Decoding the Recognition of Occluded Objects in the Human Brain

By

Courtney Elizabeth Mansfield

A thesis submitted in partial fulfilment of the requirements of the University of East Anglia for the degree of Doctor of Philosophy

Research undertaken in the School of Psychology, University of East Anglia

January 2024

Abstract

The dynamics of object recognition are intricate, particularly when under challenging visual conditions, such as occlusion. Current models of vision often fall short in explaining the human visual system's remarkable ability to represent occluded objects. Previous studies have predominantly employed simple shapes as occluders, limiting the understanding of real-world occlusion scenarios. Chapter 2 delves into neural representations by investigating occlusion with realistic stimuli—objects occluding other objects. Using event-related fMRI, participants engaged in a one-back task while viewing objects in isolation, occluded by other objects, or cut out by silhouettes. Decoding analyses in early visual cortex (EVC) revealed a reliance on visible features, while inferotemporal cortex (IT) exhibited robust representations, incorporating both visible and inferred features. Competition effects across multiple objects were evident in EVC but notably weaker in IT, highlighting IT's capacity to disentangle neural responses amidst competing stimuli. Chapter 3 expands the exploration to behavioural aspects, unveiling the impact of occlusion magnitude on recognition difficulty. IT displayed a linear increase of beta weights in processing allocation with recognition difficulty. Behaviourally, unoccluded conditions showed enhanced accuracy and faster response times, with unique recognition patterns emerging when objects served as both occluders and occluded objects. Chapter 4 uses fMRI to examine the theoretical perspective of predictive processing, employing expectation suppression in EVC during occlusion, motivated by high-level occlusion responses found previously. Multivariate pattern analysis indicated an expectation suppression effect aligning with the sharpening account of predictive processing. The concluding chapter synthesises these findings, emphasising the practical and theoretical implications. Notably, the thesis underscores the importance of utilising ecological visual information in visual neuroscience studies and highlights the differing capabilities of EVC and IT. Collectively, our research contributes valuable insights into the neural mechanisms underlying object recognition in challenging visual conditions, paving the way for future research avenues.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank Dr. Fraser Smith; your guidance and support has been priceless. You have supported my academic growth since I was an undergraduate and I am grateful for the opportunities that this process has afforded me and the skills I have been able to develop as a result. I would also like to thank my second supervisor Prof. Will Penny for his insight and feedback during supervisory review meetings, Dr. Jon Brooks for the advice on all things fMRI, and special thanks to the tech/admin team, without whom nothing would run!

To the collaboration team, Prof. Nikolaus Kriegeskorte, Prof. Tim Kietzmann, Dr. Jasper van den Bosch, Dr. Ian Charest and Dr. Marieke Mur, thank you for allowing me to delve into visual occlusion using real-world images of objects. I have learned so much from this process and fallen even more in love with research.

To my fellow PhDs; Lizzie, Annie, Alice, Delyth, and everyone in the office. I am so grateful for the friendships we created despite the start of the PhD being so isolating (thanks covid). You have made this experience exponentially better with your encouragement, wisdom and walks for little treats/coffee.

To the family I found during my years at UEA, thank you for listening to me talk about brains and handing me cups of tea when I couldn’t think any more. I am eternally grateful for you all. To Meg, your kindness and patience is such a gift. Thank you for listening to a million iterations of every presentation I’ve had to give. I love you!

Finally, I want to thank my wonderful family; my Mum, Saxon and my late grandparents for their unwavering belief in me. I would not be the person I am today without the support and love I have always received from them.

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

The research presented in Chapters 2 and 3 have been presented at conferences in oral and poster formats:

Oral presentations

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. W. (2023). Abstract. Neural representation of occluded objects in visual cortex. *Journal of Vision*, *23*(9), 4594–4594. https://doi.org/10.1167/jov.23.9.4594. Oral presentation at the Vision Science Society annual meeting in St. Pete Beach, Florida, USA.

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. W. (2022). Neural representation of occluded objects in visual cortex. Data Blitz short oral presentation at the British Association of Cognitive Neuroscience meeting in Birmingham, UK.

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. W. (2022). Neural representation of occluded objects in visual cortex. Talk at the University of East Anglia departmental postgraduate conference, Norwich, UK.

Poster presentations

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. W. (2022). Neural representation of occluded objects in visual cortex. Poster session presented at the British Association of Cognitive Neuroscience meeting in Birmingham, UK. Student Poster Prize Winner.

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. W. (2022). Neural representation of occluded objects in visual cortex. Poster at the University of East Anglia departmental postgraduate conference, Norwich, UK.

Any ethical clearance for the research presented in this thesis has been approved. Approval has been sought and granted by the School of Psychology Ethics Committee at the University of East Anglia.

Chapter 1. Decoding the recognition of occluded objects in the human brain

1.1. Object recognition

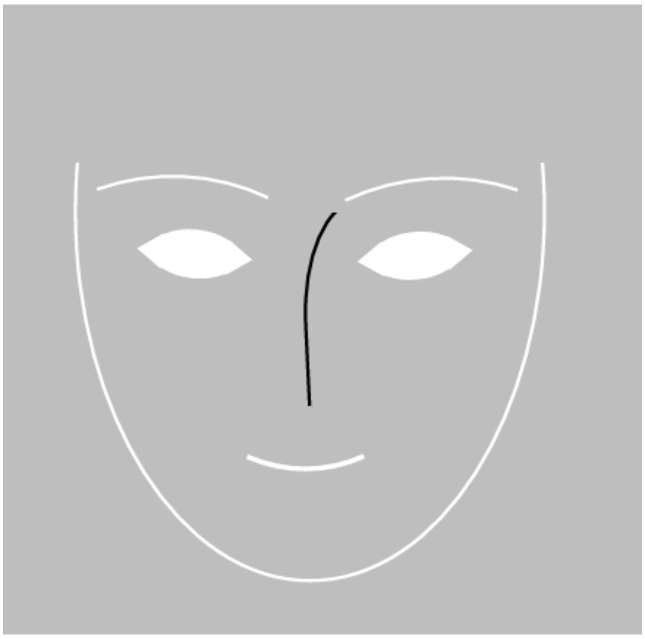
Visual object recognition is defined as the ability to accurately discriminate named objects across a range of materials, textures, and other visual stimuli (DiCarlo & Cox, 2007). This process is computationally taxing, considering the multitude of potential variations across object position, scale, illumination and visual clutter identified by the early visual cortex (EVC) where we will very rarely see the same image of an object twice (Carlson et al., 2011; Spratling, 2016). Despite this, the visual system is highly equipped to deal with this task. Neural signals are able to reflect object recognition within 150ms of initial presentation (DiCarlo, Zoccolan & Rust, 2012). Given the complexity of the processing required by the brain, multidisciplinary collaboration has been required to attempt this feat of understanding the visual system, thus combined work from psychophysics, cognitive neuroscience, computer vision and machine learning, among others have approached the question of how object recognition occurs (DiCarlo et al., 2012).

The propagation of visual signals in the visual stream work to increase the specificity of object understanding as signals move into higher visual areas, or the inferotemporal (IT) cortex. The primary visual cortex (V1) is one of the most widely studied parts of the cortex, and like much of the cortex, it is divided into six distinct laminar layers, each comprising different cell-types and functions. Layer I is known to be the most superficial layer. Layer IV is thought to contain the highest concentration of simple cells and is responsible for receiving information from the lateral geniculate (LGN; Jia et al., 2023; Lawrence et al., 2019). Layer V provides the main output of the cortex, projecting information to other areas of the cortex via long neuron axons (Ramachandran, 2002). The primary visual cortex responds to simple visual components such as orientation and direction. However the summation of this information allows for the basis of much more complicated pattern analysis to occur further along the visual stream (Horwitz & Hass, 2012; Hubel & Wiesel, 1959). V2 is known to respond more to colour, spatial frequency, and object orientation, sending feedback connections to V1 as well as maintaining feedforward connections with V3-5 and onwards along the visual system (Eickenberg et al., 2017). Together, these areas help to identify features of an object, leading to fast and accurate recognition along this well-equipped pathway. These brain regions then lead into higher visual regions which have been collated by researchers into a more overarching object-selective cortex containing a number of functionally defined regions of interest (ROIs), including the lateral occipital complex, fusiform regions and IT regions (Haushofer et al., 2008; Kaiser et al., 2019; Sayres & Grill-Spector, 2006; Wischnewski & Peelen, 2021).

Beyond basic visual features, semantic categories are a prominent way in which people classify objects, especially when the exact object is novel. For example, people have been shown to infer the full object 'car' from just composite parts such as a wheel, which can be attributed to the internal representations predicted from previous experiences (Wang et al., 2017). Rosch et al. (1976) first introduced the concept of levels of abstractions across categories. These range from superordinate (i.e., animate), to basic (i.e., faces) and subordinate (i.e., female face). This concept cemented the belief that hierarchical categorisation is vital in visual object recognition (Peelen & Downing, 2022). Contextual cues are also incredibly important for object recognition, as demonstrated by Liang and Hu (2015) who showed one black curved line, along with white lines which added context (Figure 1.1). When the extra lines were added, it became clear that a face was created and the black curved line represented a nose, yet without these contextual factors, participants were unable to categorise the black curved line correctly. Contextual cues are vitally important in object recognition, with prior knowledge shaping the categorical groups people utilise during perception (O'Reilly et al., 2013).

Figure 1.1.

Additional lines adding context to the initial black curved line to allow the correct representation. From Liang and Hu (2015).



To understand the superordinate categorisation of objects, studies with both humans and primates have frequently utilised animate and inanimate objects (Kriegeskorte et al., 2008; Mur et al., 2013; Pollicina et al., 2022). This shared top-level division relates to findings supporting the evolutionary benefits of recognising living beings from inanimate objects, as well as predator or prey to ensure survival (Conway, 2018; Mur et al., 2013). Clusters created from activity patterns found in the inferior temporal (IT) cortex represent these well-known object categories including animate and inanimate objects (Mur et al., 2013). Within the animate category, additional distinctions can be observed between human and non-human objects or images, with the human category further grouped into bodies and faces. Within the inanimate category there are subclusters of natural and artificial objects (Kriegeskorte et al., 2008). Mur et al. (2013) posited that the IT cortex categorised groupings in this way due to a fundamental drive to survive and reproduce across all human and primate species. The IT has been shown to process higher-level information relative to V1 and is located along the ventral stream of perception (Goodale & Milner, 1992). There is, however, more to be done to comprehend additional categorical and the areas of increasing processing complexity that information propagates through along the ventral stream between early visual and higher visual areas. Exploring categorical patterns of object recognition may provide useful insight into how objects are recognised overall.

The representation and categorisation of objects has been identified in the IT cortex (Bao & Tsao, 2018). Measuring single cell responses has allowed a clearer view of how receptive fields differ across the ventral, or perceptual, visual areas. Larger receptive field sizes and more complex feature preferences as well as longer latencies to visual presentation occur in the IT compared to V1. Where V1 receptive fields are responsive to orientation, IT responses show high selectivity to complex objects (Kreiman, 2008). It is unknown exactly which features are preferred by IT neurons as they are known to remain relatively invariant to changes in scale and position (Kreiman, 2008). Primate studies have been used to observe the role of IT in object recognition, with one study into macaque IT indicating that category selectivity was successful up until 60 per-cent of the object was obscured (Emadi & Esteky, 2013). Though Emadi and Esteky's (2013) stimuli did not contain large amounts of 'clutter', instead using noise to obscure varying percentages of the stimulus image, the results still support the robust nature of object recognition and the important role of IT in this process.

The processing speed of visual object recognition has been evidenced to be incredibly fast, with Agam et al. (2010) using single-cell recording on patients with epilepsy, finding evidence of rapid feed-forward visual recognition within 100-200ms of stimulus presentation. This was identified by presenting one single object or two objects from the same or different categories (across animals, chairs, human faces, cars and houses). They report that IT can support object recognition even in the presence of a second image. This finding has been complemented by additional visual work in primates. Even over very short time intervals IT was able to determine identity and category information robustly and accurately on object information ranging from position to scale, with classifiers able to decode even novel stimuli (Hung et al., 2005). In humans, magnetoencephalography (MEG) has been used to demonstrate that recognition of intact objects could occur from as quickly as 70ms, while decoding the category identity of stimuli (for example faces versus cars) could still be reliably determined by 135ms (Carlson et al., 2011). These studies demonstrate that object recognition and representation occur rapidly within the visual cortex, but the exact temporal dimensions vary dependent on specific task constraints.

The visual system exhibits a hierarchical organisation, with distinct anatomical areas each serving specific visual functions (Felleman & Van Essen, 1991; Rao & Ballard, 1999). These areas connect through multiple projections; ascending feedforward, descending feedback, and those from the same hierarchical level (Kafaligonul et al., 2015). One account for top-down feedback is previous experience driving effects, with expectations facilitating prediction with which external stimuli are judged and categorised (Kafaligonul et al., 2015; Layher et al., 2014). However, other accounts of feedback suggest that it plays a modulatory role based on attention rather than past experience, as attention is critical to high-level cognition (Thiele & Bellgrove, 2018). Bottom-up mechanisms are thought to be responsible for processing external outputs, automatically and involuntarily selecting the most visually salient stimuli, presumably as they are more likely to require immediate consideration (Sobel et al., 2007). Though the review paper by Khorsand et al. (2015) postulates that stimulus-dependent processing requires connections from top-down signals in addition to bottom-up mechanisms. These bottom-up signals relate to top-down modulatory effects in predictive processing through the ability to pass predictions and errors, thus saving mental resources as only the error is passed on instead of the whole representation.

Still, there are conflicting theories of these mechanisms within the literature. Clarke et al. (2014) argued that traditional distinctions of feedback or feedforward are too simplistic, and instead the focus should be on understanding the details that the visual system is representing. A visual crowding experiment was used to explore this, where stimuli of lines, rectangles and lines with X shapes were placed around a central target line with an offset line that participants had to identify the offset direction of. The different conditions provided differing levels of crowding, with the results highlighting the limitations of separating local versus global as well as feedback versus feedforward processes. Clarke and colleagues' (2014) results demonstrated that these distinctions may not fully capture the complexities of visual perception, with their discrimination tasks showing that even when holding local information constant, global stimulus information influenced thresholds. Their findings suggest that local information must be processed globally to achieve this, suggesting that feedback and recurrent processes played a role in this, and that the inherent connectivity of the brain representation should not be underestimated.

Visual clutter is more reminiscent of real-world object variation and thus a good way to measure visual processes (DiCarlo & Cox, 2007), though there are still improvements to be made. Additionally, Wyatte et al. (2014) proposed that object recognition requires early short-distance recurrent processing as well as later attention-related processing. Therefore, it has become clear that viewing object recognition as a purely feedforward process is limited, with visual recognition requiring interactive connections from both feedforward and feedback mechanisms (Bracci & Op de Beeck, 2023; Keshvari & Rosenholtz, 2016; Khorsand et al., 2015; Kietzmann et al., 2019; Thorat et al., 2023).

The dorsal and ventral streams are proposed to be visual pathways that branch off from the occipital cortex, each processing distinct types of visual information (Goodale & Milner, 1992). The dorsal stream is thought to be aligned towards vision for action, including how to carry out actions, judge spatial locations of objects and tools as well as pantomiming and using tools. This pathway is thought to be implicated in neurodevelopmental disorders like dyslexia and dyspraxia (Grinter et al., 2010) and encompasses the occipitoparietal cortex and posterior inferior parietal lobule (Sakreida et al., 2016). Conversely, the ventral visual stream is aligned to vision for perception and is involved in specific object recognition, concerning shape, size, colour and texture judgements, with specificity for these factors evident by 6 months of age (Gazzaniga et al., 2018). This pathway, like the dorsal stream, starts in V1, before diverging towards the IT cortex and is thought to be impacted in disorders such as Attention Deficit Hyperactivity Disorder (ADHD; Corbetta & Shulman, 2002; Helenius et al., 2011). Object recognition as a process is highly dynamic and reliant on the coordination of numerous brain areas throughout the ventral visual system (Wyatte et al., 2012). This intricate system is thought to house critical circuitry for core object recognition (Sorooshyari et al., 2020). Organised hierarchically, it contains huge amounts of both feedforward and feedback projections (Janssen et al., 2018). This visual stream, which culminates in the IT cortex, is key for processing information on object recognition (DiCarlo & Cox, 2007).

Researchers have been attempting to understand these complex processes, with computational models being utilised to try to model the brain network. This research often draws from knowledge of the hierarchical nature of primate visual cortex. The rhesus monkey is currently the most successful non-human model of the human system, with approximately 50 percent of the neocortex being devoted to visual processing (DiCarlo & Cox, 2007; Felleman & Van Essen, 1991; Gazzaniga et al., 2018; Serre et al., 2005). Primate studies have revealed that even within 150ms of stimulus onset, neurons in areas of the ventral stream such as IT cortex have encoded object information in a form robust to differences in scale and position, which is predictive of the behavioural responses of humans in terms of both object 'category' and 'identity' (Hung et al., 2005). This research enables greater confidence in the understanding of the processes of object recognition in the human brain, where hierarchical approaches in primate visual cortex have influenced the nature of human object recognition tasks (Serre et al., 2005). This primate-centred research, alongside human behavioural studies (Wyatte et al., 2012), indicates the central role of the ventral visual stream in invariant object recognition, which facilitates the rapid and accurate recognition of objects in the presence of variations such as size, rotation and position (DiCarlo & Cox, 2007; Karimi-Rouzbahani et al., 2017). The robust ability of humans to recognise objects across states, differences in orientation, contrast, visibility and variety speaks to the importance of the process and demands additional research to further this understanding. The ability to advance our knowledge on object representation and recognition could be important in creating better computational models. As well as this, the possibility of being able to better understand the visual world of those with different visual processing needs to create interventions would be influential, but this requires a thorough understanding of the visual object recognition process.

1.2. Predictive processing

The sensory cortex has been previously presented as a unidirectional system, passively receiving sensory signals in a hierarchy that extracts increasingly complex features (Kok & De Lange, 2015). However, predictive processing theories posit that the brain is constantly constructing an internal model of the world using prior knowledge as well as sensory input (Clark, 2013; de Lange et al., 2018). The relationship between brain areas may even be thought to be heterarchical rather than hierarchical. A heterarchy does not assume a fixed or static top-down relationship between brain regions, instead processing information in a flexible, functional bi-directional manner (Lee et al., 2021).

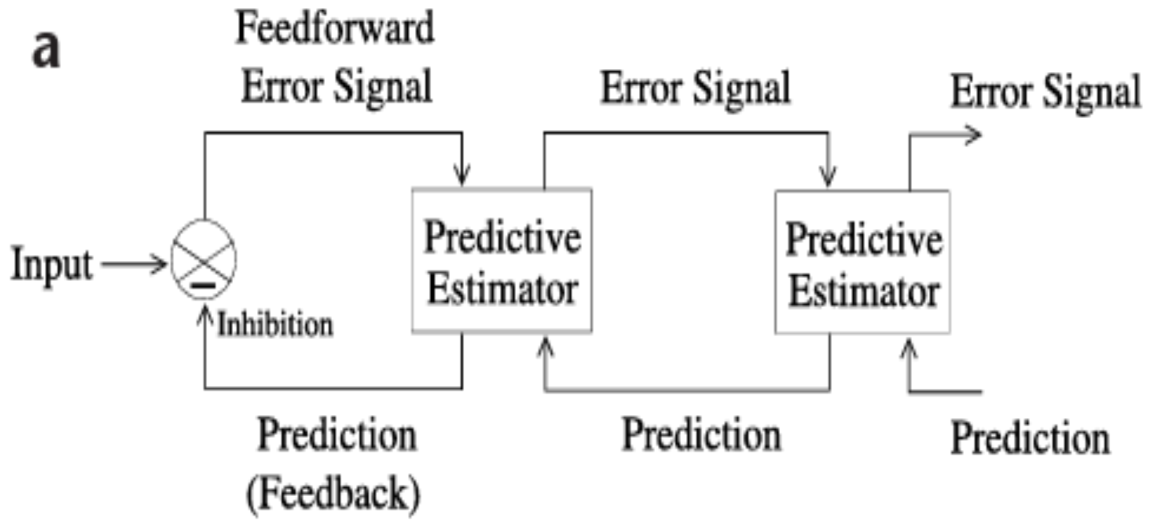
The predictive processing theory purports a unified account of sensory perception, where the overarching task of the brain is to minimise surprise, even across sensory modalities (Melloni et al., 2011; Ransom et al., 2020; Vetter et al., 2014). This theory is a synthesis of several other theories including: Bayesian decision theory (BDT), the Free Energy Principle and predictive coding (Ransom et al., 2020). BDT provides a mathematical framework for decision-making under uncertainty. The theory places the formal representation of visual information, for example a visual scene, against an internal model predicting what that scene is expected to look like based on prior expectations. This expected representation is then compared to the image noise and unexpected details to estimate the likelihood of the actual representation of the scene appearing (Knill & Richards, 1996). The Free Energy Principle was proposed to account for action, perception and learning, with expectation in reward or utility optimised against surprise (prediction error or expected cost). This principle is based on the mathematical idea that free energy within a self-organising system must be minimised to improve adaptation (Friston, 2010). Predictive coding is a strategy for minimising information transmission, whereby the difference between an input and the prediction creates a prediction error and only this is transmitted (Elias, 1955; Ransom et al., 2020). While often used interchangeably in the literature, predictive coding was initially developed as a data compression strategy, where only ‘unexpected’ variations were encoded, meaning the signal was compressed by only transmitting the prediction error (Clark, 2013). Top-down connections are proposed to convey predictions about lower-level activity while bottom-up processes transmit prediction error to higher order levels (Boutin et al., 2021).

