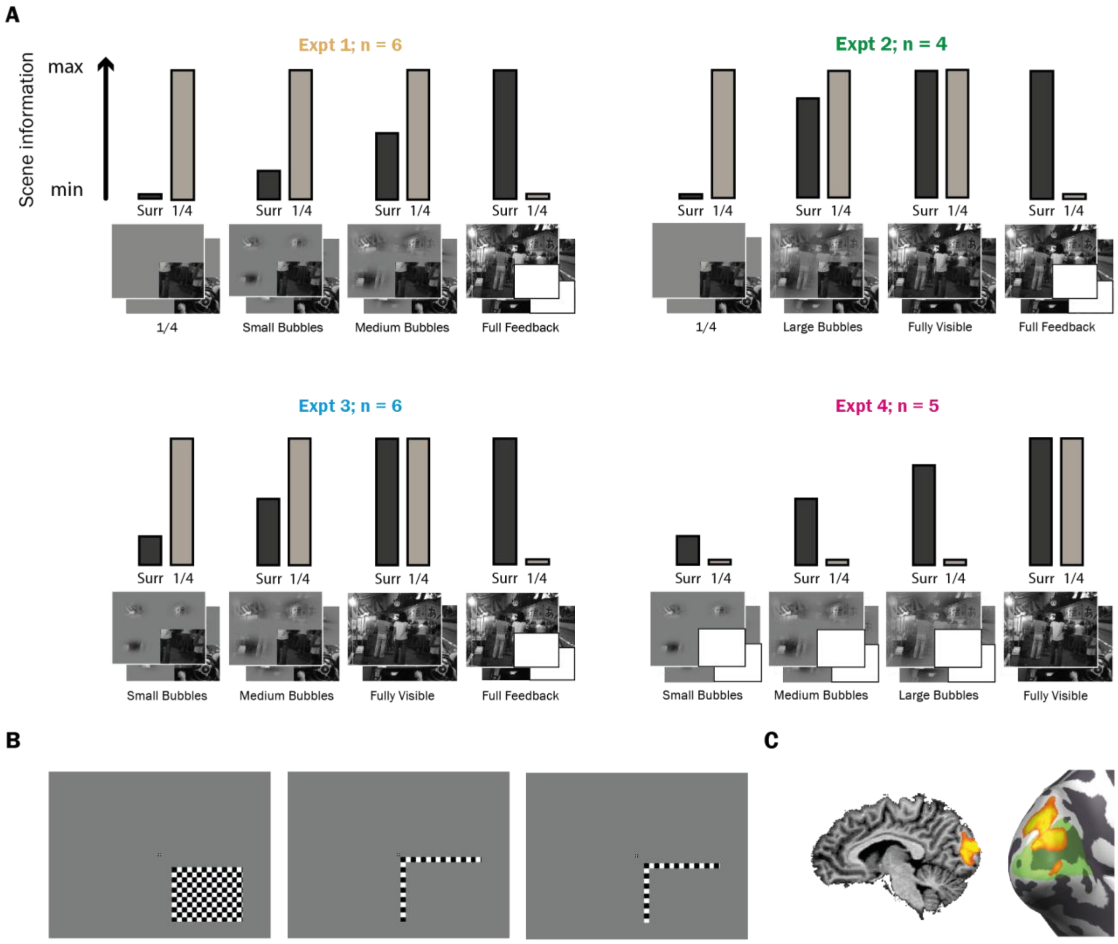


Figure 3.1 | Stimuli. A) In feedback conditions the lower right quadrant was occluded with a white rectangle, while in feedforward conditions the corresponding quadrant was visible. We manipulated scene visibility around the lower right quadrant with Gaussian apertures (“bubbles”) of various sizes (¼ condition [no surrounding scene visible], Small Bubbles [SD = 50 x 32 px], Medium Bubbles [SD = 90 x 56 px], Large Bubbles [SD = 125 x 100 px] and Full [surrounding scene fully revealed]). The dark bars labelled “Surr” represent the extent to which the surrounding ¾ of the scene was revealed, from minimum – no surround revealed, to maximum – surround fully visible. The light bars labelled ”¼” represent the extent to which the quadrant of the scene was revealed, from minimum – occluded condition, to maximum – feedforward condition. Note that the bars are not to scale. Each participant took part in one experiment only. Experiment 1, n = 6; Experiment 2, n = 4; Experiment 3, n = 6; Experiment 4, n = 5. B) Checkerboards were used to retinotopically map the occluded region in V1. Left: *Target* checkerboard (2.5° diagonally from centre, 10.2° x 7.8° visual angle), Middle: *Large Surround* (11.6° x 9.2° visual angle), Right: *Small Surround* (10.9° x 8.5° visual angle). C) The activation for the contrast of (*Target - Large Surround*) used to map the occluded region is shown on the occipital cortex, with V1 shaded in green on the inflated visualization.

3.3.3 Task & Procedure

Scenes were presented on a uniform grey background using MRI compatible goggles (NordicNeuroLab) with 800 x 600 px screen resolution, which corresponded to 32° x 24° visual angle. In each experiment there were 8 types of trial (2 scenes in 4 different conditions). In each 12 second trial the stimulus was flashed on and off (200 ms on/ 200 ms off) 28 times (11.6 secs + variable fixation to account for uncertainty in timing). This maximised the signal to noise ratio (Kay, Naselaris, Prenger, & Gallant, 2008). Each trial type was presented sequentially, with the trial order randomized in each sequence. Each sequence lasted 96 seconds (8 x 12s). A 12 second fixation period was included before and after each sequence of trials. Each experimental run lasted 10 min 48 seconds, containing four trial sequences and two mapping sequences (2 x *Target* and two *Surrounds*). There were four experimental runs in total. Thus, each stimulus was shown 16 times per experiment. Subjects' task was to fixate on a central checkerboard and report a fixation colour change with a button press. Subjects pressed a different button, depending on whether the colour change occurred during scene 1 or scene 2. The purposes of the task were to ensure that the subject paid attention to which scene was being shown and to minimize eye movements. This was important for accurate retinotopic mapping of the occluded region in the lower right quadrant of the visual field. In addition, we used eye-tracking to make sure subjects were fixating.

After the experimental runs, we performed a polar angle retinotopic mapping procedure to estimate the borders of the early visual areas V1-3. This consisted of a single checkerboard wedge which started in the right horizontal meridian and rotated clockwise (12 rotations per scan, wedge angle: 22.5°, scan time: 13 min 28 sec). For some subjects, we also performed an eccentricity mapping procedure. This consisted of an expanding ring which started at the centre and expanded towards the periphery (8 expansions per scan, ring width increased exponentially towards the periphery, scan time: 9 min 12 sec). Subjects were also familiarised with the full non-occluded scenes in a short practice run prior to going into the scanner.

3.3.4 MRI acquisition

MRI data were collected using a 3T Siemens Tim Trio System with a 12-channel head coil. Blood oxygen level dependent (BOLD) signals were measured with an echo-planar imaging sequence (echo time: 30 ms, repetition time: 1000 ms, field of view: 210 mm, flip angle: 62°, 18 axial slices). The spatial resolution for functional data was 3 x 3 x 3 mm. Each experimental run had 648 volumes. Retinotopic mapping consisted of 808 volumes (polar angle) or 552 volumes (eccentricity). The 18 slices were positioned to maximize coverage of occipital cortex. A high resolution 3D anatomical scan (3D Magnetization Prepared Rapid Gradient Echo, 1 x 1 x 1 mm resolution) was also recorded (192 volumes).

3.3.5 MRI data processing

Functional data for each experimental run and retinotopic mapping were corrected for slice time (cubic spline interpolation) and 3D motion (Trilinear/Sinc interpolation), temporally filtered (high-pass filtered at 6 cycles with GLM-Fourier, and linearly detrended), and spatially normalized into Talairach space with Brain Voyager QX 2.8 (Brain Innovation). Subsequently, the anatomical data were used to create an inflated cortical surface and functional data were overlaid.

3.3.6 Voxel selection and analysis

Excessive subject movement between runs is likely to affect correspondence between voxels from one run to another. This in turn could affect our analysis, as we selected our region of interest (ROI) based on the averaged functional data of all 4 runs. To determine whether there was good alignment between functional data covering the visual areas, we calculated an alignment value for each subject by measuring Pearson's correlation in a ROI in the visual cortex between the four functional runs. The median alignment value across subjects was 98.08% and single subject values ranged from 77.85% to 99.31%. We excluded data from further analysis if the alignment value was below 90%, which applied to three subjects. Furthermore, we excluded any subject with chance level performance in any feedforward condition in single trial analysis (significance above chance was measured using permutation analysis with 1000 trials). The feedforward conditions have bottom-up stimulation and hence there should be a difference in activity patterns. If the scenes could not be decoded in these control conditions in a subject, we excluded them from the analysis, as it suggests that the subject might not have been paying enough attention during the experiment. This excluded a further five subjects. Thus, the following analyses were performed on 21 subjects.

The occluded region was mapped using a general linear model (GLM) contrast of the *Target* region against the *Large Surround,* as described previously in Smith & Muckli (2010). The ROI was selected from activation in V1 only. To further minimize spillover activity from neighbouring stimulated areas, voxels from the ROI were then selected for analysis on the basis of the difference between *Target* and *Large* S*urround* t-values being greater than 1.

3.3.6.1 Analyses with extended safety boundary around the occluded region

To further make sure our findings of scene information in the quadrant were not due to spillover activity from the feedforward surround, we performed a separate analysis with more stringent methods of voxel selection. First of all, we selected our region of interest in BrainVoyager as the contrast of the Target mapping region being higher than both the *Large Surround* and the *Small Surround* mapping conditions. In addition, we selected voxels fitting the criteria of (*Target - Large Surround*) > 1 and (*Target - Small Surround*) > 1. This helped to restrict voxels to the more peripheral regions and to further minimize any voxels at the inner borders of the quadrant. Analysis showed the same pattern of results and significant decoding between the two scenes in all conditions except Small Bubbles Feedback and Full Feedback (Average block analysis, Experiment 1 only).

Moreover, we performed another analysis using population receptive field (pRF, Dumoulin & Wandell 2008) mapping for the subjects which had both the polar angle and eccentricity retinotopic mapping available (Expt 2: n = 4, Expt 3: n = 2, Expt 4: n = 4). Again, this was done to restrict our voxel selection to the quadrant. We only included voxels that were both within the occluded region as defined by pRF and only within our original *Target* > *Large Surround* ROI as defined in BrainVoyager.

3.3.7 Multivoxel pattern classification analysis

The voxels matching all these criteria were entered into the linear classifier (Support Vector Machine [SVM]). For classification analyses, we trained the classifier to decode between the 2 scenes in each condition. For cross-classification analyses we trained the classifier to decode between the two scenes in one condition and tested on the other. The classifier used single trial activity patterns (beta values) for training, and was then tested on either single trial (8 trials x 4 sequences = 32 trials) or average activity patterns for each of the 8 trial types. The classifier was trained on 3 of the runs and tested on the remaining run (i.e. one-run-out cross-validation).

We bootstrapped (10000 samples) the classifier performances for individual subjects, in order to estimate the single subject mean. We then bootstrapped (10000 samples) these values to estimate the group mean and associated variance. The confidence intervals (CIs) were defined with an alpha level of 0.05. Classifier performances were deemed to be significantly above chance (50%) if the 95% CIs did not intersect with 50%. We were not able to conduct an ANOVA to check for group differences as the data were not parametrically distributed. Differences between group classifier performances were thus assessed via a permutation test (1000 samples) of the differences between the group means (*p* values not corrected for multiple comparisons).

3.4 Results

Our hypothesis is that the surround stimulation drives higher visual areas with larger receptive fields to send a contextual feedback signal to voxels in V1 responding to the occluded quadrant. We can therefore modify the feedforward surround stimulation to learn more about the nature of contextual feedback.

3.4.1 Increased stimulation of the surround receptive field enhances contextual feedback

First, we tested how much surrounding scene information is needed for non-stimulated V1 to contain information about the scene. We parametrically modulated the availability of surround information and trained the SVM classifier to decode between the two scenes using voxel patterns responding to the lower right quadrant. SVM classification performance was used as an estimate of the amount of available information in the activation pattern. We used SVM on both the occluded (feedback) and the stimulated conditions (feedforward). We refer to the stimulated condition as feedforward but note that this contains a mixture of feedforward, lateral and feedback activation. In contrast, the feedback conditions contain feedback and lateral interactions but no direct feedforward input.

When the lower image quadrant was occluded, scene classification in those non-stimulated voxels improved with increasing availability of surrounding scene information (**Figure 3.2A**, left). Classification was significantly above chance for Medium Bubbles (single trial classification [ST]: 61.25%, confidence interval [CI] [0.075, 0.075]; average block classification [AB]: 70.00%, CI [0.175, 0.15]), Large Bubbles, (AB only: 82.50%, CI [0.3, 0.175]) and the Full Feedback condition (ST: 81.84%, CI [0.0723, 0.0586]; AB: 89.84%, CI [0.1328, 0.0859]), but not for the Small Bubbles. Averaging across experiments, classifier performance for the Full Feedback condition was significantly higher than Small or Medium Bubbles conditions (Small: ST: 50.62%, CI [0.0687, 0.0500], $p < 0.001$; AB:

62.5%, CI [0.15 0.15], *p* = 0.015; Medium: ST only: 61.25%, CI [0.0750, 0.0750], *p* = 0.009).

Increased surround information also improved classifier performance during feedforward processing of the scenes (**Figure 3.2A**, right), as the Fully Visible condition (ST: 88.12%, CI [0.0396, 0.0375], AB: 98.33%, CI [0.0250, 0.0167]) was significantly higher than the other feedforward conditions (Large: AB only: 93.75%, CI [0.0625 0.0625], *p* = 0.019; Medium: ST: 79.95%, CI [0.0443 0.0495], *p* = 0.028, AB: 91.67%, CI [0.0313 0.0313], *p* = 0.001; Small: ST only: 78.91%, CI [0.04127, 0.0469], *p* = 0.007; ¼: ST: 80.94%, CI [0.0469 0.05], *p* = 0.034, AB: 91.25%, CI [0.0625 0.0625], *p* = 0.017). Classification performance for all conditions and experiments is shown in **Figure 3.2B**.

3.4.2 Contextual feedback enhances feedforward processing

Classifier analyses reveal, first of all that with increased presence of the surrounding scene more scene-specific information is detected in the occluded, and hence non-stimulated, V1. Increasing input from the surround enhances the feedback signal, and the information patterns in non-stimulated V1 are richer and more classifiable. Feedforward information was also improved with increased surround stimulation. When there was feedforward scene information in the lower quadrant, revealing more image information in the surround helped classifier performance. The question arises however, do feedback and feedforward signals carry the same information? We used a cross-classification approach to test if the classifier can be trained on contextual surround information in the non-stimulated region i.e. to discriminate the two scenes in the feedback conditions and then use the feedback information to discriminate scenes in the feedforward condition (cross-classification). Successful cross-classification would suggest similar information content.

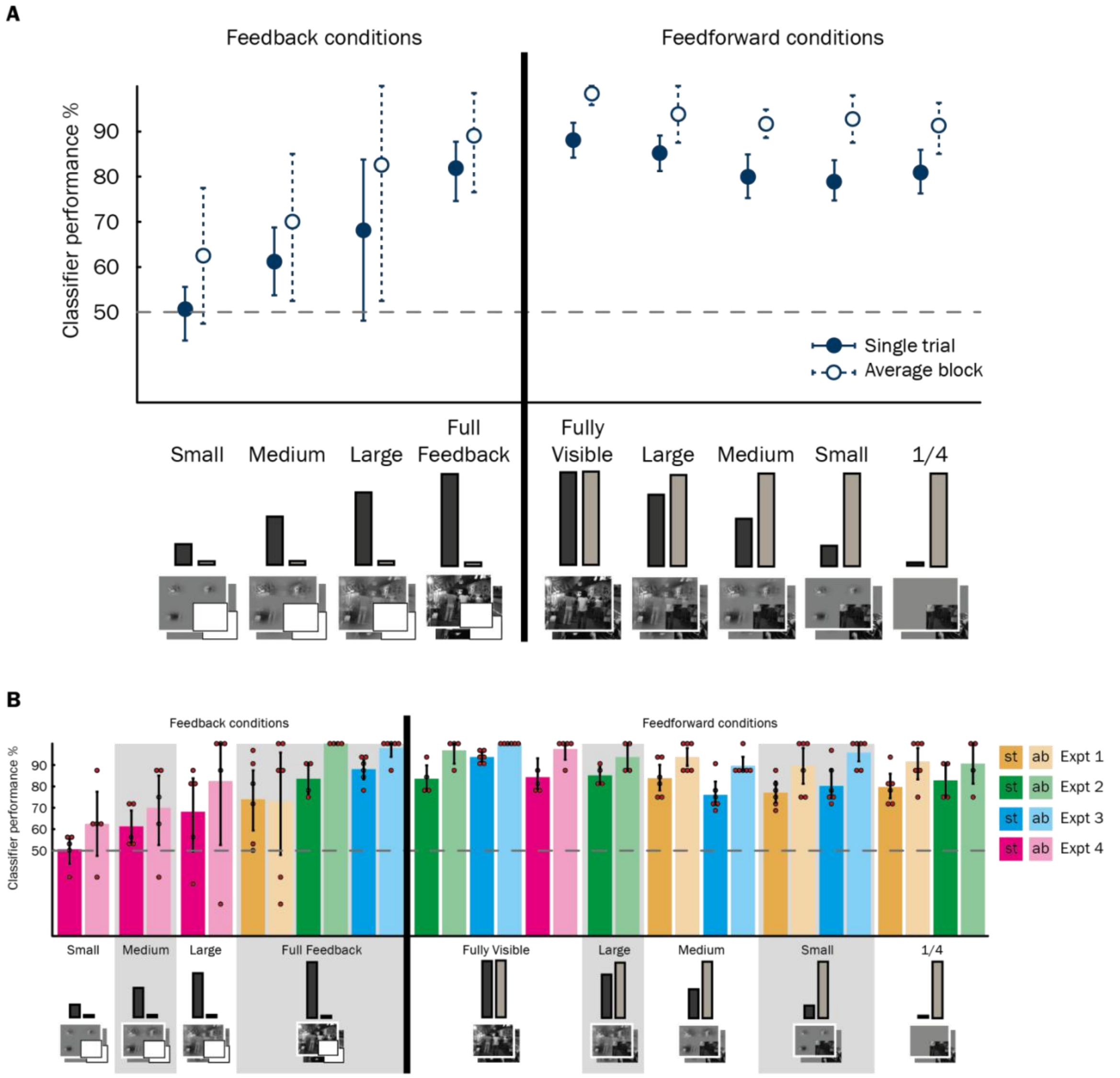


Figure 3.2 | Classification performance for decoding between the two scenes in the same condition, for feedback and feedforward stimuli. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (10000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. A) Classifier performance for each condition, averaged over the four experiments (solid line = classifier tested on single trials; dashed line = classifier tested on blocks of conditions averaged over the same type). Small, Medium and Large Feedback conditions, n = 5; Full Feedback, n = 16; Fully Visible, n = 15; Large Feedforward, n = 4, Medium and Small Feedforward, n = 12; ¼, n = 10. B) Same data as in (A) but classifier performance split by four experiments (separate colours). ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results.

3.4.2.1 How much does feedback contribute to normal visual processing?

We trained the classifier to decode between the two scenes in the Full Feedback condition (with no direct feedforward input in the quadrant) and tested on the feedforward conditions, with varying amount of feedback from the surround (**Figure 3.3**). Cross-classification performance decreased with decreasing scene information in the surround. The classifier could generalize from the Full Feedback condition to the Fully Visible (ST: 72.5%, CI [0.0781, 0.075]; AB: 75%, CI [0.1125, 0.1125]) and Large Bubbles condition (ST only: 63.28%, CI [0.0547, 0.0547]) (**Figure 3.3A**). However, cross-classification for Medium, Small Bubbles, and ¼ conditions was at chance level. Averaging across experiments (**Figure 3.3A**), the Fully Visible condition was significantly higher than the Medium Bubbles (ST: 53.91%, CI [0.0521, 0.0495], p = 0.002; AB: 57.29%, CI [0.0729 0.0833], p = 0.021), Small Bubbles (ST: 48.7%, CI [0.0495, 0.0469], p < 0.001; AB: 52.08%, CI [0.0625, 0.0833], p < 0.001) and the ¼ condition (ST: 50.31%, CI [0.05, 0.0531, p < 0.001]; AB: 50%, CI [0.05, 0.0625], p = 0.003). These results tell us that we can train on a feedback signal (that is likely to be coarse), and test on a signal that is a combination of fine-grained feedforward signal and (coarse) surround feedback, suggesting some shared scene-specific information. This common information must arise from the contextual feedback signal because when the surround stimulus is reduced to nothing (i.e. with shrinking bubbles), we learn that the content of information or its scale (coarse or fine) in feedforward and feedback signals differs.

In order to further test how much information from the surround contributes to normal visual processing, we compared the Fully Visible scene with other feedforward conditions with a reduced scene surround, as well as the feedback conditions (**Figure 3.4**), by training the classifier on the Fully Visible scene and testing on the other conditions. In a fully visible scene both parts of the information are available simultaneously and the classifier might rely more on the rich, fine-grained feedforward information. However, we found that Fully Visible feedforward to feedback cross-classification was only possible with large amounts of scene information surrounding the occluded region. Fully Visible to Full Feedback cross-classification was above chance, while Large, Medium and Small Bubbles did not reach significance in the feedback conditions. This suggests that much of the information in the activity patterns of the Fully Visible scene stems from feedback from the surround. In addition we found that although we could cross-classify above chance from the Fully Visible to all other feedforward conditions, cross-classification reduced with decreased scene information in the surround. Classifier performance was significantly higher for Large Bubbles (ST: 73.44%, CI [0.0469, 0.0469]; AB: 84.38%, CI [0.1563 0.125]) compared to Small Bubbles (ST: 58.85%, CI [0.0521, 0.0573], *p* = 0.007; AB: 60.42%, CI [0.0833 0.0833], *p* = 0.023) and ¼ (ST only: 60.16%, CI [0.0547, 0.0625], *p* = 0.028) conditions. If contextual feedback did not contribute scene-specific information to V1, we would have observed equal cross-classification across feedforward conditions, regardless of surround stimulation.