Predictive processing synthesises these theories, positing that the brain can better explain sensory input by actively minimising error in the predictions put forward (Clark, 2013; Williams, 2018). Overall, predictive processing allows the brain to process incoming information efficiently by transmitting upwards only the mismatched, unpredicted portions of the signals, filtering the predicted details (Mills et al., 2021). These mismatched prediction errors provide excitatory feedforward input, while prediction units themselves provide inhibitory feedback to minimise the error and create a more precise expectation of the visual stimuli (Walsh et al., 2020).

The predictions, according to Kok and De Lange (2015), play a pivotal role in updating the sensory hierarchy by incorporating both prior experiential expectations and discrepancies. This implies that sensory regions aren't simply passive, awaiting stimulation; instead, they engage in continuous interactions to convey information about predictions regarding the future and associated prediction errors, with each area engaging their populations of error and prediction neurons (Figure 1.2). Kok and De Lange (2015) propose that predictive processing models should prompt experimenters to carefully consider the roles of prediction and prediction error, along with temporal and cortical effects, shifting away from perceiving each trial as an independently attended event by keeping in mind the interplay of errors and signals being propagated constantly.

Figure 1.2.

A visual representation of the predictive process occurring in human cortex. At each hierarchical level, feedback pathways carry predictions of the lower levels while feedforward pathways take forward the residual errors between prediction and the actual neural activity. The prediction errors (PE) at every level are used to correct and update the estimate of the input and then generate the next prediction. Taken from Rao & Ballard (1999).



1.2.1. Expectation suppression

Expectation suppression is known to be a reduction in the measure of neural activity following the presentation of an expected, or predicted, stimulus (Feuerriegel et al., 2021). Stimuli are thought to drive feed-forward processing, while prior knowledge may be driving feedback, so expectation suppression tasks can be used to examine predictive processing mechanisms. It has been well established that there is less visual cortex activation for expected stimuli (Alink & Blank, 2021; Kok & De Lange, 2015; Williams, 2018). Kok et al. (2012) presented subjects with visual grating stimuli differing in contrast, spatial frequency, or orientation but all were preceded by an auditory cue which predicted the correct grating 75 percent of the time. Expectation based suppression in this task was present in V1, V2 and V3 when an expected stimulus was repeated due to the tone signalling the grating. This can be attributed to the prediction error being minimised, as the prediction and stimulus matched.

This expectation-suppression effect, where there are reduced stimulus-evoked responses to expected stimuli, is considered an important empirical hallmark of reduced prediction errors and offers key support to predictive processing (Alink & Blank, 2021). Researchers have used expectation suppression methods involving functional magnetic resonance imaging (fMRI; Richter et al., 2018; Summerfield & de Lange, 2014) and MEG (Todorovic & Lange, 2012) to demonstrate the presence of predictive processes, showing that expected stimuli cause less activation across both visual and auditory modalities. Additionally, though expected stimuli led to lesser activation, multivariate pattern analysis (MVPA) revealed that more decodable information about the stimuli features was available in the early visual areas (Kok & De Lange, 2015). This suggests that prior expectation may sharpen the representation of the expected stimulus in some cases (Kok et al., 2012).

Research has been conducted using fMRI to test whether predictive processing accounts for visual effects across the whole ventral visual stream using expectation suppression methods. Richter et al. (2018) found that when shown pairs of sequentially presented objects where the first was indicative of the identity of the second, there was strong suppression of the neural responses for expected compared to unexpected stimuli in areas from V1, LOC and onward. As has been previously demonstrated by de Lange and colleagues (2018), expectation strongly modulates what people perceive, with less activation when something is expected.

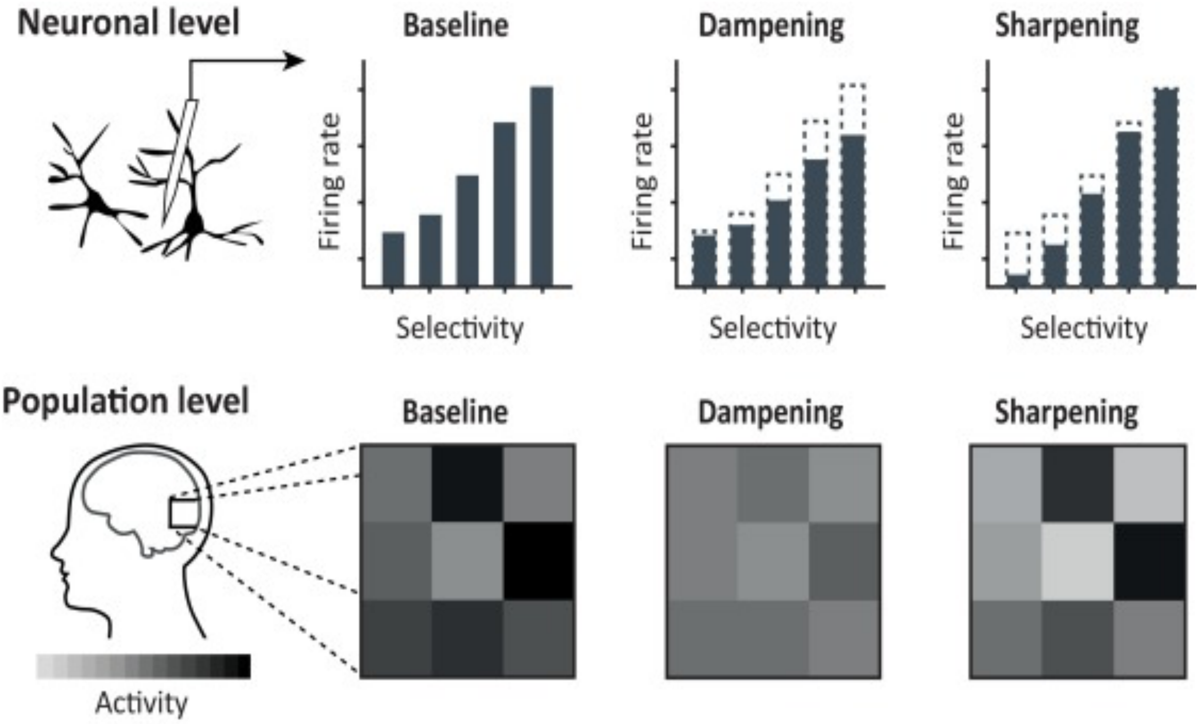
1.2.2. Dampening or sharpening theories of predictive processing

There is an ongoing debate regarding the exact neural mechanism via which predictive processing could give rise to expectation suppression, specifically within the EVC. The debate centres largely on whether the reduced stimulus-evoked responses to expected stimuli result from dampening neural representations for predictable inputs, or sharpening, where increased neural representations result from predictable stimuli (Alink & Blank, 2021).

In an expectation suppression task, the sharpening account predicts a sharper representation of the valid condition by suppressing the activity in neurons not tuned to the expected stimuli (de Lange et al., 2018). This creates a higher contrast and leads to a more distinct representation at the population level, with neurons selective to predicted features more likely to fire than the neurons without this specific selectivity. This account of predictive processing posits a benefit by decreasing the likelihood that noise in the incoming sensory stream is processed, with resources remaining focused on the key predictable elements. This leads to sharper, more detailed perceptual representations of expected features when they occur (Friston, 2005; Kok et al., 2012). Conversely, the dampening account suggests that predicted features are encoded by the inhibition and subsequent decrease in contrast due to the suppression of voxels tuned to the expected stimuli as their activity is ‘explained away’ (de Lange et al., 2018). This account would benefit the visual system by allowing an overall faster processing speed for unexpected stimuli due to fast propagation through the visual system from early visual areas to higher (Richter et al., 2018; Richter & de Lange, 2019; Walsh & McGovern, 2018). Expectation suppression in the LOC has been shown to scale positively with image preference and voxel selectivity, supporting this dampening account of expectation suppression (Richter et al., 2018). See Figure 1.3 for a visual example of these processes.

Figure 1.3.

Expectation suppression is thought to manifest in either dampening or sharpening of neural responses. Dampening predominantly suppresses neurons attuned to the expected stimulus, resulting in a reduction of contrast in activity patterns at the population level. In contrast, sharpening primarily influences neurons not aligned with the expected stimulus, leading to an increase in activity pattern contrast. Adapted from De Lange, Heilbron & Kok (2018).



Methods with high temporal resolution (e.g., EEG) have revealed both sharpening and dampening at different time points (Xu et al., 2020), and techniques like MVPA of fMRI data have been used to distinguish between these two accounts when observing expected stimuli (Kok, Jehee, et al., 2012; Richter et al., 2018). If predictive processing was utilising the sharpening method, we would expect valid conditions, where an expectation was validated, to contain a large amount of information while inhibiting unexpected features. Conversely, in the dampening account, conditions where expectations are validated would be suppressed and patterns of activation would contain only sparse information related to the stimuli, creating poorer decoding accuracy in the classifier regarding the stimuli itself. This would be the opposite in invalid conditions where expectations were violated, and decoding accuracy would be higher due to the lack of suppression of stimulus specific features. Looking at the results of a grating study at a neuronal level, where a sharpening account reflects an expectation-induced reduction of neural activity, the strongest effects would be observed in those neurons preferring the conflicting orientations, while neurons preferring the presented orientation would be unsuppressed in comparison (Kok et al., 2012). There is selectivity across individual voxels, as indicated by reduced activity for predicted stimuli in V1 – a finding that persists even when prediction is task-irrelevant (Kok & De Lange, 2015).

Moreover, Xu et al. (2020) revealed that sharpening and dampening effects are observable in EEG data at different time points. Specifically, that N1 demonstrated a sharpening effect while N2 indicated a dampening effect. They achieved this by training participants on a double-flash task. Here, an auditory cue preceded a pair of oncoming flashes of light with unpredictable stimulus onset asynchrony (SOA) between the cue and flashes (1000/1500/2000ms) and between the flashes themselves (400/600/900ms). Participants were required to hold a temporal template in their mind for each block, representative of the SOA between the flashes. They were then required to press a response button after each double-flash to make a judgement on whether the SOA time they had experienced in that trial was matched or mismatched to their prediction of the time (e.g., if asked to recall the 400ms double-flash SOA in their mind, did the double-flash they saw also have a 400ms gap, or was it 600/900ms). Their results revealed an expectation suppression effect as well as demonstrations of both sharpening and dampening effects of prediction. These effects occurred in distinct processing stages with an opposing trend, providing evidence for an opposing processing theory which provides a potential resolution for the sharpening and dampening accounts of prediction (Press et al., 2020). These findings suggest that both sharpening and dampening predictions of processing in expectation suppression co-exist, doing so in different temporal windows.

Xu et al. (2020) also claim that differences in attention cannot account for these results, addressing another key debate in the field of expectation suppression studies of predictive processing. The task itself required continued attention to both flashes in each trial. Additionally, the researchers monitored the alpha band oscillation which is relevant to the allocation of attention, and theta band oscillation which is implicated in timing prediction. They observed no changes in alpha across trials, while theta oscillations were present, indicating no differences in attention towards expected and unexpected flashes but differences in prediction for the 2nd of the flashes. These sharpening and dampening accounts of predictive processing provide a nuance that calls for careful methodological consideration. Though they have opposite effects, the accounts both imply some amount of reduced neural activity when predictions are accurate.

1.2.3. Attention versus prediction

Attention boasts a wide array of definitions, which some have argued makes the concept "ill-defined" (Alink & Blank, 2021). However, the core qualities of attention across these definitions include acting as flexible control for limited computational resources (Lindsay, 2020). Within vision, attention has been investigated at length, with processes like selective visual attention examined, which include looking out for specific visual information such as colour or shape (Lindsay, 2020). In the past, the terms 'attention' and 'expectation' have even been used interchangeably (Kok et al., 2012). There are those who suggest that as researchers, we must differentiate attention from prediction to avoid misclassifying effects (Alink & Blank, 2021; Cao, 2020). This misclassification may be the result of predictive processing theories being challenged by the view that prediction sometimes seems to amplify sensory signals rather than diminishing them, as observed in attention cueing experiments. The conventional explanation of predictive processing attributes this phenomenon to the intertwining of prediction (indicating the likelihood of a stimulus) with attention (reflecting task relevance), wherein attention can enhance these signals. However, recent research is proposing more nuance in this interplay, suggesting that attention and prediction collaboratively enhance the precision of perceptual inference (Kok et al., 2012).

According to the model by Kok et al. (2012), there exists an elevated weighting of sensory evidence influenced by attention, leading to a reversal of the typical sensory silencing associated with prediction. Kok and colleagues (2012) used predictable cued grating-word pairings in an fMRI study to analyse the interplay of attention and prediction on their neural activation results. Unattended stimuli produced a reduction in activation in EVC for predicted over unpredictable stimuli, consistent with expectation suppression and thus theories of predictive processing. In response to attended stimuli, there was a larger neural response in EVC for predicted compared with unpredicted stimuli, which is inconsistent with the main effects of prediction and attention but would be explained by a synergistic relationship between the two. Although seeming counterintuitive, these findings align with the predictive processing framework. As research demonstrates that mechanisms of predictive processing and attention can comfortably exist within the same model, with attention reflecting the precision of perceptual inference (Alink et al., 2010; Friston & Kiebel, 2009; Rao & Ballard, 1999). Under this account, attention is modulated on the synaptic gain of neurons representing sensory data (or, equivalently, prediction error), which are weighted according to the strength of the prediction. Thus this study demonstrated empirical support for predictive processing through the demonstration of attention reversing sensory silencing effects of prediction. Which may also go some way to explain the contradictory findings within the literature regarding the interplay between attention and prediction (Kok et al., 2012).

Many perceptual processes are driven by attention. Studies using visual tasks have found that attending one visual stimuli over another when a pair is presented can account for a 30 percent shift in representation showing a bias towards the attended compared to the unattended stimuli (Reddy et al., 2009). Reddy and colleagues' (2009) method was based on the biased-competition theory of attention, where targets and non-targets compete for processing capacity within visual search (Desimone & Duncan, 1995). Stimuli from various visual categories were presented in pairs from various visual categories and participants were told to attend to one, both or neither. The patterns of representation showed that when projected into a vector plane, it was possible to determine the weighted average of the pair of stimuli when averaging the sum of the two isolated objects in a linear combination. This study showed that when projected onto a weighted average line, there was approximately a 30 percent shift in weights when the stimuli was attended to.

Reddy and colleagues (2009) showed that split attention works along a continuum, which presents an interesting nuance to processing stimuli under visually challenging conditions. MacEvoy and Epstein (2009) found that under conditions of distributed attention, voxel wise patterns of activity in object selective regions of the cortex evoked by pairs of objects are the average of the patterns evoked by the individual component objects, which could be decoded with incredibly high accuracy. However, the weighted average model has recently been argued to be less effective than a normalisation model in some areas of the visual stream, such as the primary visual cortex, whose complexities may be better captured by a weighted sum model than weighted average. This is perhaps due to higher sensitivity of neurons in V1 to contrast effects, potentially triggered by an attentional shift towards higher visual contrast causing increased attention. However, findings in macaques have been previously demonstrated neuron-to-neuron variability driven by the ability to process multiple stimuli, rather than simply the differences in top-down attentional signals. Thus, it is clear that there is an attentional effect on visual representation, but not exactly how this is distinct from and interacts with prediction.

In their review exploring top-down influences on visual processing, Gilbert and Li (2013) state that attention and expectation are among the rich and varied influences carried between cortical areas. They assert that receptive fields are dynamic and adapt to carry information relevant to the perceptual demands of a specific task. In terms of attention versus prediction they find that attentional effects are more prevalent when there are multiple stimuli, particularly in V1 where contextual influences are involved. In terms of neuronal activity, neurons change their tuning in accordance with the demands of the specific task, implying that neurons are adaptive processors. They describe a primate study where neurons in V1 could alter selectivity both individually and at a population level dependent on the expectation of the observed shape (McManus et al., 2011). This finding indicates that expectation employs top-down processes in object recognition to create a set of filters more selective to the task-based stimuli, ensuring vision is an active process discernible from attention alone. Thus it becomes clear that there is a determinable difference in both computational and non-human primate studies that suggests attention and prediction are discriminable from each other (Doostani et al., 2023; McManus et al., 2011; Ni et al., 2012). This differentiation between prediction and attention was also observed in a prediction error study, while larger neural responses to surprising stimuli could be explained by attention, attention alone fails to explain why there is larger activity in omission studies (where a stimulus is predicted but withheld) due to the absence of a stimuli present to attend to (Den Ouden et al., 2012).

As touched upon previously, there is debate on whether attention or prediction is responsible for the results in expectation suppression tasks. Some suggest that predictive processing theories attempt to account for too much of a unified account of mental functioning (Ransom et al., 2020). Furthermore, this debate is extended further with the ongoing discussion of whether the neural mechanisms underlying expectation suppression could result from reduced attention to predictable stimuli (Aitchison & Lengyel, 2017; Alink & Blank, 2021). Ransom et al. (2020) suggests that a more flexible version of predictive processing, one that allows other factors to account for some effects, and incorporates more from Bayesian decision theory, may be better equipped to accommodate attention. This nuanced approach suggests that to truly grasp their roles, we must acknowledge and accommodate the unique and distinguishable contributions of each process.

Alink and Blank (2021) argue against relying solely on the prediction-based response, where attention increases the responsiveness of neurons sensitive to the discrepancy between sensory input and expected stimulus. They posit that this effect could be explained by an attention-based explanation instead. This explanation suggests expectation suppression because of reduced saliency of an expected stimuli. They state that the theory of predictive processing proposes that attention is a facilitatory mechanism of error coding (Richter & de Lange, 2019), while the attention-based explanation purports that results are a response to attention from stimulus saliency (Kanan et al., 2009). The difference in how attention is explained across these schools of thought seems to hinge on the view that attention is a cause of prediction error coding within the predictive processing framework, whereas it is seen as an effect of the saliency of stimulus in the explanation of attention.

Though this research certainly supports the need to be thorough when designing studies so that effects of prediction are not confounded with attention, there have been effects that support prediction as a parsimonious explanation. Some have demonstrated the necessity for additional research here, showing that at times of sensory input ambiguity conscious awareness is required to activate the predictive mechanisms and alter prediction and perception (Reddy et al., 2009; Vetter et al., 2014). Questions remain regarding whether the sensory attenuation process to predicted stimuli is automatic or if this is only the case when predictable stimuli are attended (Richter & de Lange, 2019). To try to understand this, research has been undertaken focusing on the interplay between attention and prediction. Repetition effects in fMRI BOLD responses have also been used to dissociate attention from prediction experimentally (Larsson & Smith, 2012; Summerfield et al., 2008).

One such fMRI study discovered that a normalisation model of object recognition could predict responses at the voxel level from primary visual cortex and across the visual hierarchy with and without the influence of attention, using conditions of isolated or cluttered stimuli (Doostani et al., 2023). Their study categorised stimuli as preferred or null in each ROI and analysed voxel responses to isolated houses and bodies. Preferred stimuli showed increased voxel responses to isolated houses or bodies compared to null stimuli. Interestingly, when attending to a preferred-category stimulus, a winner-takes-all effect was observed, reducing the impact of null stimuli on ROIs. This suggests that attention alone cannot account for recognition mechanisms, highlighting the role of categorical changes in stimulus representations.

EEG studies have also been utilised to measure predictive effects against attention effects. In one such study which involved timing precision, researchers found that N1 indicated a sharpening effect while N2 indicated a dampening effect (Xu et al., 2020). This may suggest that both accounts of predictive processing in expectation suppression co-exist but do so in different time windows. They also successfully showed that attention was not responsible for their results, as participants paid equal attention to the presented stimuli driving the effects. Additionally, the work of Kok et al. (2012) also speaks to a modulation between attention and prediction that allows both processes to occur. Consistently, we see an ability to observe effects of prediction that cannot be attributed to elevated attention effects.

Additionally, the review of Schröger et al. (2015) examined the attention and prediction literature around EEG auditory studies. They confirmed that while attention often increases various parameters of brain activity and prediction often results in the attenuation of brain activity, the predictive processing framework allows for the two to be related by noting their differing effects, rather than trying to ignore the presence of one or the other. To account for this within studies the authors suggest ensuring methodological rigor when designing manipulations of tasks measuring these effects. This will enable a better understanding of the interplay of these two key concepts and allow predictive processing to be successfully measured.

1.3. Recurrence and computational methods

Recurrent processing, where neurons influence each other through direct, bidirectional interactions, is essential to visual understanding (O'Reilly et al., 2013). This type of processing is said to be responsible for the ability to solve certain visual tasks when a feedforward sweep, or feedback connections alone are not sufficient, and it is recurrent processing that determines visual awareness of the features of an object (Lamme & Roelfsema, 2000). Recurrent processing occurs where interconnected sensory systems involve both feedforward and feedback connections. These connections enable adjacent layers to interact locally and recurrently to refine representations and give rise to a dynamically extended network that can activate both local and widespread areas (Han et al., 2018; Wu et al., 2020; Wyatte et al., 2014). Flexible communication of top-down and bottom-up influences has been found to allow enhanced representations of objects, bringing them into conscious perception (Lamme & Roelfsema, 2000; Yan et al., 2023). This recurrent processing can strengthen bottom-up signals when objects are viewed under conditions of occlusion, poor lighting, shadows, and other variable factors. Thus, facilitating the creation of a strong, stable representation of an object that allows robust recognition (Wyatte et al., 2012).

Recent research has further investigated recurrent processing networks, for example, using ultra-high field (7T) functional imaging to look at sub-millimetre resolution, Jia et al. (2023), determined that the multiple laminar layers play different roles in the orientation-specific representations of visual recognition. They showed that the superficial layers of V1 are more influenced by recurrent plasticity mechanisms connected horizontally. Prior expectations have also been found to show selective activation in deeper layers of V1, consistent with feedback processing in the context of predictive processing (Aitken et al., 2020). This combination of multiple processing mechanisms speaks to the incredible complexity of the visual system encompassing parallel as well as hierarchical features (Lamme & Roelfsema, 2000; Wyatte et al., 2014).

Furthermore, vision is a highly dynamic process reliant on multiple areas of the ventral stream for different perceptual processes across the massive amount of information input (Li et al., 2023). Accordingly, recurrent connections have become a plausible way to predict visual features of heavily occluded objects. These connections have been shown to successfully capture the physiological delays observed throughout the ventral visual stream, suggesting indirect evidence for object completion occurring in the IT cortex (Tang et al., 2018).

Anatomical findings indicate the massive prevalence of recurrent and feedback connectivity throughout the visual streams (Felleman & Van Essen, 1991; Gurney, 2003; Sporns & Zwi, 2004).

Employing recurrent networks to examine processing within digit clutter and crowding scenarios has informed computational models, with these networks outperforming simpler feedforward models when recognising occluded stimuli. Spoerer et al. (2017), used images of digits occluding each other in cluttered scenes to effectively validate that occluding objects in more realistic ways was more applicable to real-world situations. Using recurrent methods to emphasise this they found more effective representations using recurrence than purely feedforward models, however there is still a way to go in creating tasks which occlude completely naturalistically. Digits themselves are simplistic stimuli to use, but a valuable starting point for future development. The authors have since scaled up the size of the dataset using ImageNet and ecoset to train deep neural networks (DNNs) on more ecologically relevant categories to see whether recurrence improves recognition accuracy, successfully observing a better explanation for representations of complex visual objects. These studies only used isolated, whole stimuli, however they specifically note that the increased visual diet that came with the more ecologically accurate dataset was responsible for the effects found (Deng et al., 2009; Mehrer et al., 2017; Russakovsky et al., 2015; Spoerer et al., 2019; Spoerer et al., 2017). This work allows greater understanding of how visual recognition is undertaken, creating a logical next step of occlusion examined using this breadth of more ecologically valid and varied stimuli.

There is a known interaction between recurrent connectivity and learning, which predicts that high-level visual representations could be influenced by error signals from proximal brain areas throughout visual learning (O'Reilly et al., 2013). The robust way in which visual clutter or degradation is compensated for in the ventral visual stream may be attributed to recurrent connectivity, depending on the dynamic nature of the brain (Wyatte et al., 2012). Modelling of the network connections in the human ventral stream by Kietzmann et al. (2019) established that information processing is best accounted for by recurrence for recognition into object categories.

Artificial neural networks are incredibly popular methods for analysing data within computational neuroscience and beyond, being thought of as simplified models of the vast networks of neurons occurring natural in the brain (Gurney, 2003). Artificial intelligence (AI) has created incredible advances in complex problem solving in these networks, becoming a theoretical vehicle aiding in the understanding of the processing of neural information (van Gerven & Bohte, 2017). Feed-forward convolutional neural networks (CNNs) have been used in order to explore object classification, emerging as a quantitatively accurate model of primate visual cortex (Nayebi et al., 2018).

Deep convolutional neural networks (DCNNs), that mimic the major principles of the visual pathway have been able to represent how visual information is transformed from a 2D image using feed-forward layer connections to analyse the image. Research has found that early layers of these models correspond with retinotopic areas, whereas later layers capture aspects of the higher-level representations present in IT (Bracci & Op de Beeck, 2023). The formation of these layers is conceptually similar to the increase in receptive field size that characterises successive processing stages in the human visual system (Bracci & Op de Beeck, 2023; LeCun et al., 2015). When trained for object recognition, they have been successful at representing objects of the same category as being similar even when there is variation across the category, that is that they are able to generalise well across differences in low level features (Konkle & Alvarez, 2022).

These models have been generally improved by the addition of feedback, recurrent connections, and even by considering topography across the human cortex (Ali et al., 2022; Karapetian et al., 2023; Kietzmann et al., 2019; Lu et al., 2023; Spoerer et al., 2019; Thorat et al., 2023). Recurrent neural networks (RNNs) have been found to be more neurobiologically realistic than feedforward counterparts and more robust in their ability to recognise objects, especially under challenging visual conditions (Spoerer et al., 2017; Tang et al., 2018; Yan et al., 2023). From advances in these goal-driven models, convolutional recurrent neural networks (ConvRNNs) were created to explain dynamics within the visual system using layers to process and transform inputs to produce outputs (Nayebi et al., 2018). There have also been recurrent CNNs (RCNNs) that have computationally demonstrated a successful model of scene processing, effectively predicting categorisation more efficiently and accurately than feedforward models alone (Karapetian et al., 2023; Lu et al., 2023; Spoerer et al., 2019; Thorat et al., 2023). As scenes often include vast amounts of visually challenging information, the findings of these studies may be helpful in advancing the field of object recognition too.

RNNs and beyond align effectively with predictive processing models, more efficiently coding information than other kinds of models (Ali et al., 2022). Prediction errors are a key element of predictive processing and have largely been found to relay information in a bottom-up manner, establishing more computationally efficient recognition, especially when combined with top-down sensory input (Richter et al., 2023). The incorporation of prediction error in the models leads to more efficient RNNs, which may be primed by the predictions from prior visual experience and further split into subpopulations of prediction and error units respectively (Ali et al., 2022). The field of computer vision has embraced the use of AI and neural networks, including ConvRNNs which are task-optimised for sophisticated object recognition at a more human-like level (Kriegeskorte, 2015; Nayebi et al., 2018). However, ignoring the role of cross-disciplinary research would be short-sighted, as fields like neuroscience may allow for benefits in this research as well as the other way around, with respect to tackling more complex visual functions beyond categorisation. Aspects of attention, visual search and image segmentation, with the ability to align neural networks with measures of brain and behaviour data (Kriegeskorte, 2015), could paint an increasingly elucidating picture of how the visual system works and can be applied to AI.

1.4. The role of occlusion in visual processing

Occlusion is an ever-present feature in natural, three-dimensional scenes, with some degree of obstruction from one object to another generally unavoidable. The human brain can account for occlusion effectively using amodal completion (Kanizsa, 1976), where partially occluded objects are successfully and accurately recognised (Zhu et al., 2019). For example, when a cat's body is hidden behind a fence with only a head visible, humans still perceive the cat as a complete animal, being able to represent the appearance of the occluded section (Ao et al., 2023). This is done incredibly successfully and quickly, with delays of approximately 100ms observed for the recognition of occluded compared to unoccluded objects (Tang et al., 2018). Recognition capabilities of the brain for heavily occluded or even deleted objects are still extremely robust (Rajaei et al., 2019; Tang et al., 2018). As object recognition is heavily reliant on being able to 'fill in' missing information, using occlusion may be a useful measure to investigate predictive processing, as predictions are likely to be vitally important in this kind of recognition task.

The ability to recognise an occluded object is said to require a degree of prior knowledge about the object itself (Wyatte et al., 2014). The brain not only has to compensate for missing information about the object, but also process the object identity, the occluder and the scene context to create a full picture. The brain areas implicated in these processes by prior research include: V4, which in human vision is implicated in colour processing as well as luminance (Tootell & Hadjikhani, 2001); IT cortex, a higher order visual area known to be highly interconnected and associated with object categorisation and classification; as well as prefrontal cortex (PFC) which is known to play an important role in cognitive control, specifically within thought and action. It has been found that the ventrolateral region of PFC may be implicated in higher form visual processing, where there is evidence of functional interactions between this area and V4 as well as IT (Fyall et al., 2017). Consequently, there are theories that PFC may have input in perceptual processing of visual stimuli when the recognition is made more difficult by conditions such as partial occlusion. There are also interesting effects in the primary visual cortex (V1) when observing scenes with occluded quadrants. V1 is shown to contain information about the missing section, even without receiving direct sensory input (Smith & Muckli, 2010).

Tang et al. (2018) demonstrated that the brain has a robust ability to recognise an object category even when objects were heavily obscured. This work combined behavioural, neurophysiological and modelling insights to determine how recurrent connections may allow the brain to carry out pattern completion on partial information. Using a backward masking task to disrupt recognition, objects in conditions of whole, partially visible (missing sections) or partially occluded (black screen with cut out shapes showing parts of the object) were shown to participants, followed by a blank screen or a noise mask. Performance was significantly degraded by the masking as they expected, successfully disrupting the presumably recurrent processing of the original image. Overall results determined that inferences of object identity were possible when only 10-20% of the object was visible, even for novel objects. Though this study's stimuli were not occluded per se, as the occluded condition images were partially visible but not ‘blocked’ by another object. In computational modelling, performance increased with the visible percentage of the object, though this was still demonstrably lower than a human level in both RNN and Alexnet models (Krizhevsky et al., 2012). Tang et al. (2018) suggests that delays in recognition for occluded objects are a result of recurrent connections to influence pattern completion, using time to recruit lateral connections and/or top-down signals from areas higher in the ventral visual stream.

Masking tasks have been leveraged in multiple object recognition studies. Wyatte et al. (2012) manipulated objects, including cannon, car, fish, gun, key and trumpet, into conditions of control (low occlusion, full contrast), high occlusion (high occlusion, full contrast) and low contrast (low occlusion and 25 percent contrast). The mask was constructed from patches of the original images. The occlusion in this study was created using a filter made up of a circle 5 percent of the image size with edges softened with a Gaussian filter. This was applied at random locations of the image, with the amount of occlusion determined by the condition, with the control trials having a small amount of occlusion and the occlusion trials having the filter applied more often. They found that the decoding in the occluded condition was much less accurate than control and contrast conditions, with the masking condition universally less accurate across conditions. With regards to feedforward and feedback dynamics and how these give rise to object recognition, the authors noted the influence of recurrent processing in recognition of degraded stimuli. Backwards masking was determined to be a successful measure of interrupting recurrent processing, as masking creates a mismatch between feedforward and feedback responses (Lamme & Roelfsema, 2000).

It is important to develop this understanding of how the propagation of information through the ventral stream occurs in challenging visual conditions. Wyatte and colleagues (2012) successfully demonstrated that recurrent processing during object recognition creates a strong and stable representation of the object in question, even when the viewing condition of the object is degraded, which is frequently the case in real-world vision. They note that the highly interactive and dynamic visual processes depend on multiple brain areas at different levels of ventral visual hierarchy and that more needs to be done to understand exactly what mechanisms are being used during degraded visual conditions. However, the occlusion condition within this study is not very representative of true vision, with fragmented masking and blurred occlusion patches used to create the experimental stimuli instead of object occluding other objects. This may cause important processes within visual object representations to be missed when trying to compare these findings to how humans see in the real-world, necessitating further study.

Though ‘partially visible’ stimuli have been utilised in a variety of paradigms, most studies observing the effects of occlusion also occlude with masks or black ‘blocks’ which lack naturalistic relevance (Smith & Muckli, 2010; Tang et al., 2018). Johnson and Olshausen (2005) presented participants with images of real-world objects occluded or deleted with ovals obscuring an increasing percentage of missing pixels. They found that occluded trials were more easily recognised than deleted trials. However, the ovals can be argued to not be representative of the natural visual world, limiting the ecological validity and generalisation of their results. It would be important to expand this research avenue by using real-world objects as both occluders and occluded objects. This may reveal variations in how accurately people judge deleted and occluded object pairs. By introducing an extra layer of visual complexity, where two meaningful objects appear in the occluded condition as opposed to just one, a cut-out like in the deleted condition, or ovals as in the original study, we could gain insights into how our visual processing handles these situations.

Currently, much present visual object recognition research focuses on single presented objects, however this may prove unwise. Examining exactly how the brain processes the occluded, as well as occluding, object should be a critical consideration. Occlusion research tends to use either a black box or a scrambled noise mask as an occluder (DiCarlo & Cox, 2007; Johnson & Olshausen, 2005; O'Reilly et al., 2013; Smith & Muckli, 2010; Tang et al., 2018), which lacks the nuance that the visual system processes in daily life. Novel studies exploring more complex visual scenarios, such as multiple object presentations and occlusion are building upon research into single objects. This line of research has gone some way to reveal that the representation of an object may be altered by the presence of other objects, even un-occluded, though the exact mechanisms are still unknown (Baeck et al., 2013; Chelazzi et al., 1998; Li et al., 1993; Reddy & Kanwisher, 2007; Rolls & Tovee, 1995; Zoccolan et al., 2005, 2007). These works contribute to the overarching object recognition literature with added complexity and argue for researchers to take this further to explore more ecological multiple object presentations such as occlusion. It is known that most objects in natural scenes adhere to meaningful structures and locations, predictable in terms of location, viewpoints and sizes relative to one another, which has been demonstrated to enhance recognition (Kaiser & Peelen, 2018; Võ et al., 2019). Thus, it is important to consider how the representation of multiple objects, particularly when occluded may offer insight into visual processing throughout the ventral stream.

As previously mentioned, focusing on the role of the occluder, Spoerer et al. (2017) used numerical digits as stimuli for both the occluder and occluded objects. Digits, while not real-world objects themselves, are frequently observed within the natural environment and can be argued to be similar to real-world objects. Manipulating visual clutter, this study demonstrated that when multiple objects require identification at once, recurrent networks, but not feedforward mechanisms, were most successful. Spoerer et al. (2017) found that delays in processing may be caused by the identification of the occluding object in addition to the occluded object. This nuance may have been missed in trials where a non-meaningful occluder was the only occluding feature, ignoring the role of the occluding object. The digits themselves are simple stimuli that provide a useful starting point for occlusion studies which are more applicable to real-world visual experience.

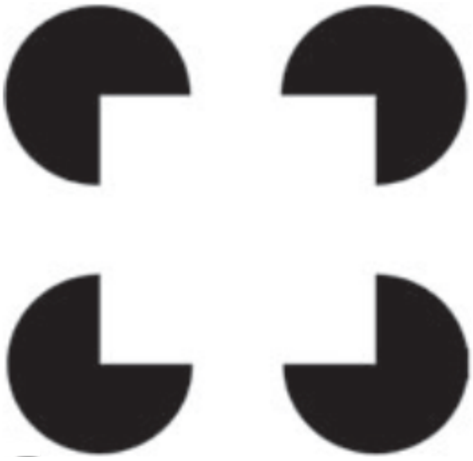
Early sensory brain areas, like V1, have been studied extensively and it was thought that V1 was primarily responsible for basic early-cortical computation by means of orientation-tuned neurons arranged in a retinotopic map (Ng et al., 2006). However, more recently there have been findings implicating V1 in higher cognitive functions, with the activity of V1 being modulated by cortical feedback from multiple brain areas (Smith & Muckli, 2010). Scenes with occluded sections have been used to demonstrate, using fMRI tasks, that even when primary visual cortex areas do not receive direct sensory input, there is internal communication in these areas that allows visual context to be known (Muckli et al., 2015).

Perceptual illusions such as the Kanizsa illusion provide an example of perceptual inference through the distinction between modal and amodal contour completion (Kanizsa, 1976; Kok & De Lange, 2015). Amodal completion is the ability to perceive an entire object despite it being occluded, by ‘filling-in’ the missing details based on prior experience of perceiving and understanding objects or scenes (Scherzer & Ekroll, 2015). The illusion seen in Figure 1.4, presents four ‘Pac-man’ figures yet gives the illusion of four black circles with an overlaid white square (Kok & De Lange, 2015), representing amodal completion as a square is represented and perceived to be occluding four whole disks. The brain infers the simplest ‘gestalt’ of the objects presented, relying on prior experience of the world. As occlusion is commonplace in the visual world, using inference in this way is key to successful perception.

These illusory figures have also been tested in primates, with evidence of neurons in V1 and V2 responding to the illusory contours (Lee & Nguyen, 2001).

Figure 1.4.

The Kanizsa square illusion. Four 'Pac-man' figures or a white square overlaying black circles? Adapted from Kok & De Lange (2015). The image appears to contain a solid white square, with well-defined contours, however this shape is subjective and lacks any physical basis.



Illusory stimuli have also been used in studies using high field fMRI methods to build a non-invasive laminar profile of feedback, feedforward and recurrent connections between the layers of cortical areas. Researchers hypothesised that the feedback mediated activity in V1 during perception of the illusory shapes would result in a distinct laminar activity profile compared to bottom-up stimulation as the illusion causes a perceived representation of a shape that is not derived from bottom-up stimulation. Thus across the laminar layers of V1 they revealed that top-down activation initiated by a perceptual illusion activates the deeper cortical layers of V1, known to be responsible for output and projection to other areas of the cortex (Ramachandran, 2002). Whilst neural responses to bottom-up stimuli are usually evident across all layers, with the strongest activation in the middle and superficial layers, which are generally associated with receiving information from the LGN (Jia et al., 2023; Lawrence et al., 2019). High field scanning affords higher resolution than typical fMRI which allows the analysis of additional fine-grained detail that contributes to our understanding of the interplay between feedback and feedforward mechanisms in the visual system and beyond. The interaction between top-down and bottom-up signals during perception can be observed clearly using techniques such as this.

Research has indicated that V1 not only holds information about visual objects but also plays a crucial role in the visual processing when these objects are occluded (Muckli et al., 2015; Smith & Muckli, 2010). These findings highlight the significance of studying V1 to unravel the mechanisms behind occluded object identification, with potential implications for enhancing computational models that currently face challenges in achieving robust recognition under occlusion (Tang et al., 2018). Morgan et al. (2019) further delved into the role of V1 in occlusion using a scene viewing task involving a partially occluded quadrant. Their fMRI study analyses then compared occluded V1 activity with line drawings of the occluded area and revealed that these drawings more successfully matched the observed activity than standard computational models. This underscores the robust processes within the visual system that facilitate the perception of occluded scenes.

1.5. Aims and objectives of this thesis

Overall, as detailed above, visual object recognition in humans is a complex multi-stage process, often outperforming artificial networks, especially under conditions of occlusion (Cichy et al., 2016; Guo et al., 2015). It remains unclear whether predictive processing best explains object recognition in humans across different levels of the visual system, particularly when objects are also used as occluders.

As consolidated by the literature from this chapter, research in recent years has established that visual object recognition uses feedback and feedforward projections to update visual information, from bottom-up stimuli information on size, shape and constancy to top-down predictions of the expected scene category and object identity (DiCarlo et al., 2012; Wyatte et al., 2012). As well as this, lateral, recurrent connections are also vital in recognising objects, particularly when the visual situation is challenging, for example when a stimuli is degraded (by blur, clutter, high contrast or occlusion among other factors that make recognition more difficult) and it is believed that amodal pattern completion is aided by these recurrent projections (Kietzmann et al., 2019; Wyatte et al., 2012).