Interestingly, we found that when the classifier was trained on the Fully Visible image it cross-classified better to Full Feedback (ST: 74.06%, CI [0.075, 0.075]; AB: 0.7875%, CI [0.125 0.1125]) than to some feedforward conditions (significantly above chance for Small Bubbles: ST: 58.85%, CI [0.0521, 0.0573], *p* = 0.013; AB: 0.6042%, CI [0.0833 0.0833], *p* = 0.035). This suggests that feedback in the occluded region from full stimulation in the surround is at least as informative about the scene as actual feedforward information in the quadrant with minimal surround stimulation. This shows that feedback is an important part of the information in V1, both when feedforward stimulation is present and when it is absent.

If feedback information interacts with feedforward processing, then increasing contextual surround information should reduce cross-classification from the ¼ feedforward condition to feedforward conditions with surround stimulation (**Figure 3.5**). Indeed, cross-classifier performance for ¼ to Small Bubbles (ST: 83.33%, CI [0.0573, 0.0625], AB: 91.67%, CI [0.0833, 0.0625]) was higher than to Large (ST only: 69.53%, CI [0.0313, 0.0391], *p* = 0.015) or the Fully Visible condition (ST: 67.97%, CI [0.0703, 0.0703], *p* = 0.021; AB: 65.63%, CI [0.125, 0.1563], *p* = 0.006). Cross-classifier performance for ¼ to Medium Bubbles (ST: 82.81%, CI [0.0677, 0.0677], AB: 0.9375%, CI [0.125 0.0625]) was also significantly higher than to Large (ST only: *p* = 0.039) or the Fully Visible condition (ST: *p* = 0.037; AB: *p* = 0.036).

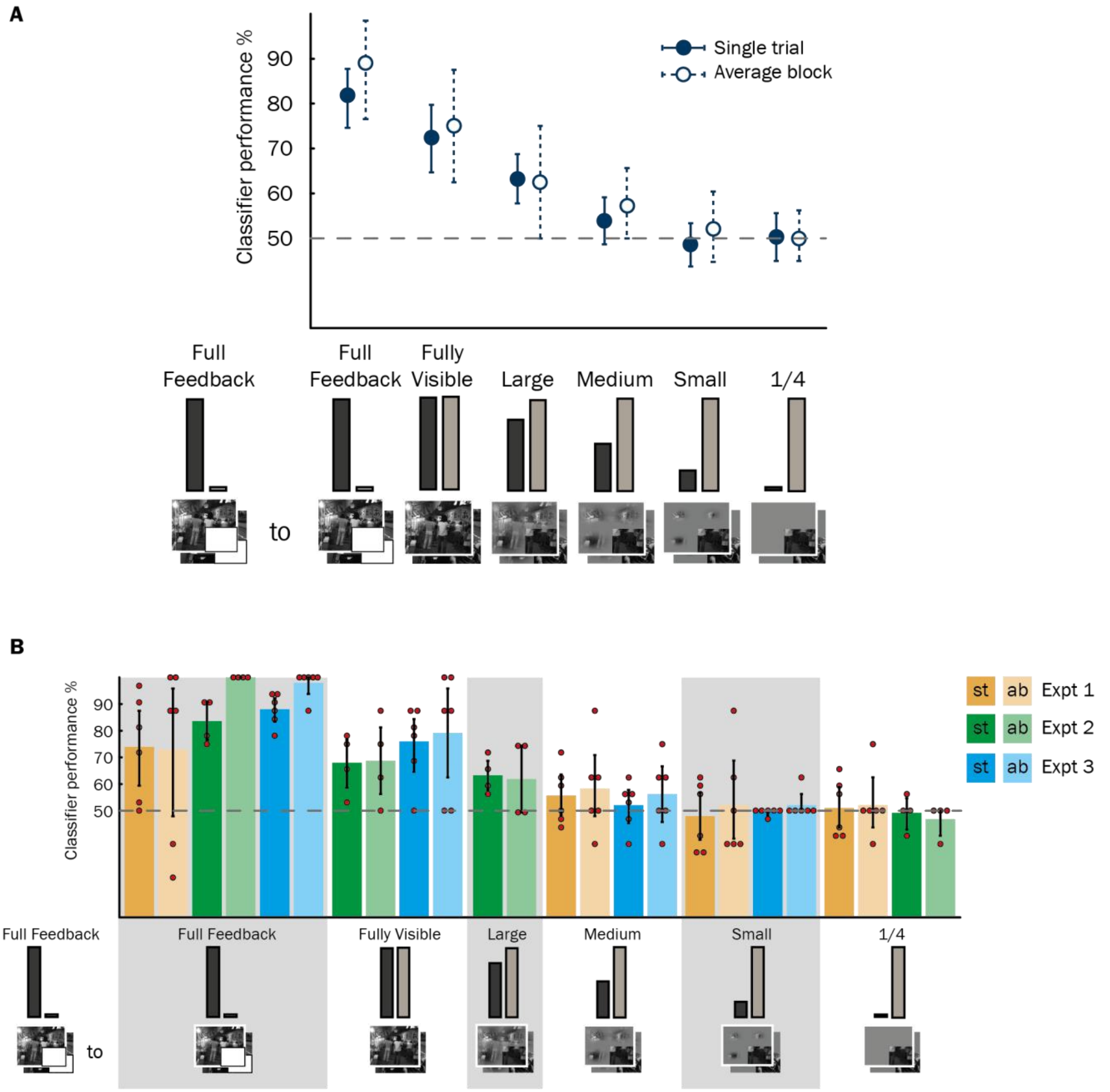


Figure 3.3 | Cross-classification performance for training on the Full Feedback condition and testing on the feedforward conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classification performance for the Full Feedback stimulus (training and testing on the same condition) is shown for comparison. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. A) Classifier performance for each condition, averaged over the four experiments (solid line = classifier tested on single trials; dashed line = classifier tested on blocks of conditions averaged over the same type). Fully Visible, n = 10; Large, n = 4; Medium and Small, n = 12; ¼, n = 10. B) Same data as in (A) but classifier performance split by four experiments (separate colours). ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results.

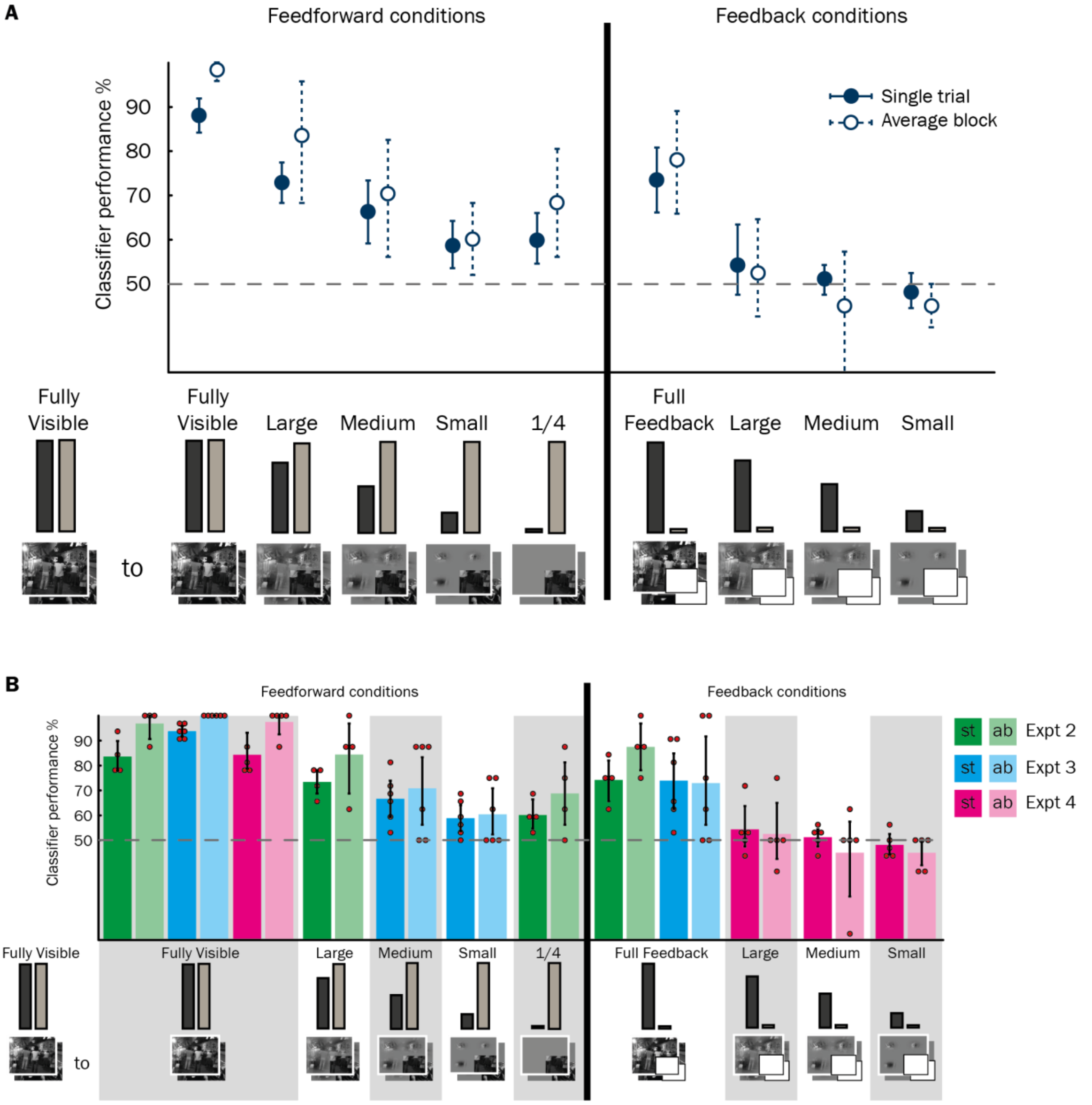


Figure 3.4 | Cross-classification performance for training the classifier on the Fully Visible scene and testing on the other feedforward and feedback conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classification performance for the Fully Visible stimulus (training and testing on the same condition) is shown for comparison. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. A) Classifier performance for each condition, averaged over the four experiments (solid line = classifier tested on single trials; dashed line = classifier tested on blocks of conditions averaged over the same type). Large Feedforward, n = 4; Medium and Small Feedforward, n = 6; ¼, n = 4; Full Feedback, n = 10; Large, Medium and Small Feedback, n = 5. B) Same data as in (A) but classifier performance split by four experiments (separate colours). ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results.

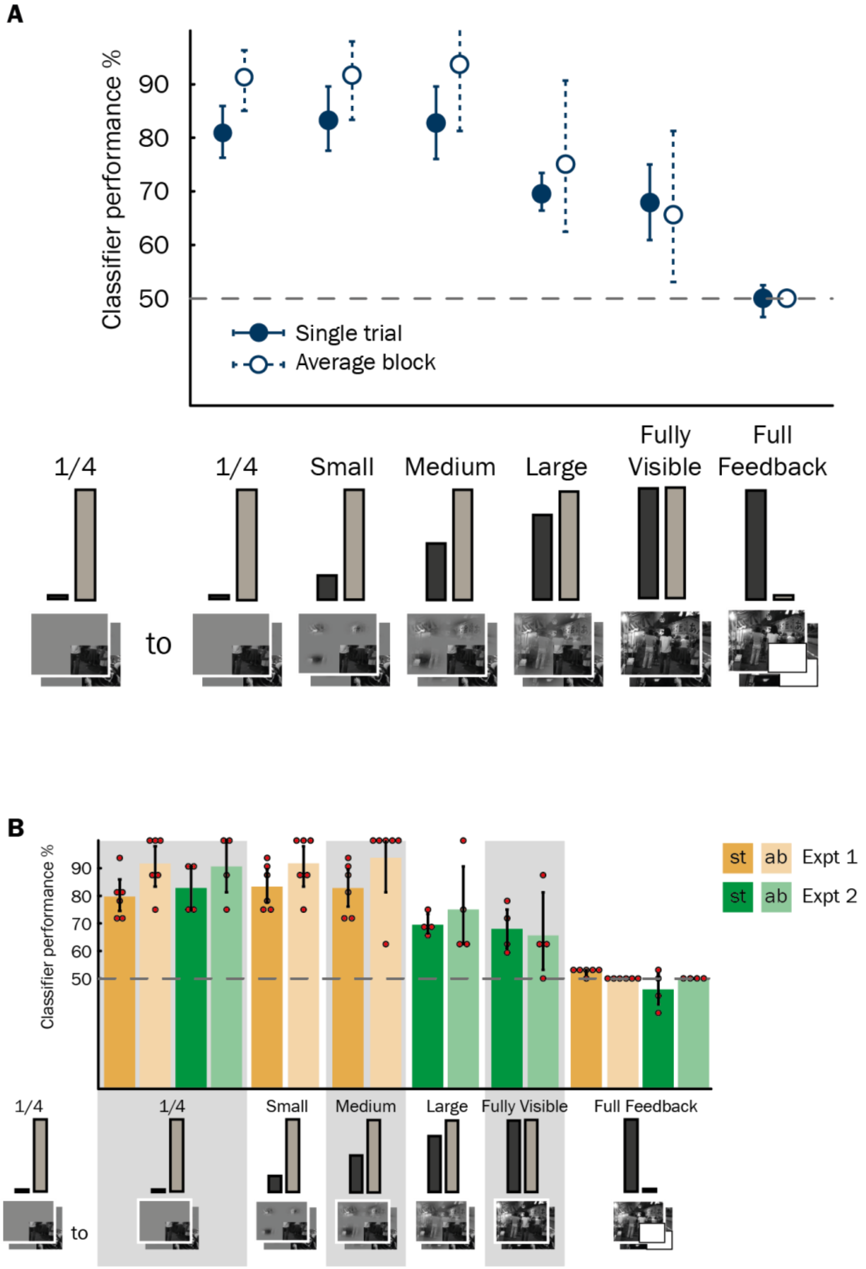


Figure 3.5 | Cross-classification performance for training the classifier on the ¼ condition and testing on the other feedforward and feedback conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classification performance for the ¼ stimulus (training and testing on the same condition) is shown for comparison. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. A) Classifier performance for each condition, averaged over the four experiments (solid line = classifier tested on single trials; dashed line = classifier tested on blocks of conditions averaged over the same type). Small and Medium, n = 6; Large and Fully Visible, n = 4; Full Feedback, n = 10. B) Same data as in (A) but classifier performance split by four experiments (separate colours). ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results.

3.4.2.2 Does increased presentation of the entire image change feedback information?

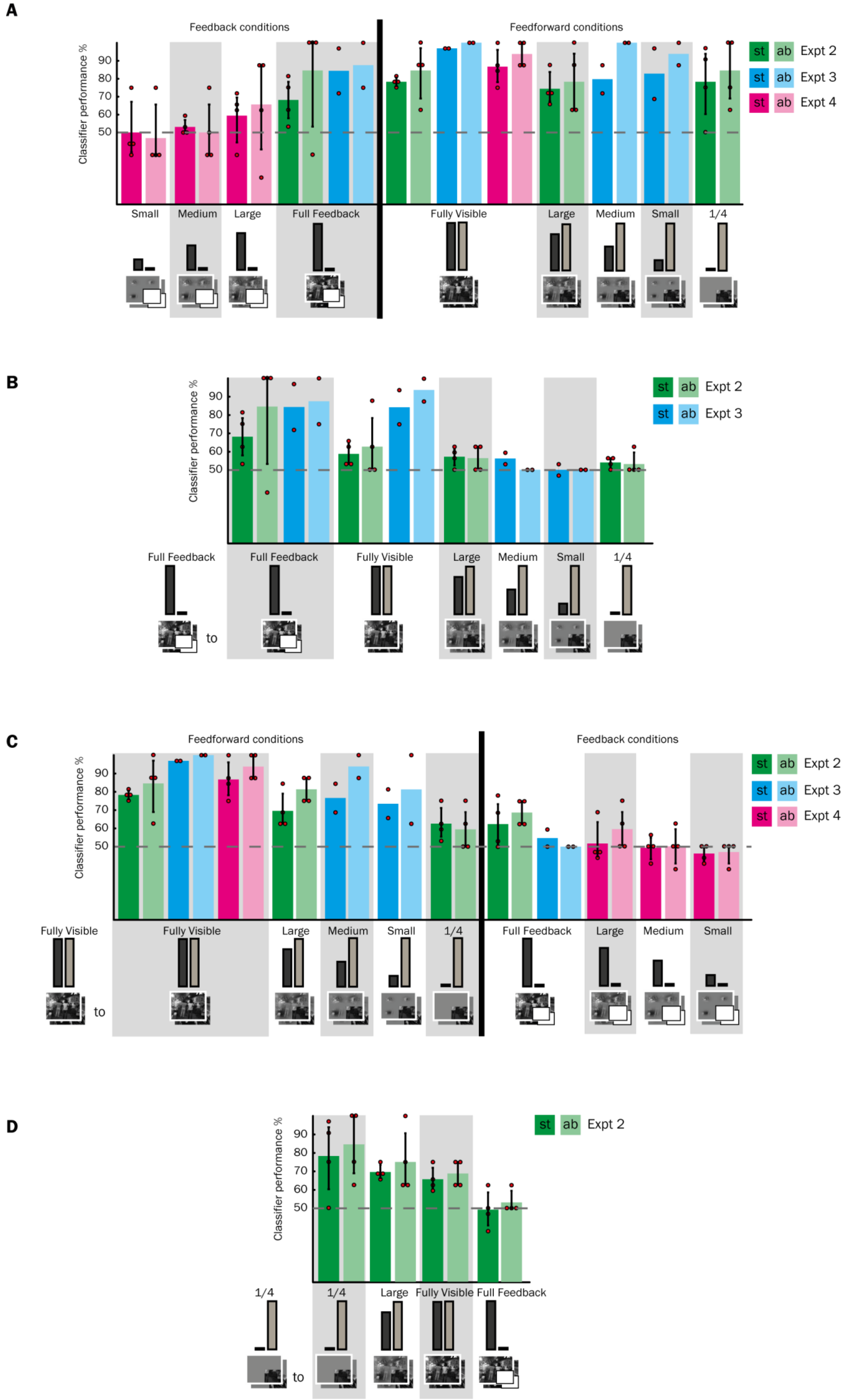
Apart from varying how much surround information is visible in a stimulus, we also wondered whether being reminded of the full scene would improve feedback in the stimuli with reduced surround. In Experiment 3, we presented the Fully Visible scenes along with the Medium and Small Bubbles stimuli, unlike in Experiment 1 where we had not presented the Fully Visible scene as one of the stimuli. Varying the frequency of the Fully Visible scene allowed us to investigate whether being reminded of the full structure of the scene would boost meaningful feedback in stimuli with reduced surround information. We found that cross-classification from Full Feedback to Small and Medium Bubbles was at chance level for both Experiment 1 and 3 (**Figure 3.3B**), suggesting that reduced feedback to the feedforward quadrant in the Small and Medium Bubbles stimuli was mainly due to the decreased contextual surround information as opposed to a reduced implicit memory of the fully visible scene.

3.4.3 Results with extended safety boundary around occluded region

We performed an additional separate analysis in order to decrease the number of voxels that are close to the boundary region and hence reduce the possibility of any feedforward stimulation “spilling over” from the surround. For the conjunction analysis using the contrast of (*Target* > *Large Surround*) & (*Target* > *Small Surround*), we found the same pattern of results and significant decoding between the two scenes in all conditions except Small Bubbles Feedback, and Full Feedback (AB analysis, Experiment 1 only, **Figure S3.1**, Appendix B).

After restricting voxels to the occluded region using pRF mapping, we saw that classifier performance decreased in some conditions, but the pattern of the results remained the same (**Figure 3.6**). Due to the low numbers of subjects in each experiment for whom we were able to perform pRF mapping, we did not calculate confidence intervals for the mean value.

[Next page] Figure 3.6 | Classification and cross-classification performance after applying population receptive field (pRF) mapping to further constrain voxels to the occluded region. Classifier performance is shown for each condition for each of the three experiments (separate colours; Expt 2: n = 4; Expt 3: n = 2; Expt 4: n = 4). ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. Chance level is 50%. A) Decoding two scenes in the same condition. B) Training on Full Feedback and testing on feedforward conditions. C) Training on the Fully Visible scene and testing on other feedforward and feedback conditions. D) Training on the ¼ condition and testing on other feedforward and feedback conditions.


A
Feedback conditions
Feedforward conditions
Classifier performance %
90
80
70
60
50
Small
Medium
Large
Full Feedback
Fully Visible
Large
Medium
Small
1/4
st ab Expt 2
st ab Expt 3
st ab Expt 4
B
Classifier performance %
90
80
70
60
50
Full Feedback
to
Full Feedback
Fully Visible
Large
Medium
Small
1/4
st ab Expt 2
st ab Expt 3
C
Feedforward conditions
Feedback conditions
Classifier performance %
90
80
70
60
50
Fully Visible
to
Fully Visible
Large
Medium
Small
1/4
Full Feedback
Large
Medium
Small
st ab Expt 2
st ab Expt 3
st ab Expt 4
D
Classifier performance %
90
80
70
60
50
1/4
to
1/4
Large
Fully Visible
Full Feedback
st ab Expt 2

3.5 Discussion

The present study aimed to establish how the contextual surround information affects cortical feedback in an occluded region of the visual field in V1. Firstly, we have replicated the findings of Smith and Muckli (2010), by showing that non-stimulated V1 receives contextual feedback from feedforward stimulation in the surrounding regions of the scene. Secondly, we have extended this by showing that this effect is modulated by how much scene contextual information is available in the regions surrounding the occluded area. We further show that feedback signals combine with feedforward information to provide a full visual representation. Using a technique sensitive to the information content in a neuronal population, we have built on previous studies by investigating how surrounding feedback changes the information pattern relating to the visual scene, rather than simply looking at increase or decrease in activation.

3.5.1 Increased stimulation of the surround receptive field enhances contextual feedback

We show that it is possible to study influence from the scene surround in populations of neurons using fMRI, complementing what we already know from single cell studies in electrophysiology. We found that increasing information in the surround, led to increased information in the occluded region of V1. This means that even without direct visual stimulation, regions of V1 are processing information about the surrounding scene.

There was no meaningful feedforward stimulation in the occluded region, and yet we could decode the two scenes using brain activity patterns corresponding to this non-stimulated region. This differential activity must originate from meaningful contextual information in the scene surround. Classical receptive fields are smaller than the surround, which carries feedback from higher areas with larger receptive fields. Hence neurons in the occluded area in V1 could receive information about the rest of the scene via cortical feedback from higher areas. V1 neurons integrate signals over a large area that is well beyond the classical receptive field (Angelucci & Bressloff, 2006; Angelucci et al., 2002). Lateral connections can modulate the response in the central receptive field over short distances. However, feedback from higher cortical areas is needed to account for the full extent of the surround modulation effects (Angelucci & Bressloff, 2006).

As we are measuring a population of neurons using fMRI, as opposed to single cells, it is hard to estimate how widespread the effect of the surround RF would be. Since V1 receives feedback from many cortical areas, which have increasing receptive field sizes as one moves to higher and more abstract processing areas, we might expect influence from the surround would be restricted to regions close to the occluded region for feedback coming from V2, for example, but transmit information from a larger area of the surrounding scene for feedback originating from higher visual areas. Williams et al. (2008) have demonstrated that feedback can indeed come from distant retinotopic regions, by showing that the fovea receives feedback about objects in the periphery.

Larger bubbles in the surround lead to more informative feedback in the occluded region. This may be because we are revealing more of the overall scene structure as we increase the bubble size. Tang et al. (2014) demonstrated top-down effects in image completion by presenting partially revealed images using bubbles. The number of bubbles was constant, but their location was changed. This suggests that revealing a certain amount of the global image structure, regardless of the specific parts, can be enough for top-down completion to take effect. Alternatively, our result could be explained by larger bubbles providing more information close to the lower right quadrant, compared to small bubbles, because our bubbles were centred in each quadrant. Since we did not specifically measure effects of bubble location, it remains to be seen whether varying proximity of surrounding information can affect feedback.

3.5.2 Interaction of feedback and feedforward signals

Stimulating the non-classical surround not only increases information in the occluded region, but also in the quadrant when it contains feedforward information. We saw that similarity between identical feedforward quadrants was reduced if the amount of information in the surround was different. If feedback signals from the surround did not combine with feedforward information or only weakly modulated it, we would have seen similar brain activity patterns relating to the feedforward quadrant regardless of the surround. This suggests that feedback from the surround combines with feedforward information and is necessary for full visual processing in V1. The feedforward signal has been traditionally considered the “loud” signal, since it drives receptive fields, while feedback has been thought of mostly modulatory and not able to trigger spikes (Bastos et al., 2012; Bullier, 2006, but see Mignard & Malpeli, 1991). By using fMRI which is sensitive to non-spiking activity (Logothetis, 2008; Muckli, 2010) we established that this modulation from feedback may be just as important as the spiking produced by stimuli in a bottom-up manner. Many studies support the notion that feedback to V1 is a crucial part of visual perception. For example, inactivation of higher areas such as V2 or MT reduces the response of neurons in the lower areas to visual stimulation in the centre RF, (Hupé et al., 1998; Hupé et al., 2001; Sandell & Schiller, 1982; Schmidt, Lomber, Payne, & Galuske, 2011) and has been shown to affect prediction in an apparent motion paradigm (Vetter, Grosbras, & Muckli, 2015) and has been implicated in awareness (Pascual-Leone & Walsh, 2001), suggesting that feedback from higher areas interacts with feedforward input. In the present study, we took the opposite approach and removed the feedforward input. In addition, we used a paradigm allowing us to investigate how feedback influences the information content in a population of neurons, rather than just spiking activity or change in overall activation. We found that not only does feedback modulate the feedforward information, as has been previously shown, but in fact provides a significant part of the information in full visual processing. The brain is now more commonly viewed as a parallel rather than serial processor (Singer, 2013) and each cortical neuron can be thought of as a “microcosm of the brain as a whole, with synapses carrying information originating from far flung regions” (Gilbert & Li, 2013). Since feedback pathways outnumber the feedforward, it may perhaps be more appropriate to think of internal processing as the important input signal, and the subsequent sensory stimulation forming the feedback signal.

3.5.3 Lateral connections

We can be sure that there is no meaningful feedforward stimulation in the occluded region. However, the information that we do record there could originate from both lateral and feedback connections. Although lateral connections might have a role, we would argue that they cannot fully explain the observed effect, as they only span a relatively small region of retinotopic space, and are not sufficient to account for the full range of surround modulatory response (Angelucci et al., 2002). Lateral connections are 2.3 times larger than the classical receptive field (Sceniak, Ringach, Hawken, & Shapley, 1999) and can project to around 2° from the classical RF in V1 at eccentricities of 2-6° (Stettler, Das, Bennett, & Gilbert, 2002).

3.5.4 Information content of feedback

Predictive coding theories (Clark, 2013; Friston, 2010; Rao & Ballard, 1999) would hypothesise that the occluded part of the scene should be represented, based on the expected scene structure behind the occluder. Several authors have demonstrated that an expected or predicted stimulus evokes activity in V1 which is similar to activity elicited by actual bottom-up stimulation (e.g. Ban et al., 2013; Gavornik & Bear, 2014; Kok, Failing, & de Lange, 2014). Therefore, it is somewhat surprising that we do not find similarity between the occluded region and the missing feedforward quadrant. This suggests that the information in feedback signals does not represent a direct filling-in of the missing feedforward input.

However, a lack of a direct filling-in of the feedforward information is possibly not so counter-intuitive since the participants do not report seeing the missing portion of the scene in the occluded trials (i.e. they do not have a hallucination), so it makes sense to suggest that feedback and feedforward information may be coded in different formats, even though both carry information about the scene. For example, it may be that the information is coarser in terms of its content because of the larger visual fields in higher visual areas or less precise retinotopically (e.g. de-Wit, Kubilius, Wagemans, & Op de Beeck, 2012) or because feedback and feedforward signals project to different cortical layers (Muckli et al., 2015; Rockland & Virga, 1989). Muckli et al. showed using high resolution fMRI that during normal visual stimulation, feedforward information peaks in mid-layers, while contextual feedback information peaks in the superficial layers.

If feedback is carrying expectations and predictions based on prior knowledge we might find that improved knowledge of the full scene structure would be important for meaningful feedback in the occluded region. However, it seems that knowledge about the particular scene being viewed is not necessary. Smith and Muckli (2010) previously found that contextual feedback in the occluded region is present even if participants never see the fully visible scene and were not familiarised with it previously. In the present study we also found that increased exposure to the full scene did not improve feedback in the conditions with reduced surround. Therefore, it appears that the contextual feedback we are seeing in this experiment stems from the scene structure available in each trial, and possibly knowledge of natural scene properties in general, but familiarity with the specific scene is not required. Natural scenes have predictable scene statistics and a lot of the information they contain is redundant (e.g. Attneave, 1954; Barlow, 1961; Torralba &

Oliva, 2003). Surround suppression has been shown to be engaged differentially based on image statistics inferred from Bayesian priors (Coen-Cagli, Kohn, & Schwartz, 2015). Therefore this suggests that prior exposure to a specific scene is not required for feedback from the surround to provide contextual information to the occluded region.

3.5.5 Which brain areas does this contextual feedback originate from?

Feedback to V1 originates from many brain regions, both visual and non-visual (Muckli & Petro, 2013). It is difficult to speculate where the feedback that we see in the occluded region originates from. However, it is plausible that because we used complex natural scenes feedback would originate from multiple areas. Parahippocampal place area (PPA) and retrosplenial cortex (RSC) could be two candidate areas, as these have been shown to preferentially process scene information (Epstein & Kanwisher, 1998; Park & Chun, 2009; Vann, Aggleton, & Maguire, 2009) and spatial and non-spatial context (Bar & Aminoff, 2003). Future studies will give more insight into where V1 receives feedback from in the case of natural scene processing.

3.5.6 Conclusion

We have demonstrated that feedback forms a large part of the brain activity under normal visual stimulation. Using a non-invasive brain imaging technique we have corroborated evidence from electrophysiology showing that stimulation in the far-surround receptive field modulates the response in the classical receptive field. We show that increased information in the scene surround results in increased scene information in both stimulated and non-stimulated visual field regions.

4 Feedback signals from the local surround combine with feedforward information in human V1

4.1 Abstract

An important aspect of visual perception is grouping local elements into global scene and object representations, and feedback signals are thought to underlie this process (e.g. Murray et al., 2002; Schmidt et al., 2011; de-Wit et al., 2012). Feedback transmits contextual information about a larger region of visual space than feedforward stimulation (Angelucci & Bressloff, 2006). Feedback may directly transmit contextual information about the global image structure (Bar, 2004) or the global scene representation may be achieved via several local mechanisms. We investigated whether feedback preferentially transmits information about neighbouring regions (local surround) or the full scene structure, transmitting contextual information from distant retinotopic regions (global surround). In two experiments, we used a partial occlusion paradigm (Smith & Muckli, 2010), functional magnetic resonance imaging (fMRI) and pattern analysis techniques to investigate activity patterns in regions of V1 which were occluded and thus not receiving bottom-up feedforward stimulation. In this way, we were able to probe feedback signals in the absence of feedforward input. In Experiment 1, we presented stimuli composed of four oriented Gabors (45° or 135°), one in each quadrant. Together the Gabors formed one of four global structures (Right, Left, Diamond or X). We showed that such simplistic shapes failed to induce contextual feedback into the occluded region. However, in the presence of feedforward information, we saw that feedback from the local surround combined with identical feedforward input to give rise to different activity patterns in that feedforward region. This suggests that feedback may be recruited differentially depending on whether feedforward stimulation is present or absent. In Experiment 2, we tested whether feedback can originate from a distant retinotopic region in natural scenes. We demonstrated that information in a distant retinotopic scene surround is not a source of feedback. These findings suggest that feedback preferentially originates from nearby regions and provides context to disambiguate local feedforward elements. Therefore context about the global scene structure may arise from a series of local surround interactions.

4.2 Introduction

An important aspect of visual perception is grouping local elements into global scene and object representations, and cortical feedback signals are thought to underlie this process (e.g. Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Schmidt, Lomber, Payne, & Galuske, 2011; de-Wit, Kubilius, Wagemans, & Op de Beeck, 2012). Feedback transmits contextual information about a larger region of visual space than feedforward stimulation (Angelucci & Bressloff, 2006). Neighbouring elements of the scene or its overall global structure can be sources of context. Thus feedback may enable a global representation via several local mechanisms such as contour completion (e.g. Gilad, Meirovithz, & Slovin, 2013; Grosof, Shapley, & Hawken, 1993; Kok & de Lange, 2014; Lee & Nguyen, 2001) or it could directly transmit contextual information about the global image structure (Bar, 2004; Oliva & Torralba, 2006, 2007). For example, Williams et al. (2008) have shown that feedback can originate from retinotopically far regions, suggesting feedback might not be constrained to only transmitting information about neighbouring elements.

We explored which regions of the scene surround contribute the most to contextual feedback signals in V1 – only local neighbouring regions or does the feedback directly contain information about the overall global image structure, taking into account distant retinotopic regions as well? A useful approach to studying feedback signals is to isolate them from the feedforward input by looking at the effect of the scene surround on the activity patterns corresponding to occluded regions of the image, which are therefore not receiving feedforward stimulation. Due to the retinotopic nature of V1 where neighbouring regions of the visual field map onto adjacent parts of the cortex (Wandell, Dumoulin, & Brewer, 2007) it is possible to investigate specific regions of a scene. Several studies have shown that occluded regions of the visual field contain information about stimulation in the surround (Ban et al., 2013; Shushruth, 2011; Smith & Muckli, 2010; Sugita, 1999).

In two experiments, using functional magnetic resonance imaging (fMRI), multivoxel pattern analysis (MVPA, see Mur, Bandettini, & Kriegeskorte, 2009; Norman, Polyn, Detre, & Haxby, 2006) and a partial occlusion paradigm (Smith & Muckli, 2010), we investigated whether the neighbouring local surround contributes more to the activity pattern in the occluded region or whether feedback signals code for the global stimulus structure. In Experiment 1, we presented stimuli composed of four oriented Gabors (45° or 135°), one in each quadrant. Together the Gabors formed one of four global structures (Right, Left, Diamond or X). We investigated whether feedback transmitted information about the orientations of the Gabors in the neighbouring quadrants (local hypothesis) or about the orientation compatible with the overall stimulus structure (global hypothesis). We showed that such simplistic shapes failed to induce contextual feedback into the occluded region, unlike natural scenes (**Chapter 2**, **Chapter 3**, Smith & Muckli, 2010, Muckli et al., 2015). However, in the presence of feedforward information, we saw that feedback from the local surround combined with identical feedforward input to give rise to different activity patterns in that feedforward region. This suggests that feedback may be recruited differentially depending on whether feedforward stimulation is present or absent. In order to determine whether feedback from a distant retinotopic region can give rise to meaningful feedback when the scene structure in the local surround is not informative, we performed Experiment 2 using natural scenes and informative surrounding context in the opposite visual field. We demonstrated that information in a distant retinotopic scene surround was not a source of feedback and did not combine with feedforward information, unlike feedback from the local surround. These findings suggest that feedback preferentially originates from nearby regions and provides context to disambiguate local feedforward elements. Therefore context about the global scene structure may arise from a series of local surround interactions.

4.3 Methods

4.3.1 Subjects

Sixteen subjects from the University of Glasgow participated in the experiment (n = 9 males; mean age: 25.625 years, range: 21-31 years). Subjects were paid for their participation. They provided informed written consent and the experiment was conducted in accordance with procedures and protocols approved by the local ethics committee at the University of Glasgow. Eleven subjects took part in Experiment 1 and five in Experiment 2. One subject from Experiment 1 was excluded due to excessive motion in one run. Hence the following data for Experiment 1 are reported for 10 subjects.

4.3.2 Stimuli

4.3.2.1 Feedback vs feedforward condition

To test feedback signals in the absence of feedforward stimulation, we used an occlusion paradigm previously employed by Smith and Muckli (2010). For the feedback conditions, the lower right quadrant was occluded by a white rectangle (Experiment 1) or an occluder with a more naturalistic shape in the left visual field (Experiment 2). In Experiment 1, the white rectangle was placed 0.5° of visual angle diagonally from the centre of the scene, and spanned 7.5° x 7.5° of visual angle. In the feedforward conditions, the corresponding region of the scene was shown. Thus, in these conditions, we looked at the full visual processing, i.e. both feedforward signals and feedback from the scene surround.

4.3.2.2 Images

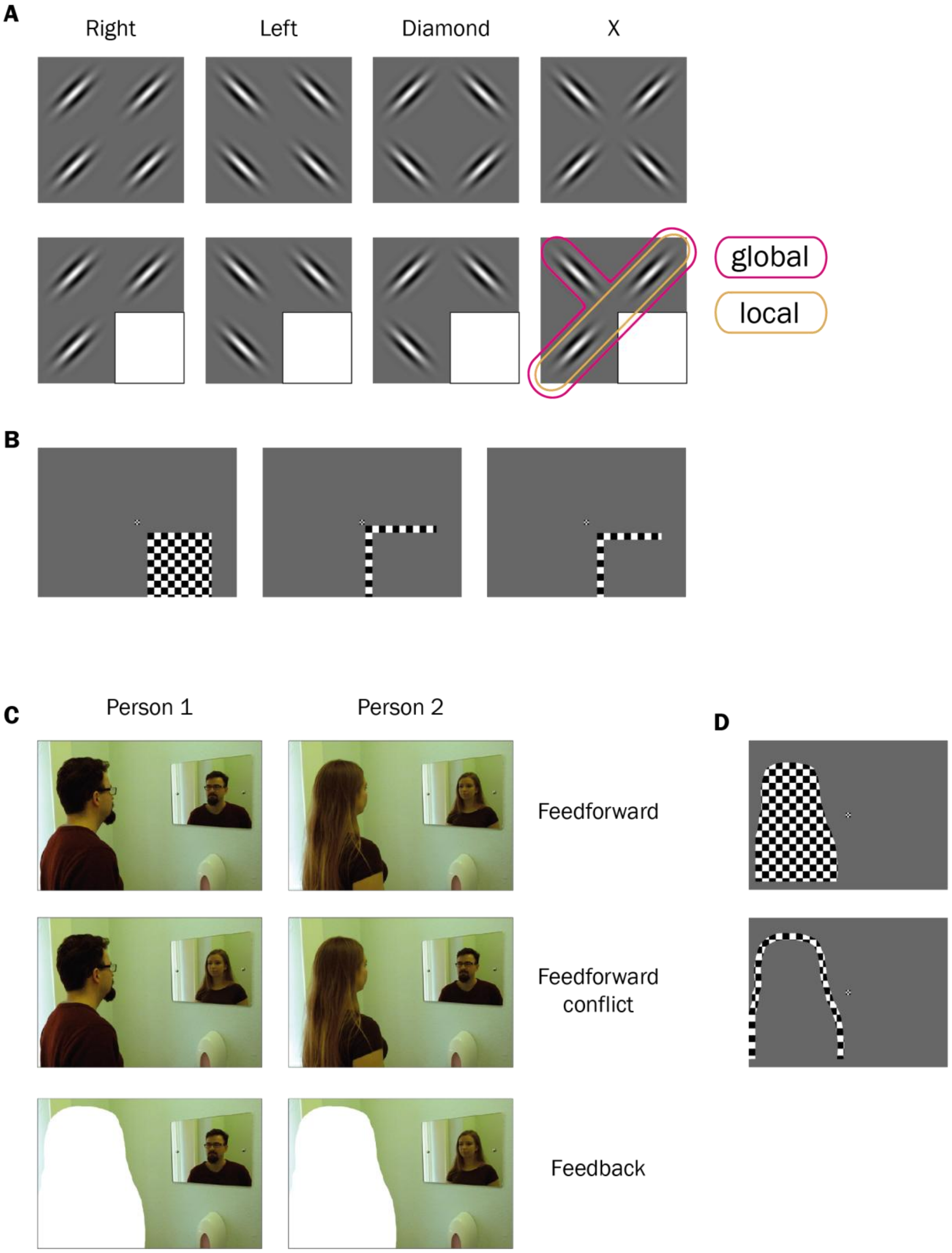
Experiment 1

We used stimuli consisting of four oriented Gabors on a grey background, one Gabor in each quadrant. Each Gabor was oriented at either 45° or 135°, with the four Gabors together forming a single global structure. There were four possible structures: Right (all Gabors at 45°), Left (all at 135°), Diamond and X-Shape (**Figure 4.1A**). We were interested in how contextual feedback to the occluded region would be influenced by the global structure, such as a diamond, versus the orientations in the two quadrants directly adjacent to the occluded quadrant (we refer to these as “local” orientations). For example, in the occluded region for the diamond, we might expect to decode a 45° orientation if the global shape is taken into account, or alternatively 135° if the local surround contributes to feedback signals the most. Right and Left slant served as control conditions since both local and global orientations would be the same. The Gabors has a spatial frequency (SF) of 0.71 cycles per degree of visual angle (cpd), as previous studies have indicated V1 preference for SFs of around 0.68-2 cpd (Haynes & Rees, 2005; Henriksson, Nurminen, Hyvärinen, & Vanni, 2008; Tong, Harrison, Dewey, & Kamitani, 2012). The particular SF of 0.71 was chosen to be optimal to activate V1 and in addition, we required that the sinusoidal pattern in each quadrant was identical, as well as aligned between quadrants in order to facilitate the perception of a uniform oriented field across the full stimulus. Gabors had a bandwidth of 0.75 and an aspect ratio of 0.5, as the four elongated shapes led to a better subjective perception of a global structure. Each quadrant was 384 x 384 pixels, corresponding to a visual angle of 7.9° x 7.9°.

Experiment 2

We used two coloured natural scenes depicting a person in the left visual field looking into a mirror located in the right visual field (**Figure 4.1C**). There were three conditions: feedback (person on the left is occluded), feedforward control (full scene visible) and feedforward conflict (full scene visible, but the reflection in the mirror does not match the identity of the person). The scenes were 1024 x 683 pixels which corresponded to 20.88° x 14.11° of visual angle. In the left visual field, the surrounding image around the occluder was kept the same for the two scenes, as we were interested in effects of global image structure rather than differences in the surround directly neighbouring the occluded region.