The theory of predictive processing has been influential in how researchers seek to 'solve' object recognition, offering a neat explanation for the recognition of highly diverse scenes and objects while conserving neural resources by only propagating the differences - prediction errors – from an expected stimuli or scene (Clark, 2013; Kok & De Lange, 2015; Rao & Ballard, 1999; Richter et al., 2018). A measure of predictive processing, particularly in neuroimaging studies is expectation suppression, which works from the assumption that a predictable stimulus requires less activation than an unpredictable one. However according to the sharpening account, when decoding, a predictable stimulus will be more accurately represented due to increased knowledge compared to an unpredictable stimulus (Alink & Blank, 2021; Feuerriegel et al., 2021), while dampening would produce a less contrasting effect. Though there are arguments throughout the literature that the predictive effects are simply due to attention, there have been successful efforts to dissociate expectation from attention to ensure the correct measure is being recorded (Doostani et al., 2023; Ransom et al., 2020; Xu et al., 2020).

Computational models have been increasingly used to explore this field, with access to huge visual datasets and a growing reliance on recurrent computations creating more successful models of vision (Ali et al., 2022; Kietzmann et al., 2019). However, this field would arguably benefit from continued multidisciplinary collaboration, with neuroscience aligning with computer science and beyond to link the neural processes of the human visual system with the huge computational power and speed of analysis from RNNs/DNNs. Understanding the process and mechanisms behind occlusion could particularly benefit computational models as the human ability to pattern complete objects even when incredibly occluded is remarkable. Hence more human-like models of computational vision may be able to further improve recognition of complex objects and scenes within computer science. From the literature, there are several questions left unanswered that this thesis has sought to address.

First, prior studies investigating occlusion, a core visual process, have been significantly limited by the stimuli they are using. There is a need to examine more naturalistic methods to represent real-world object occlusion. Real objects do not appear in isolation, constantly overlapping and occluding in a variety of dynamic shapes and patterns. In fact, we see more objects that are partially occluded than are not. Using only deletion or occluding using unnatural occluders, provide fascinating results (Johnson & Olshausen, 2005; Smith & Muckli, 2010; Tang et al., 2018), yet do not represent the same level of object detail present in realistic vision. It is currently unclear how V1 and higher visual areas engage with realistic occlusion, when objects are occluding other objects, inspiring my motivation to research this.

Secondly, while some researchers view predictive processing theories as a unified account of mental functioning, there is still some doubt over whether predictive processing is the best account of visual processing (Alink & Blank, 2021; Ransom et al., 2020), particularly under challenging visual conditions such as occlusion. We know from previous studies that context effects exist in V1 without visual stimulation (Smith & Muckli, 2010), however it is unclear from these results whether these effects are due to predictive processing. Using expectation suppression could help to answer these questions.

Therefore, this thesis has sought to address these gaps in the literature to extend the knowledge of visual object recognition and what mechanisms may be responsible for this process, using stimuli that have more complexity and ecological validity. Specifically, these experiments aim to add insight into mechanisms which have been neglected by studies so far by examining multiple object representations within occlusion, with particular attention paid to the EVC and IT. The experiments within this thesis make use of a novel stimuli set created to observe how single object presentations compare to multiple objects – by using objects as both occluding and occluded objects. In Chapter 2, fMRI was leveraged to investigate whether there are differences in decoding when observing objects versus when they were occluded by other objects. Chapter 3 sought to take this further, combining the fMRI data from Chapter 2 with a behavioural experiment using the same stimuli to understand how behavioural reaction time (RT) and accuracy data relate to neural representations of occluded objects. Chapter 4 takes a different approach to this, using fMRI and eye tracking with an expectation suppression design to understand whether predictive processing-like mechanisms implicated in processing of occluded objects in EVC. Finally, the general discussion will attempt to collate all evidence from the experimental chapters as well as the wider literature to speak on the findings and how the thesis has addressed the question of object recognition under occlusion.

Chapter 2. Neural representation of occluded and deleted objects in visual cortex

2.1. Abstract

The ability of the human visual system to recognize occluded objects is striking, but exactly how this is completed is unclear, particularly when multiple complex objects are presented. Previous studies investigating occlusion at both the behavioural and neural levels typically used simple shapes or cut-outs as occluders, rather than other objects. The goal of the present study was to understand what best explains neural representations of occluded objects under more realistic occlusion i.e., when objects occlude other objects. We approached this by explicitly relating activity patterns of occluded objects (e.g., a cup occluding a face) with those generated when viewing the same objects in isolation (the cup or the face). In an event-related fMRI design, participants (*N*=12) performed a one-back task while being presented with objects presented in isolation (un-occluded), occluded by another object, or cut out by a corresponding object silhouette. We defined anatomical regions of interest in EVC (V1-V3), mid-visual regions (V4/LO1-3) and IT. Decoding analyses showed that EVC responses to occluded objects were better determined by the visible features whereas in IT inferred features also explained the responses well. Our data also showed strong effects of competition across multiple object representations in EVC, although these were significantly weaker in IT. In sum our results demonstrate that IT better decouples responses to real-world occluded objects with robust representations evident across multiple competing objects. Thus, our data support the importance of investigating neural mechanisms underlying object recognition under more complex and naturalistic occlusion scenarios.

2.2. Neural representation of occluded objects in visual cortex

2.2.1. Significance of occlusion

Occlusion is constant in natural, three-dimensional scenes, with some degree of obstruction from one object to another being almost unavoidable. Partially occluded objects are recognised easily in the human brain, though Tang et al. (2014) noted there were delays of approximately 100ms for partial objects in contrast with whole objects. Evidence of robust object recognition under conditions of occlusion is not just unique to humans. There has been documented evidence in mammals, birds and even fish, possibly all inherited from early vertebrate ancestors (Sovrano & Bisazza, 2008).

The ability to identify the mechanisms of visual object recognition under occlusion will allow improvement in computer vision models. These often fall short in their ability to combat the multitude of variations occlusion involves, even when modified to handle mask occlusion (Zhu et al., 2019). Partial occlusions pose a challenge due to the reduction of visual evidence available. The human visual system is required to not only process the identity of the object under the condition of occlusion, but also the occluder and scene context to create a full understanding. This process is thought to require signals from the prefrontal cortex, V4 and the IT cortex (Fyall et al., 2017). The visual system is adept at compensating for the missing information in scenes that results from occlusion, but how this is done is not fully understood. In particular, the role of the occluding object in visual processing and its effect on the subsequent recognition of the occluded object is unclear. These ventral stream areas (Goodale & Milner, 1992), particularly the higher order visual areas, remain selective to occluded stimuli. Even when there is as little as nine percent of the original image available through occluding Gaussian 'bubbles' (Tang et al., 2014), incredibly robust neural selectivity is demonstrated.

However, many previous studies investigating occlusion at both the behavioural and neural levels (Johnson & Olshausen, 2005; Smith & Muckli, 2010; Tang et al., 2018) have typically used simple shapes or cut-outs as occluders rather than other objects (but see Spoerer et al., 2017), lacking some naturalistic or ecological validity. Teichmann et al. (2022) utilised more dynamic object occlusion methods, where a shape seemingly disappears briefly behind an occluding quadrant area. They found that object identity and luminance information is less important than the object position information, which was represented during occlusion for up to 200ms in predictable and unpredictable movement conditions. Thus, suggesting that the nature of object representation during dynamic occlusion is different from static perceptual visual recognition, further enforcing that research must strive to look at more naturalistic methods of presenting stimuli as far as possible. There is a body of research now that is seeking to move from reliance on the presentation of single objects, revealing that the representation of an object may be altered by the presence of other objects, though the exact mechanisms are still unknown, particularly within visual circumstances such as occlusion (Baeck et al., 2013; Chelazzi et al., 1998; Li et al., 1993; Reddy & Kanwisher, 2007; Rolls & Tovee, 1995; Zoccolan et al., 2005, 2007).

Amodal completion, a process in which missing parts of a shape are 'completed', linking the disconnected parts to a single 'gestalt', has been suggested as a potential explanation for occlusion recognition (Nanay, 2018a; Rauschenberger et al., 2006; Thielen et al., 2019; Weigelt et al., 2007). Contour information has been associated with early visual areas, while the IT cortex has been implicated in overall recognition. Amodal completion represents the occluded parts of objects we see, with the visual system completing the objects using visual contours and prior, predictive experience, employing top-down feedback driven by previous experience as well as recurrent bidirectional influences (Kafaligonul et al., 2015; Nanay, 2018a; Tang & Kreiman, 2017). EEG evidence indicates that early amodal completion effects are observed when recognising partially visible objects, facilitated when missing object information is replaced by an occluder rather than being completely removed (Johnson & Olshausen, 2005). Amodal completion is thought to be more applicable to natural images than the artificial images used in many lab-based studies (Nanay, 2018a).

Unsurprisingly, there have been prevalent effects across the lifespan regarding occlusion, with predictive mechanisms evolving to encompass the sensory experience which with aging may lead people to rely strongly on predictive processes based on prior experience and expectations of what they are perceiving (Rossel et al., 2022). There is also additional evidence to suggest that even infants perceive that objects persist during occlusion (Teichmann et al., 2022), emphasising the prominence of object recognition under occlusion and the need to understand this important process.

2.2.2. Importance of recurrent processing models

While object identity was previously considered to be extracted in a hierarchical process along the ventral object vision pathway (Baeck et al., 2013), in their review, Wyatte, Jilk and O'Reilly (2014), claimed that object recognition would not be possible if only top-down or bottom-up processing was utilised, requiring input from recurrent processing. Recurrent processing occurs where interconnected sensory systems involve both feedforward and feedback connections. These connections enable adjacent layers to interact locally and recurrently to refine representations and give rise to a dynamically extended network that can activate both local and widespread areas (Han et al., 2018). Flexible communication of top-down and bottom-up influences has been found to allow enhanced representations of objects (Yan et al., 2023). This recurrent processing can strengthen bottom-up signals when objects are viewed under conditions of occlusion, poor lighting, shadows, and other variable factors. Thus, facilitating the creation of a strong, stable representation of an object that allows robust recognition (Wyatte et al., 2012).

Furthermore, vision has been confirmed to be a highly dynamic process reliant on multiple areas of the ventral stream for different perceptual processes. Accordingly, recurrent connections have become a plausible way by which to predict visual features of heavily occluded objects. These connections have been shown to successfully capture the physiological delays observed throughout the ventral visual stream, suggesting indirect evidence for object completion occurring in the IT cortex (Tang et al., 2018). Tang et al. (2018) demonstrated the presence of pattern completion when objects are poorly visible or occluded. Even when objects were occluded, or had sections cut-out, using gaussian bubbles until they were less than 15 percent visible, recognition was still robustly successful. However, participants were worse at recognising occluded objects when a backwards masking task was employed, presumably through the interruption of recurrent processing. In addition, the researchers determined that while standard feed-forward models were not robust to occlusion, using recurrent neural networks enabled much better recognition when objects were occluded or obscured. Their recurrent computational model, adapted from feed-forward AlexNet architecture (Krizhevsky et al., 2012), was used to determine the difficulty humans would have in recognising occluded stimuli, accounting for physiological delays along the visual stream, with or without the added difficulty of a backwards mask. There have been anatomical findings to indicate massive levels of recurrent feedback connectivity throughout the visual streams (Felleman & Van Essen, 1991; Sporns & Zwi, 2004). Therefore, recurrent connectivity should be considered when determining the mechanisms driving object recognition when pattern completion is required.

Several tasks have been developed with occlusion in mind, Spoerer, McClure and Kriegeskorte (2017), utilised a computational modelling method to create tasks demonstrating that recurrent networks outperformed feedforward control models during occlusion. They used digits occluding each other using whole or fragmented forms of meaningful digit stimuli for both occluder and occluded object, hypothesising that recurrent dynamics improved the recognition performance in conditions of occlusion. This work successfully demonstrated that recurrent networks outperformed feedforward models when performing tasks under conditions of occlusion. However, though these tasks effectively validated that occluding objects in more realistic ways was more applicable to real-world situations, there is still a way to go in creating tasks which occlude completely naturalistically. The digits themselves are simple stimuli to use, but a valuable starting point for future development. Spoerer et al., (2020) have since used ImageNet – a large dataset with natural images (Deng et al., 2009; Russakovsky et al., 2015) - to observe recurrent network model performance on this varied dataset. However, the stimuli were isolated objects, so recognition under occlusion is still yet to be determined by this.

Additionally, there are interactions between recurrent connectivity and learning, which predicts that high-level visual representations could be influenced by error signals from proximal brain areas throughout visual learning (O'Reilly et al., 2013). The robust way in which visual clutter or degradation is compensated for in the ventral visual stream may be attributed to recurrent connectivity, depending on the dynamic nature of the brain (Wyatte et al., 2012). Modelling of network connections in the human ventral stream by Kietzmann et al. (2019) established that information processing is largely affected by recurrence for the understanding of categorical divisions of objects.

Deep convolutional neural networks (DCNNs), that mimic the major principles of the visual pathway have been able to represent how the visual information is transformed from 2D image using feed-forward layer connections to analyse images. Bracci & Op de Beeck (2023) found that early layers of models correspond with retinotopic areas, whereas later layers capture aspects of higher-level representations. These models have been generally improved by the addition of feedback and recurrent connections. Recurrent neural networks (RNNs) are more neurobiologically realistic than feedforward counterparts and more robust in their ability to recognise objects, especially under challenging visual conditions (Spoerer et al., 2017; Tang et al., 2018; Yan et al., 2023). RNNs align effectively with predictive processing models, more efficiently coding information than other kinds of models (Ali et al., 2022). These networks establish more computationally efficient recognition, especially when combined with lateral connections (Richter et al., 2023), leading to more efficient RNNs. These would be primed by the predictions from prior visual experience and further split into subpopulations of prediction and error units respectively (Ali et al., 2022). The field of computer vision is arguably benefitting from continued use of recurrent convolutional connections.

2.2.3. Our motivation

Previous work demonstrated that it is possible to decode the identity of objects from the patterns of brain activity when viewed in isolation. However, objects almost never appear in isolation in daily life, therefore it is important to understand how the visual system simultaneously processes several objects. A study by MacEvoy and Epstein (2009) examined this, using MVPA to observe whether the lateral occipital cortex evoked activity patterns to pairs of objects that related to the activity patterns evoked by the same objects presented singularly. They observed, using searchlight analysis, that classifiers could significantly predict object pairs from averages of single-object patterns. Their results showed that the human lateral occipital cortex (LOC) and higher visual areas may be important in the ability to average responses and normalise them to support the coding and recognition of multiple objects simultaneously. Their ability to determine that pair patterns were decoded with high accuracy to synthetic patterns created from single objects inspires the question of whether the same would occur with occluded object pairs compared to single objects.

Thus, the results of MacEvoy and Epstein (2009) motivate the present study, as if objects were overlapping each other, the results may be even more informative about the process by which objects are recognised simultaneously in a scene. A similar pattern of results from this study may represent a process by which the hidden features of an object within an occluded pair are 'completed' by the higher visual areas, influenced by context clues from the visible objects and the prior experience and expertise of the visual system associated with the information available. Being able to evaluate the location of the activity in the visual system would shed light on whether the higher visual areas are responsible for this completion and under what conditions this is expected.

A study by Reddy et al. (2009) demonstrated that the multivoxel patterns of two objects presented simultaneously could be determined by averaging the sum of the two objects when each is presented in isolation. Specifically, that the activation seen for two objects presented as a pair was well-predicted by the summation of the activation of the two objects within the pairing each being shown separately. The biased competition theory proposes that objects compete for cortical representation in a network of mutual inhibition where there is a bias towards the attended item (Proulx & Egeth, 2008). This theory provided the basis for Reddy and colleagues' (2009) research, where they found that the overall influence of attention is largely independent of the category selectivity, but that attention can bias weightings in favour of an attended stimuli at the neural level. The researchers presented objects in four categories (faces, houses, shoes, and cars) individually or in pairs, where each category was either attended, unattended or with attention divided between the object images. They found that attention shifted the weight by approximately 30 per cent in favour of the attended stimuli, following the biased-competition framework and in line with primate literature (Fallah et al., 2007; Ramezanpour & Fallah, 2022).

Moreover, the average response to a pair of stimuli approximating the sum of the individually presented responses generates questions for the occlusion literature (Reddy et al., 2009). As a result, our study seeks to observe whether this phenomenon is present when the simultaneously presented objects are layered so that one occludes another. This recognition would require the visual system to 'fill in' gaps, necessitating increased attention. We approach this question by explicitly relating the activity patterns generated for objects under occlusion to those generated when viewing the same objects in isolation. The current study tests the assumption that when observing an occluded pair of objects, or one with a deleted 'cut out' section, that the approximate responses of the pair would correspond to the responses of the individually presented objects, as if the two images were presented side by side.

In our research, we employ a novel stimuli set comparing single object displays with occluded object pairings, aiming to shed light on the neural representation of occluded objects. Our primary idea is that when we analyse the neural activity using cross decoding, the IT cortex will represent occluded objects in a way that's more similar to how it represents unoccluded objects, having successfully utilised pattern completion. We do not expect to see this similarity in EVC, with the region being more susceptible to the competition effects of multiple objects being represented at once.

2.3. Methods

2.3.1. Participants

Self-reported right-handed healthy participants (*N* = 12; 4 Male, mean age = 28.25, *SD* = 5.34) participated in this fMRI experiment. All participants reported normal or corrected to normal vision and were deemed eligible after meeting MRI screening criteria. Informed consent was obtained in accordance with approval from the Research Ethics Committee of the MRC Cognition & Brain Sciences Unit. Participants received £22.50 for their time.

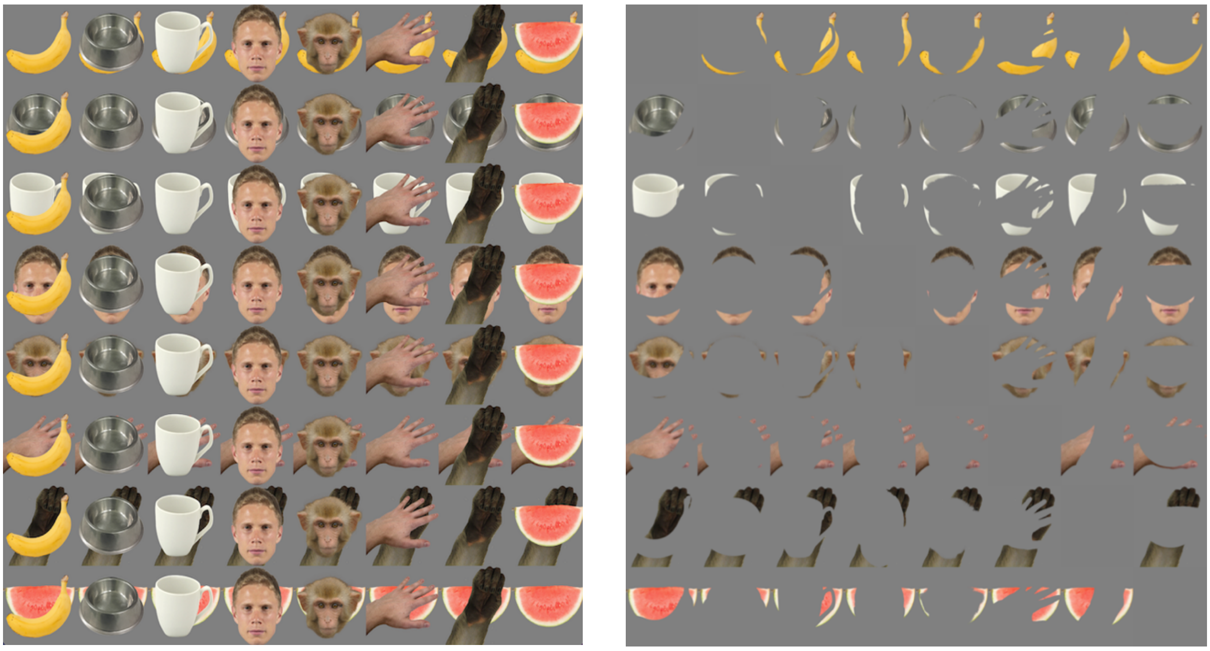
2.3.2. Stimuli and design

The study utilised a rapid event-related fMRI design where participants were presented with a set of object images presented either in isolation, occluded by another object or by a 'cut out' object silhouette (see Figure 2.1).

Building on previous research object recognition of multiple objects at once (MacEvoy & Epstein, 2009), this stimuli set has been specifically created to measure how occluded object pairs relate to single object presentations (Mansfield et al., 2023). The novel stimuli were made up of eight objects; banana, dog bowl, mug, human face, monkey face, human hand, monkey hand and watermelon, where these eight stimuli were unoccluded and represented the single object trials. These objects span various salient semantic categories (e.g., animate/inanimate, human/non-human, natural/artificial), which are easily recognised in human vision (Mur et al., 2013). There were 56 occluded and 56 deleted images comprised of all possible pairs of the eight whole objects visible in the occluded stimulus image. PNG images were presented in colour on a grey background at 799x799 pixels.

Figure 2.1.

The composite stimuli for the occluded (left) and deleted (right) conditions. The unoccluded images are visible in the diagonal of the occluded stimuli condition.



2.3.3. Procedure

Before the study commenced, participants saw an information sheet and gave informed consent to take part in the study. They each completed an MRI eligibility checklist to ensure their safety in the scanner and were talked through safety protocols and their ability to stop the scanning at any point if they were in discomfort. If they were eligible and happy to proceed, they were taken to the scanning room to begin the task. They performed some practise trials to ensure they understood the task required.

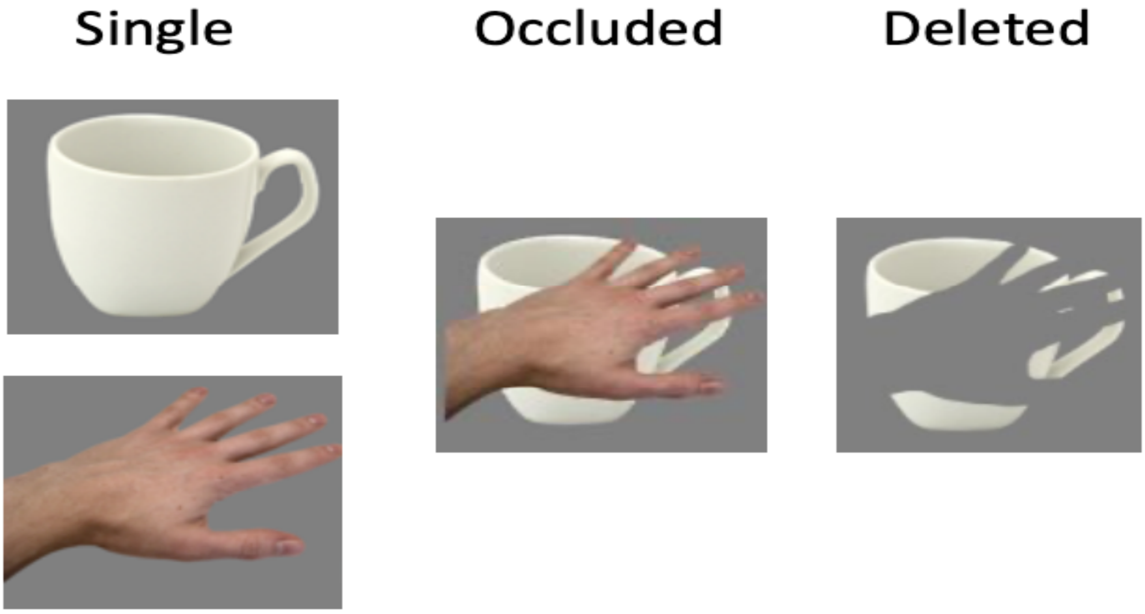
Participants saw a fixation cross which they were instructed to look at continuously. During each block, a stimulus was displayed on the screen for 1 second, followed by 2.5 seconds of fixation (3.5 ITI). The task involved a one-back repetition detection task using two response buttons on a button box, where one button should be pressed when any object in trial

N was shown on trial N-1 and the other button should be pressed when there were no repeated objects (front or back objects were both to be considered in occluded pairs). Null events accounted for 20 percent of trials.

Participants took part in up to 6 runs of the main experiment (*M* = 5.83), with each run lasting 9 minutes 44 second (128 trials per run: 56 occluded object pairs, 56 deleted pairs and 8 single objects repeated twice, see Figure 2.2). In a localiser scan, participants viewed colour images of faces, places, objects and scrambled versions in a block design. Stimuli were presented on a uniform grey background. Each block lasted 16s (444ms stimulus duration, no gap) interspersed with 16s fixation blocks. Four blocks of each stimulus type were presented within the run. This resulted in a total run time of 6 minutes 40s. Participants were also asked to lie still while an anatomical scan was run to allow clearer analysis of the areas of the cortex of interest. A 5-volume scan that lasted 30 seconds was acquired in the opposite phase encode direction (posterior to anterior) for every participant. Participants were debriefed after the completion of the scanning. The entire scanning session, including behavioural training, the localiser run and anatomical scan, lasted no more than two hours.

Figure 2.2.

An example of an occluded object pair, a deleted object pair and the two single objects both pairs are comprised of.



2.3.4. MRI data acquisition

Structural and functional MRI data was collected using a high-field 3-Tesla MR scanner (3T Siemens Prisma, MRC Cognition and Brain Sciences Unit). High resolution T1 weighted anatomical images of the brain were obtained with a three-dimensional magnetisation-prepared rapid-acquisition gradient echo (3D MPRAGE) sequence (192 Volumes, 1mm isotropic). Blood-oxygen level dependent (BOLD) signals were recorded using a multiband each-planar imaging (EPI) sequence: 471 volumes, TR = 1240ms; TE = 30ms; flip angle 74; 34 slices, matrix 78 x 78; voxel size = 2X2X2; slice thickness 2mm; no interslice gap; field of view 192; multiband factor 2, Partial Fourier = 7/8, no Grappa. The visual display was rear projected onto a screen behind the participant via an LCD projector. A 5-volume scan was acquired in the opposite phase encode direction (posterior to anterior) for every participant to allow distortion correction to be completed.

2.3.5. MRI data pre-processing

Functional data for each experimental run, in addition to localiser runs was pre-processed in Brain Voyager 20.4 (Brain Innovation, Maastricht, The Netherlands; Goebel et al., 2006), using defaults for slice scan time correction, 3D body motion correction and temporal filtering. Functional data were intra-session aligned to the pre-processed functional run closest to the anatomical scan of each participant.

Distortion correction was applied using COPE 1.0 (Breman et al., 2020; Fritz et al., 2014), using the 5-volume scan acquired for each participant. Voxel displacement maps (VDMs) were created for each participant, which were applied for EPI distortion correction to each run in turn.

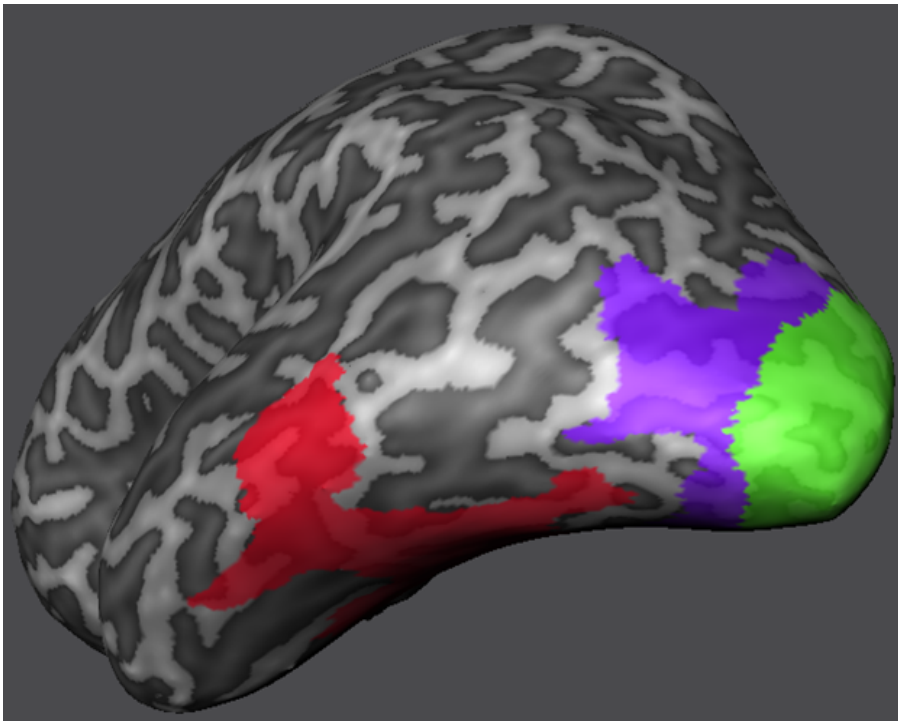
Functional data were then coregistered to the participant's ACPC anatomical scan. Note no Talairach transformations were applied, since such a transformation would remove valuable fine-grained pattern information from the data that may be useful for MVPA analysis (Argall et al., 2006; Fischl et al., 1999; Goebel et al., 2006; Kriegeskorte & Bandettini, 2007). For the main MVPA analyses (described further below) we conducted a GLM analysis independently per run per participant, with a different predictor coding stimulus onset for each trial (*N*=128: 56 occluded object pairs, 56 deleted pairs and 8 single objects repeated twice) presentation convolved with a standard double gamma model of the haemodynamic response function (see (Greening et al., 2018; Smith & Muckli, 2010). The resulting beta-weight estimates are the input to the pattern classification analyses described below (see multivariate pattern analysis). A GLM with 128 trials per run was used, with separated GLMs by run for decoding.

2.3.6. Anatomical regions of interest

Anatomical regions were created in Free Surfer using the Glasser Parcellation (Glasser et al., 2016) for early visual cortex (V1-3), mid-visual regions (V4/LO1-3) and IT as in Kietzmann et al. (2019) from each participant's anatomical MRI scan in ACPC space. The top 1600 voxels showing the strongest response from each bilateral region of interest were defined by an independent functional localiser (Faces, Places, Objects – see Charest et al., 2014) and were used in subsequent analysis. The BrainVoyager co-registration procedure was used to align the native space anatomy from FreeSurfer with that of the functional data in ACPC space (Bailey et al., 2023). See Figure 2.3 for a visual representation.

Figure 2.3.

An example of a surface from FreeSurfer with labels for the areas EVC (green), mid-visual regions (purple) and IT (red), for one hemisphere of one participant in this study.



2.3.7. Multivariate pattern analysis

Analysis involved testing hypotheses by extracting single trial response patterns in specific regions of interest (IT, EVC, mid visual regions) and analysing these patterns with multivariate pattern analysis (MVPA; e.g., Haynes, 2015) decoding. A linear support vector machine (SVM) was trained and tested on independent data, using a leave one run out cross-validation procedure (Smith & Goodale, 2015; Smith & Muckli, 2010) to decode object identity in each condition (Single Objects, Occluded Front, Occluded Back, Deleted) for a basic decoding analysis as well as in cross-decoding and synthetic decoding analyses. The classifier always received single trial brain patterns of activity (beta weights) from one of the three ROIs, and the independent test data was tested on single trial activity patterns.

The LIBSVM toolbox (Chang & Lin, 2011) was used to implement the linear SVM algorithm, using default parameters (C = 1), which uses the 1vs1 method for multiclass classification. The activity pattern estimates (beta weights) within each voxel in the training data were normalised between -1 to 1, before being used in the SVM (Bailey et al., 2023; Greening et al., 2018; Knights et al., 2021; Muckli et al., 2015).

The test data were also normalised using the same parameters as in the training set, to optimise classification performance. To test whether group level decoding accuracy was significantly above chance, non-parametric Wilcoxon signed-rank tests were performed on all MVPA analyses, against the computed empirical chance level (Formisano et al., 2008; Greening et al., 2018), with all significance values reported two-tailed. We used a permutation approach – randomly permuting the mapping between each condition and each label, independently per run, to calculate the empirical chance level for each participant and each decoding analysis separately. We note here that the average empirical chance level across all participants, regions, and analyses was 0.126, with a standard deviation of .002.

Main effects and interactions with the ROIs were tested using a permutation ANOVA which was implemented using the permuco package in R Studio (Kherad-Pajouh & Renaud, 2015; RStudio, 2021). This procedure generated a distribution of parameter averages using 10,000 permutations of individual parameter values (Avery et al., 2021). Wilcoxon tests were used to follow up significant differences from the permutation ANOVA results. Graphs of the data were created using ggplot2 (Wickham, 2016). Tables of specific values from Wilcoxon tests can be viewed in the Appendices.

2.3.8. Basic decoding

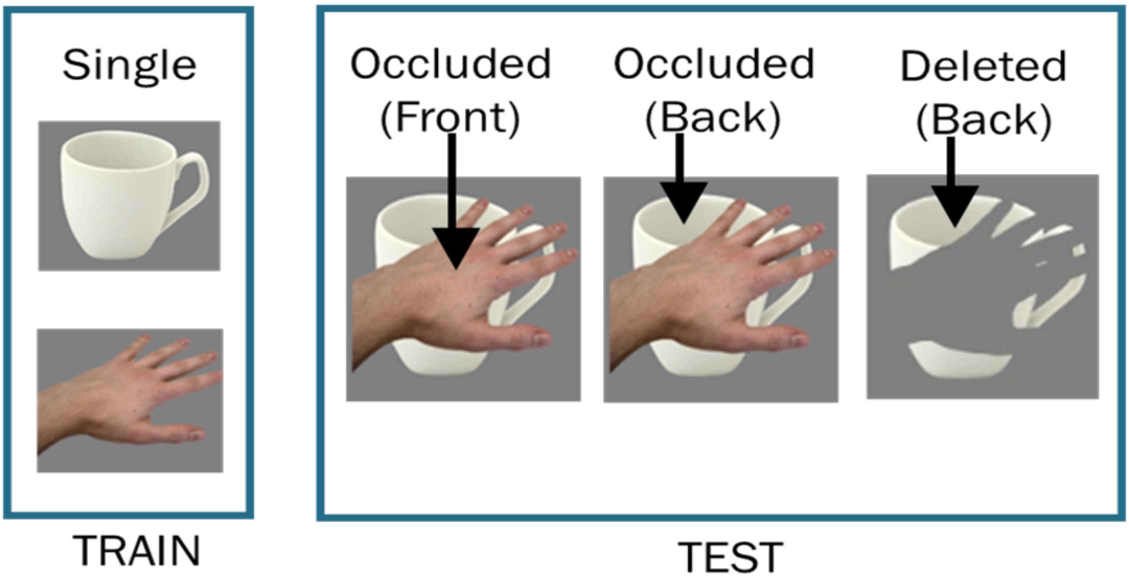
In the basic decoding analysis, the classifier was trained and tested separately for each main condition (single, occluded back, occluding front and deleted). This provides a simple gauge of how different the conditions are to the single objects presented in isolation.

2.3.9. Cross decoding 1: Comparing single presentations with remaining conditions.

The classifier was trained on single object presentations and tested on the remaining objects (front, back and deleted) (see Figure 2.4). This provides an index of how much the response present to single objects in isolation is present in each other condition. It allows an index of completion to be collected, as if there is completion occurring; for example, if the single to back condition had a higher decoding accuracy in IT than EVC it would suggest that there is more completion occurring in this area, perhaps having a higher reliance on the inferred features over the visible features.

Figure 2.4.

A visual representation of the first cross decoding condition, where the classifier was trained on single objects and tested on the remaining object conditions.

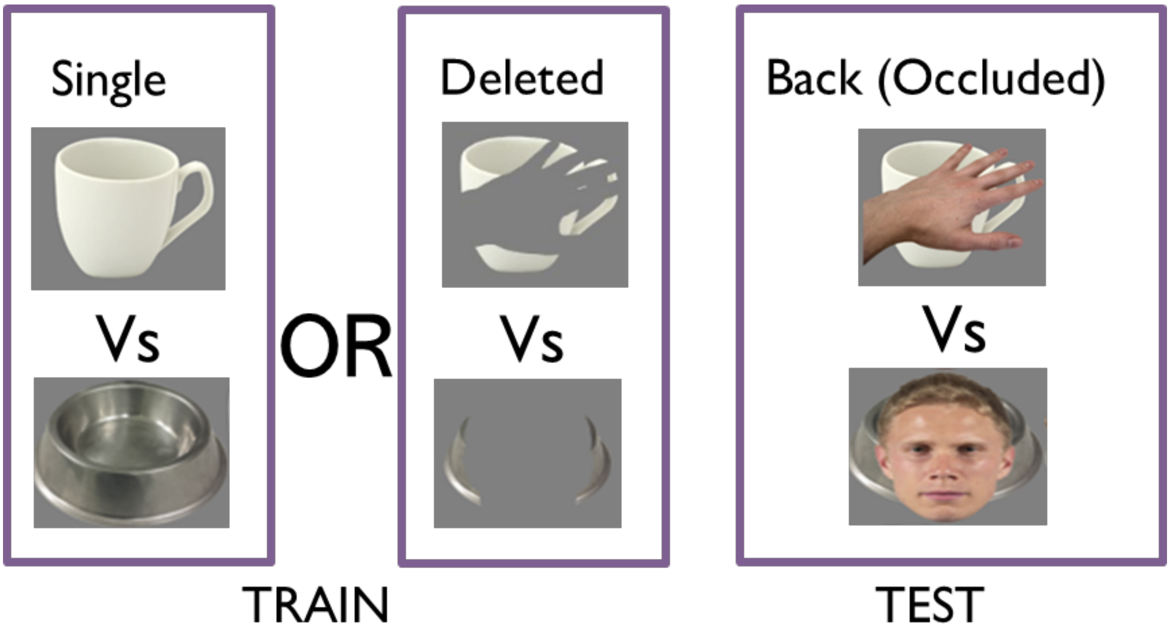


2.3.10. Cross decoding 2: Decoding the back object using single or deleted objects.

The classifier in this case was trained independently on single object presentations or deleted objects and tested on the back object for both conditions. This allows the classifier to be tested on two differing types of visual information: either the single object or the deleted pairing back object (see Figure 2.5). The rationale here is to test what best predicts the occluded object: the whole back object or just the visible part of it, as in the deleted trial. Testing this on the same information allows a demonstration of how capable the components of the visual system, from early visual to higher order, are at making these inferences of object identity when some visual object information is missing.

Figure 2.5.

A visual representation of the second cross decoding test and train data.



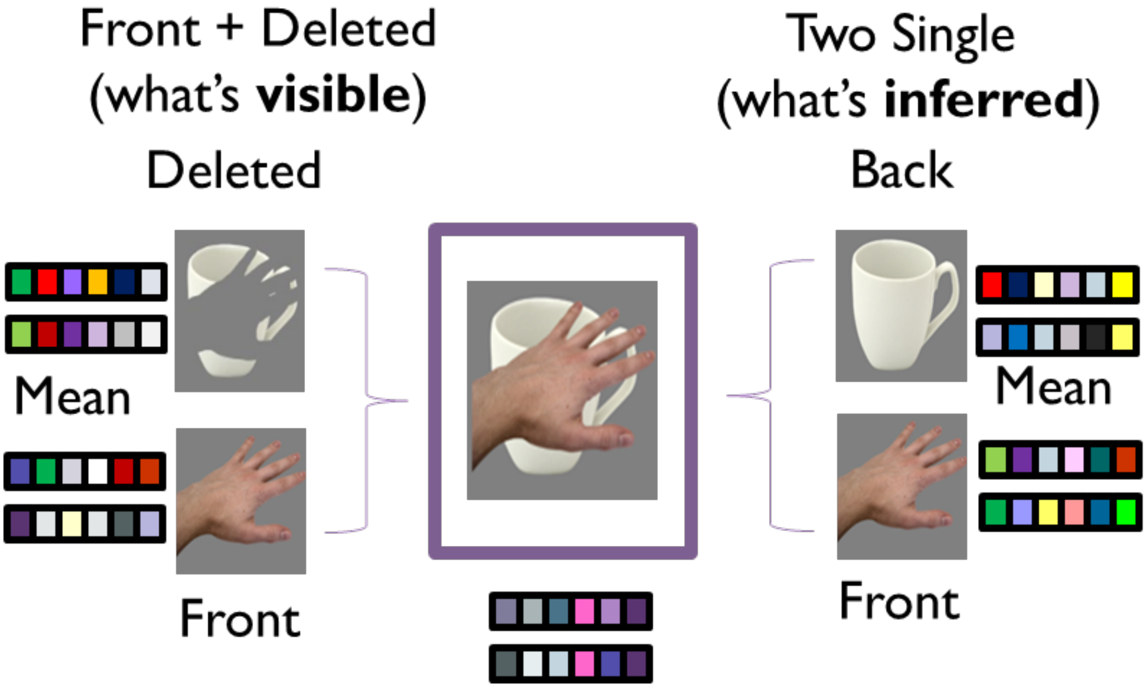
2.3.11. Synthetic decoding

For synthetic decoding, the values for either two single objects or front plus deleted object sections were used to create synthetic patterns (see Figure 2.6). Synthetic decoding analysis has been specifically motivated by the findings of MacEvoy and Epstein (2009) where decoded activation patterns of multiple simultaneous image presentations were directly comparable to those of the summed averages of the activation to the single objects (e.g. an image pair showing a cup and hand together - though not occluding each other - would have comparable activation to the summed activation of both an image of a cup and a separate image of a hand).