[Next page] Figure 4.1 | Stimuli. A) Gabor stimuli used in Experiment 1. One Gabor was placed in each quadrant and the four Gabors together formed one of four possible global structures: Right (all Gabors at 45°), Left (all at 135°), Diamond and X-Shape. In feedforward conditions the full image was shown; in feedback the lower right quadrant was occluded with a white rectangle. The "local surround" was the orientations in the two quadrants directly neighbouring the occluded region. The "global surround" was the whole surrounding image around the occluded region. B) Checkerboards used to retinotopically map the occluded region in V1 for Experiment 1. Left: *Target* checkerboard (1.5° diagonally from centre, 6.8° x 6.8° visual angle), Middle: *Large Surround* (0.5° diagonally from the centre, 7.5° x 7.5° visual angle), Right: *Small Surround* (1.5° diagonally from the centre, 6.8° x 6.8° visual angle). C) Stimuli in Experiment 2. The images consisted of a person in the left visual field looking into a mirror in the right visual field. The images were either a person looking at their own reflection (Feedforward), a person looking at another's reflection (Feedforward conflict) and the person in the left visual field occluded by a white region with a reflection in the mirror in the right visual field still present (Feedback). D) Checkerboards used to retinotopically map the occluded region in V1 for Experiment 2. Top: *Target* checkerboard (covering region inside the surround checkerboard), bottom: *Surround* checkerboard (approximately 1° wide measured from the occluder edge).


A
Right
Left
Diamond
X
global
local
B
C
Person 1
Person 2
Feedforward
Feedforward conflict
Feedback
D

4.3.2.3 Occluded region mapping

Experiment 1

Subjects were presented with three contrast-reversing checkerboards (5 Hz) twice per run. The checkerboards either covered an inner rectangular part the occluded region (*Target* mapping – 1.5° diagonally from centre, 6.8° x 6.8° visual angle) or the border between the lower right quadrant and the rest of the stimulus (*Surround* mapping). There were two types of surround stimuli (**Figure 4.1B**) – *Large Surround* (0.5° diagonally from the centre, 7.5° x 7.5° visual angle) and *Small Surround* (1.5° diagonally from the centre, 6.8° x 6.8° visual angle).

Experiment 2

Subjects were presented with two contrast-reversing checkerboards (5 Hz) twice per run (**Figure 4.1D**). The checkerboards either covered an inner part the occluded region (*Target* mapping - covering region inside the surround checkerboard) or the border between the occluded region and the rest of the stimulus (*Surround* mapping - approximately 1° wide measured from the occluder edge).

4.3.3 Task & procedure

Stimuli were projected on a screen seen through a mirror mounted on top of the head coil. The screen had a resolution of 1024 x 768 px, which corresponded to a visual angle of 20.88° x 15.87°. In each experiment there were 8 (Experiment 1) or 6 (Experiment 2) types of trial (2 scenes in 4 or 3 different conditions). In each 12 second trial the stimulus was flashed on and off (200 ms on/ 200 ms off) 30 times. This maximised the signal to noise ratio (Kay, Naselaris, Prenger, & Gallant, 2008). Each trial type was presented sequentially, with the trial order randomized in each sequence. Each sequence lasted 96 seconds (Experiment 1, 8 x 12 s) or 72 seconds (Experiment 2, 6 x 12 s). A 12 second fixation period was included before and after each sequence of trials. Each experimental run lasted 10 min 48 seconds (Experiment 1) or 10 min (Experiment 2), containing four (Experiment 1) or five (Experiment 2) trial sequences and two mapping sequences. There were four experimental runs in total. Thus, each stimulus was shown 16 times (Experiment 1) or 20 times (Experiment 2) in total. The subjects' task was to fixate on a central checkerboard and report a fixation colour change with a button press. For Experiment 2, subjects pressed a different button, depending on which person was shown in the mirror when the colour change occurred, and the buttons were counterbalanced across participants. The purposes of the task were to ensure that the subject paid attention to which scene was being shown and to minimize eye movements. This was important for accurate retinotopic mapping of the occluded region.

After the experimental runs, we performed a polar angle retinotopic mapping procedure to estimate the borders of the early visual areas V1-3. This consisted of a single checkerboard wedge which started in the right horizontal meridian and rotated clockwise (12 rotations per scan, wedge angle: 22.5°, scan time: 13 min 12 sec). For some subjects, we also performed an eccentricity mapping procedure. This consisted of an expanding ring which started at the centre and expanded towards the periphery (8 expansions per scan, ring width increased exponentially towards the periphery, scan time: 8 min 52 sec). Subjects were also familiarised with the full non-occluded scenes in a short practice run prior to going into the scanner.

4.3.4 MRI acquisition

MRI data were collected using a 3T Siemens Tim Trio System with a 32-channel head coil. Blood oxygen level dependent (BOLD) signals were measured with an echo-planar imaging sequence (echo time: 30 ms, repetition time: 1000 ms, field of view: 210 mm, flip angle: 62°, 18 axial slices). The spatial resolution for functional data was 3 x 3 x 3 mm. Each experimental run had 648 (Experiment 1) or 600 (Experiment 2) volumes. Retinotopic mapping consisted of 792 volumes (polar angle) or 532 volumes (eccentricity). The 18 slices were positioned to maximize coverage of occipital cortex. A high resolution 3D anatomical scan (3D Magnetization Prepared Rapid Gradient Echo, 1 x 1 x 1 mm resolution) was also recorded (192 volumes).

4.3.5 MRI data processing

Functional data for each experimental run and retinotopic mapping were corrected for slice time (cubic spline interpolation) and 3D motion (Trilinear/Sinc interpolation), temporally filtered (high-pass filtered at 6 cycles with GLM-Fourier, and linearly detrended), and spatially normalized into Talairach space with Brain Voyager QX 2.8 (Brain Innovation). Subsequently, the anatomical data were used to create an inflated cortical surface and functional data were overlaid.

4.3.6 Voxel selection and analysis

Excessive subject movement between runs is likely to affect correspondence between voxels from one run to another. This in turn could affect our analysis, as we selected our region of interest (ROI) in V1 based on the averaged functional data of all 4 runs. To determine whether there was good alignment between functional data covering the visual areas, we calculated an alignment value for each subject by measuring Pearson's correlation in a ROI in the visual cortex between the four functional runs. The median alignment value was 99.21% (Experiment 1, individual subjects' values ranged 98.16% – 99.64%) and 99.21% (Experiment 2, range 99.17% – 99.73%) and no subjects were further excluded.

The occluded region was mapped using a general linear model (GLM) contrast of the *Target* region against the *Large Surround*, as described previously in Smith and Muckli (2010). The ROI was selected from activation in V1 only. To further minimize spillover activity from neighbouring stimulated areas, voxels from the ROI were then selected for analysis on the basis of the difference between *Target* and *Large Surround* t-values being greater than 1. To further make sure our findings of scene information in the quadrant were not due to spillover activity from the feedforward surround, for Experiment 1 we performed population receptive field mapping (pRF, Dumoulin & Wandell, 2008). This was done to restrict our voxel selection to the quadrant. We only included voxels that were both within the occluded region as defined by pRF and only within our original *Target > Large Surround* ROI as defined in BrainVoyager. In Experiment 2, the occluded region was in a different visual field to the meaningful surround, thus ensuring feedforward spillover would not affect brain activity patterns in the occluded region. V1 has almost no direct connections from one hemisphere to the other (Van Essen, Newsome, & Bixby, 1982), and thus we assumed that any contextual influence would have been due to feedback from higher cortical areas.

4.3.7 Multivoxel pattern classification analysis

The voxels matching all these criteria were entered into the linear classifier (Support Vector Machine [SVM]). For classification analyses, we trained the classifier to decode between the different Gabor structures in either feedback or feedforward conditions. For cross-classification analyses we trained the classifier to decode two stimuli in the feedback condition and tested on the feedforward condition or vice versa. The classifier used single-trial activity patterns (beta values) for training, and was then tested on either single trial (Expt 1: 8 trials x 4 sequences = 32 trials; Expt 2: 6 trials x 5 sequences = 30 trials) or average activity patterns for each of the 8 or 6 trial types. The classifier was trained on 3 of the runs and tested on the remaining run (i.e. one-run-out cross-validation).

We bootstrapped (1000 samples) the classifier performances for individual subjects, in order to estimate the single subject mean. We then bootstrapped (1000 samples) these values to estimate the group mean and associated variance. The confidence intervals (CIs) were defined with an alpha level of 0.05. Classifier performances were deemed to be significantly above chance (50%) if the 95% CIs did not intersect with 50%.

4.4 Results

4.4.1 Local surround interacts with feedforward information

In Experiment 1, we aimed to see whether the brain activity patterns in the occluded region are more related to the local neighbouring surround orientation or more commensurate with the orientation predicted by the global shape. To test this, we decoded between all combinations of pairs of stimuli. The comparisons of particular interest were those which would predict a difference in ability to decode based on whether the local or the global surround contributes more, specifically Right vs Diamond, Right vs X, Left vs Diamond and Left vs X. For example, in the Right vs Diamond comparison, if the local orientations contribute more, we would expect to decode the two stimuli since different orientations would be predicted in the occluded region. On the other hand, if the global shape is taken into account, then we might see poor decoding as the same orientation would be expected in the occluded region. The Right vs Left comparison acted as a control for the presence of informative feedback in the occluded region in general, as the local and global surround predict the same orientation in the lower right quadrant.

In the feedback conditions, we found that we could not decode between the different stimuli apart from Diamond vs X (single trial only, 56.87%, CI [0.0563 0.0687], **Figure 4.2**, left). In this pair of stimuli there was a difference in both the local and the global surrounds and therefore this suggests that feedback might carry information about local or global surround, or a mixture of both. However, since we failed to decode above chance in the Right vs Left, which served as the control condition, and the other feedback conditions, it suggests that the surround failed to induce very meaningful stimulus specific feedback signals in the occluded region.

In the feedforward conditions, we could decode above chance in all comparisons where there was an actual feedforward difference in the orientation of the Gabor in the lower right quadrant, as would be expected (**Figure 4.2**, right). In addition, we could also decode Right vs Diamond, even though the feedforward orientations in the lower right quadrant were the same. The local surround, on the other hand, was different, suggesting that feedback from the local surround combines with otherwise identical feedforward information in the quadrant to give rise to different activity patterns for the two stimuli. Surprisingly, however, we did not see the same above chance classification for Left vs X.

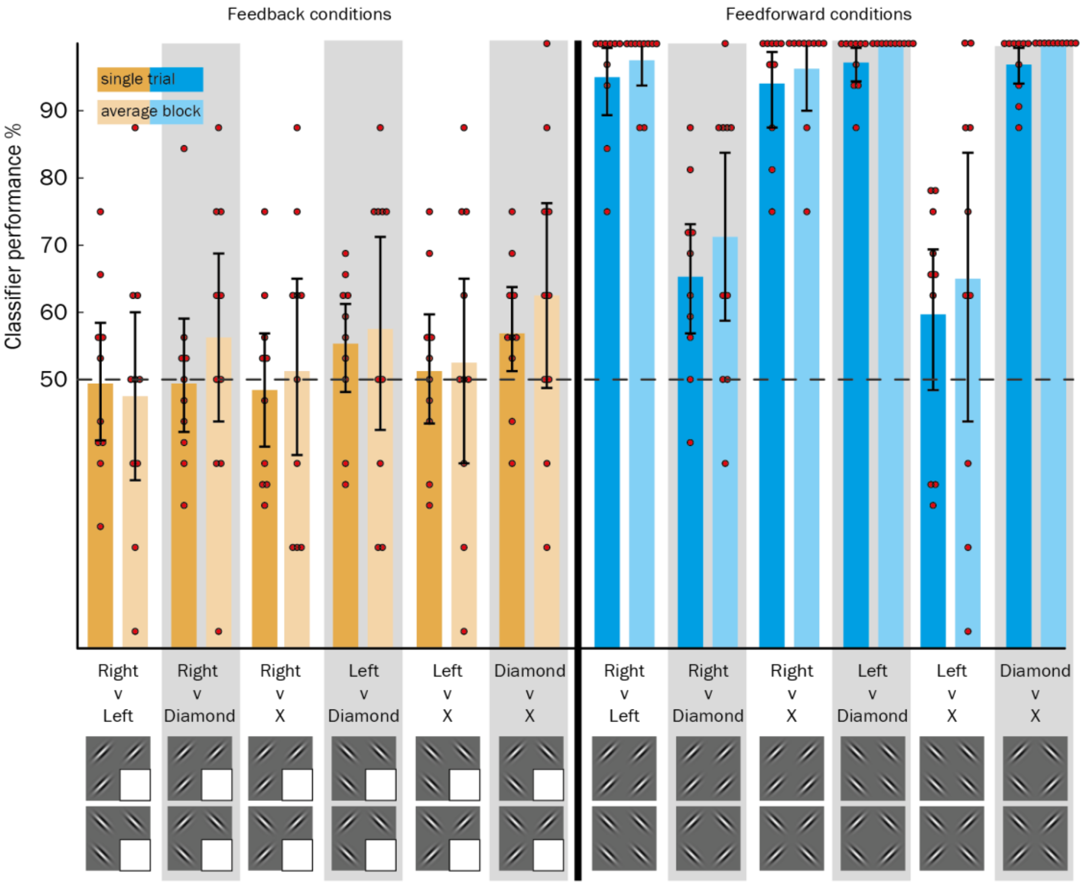


Figure 4.2 | Classification performance for decoding pairs of stimuli, for feedback and feedforward conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Single trial (dark hues) = classifier tested on single trials; average block (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. N = 10.

4.4.2 Information from the retinotopically distant surround does not contribute to feedback

In Experiment 2, we investigated whether the occluded region contains contextual information about the surrounding scene, when using more naturalistic stimuli and when the difference in the surround between the two scenes was in a retinotopically distant region. The scene behind the occluder would be different depending on the reflection in the mirror, and thus we aimed to probe the influence of global scene structure, while keeping the neighbouring surround the same. The occluded region was in the left visual field, while the informative surround (mirror) was in the right visual field. Hence we hoped to probe top-down influences of feedback, minimizing influences from lateral connections or spread of activation from the edges of the occluder. The two visual fields would be represented in different hemispheres and V1 has almost no callosal connections between the two (Van Essen et al., 1982). We found that we could not decode the scenes in the feedback conditions (**Figure 4.3**, gold bars), suggesting that information about the person in front of the mirror from the distant surround was not fed back to the occluded region.

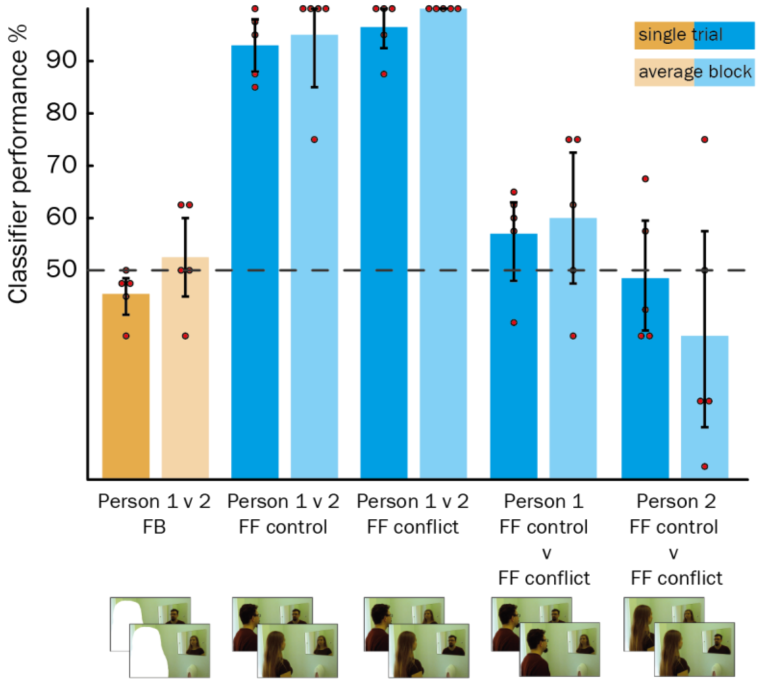


Figure 4.3 | Classifier performance for decoding the figures in feedback and feedforward conditions. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not touch the chance line. Single trial (dark hues) = classifier tested on single trials; average block (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. N = 5.

In addition, we were interested in whether feedback from the surround would combine with feedforward information, similar to the effect we found in Experiment 1 with the local surround. To test this, we decoded between the feedforward control and the feedforward conflict conditions. The feedforward information was identical (the same person) but feedback from the surround differed. Classifier performance was at chance level, suggesting that information from the distant surround about the reflection in the mirror did not influence the activity patterns representing the feedforward figure (**Figure 4.3**, rightmost two pairs of blue bars). Finally, the classifier was able to decode above chance in the feedforward conditions, both the control and conflict, as was expected (**Figure 4.3**).

4.4.3 No similarity between feedback and feedforward information

Another way to test which information is fed back to the occluded region is to compare activity patterns in the occluded region with those in the corresponding feedforward region. We performed cross-classification by training the classifier to decode two stimuli in the feedback conditions and tested its ability to decode two stimuli in the corresponding feedforward condition (and vice versa). In Experiment 1 we tested whether the information about the orientation in the lower right occluded quadrant was more compatible with the global shape or the neighbouring surround. For example, is the lower right quadrant of the occluded Diamond more similar to the lower right quadrant of feedforward Diamond/Right (global representation) or is it more similar to feedforward Left (local representation)? We did not find above chance cross-classification for any comparisons in either direction of cross-classification (**Figure 4.4A**). Classifier performance was at chance level for training and testing to decode Right vs Left, Diamond vs X, and also training on Diamond vs X in feedback and testing on Right vs Left in feedforward. This suggests there was no similarity between the information in the occluded region and the corresponding feedforward region. In Experiment 2 we also found that we could not cross-classify above chance from feedback to feedforward, or vice versa (**Figure 4.4B**, left and middle). The lack of successful feedback to feedforward cross-classification is unsurprising since we did not find meaningful feedback information in the occluded region in the first place (**Figure 4.2**, left**, Figure 4.3**). This further suggests that the person in front of the mirror was not represented in the occluded region. Finally, we could cross-classify from the feedforward control to the feedforward conflict condition, and vice versa, as expected (**Figure 4.4B**, right).

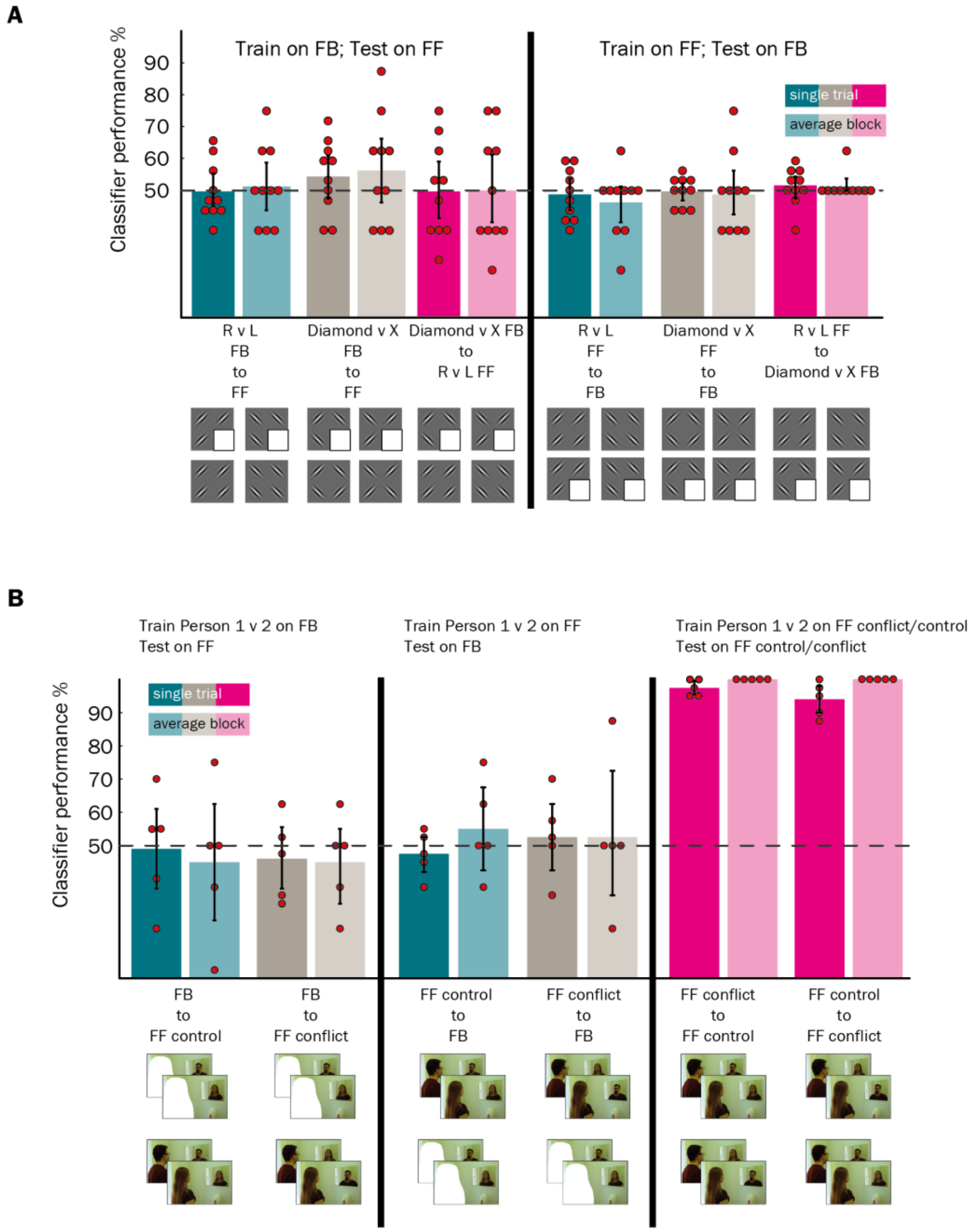


Figure 4.4 | Cross-classification performance for training on feedback conditions and testing on feedforward (and vice versa). Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Single trial (dark hues) = classifier tested on single trials; average block (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. A) Experiment 1, N = 10. B) Experiment 2, N = 5.

4.5 Discussion

In the present study we investigated whether feedback to V1 carries more information about the neighbouring surrounding regions, or whether it codes about the global image structure. We showed that simplistic shapes, as opposed to natural scenes, failed to induce contextual feedback into the occluded region. However, in the presence of feedforward information, we saw that feedback from the local surround combined with identical feedforward input to give rise to different activity patterns in that feedforward region. This suggests that feedback may be recruited differentially depending on whether feedforward stimulation is present or absent. We also demonstrated that information in a distant retinotopic scene surround was not a source of feedback. Overall, our findings suggest that feedback preferentially originates from nearby regions and provides context to disambiguate local feedforward elements. Therefore context about the global scene structure may arise from a series of local surround interactions.

4.5.1 No contextual feedback in occluded region for simple shapes

We did not find informative activity patterns in the occluded region of the Gabor structures in Experiment 1. This finding was surprising as we found contextual feedback signals previously using natural scenes (**Chapter 2, Chapter 3**; Smith & Muckli, 2010, Muckli et al., 2015). One explanation could be that these shapes are too simple and artificial, and therefore do not elicit contextual feedback from the surround. Natural scenes provide rich context and expectations (Bar, 2004), which may not have been available in our Gabor stimuli. Even though participants were familiar with the full structures in the feedforward conditions, it is possible that in the occluded conditions participants did not get a strong percept of the full structure, such as Diamond, but rather perceived the stimulus as three orientated shapes and a white square.

In natural scenes there are predictable contours, from prior knowledge of what everyday objects and scenes look like. If one role of feedback is to integrate local elements into continuous contours, then this could be the mechanism underlying informative activity patterns in the occluded region of natural scenes. In the Gabor structures of the present study, filling in information in the occluded region was less to do with contour continuation of several collinear elements arranged end to end, but rather relied more on the expectation and prediction of the missing Gabor that would complete the full structure. In addition, feedback is recruited more for low salience stimuli (Bullier, Hupé, James, & Girard, 2001; Bullier, 2006). Since our stimuli had high contrast, this could explain why contextual feedback was not strong. It may be the case that feedback is more useful in ambiguous situations or when there is a strong contextual drive from the surround, such as in natural scenes.

4.5.2 Feedback from the local surround combines with feedforward information

We found that in the feedforward conditions, even when the feedforward orientation in the lower right quadrant was identical, we could decode the different stimuli (Right vs Diamond). This shows that feedback from the local surround combines with the feedforward information in the quadrant. This is a surprising finding considering we did not find meaningful feedback in the occluded quadrant. This suggests that contextual feedback is more likely to be recruited when feedforward stimulation is available, compared to when it is missing due to occlusion. One interpretation could be that feedback is useful for modulating the response to stimuli presented in the classical RF. Several studies have highlighted the role of feedback in modulating feedforward information (Hupé et al., 1998; Hupé et al., 2001; Sandell & Schiller, 1982; Schmidt et al., 2011). A mechanism by which modulation from feedback may occur is apical amplification (Phillips, 2015) where inputs at the apical tufts of pyramidal cells amplify the response to the basal inputs (where feedforward information enters). Via this mechanism, contextual input from feedback can amplify the cell’s response to the feedforward input. Feedback may be needed in the feedforward case to help interpret how the feedforward orientation fits in with the bigger picture. In other words, is the 45° orientation a part of the Diamond or the Right slant stimulus? When information in the lower right quadrant is missing, it is less necessary to interpret and disambiguate.

We should note that the analogous comparison of Left vs X in the feedforward condition was not decodeable. The local surround was different hence we should have been able to decode. A possible explanation for the difference between Right vs Diamond and Left vs X could be the arrangement of the Gabors. In the Diamond stimulus, the ends of the Gabors were oriented towards the periphery, while not containing any information in the fovea. For the X stimulus, it was the opposite arrangement. Since receptive fields in the periphery are larger than those in the fovea (Xu, Anderson, & Casagrande, 2007), the smaller surround receptive fields in the occluded region near the fovea would not capture the informative parts of the image surround. This further suggests that the local neighbouring surround contributes the most to the information in the lower right quadrant, rather than information about the overall global structure of the stimulus.

One worry about seeing effects from neighbouring regions is that we are measuring activity from voxels receiving feedforward stimulation from the surround – a sort of “spillover” effect. However, we would argue that this is not the case. If our results could simply be explained by voxels receiving feedforward stimulation, we would have observed above chance decoding in the occluded as well as the non-occluded conditions. Since we only saw an effect of surrounding information in the feedforward conditions, this suggests that our results are more likely to be due to modulatory effects of feedback/lateral connections.

4.5.3 No contextual feedback from retinotopically distant regions

We established that there is a contextual influence of the neighbouring surround on information in the lower right quadrant. We did not find evidence of feedback about the global structure of the stimulus, stemming from the top left quadrant. In Experiment 2, we tested whether scene information in a retinotopically distant surround would affect activity patterns in the occluded region when local information was not informative. We used natural scenes, since contextual global image structure may be more important in this type of stimulus. We found that there was no meaningful scene information in the occluded region, stemming from a retinotopically distant surround. We also did not find a difference in the activity patterns of identical feedforward information, when there was a difference in the surround. This suggests that the local surround contributes more to feedback, as we demonstrated in Experiment 1.

Although our goal was to use a natural scene where contextual influences from the surround would be relevant for its interpretation, it is possible this was not the most optimal stimulus to elicit contextual effects. It may be that “filling-in” of the figure in the occluded region was a too high-level task, because it is more complex than contour or background extrapolation, for example. It could also be the case that seeing the reflection in the mirror was enough to identify the person depicted and thus interpret the scene, and contextual feedback to the occluded region was not required. Some subjects reported that the stimuli were subjectively perceived as a person standing in front of a window, or the figure was viewing a photograph on the wall. If participants did not perceive a strong relationship between the figure and the information in the surround, it could explain why contextual feedback was diminished. Feedback about the global scene structure may also be more informative when the global scene structure differs more and is behaviourally relevant. For example, when the spatial layout of the scenes is different and they provide different navigational affordances. The spatial layout of the two scenes we used was the same and we only manipulated objects in the scene.

Overall, we show that information from the nearby surround contributes more to feedback. This makes sense if we consider that feedback is involved in perceptual grouping, contour completion and figure-ground segregation (Gilad et al., 2013; Gilad, Pesoa, Ayzenshtat, & Slovin, 2014; Grosof et al., 1993; Hess & Field, 1999; Kok & de Lange, 2014; Lee & Nguyen, 2001; Scholte, Jolij, Fahrenfort, & Lamme, 2008). The surround receptive fields provide information about the neighbouring image structure to aid interpretation of local features (Angelucci & Bullier, 2003; Coen-Cagli, Kohn, & Schwartz, 2015). This could explain why increasing the size of bubbles in the surround (**Chapter 3**) led to more informative feedback signals in the lower right quadrant. Making the bubbles smaller meant that contours neighbouring the occluded region were not available, and therefore feedback was not recruited to complete and extrapolate them. However, feedback is involved in more than just local centre-surround interactions. For example, Williams et al. (2008) showed that feedback to the fovea can originate in the periphery. Tang et al. (2014) found object completion effects regardless of the location of “bubbles” containing scene information. It seems plausible therefore that feedback can be signal about the overall scene structure. This might be more likely if the global image structure is behaviourally relevant, and might be achieved via a series of more localised interactions.

4.5.4 Information content of feedback

Predictive coding theories (Clark, 2013; Friston, 2010; Rao & Ballard, 1999) would hypothesise that the occluded part of the scene should be represented, based on the expected scene structure behind the occluder. Several authors have demonstrated that an expected or predicted stimulus evokes activity in V1 which is similar to activity elicited by actual bottom-up stimulation (e.g. Ban et al., 2013; Gavornik & Bear, 2014; Kok, Failing, & de Lange, 2014). However, in the present study we did not find that the orientation of the Gabor expected in the occluded quadrant, considering the overall structure, was represented, but rather, feedback transmitted information about neighbouring orientations. In the mirror stimuli, there should have been a violation of expectations in the incongruent feedforward stimuli, compared to the control stimuli. However, we did not find a difference in contextual feedback between the congruent and the incongruent versions of each scene. This suggests that expectations did not play a major role in eliciting the contextual feedback in our case. In Experiment 1 it is possible that participants did not have strong enough expectations about the occluded Gabor, for example, if the stimulus did not lead to a strong perception of the global structure but rather a combination of several individual elements. In Experiment 2, the incongruent feedforward condition may not have been unexpected enough. On the one hand, it would be unexpected in terms of real life perception, but on the other hand, it is not unexpected as part of the experiment, since all conditions were presented an equal number of times. Overall, we demonstrate that feedback is more involved with transmitting information about the neighbouring structure, rather than "filling-in" the missing stimulus. Future studies should investigate whether filling-in may occur when there is a stronger expectation of the missing stimulus or when the scene structure behind the occluder is highly task relevant.

4.5.5 Conclusion

We show that feedback may be recruited differentially depending on whether feedforward stimulation is present or not. Simple shapes made of Gabors do not give rise to contextual feedback in an occluded region, but feedback from the nearby surrounding regions combines with corresponding feedforward information when it is available. Finally, we do not find informative feedback from a distant retinotopic region, suggesting that feedback may be preferentially transmitting information about neighbouring local elements and global scene context may be achieved via a series of local mechanisms.

5 General Discussion

5.1 The influence of scene surround on cortical feedback

This thesis aimed to investigate the influence of the scene surround on cortical feedback to non-stimulated visual cortex. Normally during vision, both feedback and feedforward signals are present. Feedforward signals act on the central region of a neuron's receptive field, whereas feedback signals carry information from higher cortical areas with larger receptive fields, and are therefore able to "inform" the central receptive field about the surrounding scene. To study feedback signals in isolation, we took advantage of the retinotopic organisation of V1 and the spatial resolution of fMRI, in order to investigate activity patterns in regions of V1 which were not receiving meaningful feedforward stimulation. In the absence of feedforward stimulation, overall activity levels in those neurons were expected to be low. However, by using decoding techniques to probe the information content of the non-stimulated region, we hoped to study the information that feedback carries about the surrounding scene. Functional MRI is a technique sensitive to non-spiking activity and therefore it allowed us to investigate feedback signals, which are considered to be mostly modulatory (Muckli, 2010).

Previously it has been shown that visually occluded regions of V1 (and thus non-stimulated in a feedforward manner) contain contextual information about the scene surround outside the occluded area (Smith & Muckli, 2010). In the present work, we asked how exactly the scene surround influences feedback signals, and investigated the information that feedback transmits.

To answer this question, in **Chapter 2** we looked at whether feedback transmits coarse or fine-grained information, represented by low and high spatial frequencies respectively. We found that both high and low spatial frequency surrounds elicited informative feedback signals in the occluded region, suggesting that feedback can transmit both coarse and fine-grained information. There was a similarity between the HSF and LSF feedback activity patterns, indicating that feedback signals are likely to be broad in their spatial frequency tuning. Another finding of **Chapter 2** was the lack of similarity between feedback activity patterns and the corresponding feedforward information which was presented in isolation, without the scene surround. We later demonstrated (**Chapter 3**) that this feedback to feedforward correspondence depends on information from the scene surround, suggesting that in feedforward conditions, surrounding feedback provides additional information about the visual stimulation, different to what is represented by the feedforward input.

In **Chapter 3**, we explored the amount of scene surround required to elicit meaningful feedback signals to the occluded region. The results showed that increasing the amount of scene information in the surround, increased meaningful feedback signals. Again we replicated the dissimilarity between information in the occluded region and the isolated region of the corresponding feedforward stimulation, which was lacking the surround. Similarity between feedback activity and activity in the visible feedforward regions increased with increasing amounts of information added in the scene surround. Thus we confirmed the original finding of Smith and Muckli (2010) who showed that the activity patterns in the occluded region were similar to the corresponding portion of the full feedforward image, suggesting that activity patterns in the full scene represent in part feedback from the surround, and in part the feedforward input. Overall, these findings point to the notion that feedback signals combine with feedforward input under normal visual processing. Isolated feedforward input in the absence of the surround provides V1 neurons with impoverished information.

Neighbouring elements of the scene or its overall global structure can be sources of context. Therefore the finding of **Chapter 3** that informative feedback reduced with a decreased scene surround could be explained either by disruption of the informative global scene structure or removal of informative surround directly next to the quadrant. In **Chapter 4** we explored which regions of the scene surround contribute the most to contextual feedback signals to V1 – are local neighbouring regions most important, or does feedback directly contain information about the overall global image structure, taking into account distant retinotopic regions as well? Using simple global structures made up of four Gabor elements, we first of all showed that such simplistic shapes failed to induce contextual feedback into the occluded region. However, in the presence of feedforward information, we saw that feedback from the local surround combined with identical feedforward input to give rise to different activity patterns in that feedforward region. This suggests that feedback may be recruited differentially depending on whether feedforward stimulation is present or absent. In a second experiment, in Chapter 4, using natural scenes, we tested whether contextual feedback can originate from a distant retinotopic region when the local scene surround was not informative. We manipulated the scene information in a distant retinotopic region (in the opposite hemisphere) while keeping the local neighbouring surround information the same. The results showed a lack of meaningful feedback in the occluded region, and feedback from the distant surround did not modulate identical feedforward information, in contrast to what we saw previously with the local surround. These findings suggest that feedback preferentially originates from nearby regions and provides context to disambiguate local feedforward elements. Therefore context about the global scene structure may arise from a series of local surround interactions.

Overall, the results of the thesis show that feedback contains both coarse and fine-grained information, is dependent on the amount of surrounding scene information available and preferentially transmits information from the nearby surrounding regions. Some overarching themes relevant to the work of this thesis are 1) the content of feedback in V1 – what does it represent, what features of the surround is it transmitting? 2) interaction of feedback and feedforward information - under normal visual processing, what is the relative contribution of feedback signals? and 3) where is the origin of the non-feedforward signals we observe in the occluded region? The following sections will discuss these themes in relation to the present findings and current literature. Towards the end of this chapter, I will review some limitations of the present data and will propose future studies for the remaining unanswered questions.

5.2 Content of feedback

We know that feedback carries information about the surrounding regions. However, the exact nature of this information is not entirely clear. Feedback has been implicated in various functions, from informing the neurons' receptive fields about nearby local elements, through to underlying higher level cognitive functions, such as memory, expectations and predictions. Therefore, it is likely that feedback has many roles. I will now discuss our results in relation to these proposed functions of feedback.

5.2.1 Eavesdropping on the neighbours

Feedback acts on the surround receptive field of neurons and thus provides the central receptive field with information about the nearby surrounding regions. This enables tasks such as perceptual grouping, figure-ground segregation and contour integration (e.g. Coen-Cagli, Kohn, & Schwartz, 2015; Hess & Field, 1999; Scholte, Jolij, Fahrenfort, & Lamme, 2008). On the one hand, on a single neuron level, these centre-surround interactions could be quite local in terms of their retinotopic extent, up to the amount that the far surround region extends. On the other hand, the size of the far surround increases with cortical distance from V1 (Angelucci & Bressloff, 2006) and hence feedback to even a single neuron may originate from a relatively large region of the scene. Effects of combining local elements into a global percept have been observed in V1 and this is proposed to arise from feedback originating from higher areas (Schmidt, Lomber, Payne, & Galuske, 2011; de-Wit, Kubilius, Wagemans, & Op de Beeck, 2012). How is this global percept achieved? It might be that V1 has direct access to the global scene representation from large surround receptive fields. Alternatively, a global percept may be achieved via several more local centre-surround interactions.

Our results are more compatible with the role of feedback in signalling about the local surrounding regions. In **Chapter 4** we found that feedback carries information about the local surrounding regions, and does not transmit information from distant retinotopic regions. The results of **Chapters 2** and **3** are also compatible with this finding. We saw informative activity patterns in the occluded region of the scene when the surrounding information was available close to the occluded region. When this information was reduced, in the conditions with the “bubbles”, we saw a reduction in contextual feedback. Our data suggest that feedback predominantly codes for information in the nearby surround. It is possible that global scene representation arises from a series of local interactions. Alternatively, it could be that the participants did not really perceive the global shapes in our experiment in **Chapter 4,** and the visual system treated them as a collection of local elements. In addition, the global scene structure in the mirror stimuli in **Chapter 4** was very similar. We showed that feedback does not directly transmit information about a distant retinotopic object, but our results are not necessarily incompatible with the role of feedback in global scene structure in other circumstances. For example, global scene context may be more evident in cases where the local information in a particular region is the same, but the overall spatial layout of the scene differs. Furthermore, Williams et al. (2008) showed that information about objects placed in the periphery could be decoded in the fovea. The occluded region in the mirror stimuli (Experiment 2, **Chapter 4**) was more peripheral, and perhaps this diminished contextual feedback relating to a peripheral object (in our case, the face in the mirror).

To summarise, it is highly unlikely that feedback is only involved in signalling about local surrounding information. This would not account for top-down modulation in cases where the feedforward information is identical, such as a different percept arising from task demands, effects of prior knowledge and expectations. Feedback is also involved in cases where feedforward stimulation is absent altogether, such as visual imagery. It could be that local surround interactions predominate and are a more automatic process, whereas global scene interpretation requires more cognitive effort. For example, in the famous image of the Dalmatian from Gregory (1970), at first the scene is parcellated automatically into many black and white patches, but prior knowledge and a willingness to identify a dog are required to interpret the global structure within those patches.

5.2.2 Blobs or edges?

On the surface, the global/local processes may seem to correspond to coarse/fine processing. However, it is important to note the distinction. For example, Oliva and Schyns (1997) suggest that global/local processing happens on the 2D image, while coarse/fine processing is orthogonal to the image plane, in a third dimension. In **Chapter 4**, we looked at global and local scene structure, in terms of which regions of the surround feedback transmits signals about. We were interested in which regions of the 2D image were important for eliciting feedback – is it only nearby local regions, or the global structure incorporating information from distant retinotopic regions as well? In **Chapter 2**, we looked at the spatial scale of the information (fine-grained or coarse) that feedback might contain, without concentrating on whether the global image structure is taken into account or if only specific retinotopic regions elicit feedback. Although some authors have suggested that LSF must be implicated if feedback is transmitting information about the global image structure (Bar et al., 2006), this might not necessarily be the case. For example, Walther, Chai, Caddigan, Beck and Fei-Fei (2011) have suggested that global scene structure can be extracted from fine-grained information. In their study this corresponded to line drawings. This means that both HSF and LSF information could be important for providing scene context.

Some studies have suggested that feedback information is less fine-grained than feedforward representations. For example, Vetter, Smith and Muckli (2014) found a similarity for feedback activity patterns within a sound category, when blindfolded participants listened to different auditory stimuli. They suggested that feedback information is rather abstract and not pictorial (as actual perception presumably would be). Muckli et al. (2015) showed that feedback information was coarse enough to withstand shifts of up to 2° of visual angle in the image surround. When trained on the original image, the decoding of a 2° shifted image was above chance. Feedforward information is more sensitive to this shift (Lucy Petro, personal communication), therefore suggesting it is more fine-grained. Feedback is likely to be coarser because of the larger RFs in the higher visual areas, which is where the information is transmitted from. A recent study which aimed to measure imagery receptive fields – that is a region of the mental visual field in which imagined stimuli evoke real brain activity – found that these were on average larger and more scattered than the corresponding visual RFs (Breedlove, St-Yves, Olman, & Naselaris, 2016). Since visual imagery is a top-down process, this suggests that feedback information is likely to be coarser than bottom-up stimulation.

In **Chapter 2** we found that both the HSF and LSF surrounds induced informative feedback signals. This may suggest that feedback contains both coarse and fine-grained information. However, it is difficult to know for certain from the current data, how different the feedback information in the HSF and LSF stimuli was and how coarse the feedback representations are in relation to the corresponding feedforward input. We could decode between HSF and LSF feedback, suggesting there is some difference between the two. However, the spatial scale of these representations may be shifted to be coarser, in comparison to the feedforward HSF and LSF. With the paradigm that we used, it is not possible to tell what precise SF band feedback represents. Comparing feedback to feedforward activity was not useful since we found (**Chapter 3**) that this was driven by the similarities in the scene surround, and the shared surround feedback in the stimuli. Perhaps studies of imagery and working memory could shed more light on how fine-grained feedback information could be, since visual imagery has been found to elicit similar activity as actual perception (e.g. Albers, Kok, Toni, Dijkerman, & de Lange, 2013).

Since we find that the HSF image surrounds could induce meaningful information in the occluded region, we can at least say that feedback transmits information *about* fine-grained image features, even if we cannot say how coarse this transmitted information is. At first glance, this speaks against previous suggestions that contextual scene information stems predominantly from LSF (Bar et al., 2006). However, our finding might not be completely incompatible. First of all, we used filtered images, where the image surround was only informative in one SF band. When the LSF information about the scene was unavailable, it is possible that HSF was used instead. Schyns & Oliva (1997) suggest that subjects use the most informative scale that is applicable for a task, be it LSF or HSF. Secondly, the dominance of LSF stems mostly from the finding that LSF is processed faster than HSF, and hence is able to be processed quickly enough to be useful for biasing interpretations of bottom-up input. Due to the low temporal resolution of fMRI, this advantage should not be evident in our studies. Indeed Schyns and Oliva (1994) have shown that with longer exposure, HSF stimuli become more important.

There is evidence for feedback being both precise (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Murray, Kersten, Olshausen, Schrater, & Woods, 2002) and diffuse (Muckli et al., 2005; de-Wit et al., 2012). Although, precision is not exactly the same as spatial scale, this nevertheless suggests that feedback may transmit coarse, as well as more fine-grained, more precise information. The more coarse feedback could gain back some precision by combining with lateral connectivity (Muckli & Petro, 2013).