As well as this, MacEvoy and Epstein (2009) used synthetic decoding to great effect. They found that if they replaced pair pattern classification (which had previously demonstrated robust ability to predict multiple objects from their single objects) for half of the data with synthetic patterns created through the averages of the corresponding single objects, these synthetic patterns for pairs of objects were successfully recognised, particularly when using the mean of the composite objects. Therefore, to build on this within occlusion, rather than two simultaneously presented but distinct objects, synthetic decoding allows us to target what best explains the activity patterns that account for occluded pairs. The front plus deleted condition represents what specific visual features are visible while the two single objects represent both the visible parts of each object, and what is potentially being inferred to 'complete' the object recognition.

Figure 2.6.

Visual representation of the synthetic decoding.



2.4. Results

2.4.1. Basic decoding

To assess the baseline decoding performance in each condition, the classifier was trained and tested separately for each main condition.

2.4.1.1. Wilcoxon

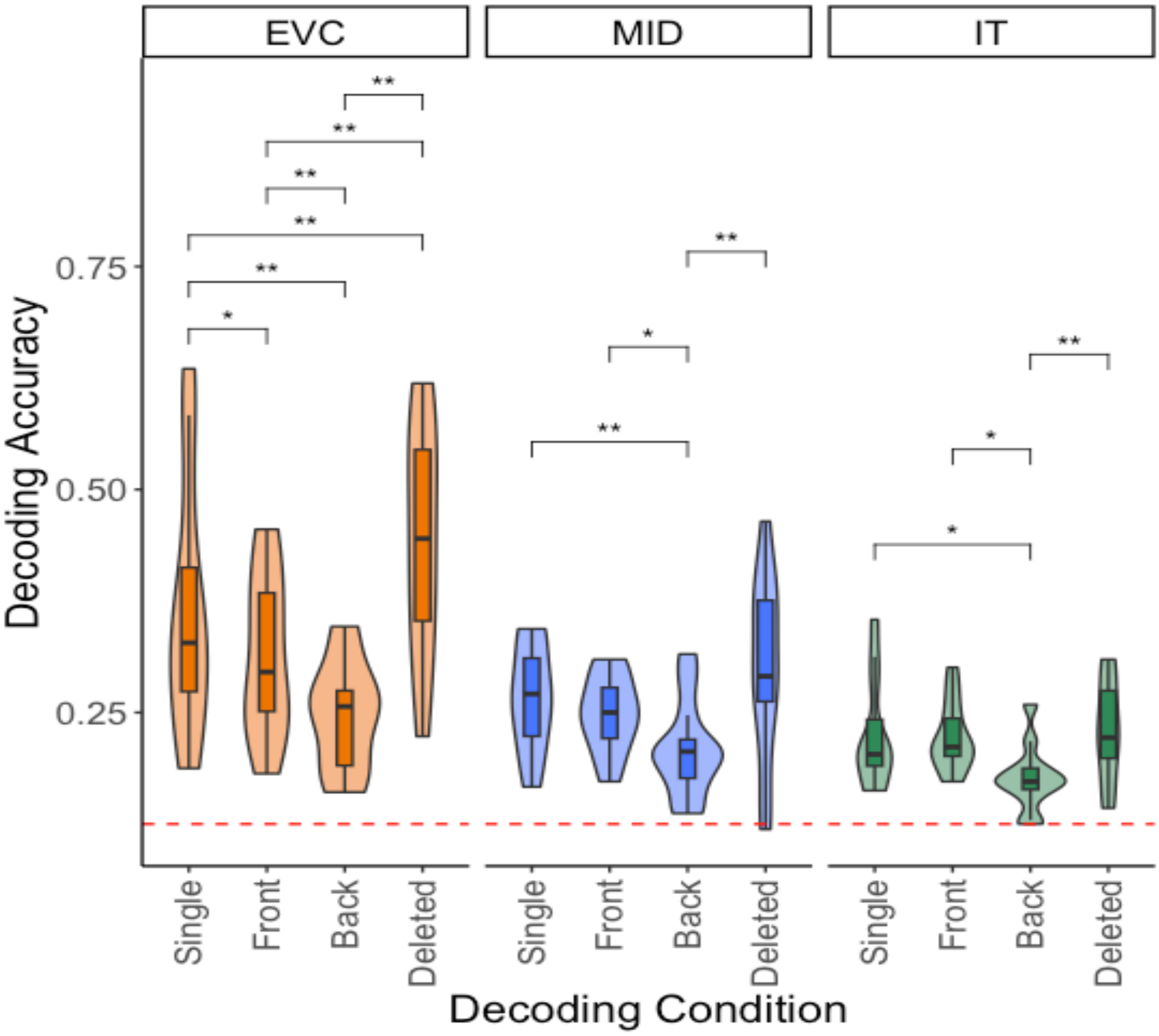
To test whether group-level decoding accuracy was significantly above chance we used Wilcoxon signed-rank tests against empirically derived chance levels (Bailey et al., 2023; Formisano et al., 2008; Greening et al., 2018), all significance levels two-tailed. This revealed decoding of object identity as significantly different from chance in each condition (single, occluded front, occluded back and deleted) in each brain region (EVC, MID, IT), with all *p*'s ≤ .001, all *d's* >= .884, using signed rank two-tailed tests versus subject-specific empirical chance levels.

2.4.1.2. ANOVA

A permutation ANOVA on condition (single, occluded front, occluded back and deleted) and region (EVC, Mid-visual regions, IT) was run (see Figure 2.7 for visual representation of each condition across regions). This revealed a significant effect of condition *F*(3,33) = 27.65, *p* < .001, as well as a significant effect of region, *F*(2,22) = 32.13, *p* < .001. There was also a significant interaction between brain region and condition, *F*(6,66) = 9.77, *p* < .001.

Figure 2.7.

Violin plot showing the basic decoding accuracy across either single objects, the occluding front, occluded back or deleted back objects in areas of visual cortex. The classifier was trained and tested separately for each main condition. Lines between conditions represent significant differences within regions as determined by pairwise wilcoxon tests, FDR corrected.



Note. * *p*<.05, ** *p*<.01, *** p<.001, ****p<.0001.

2.4.1.3. Post-hoc tests

After observing the results from the main analysis, post-hoc tests were conducted to further examine the differences between decoding conditions across brain regions. Wilcoxon signed rank tests were performed for each pair of decoding conditions, and the significant pairings, were corrected for multiple comparisons using FDR adjustments. The non-parametric Wilcoxon test was used as a result of completing the permutation ANOVA.

Among the significantly different pairs, the pairing of Occluded Back – Deleted remained significant in each brain region, as was Occluded Back – Occluded Front. Additionally, the Occluded Front – Deleted pairing was found to be significant in the EVC, but not in the higher visual region of IT.

2.4.2. Cross decoding 1: Comparing single presentations with remaining conditions.

The first cross decoding used the classifier was trained on single image presentations and tested on each of the remaining conditions (back, front and deleted).

2.4.2.1. Wilcoxon

Wilcoxon signed rank tests against empirically derived chance levels revealed decoding of object identity as significantly different from chance in each cross-decoding condition (single to front, single to back, single to deleted (visible) object) in across brain regions. This revealed decoding of object identity as significantly different from chance in each condition, across each brain region, with all *p* values ≤ .001, all *d's* >= 2.12; signed rank two-tailed test versus subject-specific empirical chance level, FDR corrected

2.4.2.2. ANOVA

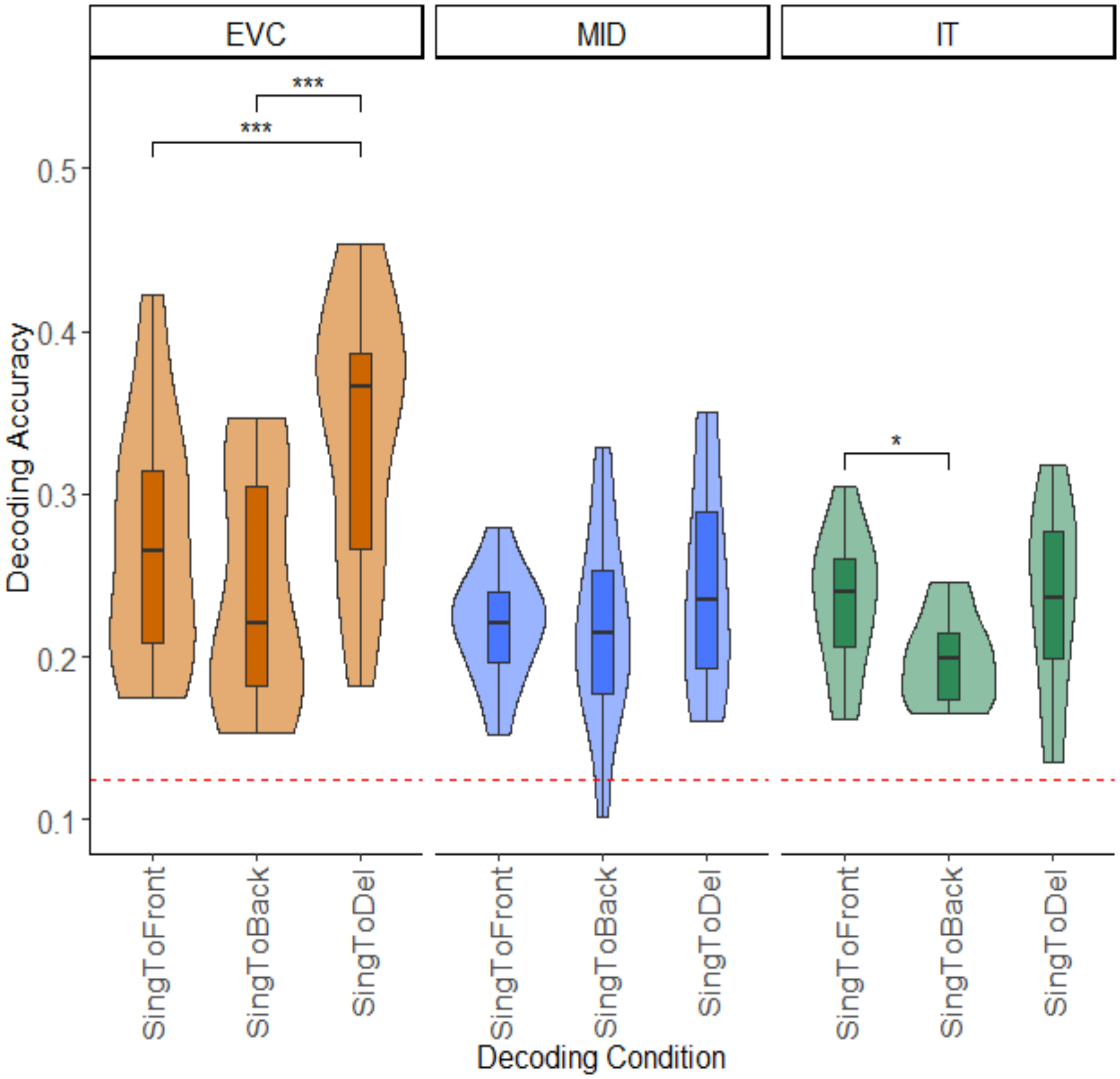
A permutation ANOVA on condition (single to front, single to back, single to deleted back object) region (EVC, Mid-visual regions, IT) was run (see Figure 2.8 for visual representation of each condition across regions). This revealed a significant effect of condition overall, $F(2,22) = 20.13, p < .001$. There was also a significant effect of region found, $F(2,22) = 11.35, p < .001$. There is a significant interaction between brain region and condition overall, $F(4,44) = 6.28, p < .001$.

2.4.2.3. Post-hoc tests

The statistical analysis involved conducting Wilcoxon signed rank tests for each pair of decoding conditions across brain regions. The results, presented in Figure 2.8 show that specifically in EVC, the pairings of single to deleted – single to front and single to back – single to deleted were both found to be significant. This may be due to a competition effect where in the occluded pairings there are multiple object representations to reconcile, causing lower decoding accuracy, whereas the deleted condition only has one visible object and thus less competition visual information to be recognised. In the IT region, there are no such significant pairings from the deleted object condition, suggesting a higher tolerance to multiple object representations. The only significant pairing was single to back – single to front, perhaps simply due to the easier recognition of the front – unoccluded – object compared to the occluded back object. No significant pairings were observed in the mid-visual region.

Figure 2.8.

Violin plot showing the cross-decoding accuracy when the classifier was trained on single objects and tested on either front object, back object, or deleted object presentations. These are split across areas of visual cortex. Lines between conditions represent significant differences within regions.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

2.4.3. Cross decoding 2: Decoding the back object using single or deleted objects

In the second cross-decoding, the classifier was trained on the single or deleted objects and tested on the back objects.

2.4.3.1. Wilcoxon

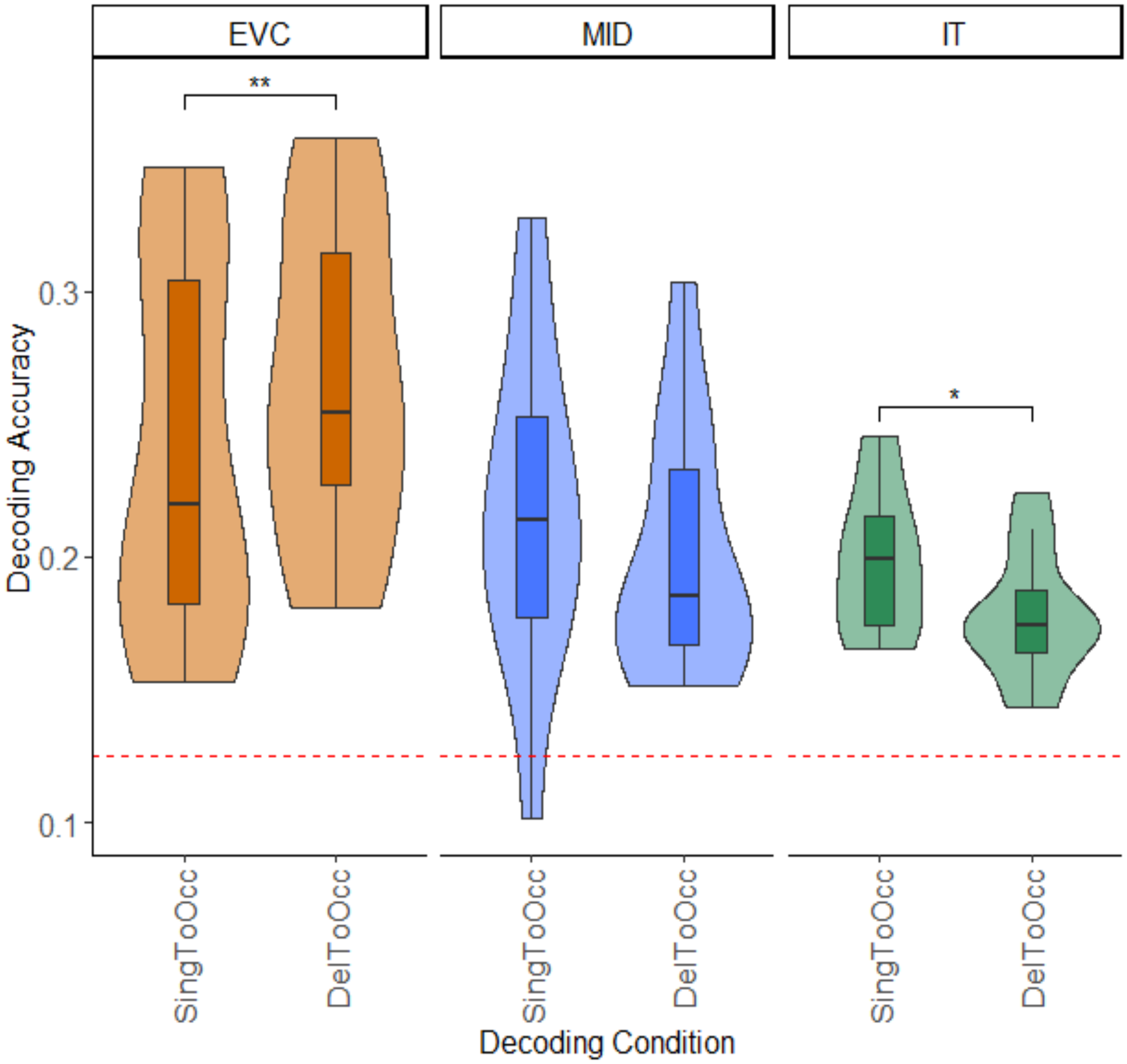
Wilcoxon signed rank tests against empirically derived chance levels revealed decoding of object identity as significantly different from chance in each cross-decoding condition (single to back and deleted to back (visible) object) across brain regions. This revealed decoding of object identity as significantly different from chance in each condition, across each brain region, with all p values $\leq .001$ and all *d's* >= 2.12; signed rank two-tailed test versus subject-specific empirical chance level.

2.4.3.2. ANOVA

A permutation ANOVA on condition (single to deleted back object and deleted to back object) and region (EVC, Mid-visual regions, IT) was run (see Figure 2.9 for visual representation of each condition across regions). This revealed no significant effect of condition overall, $F(1,11) = .768$, $p = .400$. There was a significant effect of region found, $F(2,22) = 10.62$, $p < .001$. Results revealed a significant interaction between brain region and condition overall $F(2,22) = 10.00$, $p < .001$.

Figure 2.9.

Violin plot showing the cross-decoding accuracy when the classifier was trained on single or deleted object images and tested on occluded back objects to predict back object identity. They are split across areas of visual cortex. Lines between conditions represent significant differences within regions from Wilcoxon tests, FDR corrected for pairwise errors.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

2.4.3.3. Post-hoc tests

Pairwise Wilcoxon tests were conducted for each pair of decoding conditions across brain regions, with significant pairings expanded in the Appendices. The significantly different pairings were observed between the conditions of single to occluded and deleted to occluded in EVC and in IT. These results emerge in opposing patterns between the regions. In EVC there is a higher decoding accuracy in the deleted to occluded condition, where the classifier was trained on only the visible visual information from the back, occluded object. Whereas this pattern is flipped for the higher visual regions. In IT we discover the single to occluded condition boasts higher accuracy, suggesting that this region is better able to activate the full object representation from a partially occluded stimulus, using the inferred details successfully.

2.4.4. Synthetic decoding

This decoding condition used either two single objects or front plus deleted object sections. This allowed the analysis of whether analysis of single objects or the front plus deleted sections that made up the occluded pair yielded more accurate representations of the occluded back object.

2.4.4.1. Wilcoxon

Wilcoxon signed rank tests against empirically derived chance levels revealed decoding of object identity as significantly different from chance in each synthetic-decoding condition (two single objects and front plus deleted objects) in each brain region (EVC, MID, IT). All *p*'s ≤ .012, all *d's* >= 1.34; signed rank two-tailed test versus subject-specific empirical chance level.