5.2.3 Completing the jigsaw?

In the natural environment occlusion is ubiquitous. Despite this, we often do not notice objects and scenes as being incomplete. Is one role of feedback to “fill-in” the missing visual information, in order to give us a more coherent representation of the world? Theories of predictive coding propose that the brain has internal models of the world based on previous experience (Clark, 2013; Friston, 2010; Rao & Ballard, 1999). These predictions and expectations about incoming input, based on internal models, are sent back to V1 to be compared with the current sensory stimulation (Bastos et al. 2012). Any mismatch is sent to higher areas as a prediction error to update the internal models. Are these internally generated models represented in V1 in the same way as actual bottom-up stimulation?

Several previous studies have suggested that V1 represents the expected, occluded or imagined object, and this activity is similar to the actual bottom-up stimulation (Albers et al., 2013; Ban et al., 2013; Gavornik & Bear, 2014; Kok, Failing, & de Lange, 2014). Using fMRI, Kok et al. showed that prior expectation of a specific visual stimulus evokes a pattern of activity in V1 which is similar to the activity elicited by the corresponding bottom-up stimulus. When an expected stimulus was omitted, stronger BOLD activation was seen in voxels which had a preference for that stimulus, suggesting a similarity between activity for expected but omitted stimuli and the ones actually presented. Gavornik and Bear also found a similar result in mice. When a stimulus from a trained spatiotemporal sequence was omitted, V1 nevertheless showed activity commensurate with the full sequence. Ban et al. showed that a target behind an occluder evoked a similar response in V1 compared to a corresponding non-occluded target. Albers et al.

demonstrated that working memory, imagery and actual stimulus representations were similar in early visual areas. A classifier could be trained on one type of representation and successfully tested on the other. However, it has also been suggested that imagery and bottom-up perception, although showing similarities, involve different dynamics in the ventral stream (Lee, Kravitz, & Baker, 2012). In addition, these authors showed that the representations of the different object stimuli were more similar during imagery than perception.

Using decoding methods, sensitive to the fine-grained pattern of activation in V1, we found that the activity patterns in the occluded region are not a direct representation of the missing feedforward information. When we presented the feedforward portion of the image in isolation, without the image surround, we found that it was not similar to the occluded region. We only found a feedback/feedforward similarity when using feedforward stimuli with a large amount of scene information in the surround. This is presumably because in the full feedforward image, a part of the information comes from the feedforward stimulation, and the other part from feedback from the surround. Hence the activity pattern in the completed scene is then similar to the occluded scene because both stimuli have surrounding feedback.

On the one hand, it makes sense that the representation would not be exactly the same, because participants do not report *seeing* the missing quadrant of the scene (i.e. they do not have a hallucination). So even if feedback is coding for something specific to the image, the representation does not have to exactly match the activity evoked by bottom-up stimulation (Lee et al., 2012). It may also be that information from feedback is coarser because of the larger receptive fields of higher areas, or is less precise retinotopically. For example, the Vetter et al. (2014) study found that there was a similarity between the V1 representations of the auditory stimuli within a particular category, suggesting that the feedback representation was not fine-grained and pictorial, but perhaps something more abstract. The representation in terms of the pattern of activity might be different also because feedback and feedforward signals project to different cortical layers (Muckli et al., 2015; Rockland & Pandya, 1979). Muckli et al. found using high resolution fMRI that during normal visual stimulation, feedforward information peaks in mid-layers, while feedback peaks in superficial layers. This anatomical difference is unlikely to be a major factor in our studies, as we used a 3T MRI scanner and hence we had a relatively coarse resolution. However, it is something to bear in mind.

Another reason we find a dissimilarity between feedback and feedforward information could be because using decoding methods allowed us to compare the fine-grained pattern of information, rather than just overall activation, as some of the other studies did (Kok, Failing, & de Lange, 2014; Ban et al., 2013). Albers et al. (2013) used a decoding method and did find a similarity between top-down generated information and bottom-up stimulation. However, the crucial difference to our studies might be that Albers et al. specifically measured imagery and working memory. In other words, these were active conditions where participants had to generate and keep a visual representation in their mind. In our studies participants were passively viewing the stimuli, and were not instructed to mentally fill-in the missing portion of the occluded region, even though some participants may have unintentionally imagined the missing quadrant. However, it is possible that more active imagery generation may be important for eliciting V1 representations which are similar to the feedforward input. Other work in our lab (Paton, Petro, & Muckli, 2016) has built on the Vetter et al. (2014) study, and showed that V1 activity for sounds was similar to that of feedforward visual stimulation with the corresponding visual scene, suggesting that perhaps the auditory stimulus triggered involuntary visual imagery. In addition, Vetter et al. showed that a cognitively demanding visuospatial task disrupted these representations in V1, further suggesting that visual imagery may be involved. Interestingly, Paton et al. showed that there was no similarity between activity patterns for the sound and those in the occluded region of a scene (feedback). This further highlights that the activity patterns we recorded in the occluded region were different from visual imagery, and for this reason did not closely match bottom-up stimulation.

If expectations and prior knowledge evoked corresponding activity in V1 of the expected stimulus, we might have seen activity patterns corresponding to the predicted Gabor orientation corresponding to the global structure of either Diamond or X (Experiment 1, **Chapter 4**). However, this was not the case. We also did not see any informative activity patterns in the occluded region of the mirror stimuli. One explanation could be that the participants did not have a strong expectation to see something in the occluded region. For example, they might not have perceived the stimulus as occluded, but as “three Gabors with a white square”. If the fourth Gabor was perceived as missing rather than occluded, this could have changed the participant’s expectations about the lower right quadrant. Johnson and Olshausen (2005) suggest that occlusion and deletion modulate perception in slightly different ways. Occluded objects are recognised more accurately than objects with the same parts simply deleted. The Gabor stimuli may have been too artificial to evoke any strong expectation of scene structure behind the occluder, compared to natural scene stimuli which may evoke stronger contextual associations (Bar 2004). In Experiment 2, **Chapter 4,** where we used the mirror stimuli, the task was to report the identity of the face in the mirror, which does not necessarily require forming expectations about the identity of the person under the occluder. In the incongruent feedforward conditions, there should have been a violation of expectations, compared to the congruent feedforward conditions. However, it is possible that participants did not perceive the incongruent condition as unexpected enough. Each condition in the experiment was presented an equal number of times, and so participants came to expect the incongruent stimulus just as much as the congruent one. On the one hand, it is unexpected in terms of real life perception; on the other hand, it is fully expected as part of an experiment, which is in itself an artificial setting. It might be easier to find effects of expectation in predictable spatiotemporal sequences, such as A, B, C … (D), or when expecting a particular object in a natural scene is somehow behaviourally relevant.

Overall, we find that our results are more compatible with the notion that feedback is mainly transmitting information about image statistics, rather than something higher-level or more semantic. We found similarities between feedback signals of the same spatial scale, across different scenes. In other words, HSF feedback from the car scene was similar to HSF feedback from the people scene. This means feedback contains information about the spatial scale that is not specific to a particular scene. In addition, we did not find that feedback represents a direct filling-in of the feedforward information, as theories of predictive coding might postulate. For example, the orientation compatible with the global shape made up of Gabors was not filled-in. In the mirror stimuli, the person in front of the mirror was also not filled in. This may be surprising if we suppose that feedback is transmitting knowledge about the scene or some semantic information, but less unexpected if feedback is more concerned with coding for image statistics. Further arguing against prior knowledge and expectations playing a large role in the information we decode from the occluded region, we found that increased exposure to the full scene did not increase meaningful feedback (**Chapter 3**). It appears that feedback information depends more on the features of the image available on each trial, and perhaps natural scene statistics in general. However, familiarity with a specific scene is not required. Indeed if a large role of feedback is to transmit information from the (predominantly local) surround, predictions about the precise structure of the missing scene quadrant are less necessary.

5.3 Where is the information in the occluded region coming from?

Now that we have established what the activity patterns in the occluded region might be representing, the next question of interest is regarding the origin of this information. We can be sure that there is no meaningful feedforward stimulation in the occluded area. However, the information that we record there could originate from both lateral and feedback connections. Although it is likely that lateral connections do have a role, we would argue that they are not sufficient to explain the effect. Lateral connections span a smaller part of a neuron's RF compared to feedback connections (Angelucci & Bressloff, 2006), and are not sufficient to account for the full range of surround modulatory response (Angelucci et al., 2002). Lateral connections are about 2.3 times larger than the classical RF (Sceniak, Ringach, Hawken, & Shapley, 1999) and can project to around 2° from the classical RF in V1 at eccentricities of 2-6° (Stettler, Das, Bennett, & Gilbert, 2002). Since our scene surround spanned a large region of the visual field, we would have stimulated the far surround RF of neurons in the occluded region, as well as possibly the near surround of some neurons very close to the quadrant border. Therefore, it is highly likely our contextual surround effects would come mostly from feedback, as well as some minimal influence directly from lateral connectivity. Nevertheless, we tried to minimize the selection of voxels close to the occluded region border in various ways. First of all, we reduced the number of voxels in the 1° region directly next to the border, as measured diagonally from the centre, by choosing our region of interest as responding higher to the *Target* mapping checkerboard than the *Surround* mapping checkerboard (which was 1° wide, diagonally). This should have ideally eliminated any "spillover" voxels receiving feedforward stimulation from the surround and minimized the influence of lateral spread. We also repeated our analysis with population receptive field mapping (pRF) for subjects for whom both eccentricity and polar angle retinotopic mapping were available, where we removed any voxels whose estimated RF was not fully within the occluded quadrant. This technique is likely to be more stringent than the *Target-Surround* mapping, as fewer voxels remained for analysis. Although the classifier performance did reduce somewhat, we still saw the same pattern of results. This suggests that any unintentional selection of voxels responding to the feedforward surround that we might have not accounted for is unlikely to fully explain the observed result. In **Chapter 3** we found an intriguing result that the activity in the lower right quadrant of the full scene seemed to be as similar to the occluded region with the full surround, as it was to actual feedforward information without the surround. It seems unlikely that a few "stray" voxels from the surround would contain just as much information as a feedforward quadrant about the corresponding full feedforward scene. We also found some modulation from the surround in conditions where we had reduced feedforward stimulation directly next to the quadrant (“bubbled” conditions), thus arguing against a “spillover” effect and making the result solely due to lateral connections unlikely.

Other studies have shown that information in the occluded region is likely to stem from the far surround RFs. Smith and Muckli (2010) showed that the most informative voxels were correlated more with the inner region of the occluded area, rather than the border region. Shushruth (2011) used a more precise method of electrophysiology and recorded local field potentials (LFPs) in monkey V1, while the animals viewed partially occluded natural scenes. Similar to our findings, the author found informative activity patterns in the occluded region. Each LFP represented the activity of around 0.25° of visual angle, while the border of the occluded region was at least 3° away from the recorded RF. Therefore, it is highly likely that this result was due to the modulatory signals from the far surround, rather than feedforward information at the border or lateral activation. This study also showed a very similar pattern of results in comparison with our own findings. For example, when trained on the full scene, classifier performance was higher for testing on the occluded region than on an isolated feedforward part of the scene (similar to our ¼ condition). This provides support to our finding that feedback from the surround is an important component of the information in the full scene. In addition, there was a low similarity between the occluded region and the isolated feedforward section, mirroring the results presented in this thesis. Information in the occluded regions can also be modulated by prior knowledge (Ban et al., 2013), arguing against a simple spread of lateral activity from the feedforward information near the border of the occluded region.

We found that surrounds closer to the occluded region influence the activity patterns the most, and scene surround information in the opposite visual field did not have an effect. Therefore we cannot fully rule out the contribution of lateral connections in our studies, although this result can also be compatible with contributions of feedback. Most surround RFs would have been near the occluded region since we were measuring populations of neurons within the occluded region, with a few very large surround RFs perhaps spanning most of the scene surround. Therefore, the finding that neighbouring surround information is important, does not speak against the role of feedback. In addition, since previous studies from other researchers have suggested that top-down feedback is needed to account for similar findings, we would argue that feedback from higher areas is highly likely to be involved. Feedback may act on lateral connections which in turn act on the classical RFs of neurons (Muckli & Petro, 2013), so lateral connections are likely to be important as well.

If the information in the occluded region is transmitted via feedback from higher areas, then the question arises of which higher cortical areas contribute to the information in the occluded region in our case? V1 receives inputs from many visual and non-visual areas (Muckli & Petro, 2013). We cannot tell from our paradigm where the feedback signals are coming from. However, I would speculate that the majority of feedback would originate from other early visual areas, such V2 and V3, as well as scene processing regions like the parahippocampal area (PPA) and the retrosplenial cortex (RSC), which have been shown to code for spatial, as well as non-spatial context (Bar & Aminoff, 2003).

5.4 How much does feedback contribute to normal visual processing?

Our results are compatible with the many studies showing that feedback signals combine with the feedforward input (e.g. Hupé et al., 1998; Sandell & Schiller, 1982; Schmidt et al., 2011). We found that the activity pattern for otherwise identical feedforward stimulation is modulated by information in the surround (**Chapters 3** and **4**). This means that feedforward input alone is not enough to fully explain the response in a given retinotopic region. The activity pattern for the feedforward quadrant of the full scene is a combination of feedforward input and feedback from the surround. Haslinger et al. (2012) showed that under natural scene viewing, the surround, spike history, and LFPs account for almost as much of the V1 response as the classical RF, suggesting that feedforward stimulation alone cannot fully account for the activity in V1.

It is difficult to quantify precisely how much feedback contributes to explaining the neural activity under normal visual processing. However, feedback must contribute a significant amount, as non-feedforward inputs account for most of the connections in the visual cortex and models mostly based on feedforward processing are unable to account for the majority of the response variance (Carandini et al., 2005). Many authors suggest that inputs from feedback are necessary for fully-fledged visual processing (Angelucci & Bressloff, 2006; Angelucci & Bullier, 2003; Kayser, Körding, & König, 2004; Muckli, 2010). The brain may be better thought of as a parallel processor (Singer, 2013) and with both feedback and feedforward streams being (equally?) important. In fact, the function of the brain may be to mostly process internal input, while occasionally sampling information from the external world (Bullier, 2006; Muckli, 2010). Predictive coding theories, for example, might propose such a hypothesis. The important computations are the internal models, and the external stimulus is only needed to verify or refute the models. Neurons in the brain may be coding an error or a deviation from the expected input, rather than a specific piece of content (Rao & Sejnowski, 2002). For example, Egner, Monti and Summerfield (2010) found that the fusiform face area (FFA) also responds strongly to houses as well as faces, when the expectation of seeing a face is high. The authors suggest that feature expectation and surprise, rather than actual input per se, are more important factors in explaining the response of that cortical region. Dopaminergic neurons in the striatum have also been shown to code for unexpected information about the value of an upcoming reward, rather than the properties of the present stimulus (Koster-Hale & Saxe, 2013). The difficulty in assessing the relative contribution of feedback vs feedforward signals is also due to the case that how much feedback contributes to the response might not be set in stone. The influence of feedback signals may be gated by attention (Clark, 2013) and subcortical mechanisms (Pafundo, Nicholas, Zhang, & Kuhlman, 2016).

In **Chapter 4** we found an influence of feedback from the surround when feedforward information is available, but not when the quadrant is occluded and meaningful feedforward stimulation is missing. Although the role of context in modulating neuronal response has been demonstrated in the absence of feedforward input, an important role for feedback is acting on the surround receptive field to modulate the response to stimuli presented in the classical receptive field. In the mechanism of apical amplification (Phillips, 2015), when apical and basal inputs coincide (corresponding to the feedback and feedforward inputs respectively), the cell’s response to its basal inputs is amplified. This suggests that the presence of feedforward input may be important for contextual modulation to take place, at least in some circumstances. For example, in our stimuli in **Chapter 4**, feedback from the surround may have been important in disambiguating which wider structure the identical feedforward orientation belongs to, or signalling about the neighbouring orientations – important for detecting contours and identifying which local elements belong to the same object. When the feedforward information in the quadrant was missing, feedback from the surround may have been less useful, as there was nothing to disambiguate.

5.5 Future directions

In the following section, I will briefly review some limitations of the studies we have conducted so far and will propose potential future experiments which could investigate some of the remaining questions.

5.5.1 Methodology

There are some aspects of the methods and the paradigm we used which limit what we can conclude from the data. First of all, using fMRI allowed us to have a good spatial resolution and record activity patterns from specific retinotopic regions. However, due to the low temporal resolution, we are not able to say much about the timecourse of the contextual feedback signals. They are likely to be in the range of hundreds of milliseconds, which we cannot differentiate with fMRI. It would be interesting to see in future studies how fast the contextual feedback from the surround propagates to the occluded region. Feedback connections are faster than lateral (Girard, Hupé, & Bullier, 2001), so it would be interesting to investigate the timescale of the effects, are they rapid and more commensurate with feedback or slow and more in line with lateral signals? Perhaps both initial fast effects and later slower effects would be seen.

MVPA is a useful method for examining the pattern of brain activity in a given region of interest, compared with looking at overall activation or deactivation which may miss subtle differences between conditions due to averaging over many voxels. However, a limitation of using decoding methods is that we can often only say that there is a difference between the conditions and it is difficult to determine what that difference stems from or what the activity patterns represent about the stimulus. For example, in **Chapter 2**, we could decode a difference between the HSF and LSF occluded stimuli. However, we cannot say with certainty if the information in the occluded region of the HSF scene represented high spatial frequencies in the same SF band as the scene surround. For example, the information may have been coarser for the HSF stimulus and more fine-grained for the LSF one (or vice versa). However, as long as there was still a difference between the two types of stimuli, we would have been able to decode them. An extreme example would be that HSF and LSF surround elicit LSF and HSF feedback, respectively. This is highly unlikely; however, it illustrates the point that being able to differentiate two conditions does not tell us exactly which stimuli features the activity pattern in each condition represents.

One way to probe the nature of the content is to compare one set of conditions to another set of conditions. For example, we can train the classifier on the feedback conditions and test whether it can use the same information to decode the stimuli in the feedforward conditions. If the classifier can do this successfully, we can say that there is a similarity between the stimuli representation for feedback and feedforward conditions. However, in our studies we did not find a generalisation between feedback and feedforward conditions. From this, we can conclude that the representation in the occluded region is dissimilar to that of the corresponding actual stimulation, but still leaves us somewhat in the dark about what the occluded activity pattern represents. To bring us closer to being able to reconstruct the occluded representations, future work should compare the occluded activity patterns to a wider range of feedforward features. In addition, future experiments could use encoding methods (as opposed to decoding). Encoding and decoding are complementary methods, but the direction of the operations is the opposite. Encoding methods use stimuli to predict activity patterns and vice versa for decoding. For example, the activity patterns of several feedforward stimuli could be used to predict the activity patterns in the occluded region. Although decoding methods can be very useful, only encoding methods are able to obtain a complete description of the features represented in a ROI (Naselaris, Kay, Nishimoto, & Gallant, 2011).

5.5.2 Spatial frequency and feedback

There are several unanswered questions about how the spatial frequency of the surround affects information in the occluded region, such as hemisphere differences, eccentricity differences and task constraints. For example, our occluded quadrant was always in the right visual field. However, some studies have suggested hemisphere differences in spatial frequency processing (Peyrin, Mermillod, Chokron, & Marendaz, 2006; Peyrin et al., 2005; Peyrin, Chauvin, Chokron, & Marendaz, 2003; Peyrin, Chokron, et al., 2006; dos Santos, Andrade, & Fernandez Calvo, 2013). These studies found that the left hemisphere preferentially deals with HSFs, while the right hemisphere processes LSFs. Future studies can investigate whether this association exists in the feedback from higher areas as well. For example, can we decode LSF stimuli better in the right hemisphere compared to the left?

Another potential follow-up experiment concerns how the information in the occluded region varies by eccentricity. Neurons in the fovea have a preference for higher SFs compared to cells in the periphery (Xu, Anderson, & Casagrande, 2007). Future studies could investigate whether HSF feedback is stronger closer to the fovea and feedback concerning LSF is predominantly in the periphery. Does feedback follow the same pattern as feedforward processing or is the information from feedback more uniform?

SF processing can also be modulated by task constraints or by sensitization to a particular SF (Schyns & Oliva, 1997, 1999; Sowden, Özgen, Schyns, & Daoutis, 2003; Özgen, Sowden, Schyns, & Daoutis, 2005). Future studies should investigate whether contextual feedback about SF to the occluded region can also be modulated by task demands or whether the SF feedback represents more automatic processes which cannot be influenced much by top-down cognitive states. For example, future work could use a hybrid stimulus containing one scene in LSF superimposed with another scene in HSF. Participants could be made to attend to a specific SF band, either via cueing or sensitization, for example. We could then measure whether the information in the occluded region of the hybrid scene is more similar to one SF or another depending on task demands by cross-classifying to non-hybrid stimuli presented in only one SF band. The feedforward information in the surround would be exactly the same, so any potential difference can be attributed to top-down processes, rather than lateral spread of activation or any “spillover” from the feedforward surround voxels.

5.5.3 Global vs local effects

In the experiments reported in this thesis we have recorded brain activity of V1 during passive viewing. However, we did not establish what participants perceived in these stimuli and whether the contextual surround actually elicited any perceptual differences which could be measured in a behavioural study. For example, in Experiment 1 of **Chapter 4**, participants may simply not have had a strong perception of the global shape, and hence we did not observe contextual feedback about the global scene.

Stimuli embedded in a congruent contextual surround are recognised faster and more accurately (e.g. Palmer 1975). This means that it may be easier to accurately report the orientation of the Gabor in the lower right quadrant (in a non-occluded stimulus) that is compatible with the global structure, if the global structure indeed elicits context in our set of stimuli. A follow-up study could look at behavioural measures such as reaction times and accuracy of reporting the Gabor depending on whether it is part of a compatible global structure (Diamond or X) or not. Another measure could be the probability of reporting an orientation compatible with the global shape in conditions where an actual stimulus is not shown in the lower right quadrant (just the noise background). Participants would not be told about these "catch" trials and we could measure if there is a bias in reporting a globally congruent orientation. If the participants are just guessing, we should see an equal probability of reporting either orientation. However, top-down expectations are powerful enough to create percepts out of noise (Smith, Gosselin, & Schyns, 2012). If the stimuli that were used in **Chapter 4** do not elicit strong contextual effects behaviourally, we could find stimuli that do and then test those with our fMRI paradigm. At the moment, we cannot say whether the lack of differential information in the occluded region of V1 was due to decreased contextual feedback from the surround for these simplistic shapes or more so because participants did not perceive strong surrounding context in the first place.