2.4.4.2. ANOVA

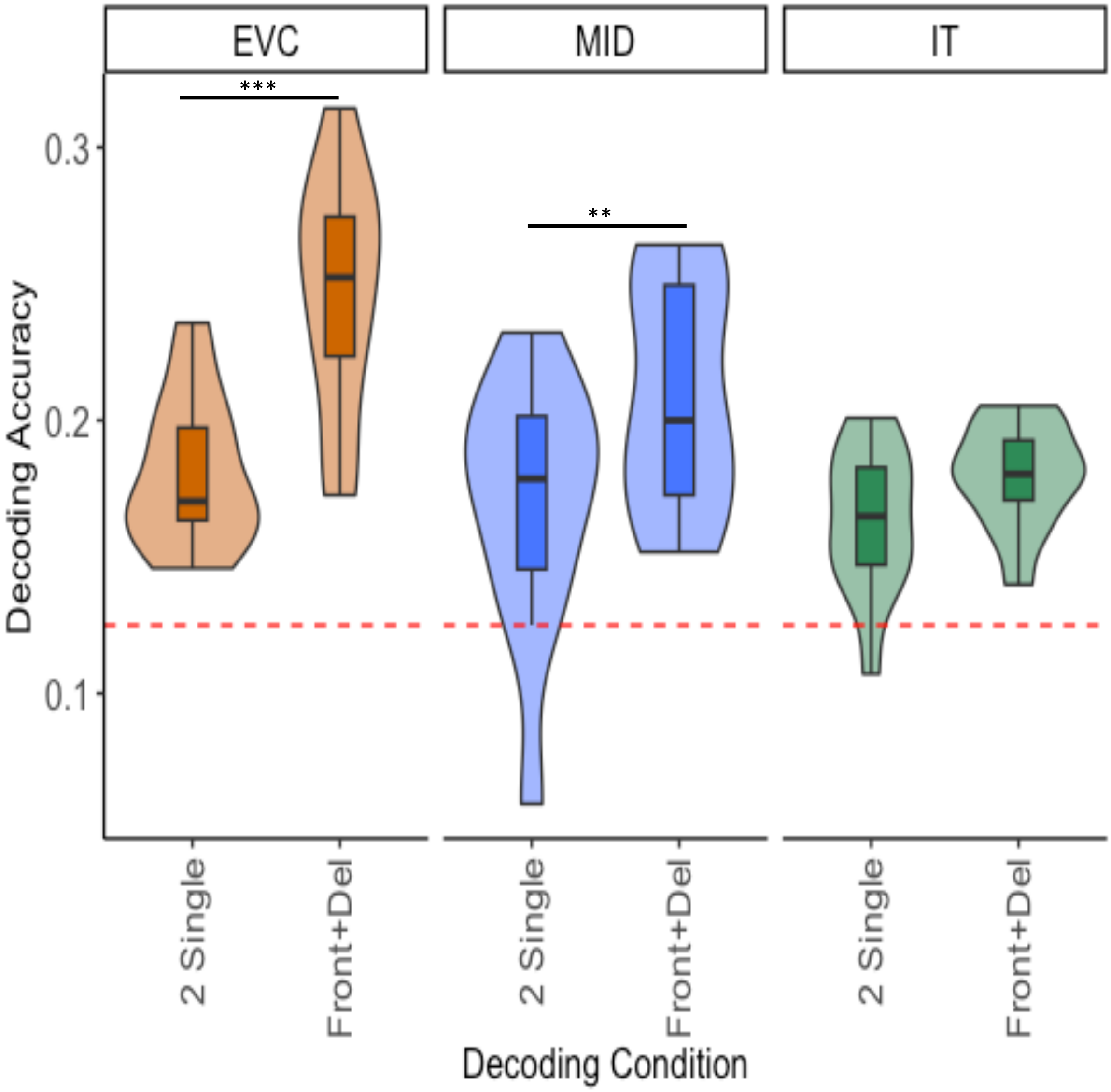
A permutation ANOVA was conducted to examine the effects of condition (two single objects and front plus deleted objects) and region (EVC, Mid-visual regions, IT). The visual representation of each condition across regions can be found in Figure 2.10. The results revealed a significant main effect of condition, F(1, 11) = 58.32, p < .001, and region, F(2, 22) = 9.28, p = .001. Additionally, there was a significant interaction between condition and region, F(2,22) = 7.71, p = .003.

2.4.4.3. Post-hoc tests

The statistical analysis involved conducting pairwise t tests for each pair of decoding conditions across brain regions. The results show significant differences (Figure 2.10). Specifically, there were significant differences in the EVC and mid-visual regions where the front plus deleted condition was more accurately decoded than the two single condition. There was no difference in the IT region. These results demonstrate the largest differences between the accuracy in conditions are in the early visual areas, with these differences diminishing by the time they reach IT. This may be indicative of the visual visible features being crucial to EVC, while in IT the higher-level models can also predict occluded object identity well, with decoding relatively stable in this ROI whether there were inferred features to contend with or not.

Figure 2.10.

Violin plot showing the synthetic decoding accuracy when training a model on two single objects as well as the front occluding object plus the back object in the deleted condition. These are split across the early visual cortex, mid-visual regions, and higher visual regions. Lines between conditions represent significant differences within regions from Wilcoxon tests, FDR corrected.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

2.5. Discussion

In the present study, we reveal how visual object representation takes place under conditions of occlusion and deletion. The results first demonstrated using Wilcoxon signed rank tests that each decoding condition was significantly different from chance. In line with our hypothesis, we found differences between EVC and IT visual areas, where IT was more tolerant of the presence of multiple objects than EVC. As determined by our first cross decoding analysis, the differences in accuracy in EVC when multiple objects were present were not observed in IT. Supporting our second hypothesis, we show that in IT, it is possible to decode the identity of multiple objects using both visible and inferred features. This finding was displayed using our second cross decoding condition, where IT demonstrated higher decoding accuracy after training on the single objects, which included inferred features, as opposed to the deleted condition that showed only the visible features from the back, occluded object. We also see this distinction in differences between visible and inferred features in synthetic decoding, where EVC could decode significantly more accurately when the pattern was created using only the visible visual information that would be seen in the occluded pairing. The pattern of results in IT differed from EVC, showing a relatively stable decoding accuracy across both the inferred and visible features.

The basic condition, where the classifier was trained and tested separately for each main condition on the occluded and deleted objects, demonstrated an effect of condition on decoding. The occluded back condition was significantly less accurate than any other condition, followed by occluded front and single then deleted back, particularly in EVC. There is a significant difference across brain regions, with EVC showing much higher decoding than higher visual areas which are more uniform and show smaller differences between conditions. The significantly higher decoding accuracy for the deleted condition compared to the occluded back condition in each region of interest, which both seek identification for the same amount of visible back object, represent the effect the visual clutter has on the computational effects of the brain when faced with multiple objects. The most significant difference between these two conditions was present in EVC, suggesting a greater reliance on lower-level visual components and less ability to segregate the two objects in the occluded pairing. The difference in the occluded pairing front object and deleted object condition is present in EVC but non-significant in IT. We posit that this is due to the additional information present in the occluded pairing, regardless of whether the front or back object is being recognised, is more computationally taxing for the EVC to process, leading to this lower accuracy in the front compared to deleted condition. In IT, this is not the case, with the ability to process multiple object representations at one time being more stable across conditions in IT.

This finding is partially in line with Johnson & Olshausen (2005), where occluders reduce the ability to accurately decode object identity. However, they found that the deleted condition was less accurate than the occluded condition. While their method still utilised the depth cue of having multiple shapes presented at once, they used ovals as the occluders and cut-out sections. Therefore, lacking meaningful visual information that an additional object brings to the object pairings. Our findings present multiple objects, increasing the amount of visual object data, causing competition effects across multiple object identities to emerge. Perhaps playing a critical role in the recognition of partially visible objects in challenging visual conditions across brain areas.

Using cross-decoding, we found that when comparing the single object condition to the conditions of occluding front, occluded back, and deleted back, EVC showed better decoding for the single to deleted condition than either of the others. This may reveal that the early visual areas were more affected by the presence of multiple object categories, as the deleted object condition was the only condition here showing a single object. This could represent impaired processing facilitated by the presence of additional object information. This effect is not seen across the higher visual areas, where only the front to back occluded conditions is significantly different in IT. This may represent a more robust tolerance to the competition effect of multiple object representations in the higher visual compared to the early visual areas, where in the higher areas there is the successful representation of higher level shape information rather than simple image features (Kourtzi & Kanwisher, 2001). This difference in IT between the front and back conditions may simply reflect the occluded front condition being easier to process than the back object due to the lack of occlusion over the front object.

The second cross decoding condition, investigated if occluded trial object identity was best predicted by the visible stimulus features or inferred features, demonstrating significant differences between conditions in EVC and IT. However, the direction of the effects differed, with higher decoding in the deleted to back condition – representing the visible information – in the early visual areas. While in the IT region, higher decoding is revealed in the single to occluded condition. In line with research on simple object recognition (Haushofer et al., 2008; Kaiser et al., 2019; Sayres & Grill-Spector, 2006; Wischnewski & Peelen, 2021), as the information is processed further along the ventral stream, higher visual areas are better able to activate the full object representation from a partially occluded stimulus. Leveraging the present occlusion results against prior studies focusing on object recognition of separate objects is beneficial to ensure our results capture the complexity of object recognition as well as adding additional knowledge regarding challenging visual conditions.

Synthetic decoding revealed significant differences between conditions in EVC, with the front plus deleted condition being more accurately decoded than the sum of the two single objects. This is not the case in the higher visual areas, with IT having a nearly identical decoding accuracy for both conditions. This may imply that in EVC there is a greater reliance on visible features, whereas in IT the higher-level model involving the implied or 'hidden' visual features from the occluded pair was also found to predict occluded object identity well. Weigelt et al.'s (2007) research suggested that local contour information is processed in the EVC, while regions of IT cortex represent a more completed shape. Though their research primarily used 2D line drawings of shapes, our findings are comparable, with amodal completion potentially being evoked in IT to allow back object recognition despite the complexity of the occluding front object.

This suggests that, in line with MacEvoy & Epstein (2009), single object responses can be used to model occluded object pairs, specifically in higher ventral visual areas. This confirms that occlusion is robustly accounted for, with higher visual areas able to 'complete' occluded objects despite not having all the visual information present. These findings also align with the work of Reddy et al. (2009), where the ability to share attention between the two simultaneously presented object categories reflected an average pattern between the two individual object activity patterns. However, our data show that when predicting the identity of the back object of the occluded pair, this ability is impaired by the presence of additional, inferred, visual information in early and mid-visual regions.

In addition, as in Tang et al. (2018), we find that pattern completion – here the ability to recognise the occluded and deleted objects despite incomplete visual information – occurs throughout the mid visual and higher visual areas. This can relate to recurrent connections within the visual system, necessitating the combination of top-down predictions as well as bottom-up visual features and lateral connections. The shape representation of a completed object is better processed by the IT cortex, with richer responses completed across multiple objects. This competition within the recognition of multiple objects presented simultaneously requires the visual system to detect multiple overlapping contours, determining what contours belong to any one object (Ao et al., 2023).

The results found in the present study may benefit from future research which adds in an additional condition relating specifically to the missing part of the occluded object in isolation, e.g., the hidden section of the occluded object. Currently, without this condition, inferences are being made regarding the completion of objects and while they make logical sense, having a condition to support this experimentally would allow greater understanding of the visual system response to the missing section of an occluded object (Smith & Muckli, 2010). In addition, to add to the naturalistic method of stimuli presentation, having 3D images or videos displayed would aid in the ability to determine additional cues of depth, luminance and motion which have been found to represent important factors in recognition (Johnson & Olshausen, 2005; Teichmann et al., 2022).

Amodal completion requires a mix of feedforward, recurrent and feedback processes (Thielen et al., 2019). Therefore, adding in computational RNN models to this dataset would allow the greater knowledge of the ability of the visual system to recognise these challenging visual conditions to extend and improve computational methods when engaging in object completion (Ao et al., 2023; Wyatte et al., 2014). This could be achieved by training networks to both propagate information from layer to layer whilst also utilising lateral connections within convolutional layers to improve the proportion of correctly recognised objects as has been done previously by Spoerer et al. (2020). Creating a larger dataset of occluded pairs, perhaps created from image databases such as ImageNet could also add extra utility to this method, facilitating much greater numbers of stimuli to be processed than a human sample would have the attention for (Deng et al., 2009; Russakovsky et al., 2015). The field of object recognition under conditions of occlusion is key in the improvement of computational models of object recognition, having potential impact on technological advancements such as self-driving cars (Cheng et al., 2020; de Oliveira et al., 2023; Tang et al., 2014; Wu et al., 2020).

2.5.1. Conclusion

Overall, our results expose much poorer decoding of object identity when objects are occluded by other realistic objects compared to when the same information is rendered absent via deletion. Analyses reveal that EVC responses were better determined by visible features, whereas in IT the inferred features also explained responses well, a result particularly visible in the second cross decoding analyses where opposing patterns of decoding accuracy occur from EVC to IT. Our data for EVC demonstrates higher similarity between isolated single object presentations and deleted object presentations, rather than those under occlusion. These results reflect effects of competition across multiple object representations in EVC, though these are significantly weaker in IT. In sum our results demonstrate that IT better decouples responses to real-world occluded objects with robust representations evident across multiple competing objects, relying less on the lower-level visual features that seem to drive recognition in EVC. Recognising multiple objects presented simultaneously suggests an enhanced capacity to interpret and predict complex visual information, drawn from previous understanding. Our data support the importance of investigating neural mechanisms underlying object recognition under more naturalistic occlusion scenarios where complex visual processing is required.

Chapter 3. Determining the effects of occlusion and deletion on object recognition

3.1. Abstract

Occlusion is unavoidable in the visual world. Previous work using occluded stimuli has missed some of the nuance of viewing multiple objects simultaneously. Object recognition is known to be facilitated by the ventral stream. As objects are almost never shown in isolation, it is important to consider how simultaneously presented objects are encoded and recognised, particularly when one object occluded another. This study sought to use behavioural and neuroimaging methods to investigate how perception of occluded objects relates to neural activity evoked from single objects. We aimed to measure behavioural recognition as well as tying this to neuroimaging data to gain greater insight in to the process. Using an online study ($N = 33$) measuring RT and accuracy regarding recognition of occluded objects revealed a cost of processing multiple objects at once, where performance was significantly worse when multiple objects were present compared to the deleted condition, where the same features were instead cut out. Using linear regression to further expand on previous fMRI data ($N = 12$) combined with the results of the behavioural study demonstrated differences in early and higher visual stream areas. Where beta weights to occluded objects in EVC scale with the amount of the occluded object visible, IT areas are demonstrably better equipped to having multiple objects presented at once, correlating greater weights with more difficult recognition conditions, while this is not the case in EVC. This provides interesting insight into how multiple-object occlusion and recognition is processed in the visual system that future research could build upon.

3.2. Determining the effects of occlusion and deletion on object recognition

3.2.1. Visual object recognition

Visual object recognition, the ability to accurately discriminate named objects across a range of materials, size, positions, textures, and in the presence of other visual stimuli, is an important characteristic of human vision, with objects virtually never appearing in isolation (DiCarlo & Cox, 2007; MacEvoy & Epstein, 2009). This process is computationally taxing, with infinite possibilities of position, scale, illumination, and visual clutter to account for (Spratling, 2016). Despite this, the visual system is highly equipped to deal with this task, correctly identifying objects within 150ms of initial presentation (DiCarlo et al., 2012).

The visual system is hierarchically organised in distinct anatomical areas each functioning differently for specific roles (Felleman & Van Essen, 1991). The connections between these areas occur in several ways, using ascending feedforward, descending feedback, and lateral connections from the same hierarchical level (Kafaligonul et al., 2015). Although object processing has been commonly regarded as a feedforward process (Kietzmann et al., 2019), Wyatte et al.'s (2014) review claims that object recognition would not be possible if only top-down or bottom-up processing was utilised. Wyatte et al. (2012) also suggest that the incredible ability of the brain to recognise stimuli even when they are degraded, for example by occlusion or low contrast, stems from the recurrent connectivity of the ventral visual stream. Amodal completion is the ability to represent parts of a perceived object that have no sensory stimulation, for example when parts of an object are occluded (Nanay, 2018b). Using object stimuli spanning identities such as keys, cars, cannons, and fish, that were visually degraded using Gaussian filtering, Wyatte et al. (2012) carried out a visual study illustrating the limits of feedforward processing during object recognition. The completion effects harnessed in their study suggest that recurrent processing effects aid in identification of objects, as purely feedforward models do not explain effects in this manner (Ernst et al., 2019; Kietzmann et al., 2019).

Recent advances in machine learning have allowed DNNs to become more reliable models of object recognition, rivalling the representational performance of the IT cortex (Cadieu et al., 2014). The models are known to be the best current models of biological vision, inspired by the primate brain (Spoerer et al., 2019). There are discrepancies in the methods by which computer models and the human or primate brain recognises an image, with DNNs relying more on texture while humans rely on shape information (Kubilius et al., 2016). DNNs are also commonly feedforward and trained on a huge array of labelled images. RCNNs have been inspired by recent work on recurrent processes within the visual system and more similarly match the biological visual systems. They outperform purely feedforward models, with reaction times much more like primate visual cortex reactions in terms of the trade-off between accuracy and speed (Nayebi et al., 2018; Spoerer et al., 2019). Additionally, when looking into laminar brain circuits, it becomes clear that recurrent processing is a huge driver of learning and perceptual understanding (Jia et al., 2023).

In an MEG study by Kietzmann et al. (2019), where participants viewed a diverse set of object categories (human and non-human, faces and bodies, natural and manmade inanimate objects), results showed using both representational dissimilarity matrices (RDM) and DNN models that ventral stream visual dynamics arose from recurrent connections. They sought to utilise this technology to understand ventral stream dynamics, discovering that recurrent DNNs significantly outperformed feedforward architectures across all levels of the ventral stream, from early visual areas including V1-3, to LO and IT. Other works involving DNNs have maintained that even in rapid object identification, recurrent connections are critical (Kar et al., 2019; Kar & DiCarlo, 2021).

3.2.2. Occlusion

Occlusion is an ever-present feature in natural, three-dimensional scenes, with some degree of obstruction from one object to another generally unavoidable. The human brain can account for occlusion effectively, with studies showing partially occluded objects are successfully and accurately recognised (Zhu et al., 2019). There are delays of approximately 100ms for the recognition of partially visible objects compared to unoccluded objects (Tang et al., 2018). Though recognition capabilities of the brain for heavily occluded or even deleted objects are still extremely robust (Johnson & Olshausen, 2005; Rajaei et al., 2019; Tang et al., 2014; Tang et al., 2018).

The ability to recognise an occluded object or scene is said to require a degree of prior knowledge about the object itself (Wyatte et al., 2014). The visual system not only has to compensate for missing information about the object, but also process the object identity, occluder identity (dependent on whether an occluded object has been used or if information has been cut-out by simple deletion) and the scene context to create a fuller understanding. The brain areas implicated in these processes by prior research include the prefrontal cortex, V4 and IT cortex (Fyall et al., 2017). There are also interesting effects in V1 when observing scenes with occluded quadrants, where information regarding the missing section is observable, even without receiving direct feedforward input from that area (Morgan et al., 2019; Smith & Muckli, 2010).

Current research often focuses on objects presented in isolation, however this may prove unwise. Tang et al. (2018) demonstrated that the brain has a robust ability to recognise an object category even when objects were heavily obscured. Though this was not occlusion per se, as the images were partially visible but not 'blocked' by another object. Examining exactly how the brain processes the occluded, as well as occluding, object may be a critical consideration. Occlusion research tends to use either a black box or a scrambled noise mask as an occluder (DiCarlo & Cox, 2007; Johnson & Olshausen, 2005; O'Reilly et al., 2013; Smith & Muckli, 2010; Tang et al., 2018), which lacks some of the nuance that the visual system processes in daily life.

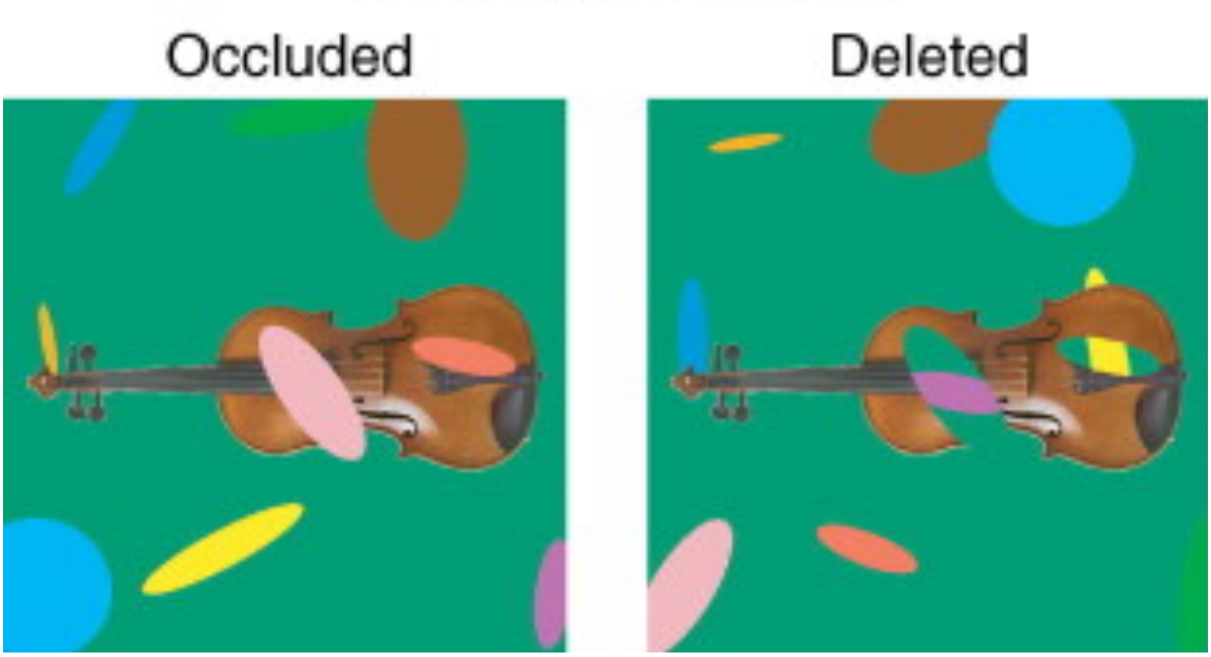
Consequently, there is a need to examine more naturalistic ways to represent real-world object occlusion. Real objects are rarely – if ever – seen in isolation, constantly overlapping and occluding in a variety of dynamic shapes and patterns. Using only deletion, or occluding using meaningless occluders, has provided important insights (Johnson & Olshausen, 2005; Smith & Muckli, 2010; Tang et al., 2018), yet does not represent realistic vision, generating limitations. It has been unclear how V1 and higher visual areas deal with realistic occlusion scenarios, incentivising research in this area. In Chapter 2, we demonstrated that occlusion was robustly processed in IT, but that in EVC there was a much larger cost of processing multiple object representations. In IT, we found that the visible features as well as the inferred features from multiple object representations predicted the occluded pairs well.