Future experiments could also design better stimuli to elicit global scene structure context. The mirror stimuli used in Experiment 2 of **Chapter 4** might not have been optimal as some subjects reported subjectively perceiving the mirror as a window or a picture on the wall. Stimuli designed to elicit a strong contextual effect of the global surround, especially where it is highly relevant for interpreting the scene, might be better posed to investigate contextual feedback about the global image structure. For example, stimuli where the spatial layout or the navigational affordances are different may elicit more global contextual effects. In addition, follow-up experiments could look at the influence of surrounding feedback on the activity patterns representing the face in the mirror. There may be more modulation for the face stimulus, rather than for the person viewed from the back. The face is a more important stimulus, so the direction of modulation by feedback maybe reversed from the one we tested in our study.

5.5.4 Visual imagery

Feedback signals in the absence of feedforward stimulation may represent visual imagery and working memory in some paradigms (e.g. Albers et al., 2013; Harrison & Tong, 2009; Vetter et al., 2014), but not necessarily on other studies on occlusion (e.g. Ban et al., 2013; Sugita, 1999). However, it would be interesting to establish whether feedback in the occluded region is related to the same mechanisms as imagery/working memory, even when participants are not explicitly instructed to visualise and do not report vivid imagery in the occluded/non-stimulated regions. For example, feedback signals in the non-stimulated regions might lie on a continuum of imagery, from extremely weak to very vivid, or be part of an altogether different mechanism. Imagery strength should be modulated by instructing participants to either actively visualise the missing quadrant or just passively view the stimuli. If the activity patterns in the occluded region are related to imagery, we might observe better decoding in the imagery conditions, compared to passive viewing. In the passive viewing conditions it would also be interesting to correlate the strength of decoding in the occluded region with individual differences in visual imagery strength. For example, some people are better at visualising than others, and there are people who appear to lack the ability to produce visual imagery entirely (Zeman, Dewar, & Della Sala, 2015). Therefore, it would be useful to see if people who are naturally more prone to visualising would have more informative feedback in the occluded region.

5.5.5 Occlusion

Missing feedforward information in the natural world can come from various sources, such as, occlusion or strong shadow. In addition, some scene features may fall onto the retina in the position of the blind spot. Finally, some parts of the object may simply be missing. It is unclear if the contextual feedback in these non-stimulated regions would be similar or different depending on the reason behind the missing information. For example, Johnson and Olshausen (2005) showed that objects which are partially occluded are recognised more accurately compared to objects which are partially deleted in the same regions. Future studies can investigate if feedback is stronger in cases of occlusion vs deletion. Results from some previous studies may suggest that this would be the case, as they found that V1 activity was different depending on whether the stimulus was perceived as occluded or deleted (Ban et al., 2013; Sugita, 1999). In the former case, the occluded part of the stimulus was still represented in V1, while not in the latter case.

5.5.6 Contextual feedback in patients

In the experiments described in this work we have only tested healthy participants. However, our aim as cognitive neuroscientists is to understand the brain in both health and disease. Future studies could extend the paradigm to patient groups. For example, people who have schizophrenia have been shown to have deficits in predictive mechanisms (Clark, 2013), which is one of the functions of feedback. It would be useful to study how contextual feedback is affected when predictive mechanisms are impaired. Another potential patient group of interest could be patients with hemispatial neglect. These patients allocate their attention predominantly to the right and “neglect” their left visual field. To investigate how feedback is modulated by attention future studies could compare feedback in the occluded region in the left vs the right visual field in patients with hemispatial neglect.

5.6 General conclusions

In the studies described in this thesis we used a partial occlusion paradigm to investigate feedback in the absence of meaningful feedforward stimulation. We studied how the scene information in the image surround affects feedback signals to a particular retinotopic region of V1. We lend support to previous studies suggesting that information presented in the surround receptive field can modulate neuronal activity in the central receptive field, both in the presence and absence of bottom-up visual stimulation. In addition, we found that feedback combines with feedforward information during full visual processing. To build on the previous findings, we tested which aspects of the surround are important for eliciting these contextual effects. First of all, we showed that both coarse and fine-grained surrounds can elicit contextual feedback about the scene. Second, we demonstrated that the strength of the contextual feedback depends on the amount of scene information in the surround and does not appear to be modulated by the familiarity with the scene. Finally, we found that this feedback originates predominantly from the local surrounding regions and it does not send information about objects located in a distant retinotopic region. Together these results bring us closer to understanding how neuronal activity is modulated by contextual information from the surround receptive fields.

Appendix A - Supplementary Methods & Results for Chapter 2

Methods

Stimuli

Scenes

Twenty-three subjects saw the car and people scenes, while ten subjects saw the concert and New York images. Concert and New York images were 800 x 600 px, which corresponded to 32° x 24° visual angle. Two subjects saw the concert and New York images in Small 0.65/1.30 cpd group; three subjects saw them in Small 0.81/1.62 cpd group; two subjects saw them in Large 0.81/2.03 cpd group; and three subjects saw them in Large 0.97/2.43 cpd group. All the images used for all conditions are shown in **Figures S2.1 and S2.2**. SF cut-offs used for each pair and the overlap conditions are shown in graphical representation in **Figure S2.3**.

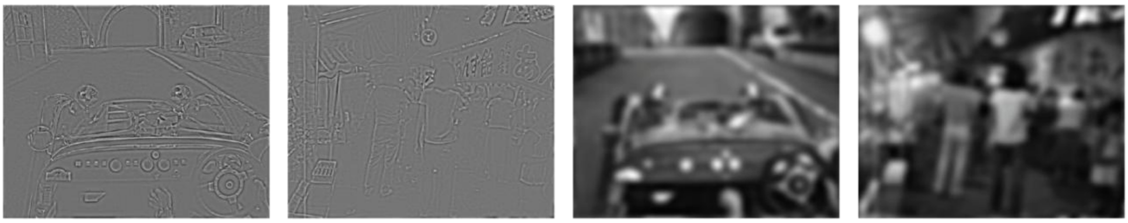
Occluded region mapping

The mapping stimuli used for the New York and concert stimuli are shown in **Figure S2.4**. The occluded region spanned 15.6° x 11.6° visual angle. Target mapping spanned 14.9° x 10.9°. Large surround spanned 15.6° x 11.6°, while Small Surround spanned 14.2° x 10.2°. The occluded region and the Large surround were presented 0.5° diagonally from the centre. Target mapping was presented 1.5° from centre, while Small surround was 2.5° from centre.

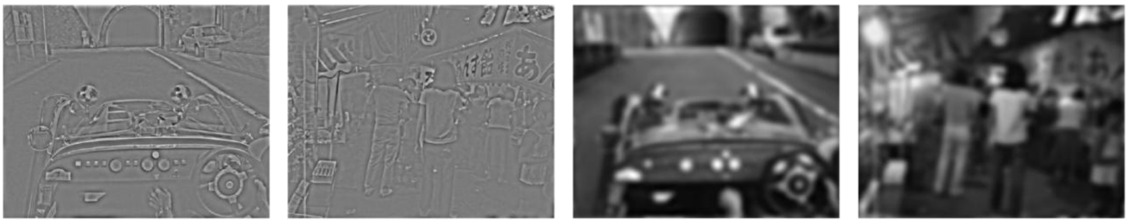
Task & Procedure

For subjects presented with the concert and New York images, in each 12s trial the stimulus was flashed on and off (200ms on/ 200ms off) 28 times (11.6s + variable fixation to account in uncertainty in timing).

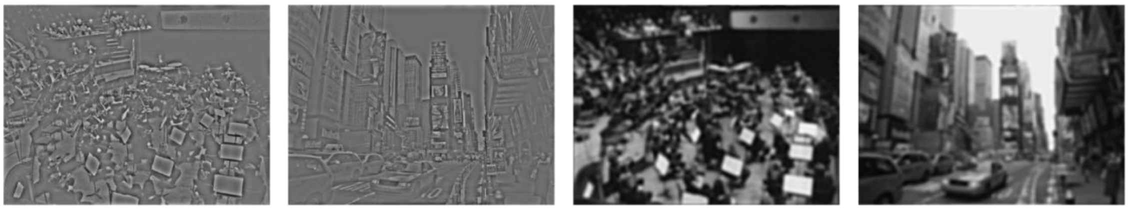
Gap



No overlap



Small 0.65/1.30 cpd



Small 0.81/1.62 cpd

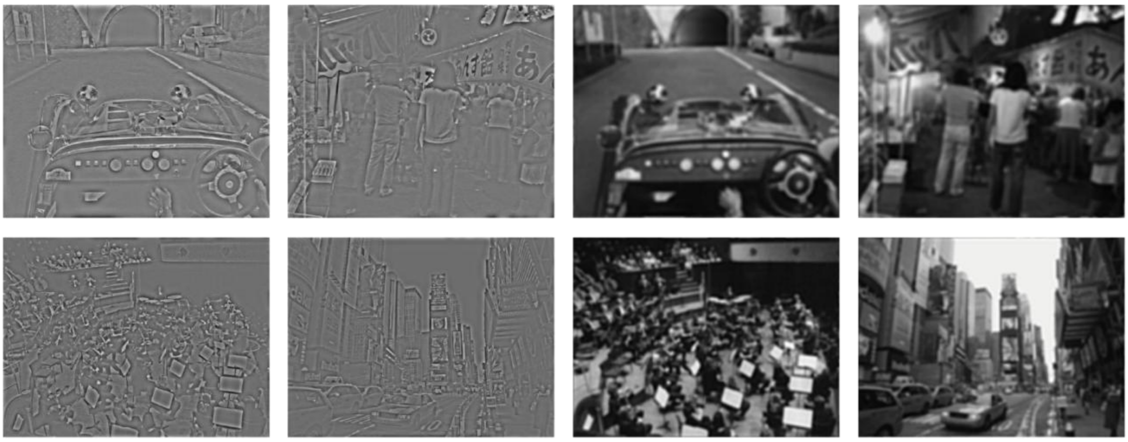


Figure S2.1 | Images used for Gap, No Overlap and Small Overlap groups.

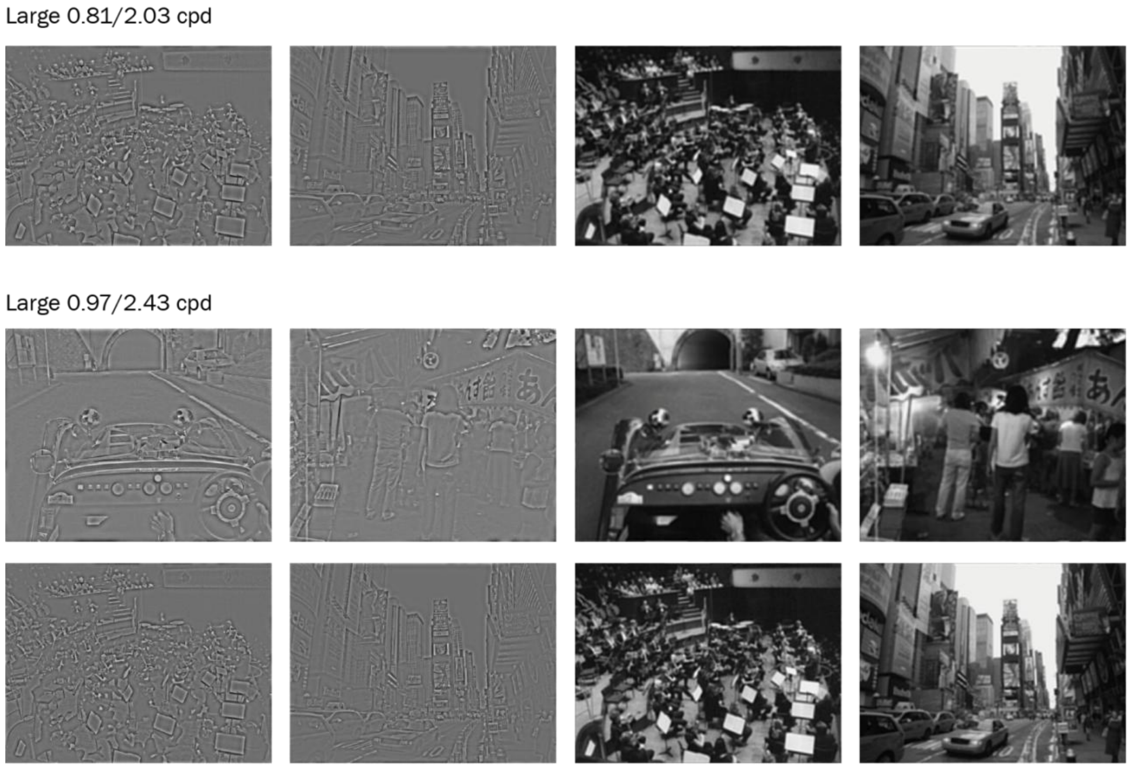


Figure S2.2 | Images used for Large Overlap groups.

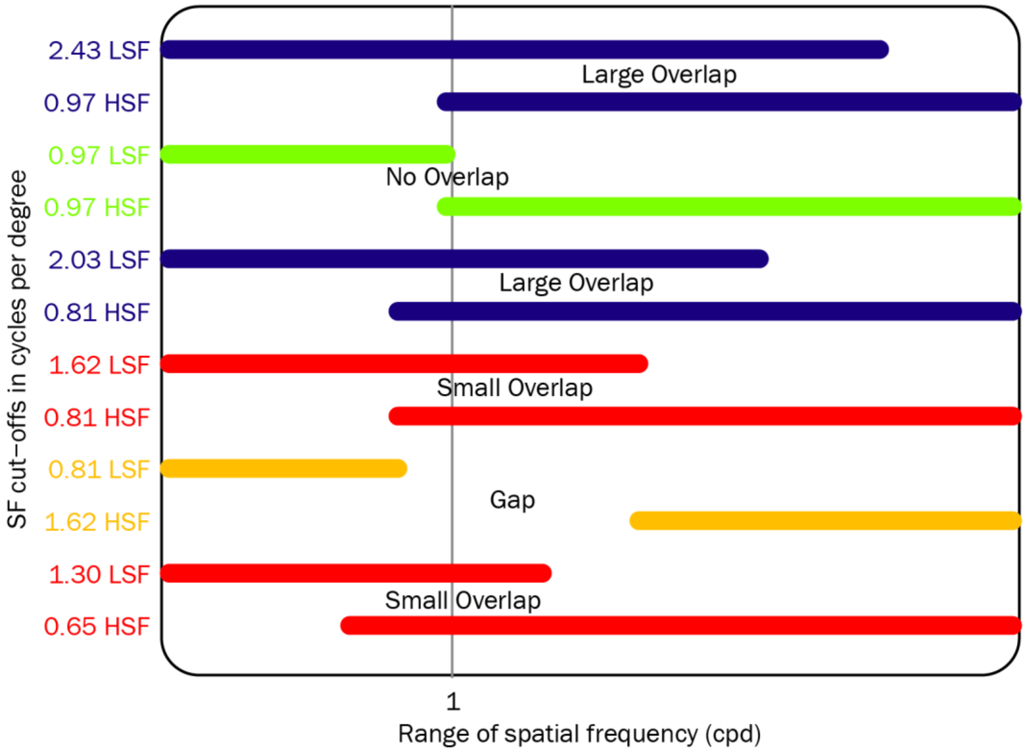


Figure S2.3 | SF cut-offs used for each pair of stimuli and the Overlap groups these corresponded to.

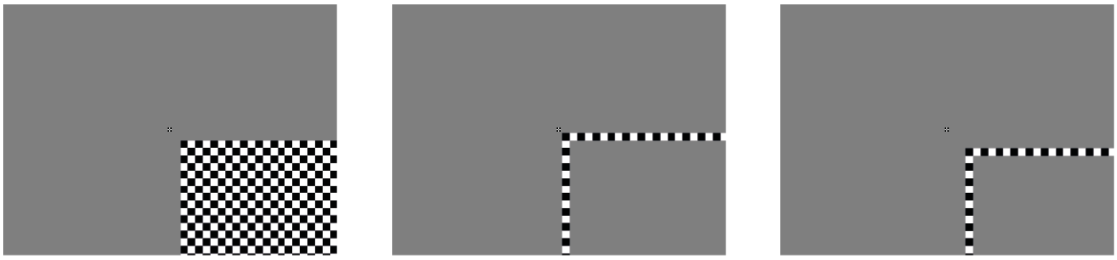


Figure S2.4 | Mapping stimuli used for subjects who saw the concert and New York images. *Target*, left; *Large Surround*, middle, *Small Surround*, right.

Results

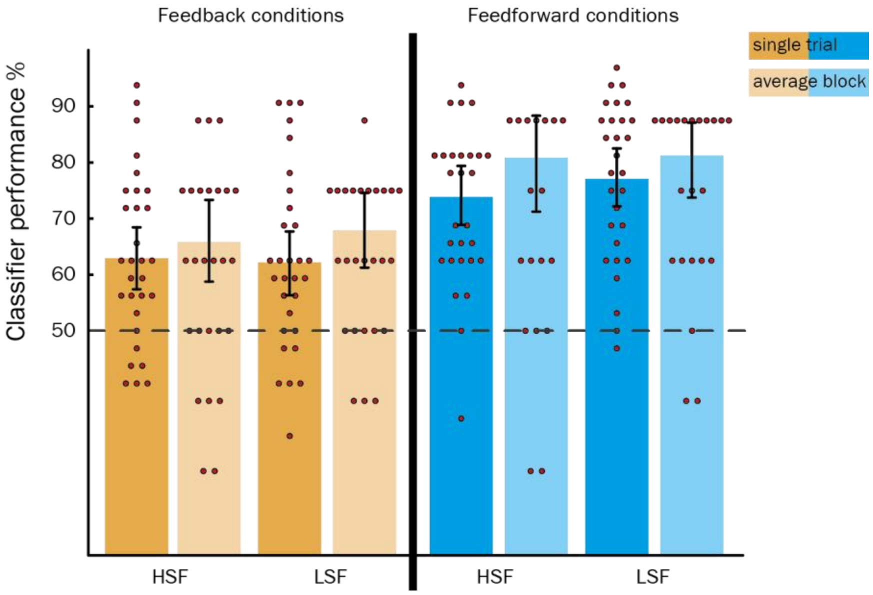


Figure S2.5 | Classification performance for decoding the two scenes in HSF and LSF conditions, for feedback and feedforward stimuli using both surround mappings for region of interest selection. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. A) Classifier performance for HSF and LSF conditions, collapsed over different SF cut-offs. N = 30, three subjects were removed from this analysis due to not having above threshold activation in V1 using the more conservative contrast.

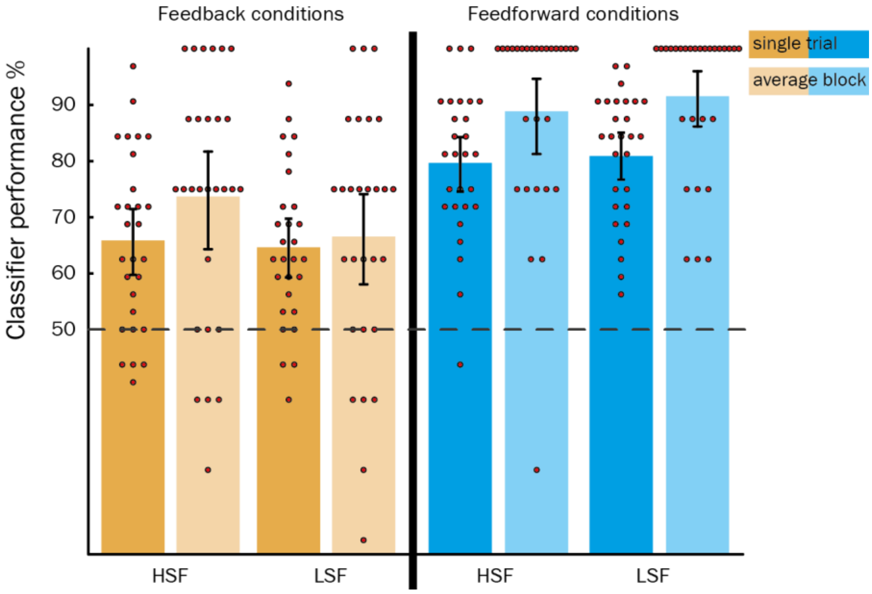


Figure S2.6 | Classification performance for decoding the two scenes in HSF and LSF conditions, for feedback and feedforward stimuli using only subjects who had at least one feedforward condition classifying above chance. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (1000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. A) Classifier performance for HSF and LSF conditions, collapsed over different SF cut-offs. N = 28.

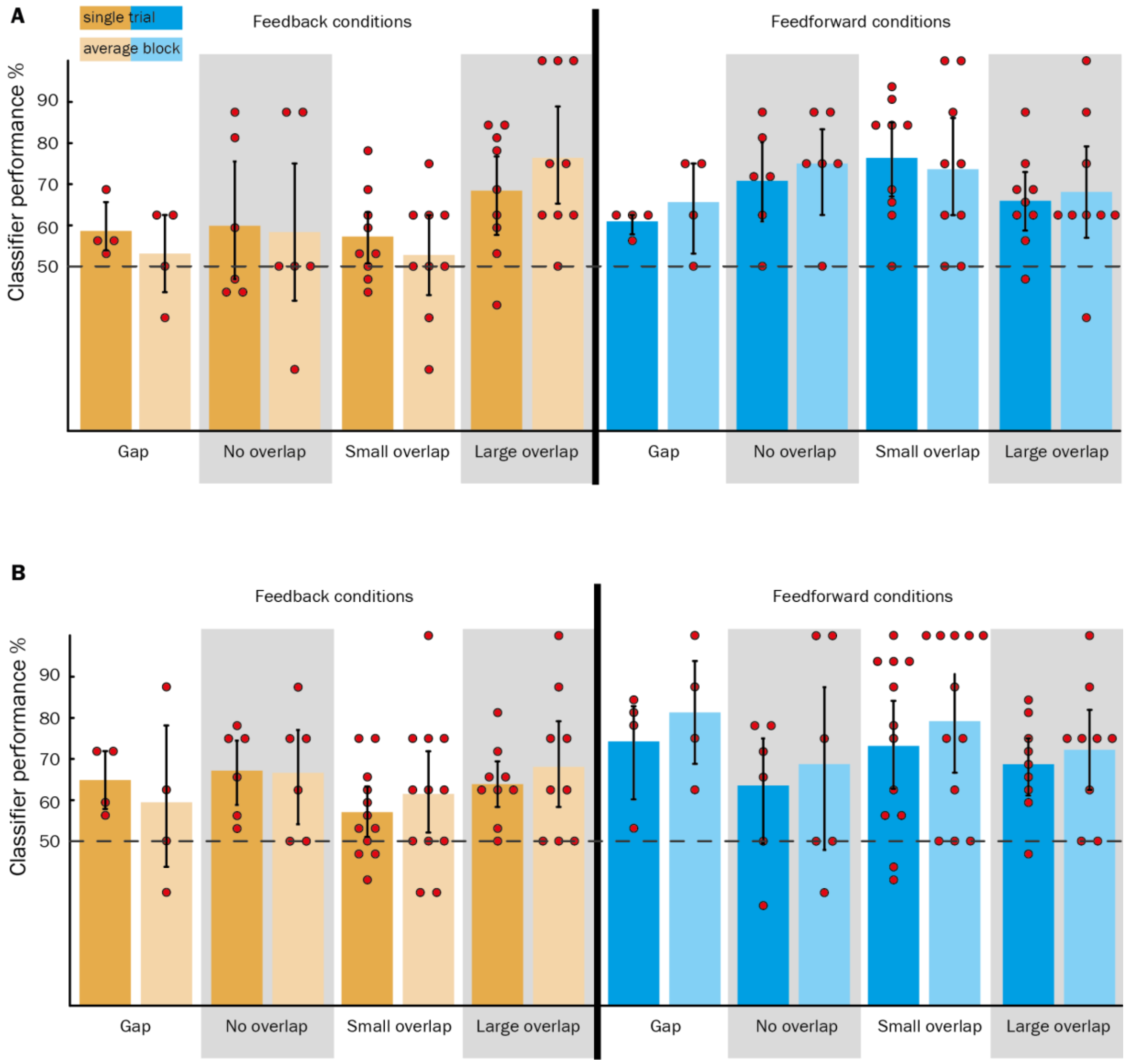


Figure S2.7 | Cross-classification performance for training to decode the two scenes in one SF and testing in the other, for different Overlap groups, using only subjects who had at least one feedforward condition classifying above chance. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean. Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Dark hues = classifier tested on single trials; light hues = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. Gap: n = 4; No Overlap, n = 6; Small Overlap, n = 9; Large Overlap, n = 9. A) Classifier performance for training on HSF and testing on LSF. B) Classifier performance for training on LSF and testing on HSF.

Appendix B – Supplementary Figure for Chapter 3

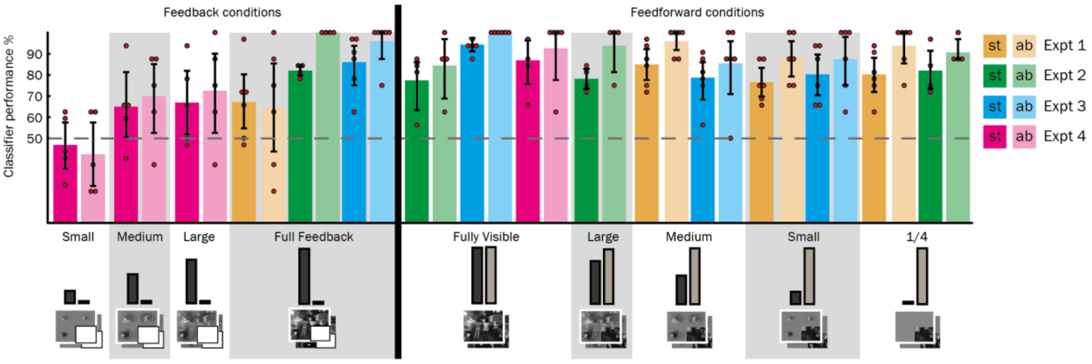


Figure S3.1 | Classification performance for decoding between the two scenes, with an extended safety boundary around the occluded region. We selected our ROI in BrainVoyager as the contrast of the Target mapping region being higher than both the Large Surround and the Small Surround mapping conditions. In addition, we selected voxels fitting the criteria of (*Target - Large Surround*) > 1 and (*Target - Small Surround*) > 1. Chance level is 50%. Lines represent 95% confidence intervals around the bootstrapped mean (10000 bootstrap samples of individual subjects' performances). Classifier performance is significantly above chance at α = 0.05 (not corrected for multiple comparisons) if the confidence intervals do not intersect with the chance line. Classifier performance for each condition, split by four experiments. ST (dark hues) = classifier tested on single trials; AB (light hues) = classifier tested on blocks of conditions averaged over the same type. The small red circles represent individual subjects' results. Expt 1, n = 6; Expt 2, n = 4; Expt 3, n = 6; Expt 4, n = 5.

References

Ahmed, B., Hanazawa, A., Undeman, C., Eriksson, D., Valentiniene, S., & Roland, P. E. (2008). Cortical dynamics subserving visual apparent motion. *Cerebral Cortex*, *18*(12), 2796–2810.

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.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of Neuroscience*, *30*(8), 2960–2966.

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.

Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *Journal of Physiology Paris*, *97*(2-3), 141–154.

Angelucci, A., Levitt, J. B., Walton, E. J., Hupé, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, *22*(19), 8633–8646.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*(3), 183–193.

Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: neuromagnetic and behavioral findings. *Frontiers in Human Neuroscience*, *7*(91), 1–11.

Ban, H., Yamamoto, H., Hanakawa, T., Urayama, S., Aso, T., Fukuyama, H., & Ejima, Y. (2013). Topographic representation of an occluded object and the effects of spatiotemporal context in human early visual areas. *The Journal of Neuroscience*, *33*(43), 16992–17007.

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

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

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

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D.L., Rosen, B.R. & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343–352.

Barlow, H. B. (1961). The Coding of Sensory Messages. In: *Current Problems in Animal Behaviour*, W.H. Thorpe and O. Zangwill (Eds.). Cambridge, UK: Cambridge University Press, 331–360.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron*, *76*(4), 695–711.

Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwadry, F., Carson, R. E., Varga, M., Baldwin, P., Belenky, G. Herscovitch, P. (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science*, *279*(5347), 91–95.

Breedlove, J., St-Yves, G., Olman, C., & Naselaris, T. (2016). Imagery receptive fields. *Journal of Vision*, *16*(12), 126, Abstract presented at the VSS conference, FL, USA.

Breitmeyer, B. G. (2014). Contributions of magno- and parvocellular channels to conscious and non-conscious vision. *Philosophical Transactions of the Royal Society B. Biological Sciences*, *369*(1641), 20130213.

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society B. Biological Sciences*, *265*(1400), 1037–1044.

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, *108*(27), 11262–11267.

Bullier, J. (2006). What Is Fed Back? In: *23 Problems in Systems Neuroscience*, J. L. van Hemmen and T. J. Sejnowski (Eds.). New York: Oxford University Press, 103-132.

Bullier, J., Hupé, J.-M., James, A. C., & Girard, P. (2001). The role of feedback connections in shaping the responses of visual cortical neurons. *Progress in Brain Research*, *134*, 193–204.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N.C. (2005). Do we know what the early visual system does? *The Journal of Neuroscience*, *25*(46), 10577–10597.

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, *22*(7), 622–626.

Ciaramelli, E., Leo, F., Del Viva, M. M., Burr, D. C., & Ladavas, E. (2007). The contribution of prefrontal cortex to global perception. *Experimental Brain Research*, *181*(3), 427–434.

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

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience*, *18*(11), 1648–1655.

Covic, E. N., & Sherman, S. M. (2011). Synaptic properties of connections between the primary and secondary auditory cortices in mice. *Cerebral Cortex*, *21*(11), 2425–2441.

Davis, E. T., & Graham, N. (1981). Spatial frequency uncertainty effects in the detection of sinusoidal gratings. *Vision Research*, *21*(5), 705–712.

De Pasquale, R., & Sherman, S. M. (2011). Synaptic properties of corticocortical connections between the primary and secondary visual cortical areas in the mouse. *The Journal of Neuroscience*, *31*(46), 16494–16506.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559.

de-Wit, L. H., Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2012). Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically predicted parts. *Journal of Vision*, *12*(11:12), 1-14.

dos Santos, N. A., Andrade, S. M., & Fernandez Calvo, B. (2013). Detection of spatial frequency in brain-damaged patients: influence of hemispheric asymmetries and hemineglect. *Frontiers in Human Neuroscience*, *7*(92), 1–6.

Douglas, R. J., & Martin, K. A. (2007). Mapping the matrix: the ways of neocortex. *Neuron*, *56*(2), 226–238.

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

Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by fMRI. *Cerebral Cortex*, *17*(9), 2123–2133.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*(49), 16601–16608..

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *The Journal of Neuroscience*, *29*(48), 15258–15265.

Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, *19*(9), 1488–1497.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47.

ffytche, D. H., Howard, R. J., Brammer, M. J., David, A., Woodruff, P., & Williams, S. (1998). The anatomy of conscious vision: an fMRI study of visual hallucinations. *Nature Neuroscience*, *1*(8), 738–742.

Flevaris, A., & Murray, S. (2015). Attention determines contextual enhancement versus suppression in human primary visual cortex. *Journal of Neuroscience*, *35*(35), 12273–12280.

Fries, P. (2015). Rhythms for cognition: communication through coherence. *Neuron*, *88*(1), 220–235.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Gavornik, J. P., & Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nature Neuroscience, 17*(5), 732–737.

Gilad, A., Meirovithz, E., & Slovin, H. (2013). Population responses to contour integration: early encoding of discrete elements and late perceptual grouping. *Neuron*, *78*(2), 389–402.

Gilad, A., Pesoa, Y., Ayzenshtat, I., & Slovin, H. (2014). Figure-ground processing during fixational saccades in V1: indication for higher-order stability. *The Journal of Neuroscience*, *34*(9), 3247–3252.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363.

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*(5), 677–696.

Girard, P., & Bullier, J. (1989). Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *Journal of Neurophysiology*, *62*(6), 1287–1302.

Girard, P., Hupé, J.-M., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology*, *85*(3), 1328–31.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception, 34*(1), 77–86.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*(21), 3537–3560.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Research*, *41*(17), 2261–2271.

Gregory, R. L. (1970). *The Intelligent Eye.* London: Weidenfeld and Nicolson.

Grosof, D., Shapley, R., & Hawken, M. (1993). Macaque V1 neurons can signal “illusory” contours. *Nature*, *365*(6446), 550–552.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Haslinger, R., Pipa, G., Lima, B., Singer, W., Brown, E. N., & Neuenschwander, S. (2012). Context matters: the illusive simplicity of macaque V1 receptive fields. *PLoS ONE*, *7*(7), e39699.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691.

Henriksson, L., Nurminen, L., Hyvärinen, A., & Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. *Journal of Vision*, *8*(10:5), 1–13.

Hess, R., & Field, D. (1999). Integration of contours: new insights. *Trends in Cognitive Sciences*, *3*(12), 480–486.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665–667.

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: an epistemological review. *Cognition*, *108*(3), 687–701.

Holmes, G. (1918). Disturbances of vision by cerebral lesions. *The British Journal of Ophthalmology*, *2*(7), 353–384.

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.

Hupé, J.-M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*(6695), 784–787.

Hupé, J.-M., James, A. C., Girard, P., Lomber, S. G., Payne, B. R., & Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *Journal of Neurophysiology*, *85*(1), 134–145.

Hübner, R. (1996). Specific effects of spatial-frequency uncertainty and different cue types on contrast detection: Data and models. *Vision Research*, *36*(21), 3429–3439.

Ichida, J. M., Schwabe, L., Bressloff, P. C., & Angelucci, A. (2005). Feedback-mediated facilitation and suppression from the receptive field surround of macaque V1 neurons. Program No. 820.4. Abstract presented at the SfN conference, Washington DC, USA.

Jehee, J. F., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *The Journal of Neuroscience*, *31*(22), 8210–8219.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*(1), 100–107.

Johnson, J. S., & Olshausen, B. A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Research*, *45*(25-26), 3262–3676.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–355.

Kayser, C., Körding, K., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current Opinion in Neurobiology*, *14*(4), 468–473.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554.

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.

Kok, P., & de Lange, F. P. (2014). Shape Perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Current Biology*, *24*, 1531–1535.

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, *79*(5), 836–848.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience*, *27*(48), 13232–13240.

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

Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., & Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science*, *325*(5941), 756–760.

Lee, J. H., Whittington, M. A., & Kopell, N. J. (2013). Top-down beta rhythms support selective attention via interlaminar interaction: a model. *PLoS Computational Biology*, *9*(8), e1003164.

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. *Journal of the Optical Society of America A. Optics, Image Science, and Vision*, *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.

Lerner, Y., Hendler, T., & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex*, *12*(2), 163–177.

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.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878.

Maus, G. W., & Nijhawan, R. (2008). Motion extrapolation into the blind spot. *Psychological Science*, *19*(11), 1087–1091.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron*, *89*(2), 384–397.

Mignard, M., & Malpeli, J. G. (1991). Paths of information flow through visual cortex. *Science*, *251*(4998), 1249–1251.

Morgan, A. T., Petro, L. S., & Muckli, L. (2016). Cortical feedback to V1 and V2 contains unique information about high-level scene structure. *bioRxiv, 041186.*

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*, 131–139.

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

Muckli, L., & Petro, L. S. (2013). Network interactions: non-geniculate input to V1. *Current Opinion in Neurobiology*, *23*(2), 195–201.

Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. *Biological Cybernetics*, *65*(2), 135–145.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109.

Murphy, P., & Sillito, A. (1987). Corticofugal feedback influences the generation of length tuning in the visual pathway. *Nature*, *329*(6141), 727–729.

Murray, S. O., Kersten, D., Olshausen, B., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *99*(23), 15164–15169.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.

Nassi, J. J., Lomber, S. G., & Born, R. T. (2013). Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *The Journal of Neuroscience*, *33*(19), 8504–8517.

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.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences*, *11*(12), 520–527.

Orpwood, R. (2013). Qualia could arise from information processing in local cortical networks. *Frontiers in Psychology*, *4*(121), 1–10.

Özgen, E., Sowden, P., Schyns, P., & Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition*, *12*(6), 925–937.

Pafundo, D. E., Nicholas, M. A., Zhang, R., & Kuhlman, S. J. (2016). Top-down-mediated facilitation in the visual cortex is gated by subcortical neuromodulation. *The Journal of Neuroscience, 36*(10), 2904–2914.

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

Panichello, M. F., Cheung, O. S., & Bar, M. (2013). Predictive feedback and conscious visual experience. *Frontiers in Psychology, 3*(620), 1–8.

Park, S., & Chun, M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, *47*(4), 1747–1756.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1448–1466.

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, *292*(5516), 510–512.

Paton, A., Petro, L., & Muckli, L. (2016). An investigation of sound content in early visual areas. *Journal of Vision*, *16*(12), 153, Abstract presented at the VSS conference, FL, USA.

Petro, L., & Muckli, L. (2016). The laminar integration of sensory inputs with feedback signals in human cortex. *Brain and Cognition*, *112*, 54-57.

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

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in Psychology*, *5*(1223), 1–8.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition, 53*(2), 278–282.

Peyrin, C., Chokron, S., Guyader, N., Gout, O., Moret, J., & Marendaz, C. (2006). Neural correlates of spatial frequency processing: A neuropsychological approach. *Brain Research*, *1073-1074*, 1–10.

Peyrin, C., Mermillod, M., Chokron, S., & Marendaz, C. (2006). Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. *Brain and Cognition*, *62*(3), 214–220.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage*, *28*(2), 464–473.

Phillips, W. A. (2015). Cognitive functions of intracellular mechanisms for contextual amplification. *Brain and Cognition*, in press.

Pratte, M. S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of Vision*, *14*(3:22), 1–12.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biology, 9*(4), e1000608.

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), 79–87.

Rao, R. P., & Sejnowski, T. J. (2002). Predictive coding, cortical feedback, and spike-timing dependent plasticity. In: *Probabilistic Models of the Brain: Perception and Neural Function*, R. P. N. Rao, B. A. Olshausen and M. S. Lewicki, (Eds.). Cambridge, MA: MIT Press, 297-315.

Rees, G. (2007). Neural correlates of the contents of visual awareness in humans. *Philosophical Transactions of the Royal Society B. Biological Sciences*, *362*(1481), 877–886.

Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C., & Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, *123*(8), 1624–1633.

Rennig, J., Karnath, H.-O., & Huberle, E. (2013). The role of size constancy for the integration of local elements into a global shape. *Frontiers in Human Neuroscience*, *7*(342), 1–8.

Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., & Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Current Biology, 13(12)*, 1038–1041.

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research*, *179*(1), 3–20.

Rockland, K. S., & Virga, A. (1989). Terminal arbors of individual “feedback” axons projecting from area V2 to V1 in the macaque monkey: a study using immunohistochemistry of anterogradely transported Phaseolus vulgaris-leucoagglutinin. *The Journal of Comparative Neurology*, *285*(1), 54–72.

Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, *56*(5), 785–792.

Sandell, J. H., & Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *Journal of Neurophysiology*, *48*(1), 38–48.

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast’s effect on spatial summation by macaque V1 neurons. *Nature Neuroscience, 2*(8), 733–739.

Schmidt, K., Lomber, S., Payne, B., & Galuske, R. (2011). Pattern motion representation in primary visual cortex is mediated by transcortical feedback. *NeuroImage*, *54*(1), 474–484.

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.

Schwiedrzik, C. M., Alink, A., Kohler, A., Singer, W., & Muckli, L. (2007). A spatio-temporal interaction on the apparent motion trace. *Vision Research*, *47*(28), 3424–3433.

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

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, *26*(8), 1027–1038.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–265.

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

Seth, A. K., Suzuki, K., & Critchley, H. D. (2012). An interoceptive predictive coding model of conscious presence. *Frontiers in Psychology*, *2*(395), 1–16.

Shushruth, S. (2011). *The role of extrastriate feedback in contextual computations in the primate primary visual cortex*. Doctoral Thesis. University of Utah.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of Neurophysiology*, *97*(1), 229–237.

Singer, W. (2013). Cortical dynamics revisited. *Trends in Cognitive Sciences*, *17*(12), 616–626.

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.

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

Sowden, P. T., Özgen, E., Schyns, P. G., & Daoutis, C. (2003). Expectancy effects on spatial frequency processing. *Vision Research*, *43*(26), 2759–2772.

Stettler, D., Das, A., Bennett, J., & Gilbert, C. (2002). Lateral Connectivity and Contextual Interactions in Macaque Primary Visual Cortex. *Neuron*, *36*(4), 739–750.

Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature*, *401*(6750), 269–272.

Tang, H., Buia, C., Madhavan, R., Crone, N. E., Madsen, J. R., Anderson, W. S., & Kreiman, G. (2014). Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron, 83*(3), 736–748.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience*, *4*(3), 219–229.

Tong, F. (2013). Imagery and visual working memory: one and the same? *Trends in Cognitive Sciences*, *17*(10), 489–490.

Tong, F., Harrison, S. A., Dewey, J. A., & Kamitani, Y. (2012). Relationship between BOLD amplitude and pattern classification of orientation-selective activity in the human visual cortex. *NeuroImage*, *63*(3), 1212–1222.

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

Van Essen, D. C. (2004). Organization of visual areas in macaque and human cerebral cortex. In: *Visual Neurosciences*. L. Chalupa and J. Werner (Eds.). Cambridge MA: MIT Press, 507–521.

Van Essen, D. C., Newsome, W. T., & Bixby, J. L. (1982). The pattern of interhemispheric connections and its relationship to extrastriate visual areas in the macaque monkey. *The Journal of Neuroscience*, *2*(3), 265–283.

Vann, S., Aggleton, J., & Maguire, E. (2009). What does the retrosplenial cortex do? *Nature Review Neuroscience*, *10*(11), 792–802.

Vetter, P., Grosbras, M.-H., & Muckli, L. (2015). TMS over V5 disrupts motion prediction. *Cerebral Cortex*, *25*(4), 1052–1059.

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

Volberg, G., Wutz, A., & Greenlee, M. W. (2013). Top-down control in contour grouping. *PloS ONE*, *8*(1), e54085.

Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences, 108*(23), 9661–9666.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience, 11*(12), 1439–1445.

Wilson, H. R., & Bergen, J. R. (1979). A four mechanism model for threshold spatial vision. *Vision Research*, *19*(1), 19–32.

Xu, X., Anderson, T. J., & Casagrande, V. A. (2007). How do functional maps in primary visual cortex vary with eccentricity? *The Journal of Comparative Neurology*, *501*(5), 741–755.

Zeman, A., Dewar, M., & Della Sala, S. D. (2015). Lives without imagery – Congenital aphantasia. *Cortex*, *73*, 378–380.