Johnson and Olshausen (2005) presented participants with images of real-world objects occluded or deleted with ovals obscuring an increasing percentage of missing pixels (see Figure 3.1). The occluded trials involved a cut-out object placed behind ovals that occluded a specified percentage of the image pixels. The deleted trials involved a similar partially visible object placed in front of colourful ovals while other oval shapes had been cut-out of the visible object. One occluded and one deleted version of each source object was created for their experiment.

The salient difference in the makeup of trials was created by the inferred depth of the occluded and deleted trials. They found that occluded trials were more easily recognised than deleted trials. Consequently, the results of this study suggest that the depth effect created by these inferences may be playing a critical role in the recognition of partially occluded objects, perhaps guided by amodal completion (Rauschenberger et al., 2006). Thus, researchers have suggested that depth cues may enable amodal completion prior to recognition, with figure-ground segregation being facilitated by grouping elements in the visual scene by proximity, common regions and connectedness (Rashal & Wagemans, 2022). However, the ovals Johnson and Olshausen (2005) employed as occluders are not fully representative of the natural visual world, being two dimensional, monochromatic, and meaningless, where there is ordinarily a great deal of competing visual information to perceive simultaneously.

Figure 3.1.

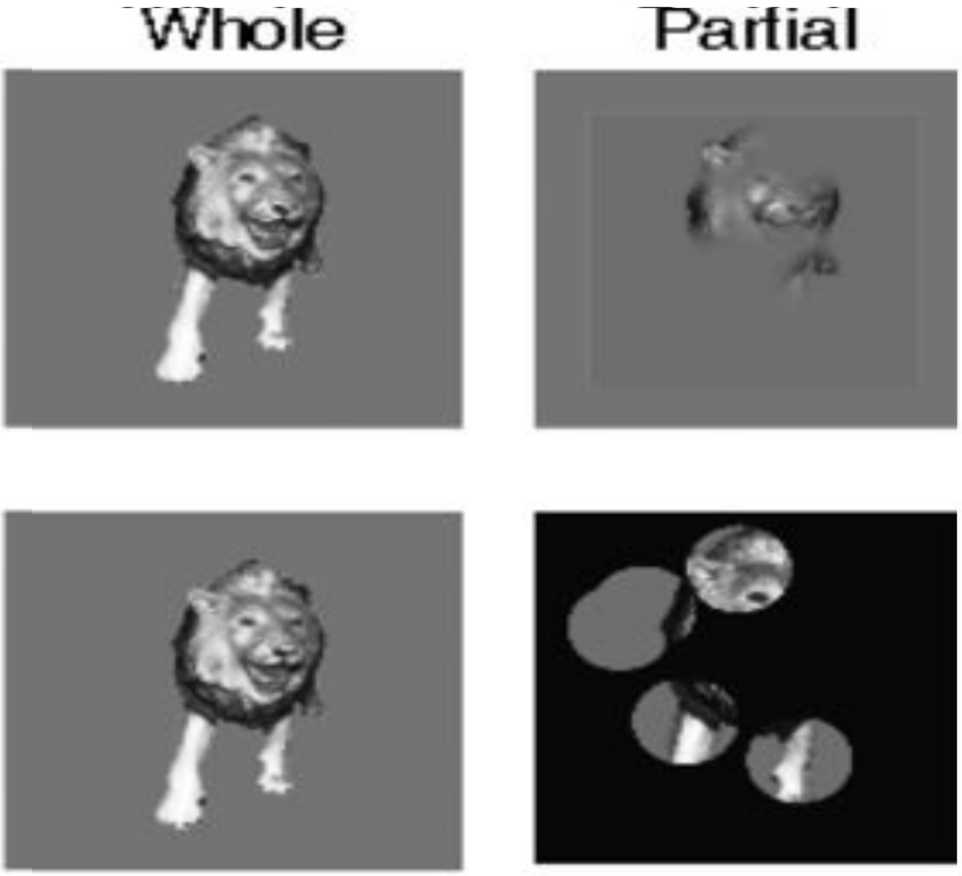
Examples of the stimuli used by Johnson & Olshausen (2005) for their experiments looking at occlusion and deletion.



By using masking to interrupt recurrent processing, Wyatte et al. (2012) found that occluded trials were significantly less accurately recognised than the less-occluded control condition, particularly when a mask was applied. The authors suggest that recurrent processing is important in recognising degraded stimuli and backwards masking acted as a successful measure of interruption for this process due to the mismatch that is created between feedforward and feedback responses (Lamme & Roelfsema, 2000). Whilst investigating category recognition in partially visible stimuli, Tang and colleagues (2018) used deletion of sections of objects, as well as a condition of occlusion (see Figure 3.2). They also used backwards masking to interrupt recurrent processing, seeing significantly decreased recognition speed in masking conditions. Even when 80 percent of an item was occluded, recognition was still found to occur in human participants, whereas using computational models, they saw an increase in performance that aligned with an increase in the visible percentage of an object. Between the deleted and occluded conditions, they found that performance was higher for the occlusion condition than the deleted, both with and without a mask. These studies demonstrate masking effects and show how pattern completion may be employed by the visual system using recurrent connections. While the occlusion in these studies was again very simplistic, it is useful to note the potential effects of recurrence and how masking could be reliably employed to interrupt this.

Figure 3.2.

Tang and colleagues (2018) presented stimuli presented whole and unaltered, rendered partially visible using deletion, or partially visible by occlusions using a black square with cut-outs to reveal some of the image.



Therefore, expanding this research using real-world objects as both occluders and occluded objects is key. This may present some differences in the accuracy of the judgements people make between deleted and occluded object pairs. Here the visual system must process two full objects in the occluded condition instead of just one and a cut-out created using ovals or Gaussian noise as in previous studies (Johnson & Olshausen, 2005; Tang et al., 2018; Wyatte et al., 2012).

In a study attempting to focus on the role of the occluder Spoerer et al. (2017) used numerical digits as stimuli for both the occluder and occluded objects. Digits, while not real-world objects themselves, are constantly observed and can be argued to be a useful comparison to real-world objects due to their frequency. A visual clutter study, where a digit was occluded by another digit, demonstrated that recurrent networks, but not feedforward mechanisms, were most successful when multiple object identification was required. Spoerer and colleagues (2017) found that delays in processing may be due to the necessity of identifying the occluding object as well as the occluded object. This nuance may have been missed in scenes where a meaningless occluder was used, ignoring the role of the occluding object. Therefore, while previous occlusion research has yielded interesting results (Johnson & Olshausen, 2005; Tang et al., 2018; Wyatte et al., 2012), the findings, particularly those regarding the differences between deletion and occlusion, may not be accurately representing visual capabilities in ecological conditions.

3.2.3. Our motivation

Here, we extend this type of paradigm to real-world objects as both occluded and occluding objects. Behaviourally, our goal was to determine whether performance was better for occluded or deleted trials when real objects were both the occluded and occluding objects, building on the work of Spoerer et al. (2017). The present study will use a recognition backwards masking study with the goal of interrupting recurrent processing to test whether recognition is superior under occlusion or deletion when images of real-world objects act as both the occluders and occluded. Using Gorilla, an online experiment builder with highly sensitive timing effects (Anwyl-Irvine et al., 2021), met the restrictions of the COVID-19 pandemic while still allowing a high-fidelity study to be carried out.

Our second goal was to determine how behaviour relates to neural representation of occluded objects by relating our data to that from Chapter 2 (Mansfield et al., 2023), expanding these analyses using linear regression. Reddy and colleagues (2009) utilised a method of combining the patterns of responses to each of two object categories presented in isolation across a plane in multidimensional space to predict the response when a pair of objects are presented simultaneously. They found that the paired object presentations could be expressed as a linear combination of the existing patterns from the original two stimuli. The use of this approach to model the occluded trials in terms of the constituent objects has been influential in the present study. Here we sought to predict each occluded trial activity pattern from the linear combination of the activity patterns of each of the constituent objects that made up that trial in isolation to get a beta weight value. We then utilised this to correlate with behavioural data for RT and accuracy to determine the effects on occluded object pairs.

3.3. Experiment 1

To address the question of how occlusion and deletion affect object recognition we presented participants with the same stimuli as in Chapter 2, but instead asked them to recognise object identity across conditions (single object, occluded pair front object, occluded pair back object or deleted pair back object). We measured both the reaction time (RT) and the accuracy of the object judgements. The stimuli were presented at presentation speeds of 33ms, 50ms and 100ms to maximise the effects of the masking design (Tang et al., 2018). We predict that the single and occluded pair front object conditions will be faster and more accurately recognised, as here there is not information to 'fill in' in order to recognise the object as is the case in the occluded back object and deleted conditions. It is also anticipated that increased presentation time will be associated with greater accuracy and faster RTs.

Specifically informed by the work of Johnson and Olshausen (2005) and Spoerer et al. (2017), we will test whether performance is better for occluded back object or deleted back object trials, to see whether adding multiple object representations changes the pattern of results. Prior research found that deleted images were less accurately recognised than occluded images, though these studies used meaningless occluders, hence the present study seeks to explore if this is the case when processing two types of objects simultaneously in object pairs. If there is in fact a cost to processing multiple objects at once (Spoerer et al., 2017), demonstrated through slower RT and lower accuracy, we would expect greater accuracy and speed in deleted over occluded trials.

3.3.1. Methods

3.3.1.1. Design and participants

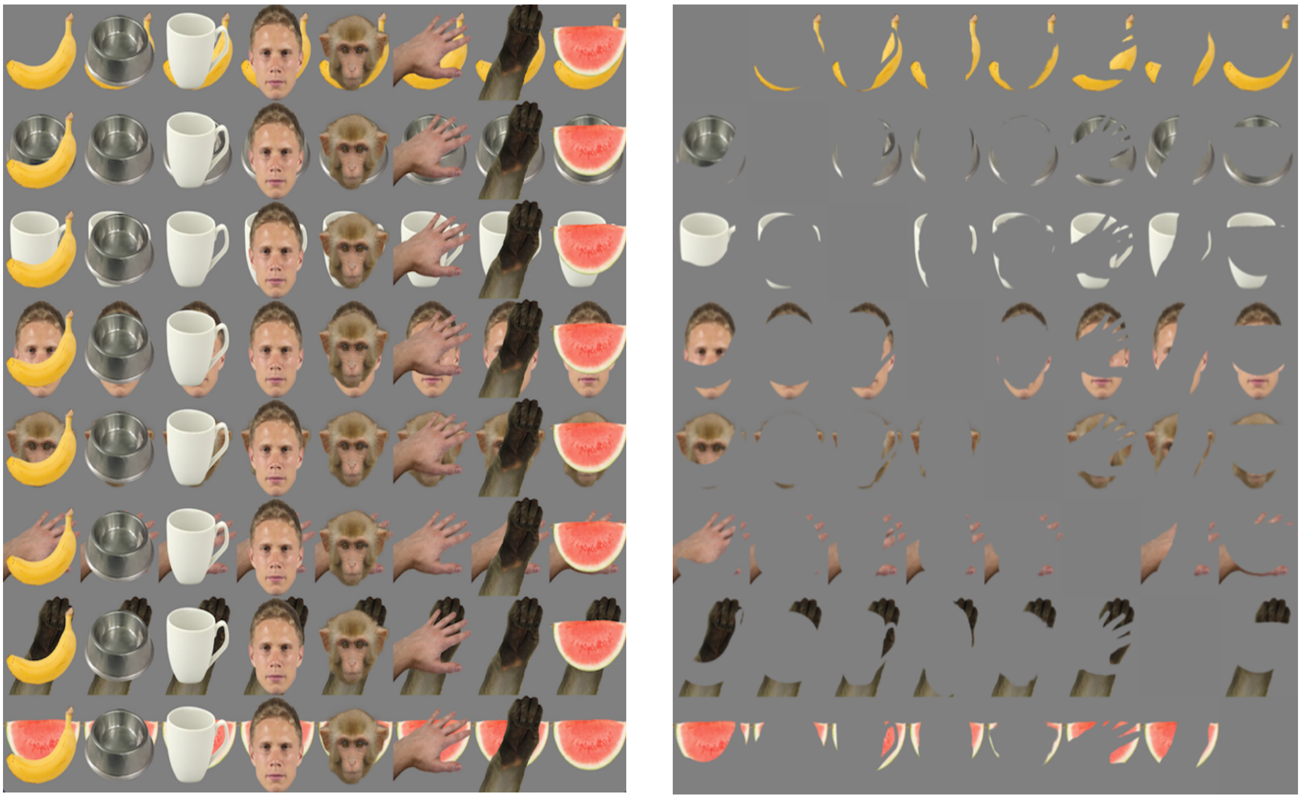
Data collection was completed using Gorilla, an online experiment builder with highly sensitive timing effects (Anwyl-Irvine et al., 2021). Single, occluded, and deleted stimuli were presented in four separate blocks (where occluded pairs were used in two separate blocks, one where the front object was the focus and one where the back object was the focus). Participants (*N* = 40) were told what object was of interest at the start of each block and reminded at each response screen. The participants were recruited through Prolific, where they received £8 an hour. Data was excluded for participants with an overall accuracy of less than 75 per cent across all trials (*N* = 7). The final sample included 33 participants (Mean age = 26.82, SD = 9.24, 17 Female). The study received full ethical approval from the UEA Psychology Ethics Committee.

3.3.1.2. Stimuli

The stimuli from Chapter 2 were used again for the online study, presented in colour (see Figure 3.3). A masking stimulus was created as in Tang et al. (2018) by scrambling the phase of these images, while retaining spectral coefficients.

Figure 3.3.

The stimuli in the occluded (left) and deleted (right) conditions. The eight images making up the single condition are seen in the diagonal of the occluded objects square.



3.3.1.3. Procedure

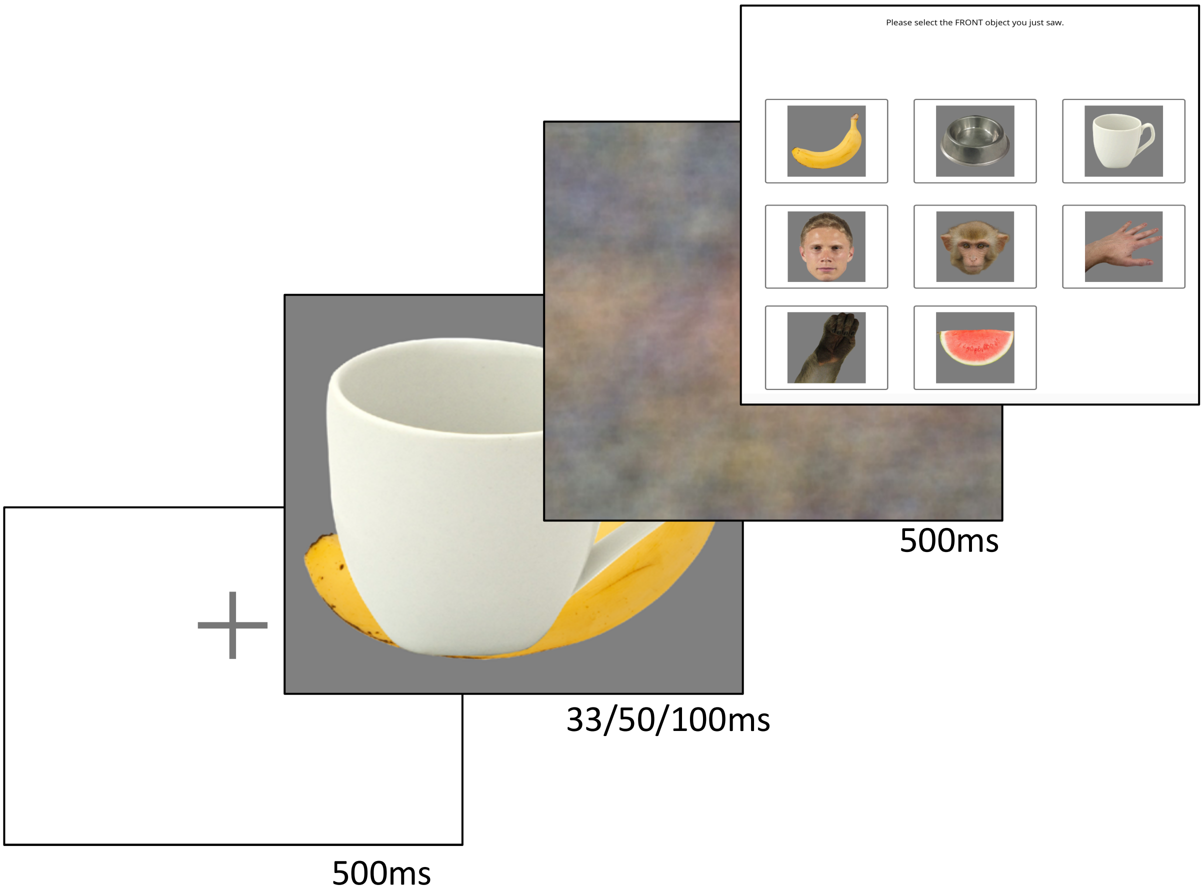
Using Gorilla experiment builder, the participants first read through an information sheet and clicked through a consent form before starting the task. The experiment involved three different configurations of stimuli (single, occluded and deleted). There were four different counterbalanced blocks in this study, one for every task condition where the occluded back and front objects were presented as the focus in two different blocks. Each object in every condition was presented three times, once for each of the three presentation speeds. The single condition block contained 24 randomised- order trials (eight objects at each of three presentation speeds) and required participants to select the object they had just seen. The occluded configuration images were used in two separate blocks, both containing the same 56 images. The occluded front block required participants to select the object they recognised at the front of the object pair (168 trials, 56 objects at each of the three presentation speeds), the occluded back block required participants to select the occluded, back object from the object pair presented (168 trials). The deleted block required participants to look at the object pair with the front object 'cut out' of the back object and select the identity of the back object (168 trials).

For each trial, a fixation cross was presented for 500ms, followed by a stimulus, presented for either 33, 50 or 100ms (see Figure 3.4). Informed by Tang et al. (2018), 33ms was chosen as the fastest speed due to suggestions that conscious awareness of a stimulus takes at least 30ms (Schräder et al., 2023). Tang et al. (2018) successfully found effects using variable image presentation times (25ms, 50ms, 75ms, 100ms and 150ms). Therefore, we were confident in our choices that there would not be significant floor or ceiling effects at these times. The scrambled masking stimuli was presented for 500ms. Participants were asked after each trial to select the object that they had seen in accordance with the instructions provided at the start of each block regarding which condition they were completing. There were 528 trials over the course of the experiment to show every object in each condition and speed.

After completing these trials, participants were asked to complete the Autism Quotient questionnaire (Baron-Cohen et al., 2001) and the Schizotypal Personality questionnaire (Raine, 1991) to measure traits of ASD and SPD. It has been suggested in the literature that these groups may have differences in their ability to undertake predictive processing and object encoding (Sterzer et al., 2018; Van de Cruys et al., 2014), hence there are differences expected in the results of backward masking interrupting recurrent processing (Sterzer et al., 2018; Van de Cruys et al., 2014). However, these questionnaires were not analysed further in this study.

Figure 3.4.

An example of a trial in the occluded front condition block. Participants were instructed to select the front object of this pairing.



Please select the FRONT object you just saw.

3.3.1.4. Analysis software

Data were analysed using SPSS statistics (29.0) as well as R (RStudio 2022.02.1), where the latter was also used for data visualisation and graphs. The mean RTs and accuracy values for each condition were analysed. Extreme outliers (>±2SD of the mean per condition) were removed from further analyses. Analysis of variance was run on presentation speed and display condition to observe interactions with reaction time and accuracy respectively. Pairwise t tests were used for post hoc testing.

3.3.2. Results

3.3.2.1. Accuracy

ANOVA

A two-way ANOVA was performed to analyse the effect of presentation speed (33/50/100ms) and display condition (single, occluded pair front, occluded pair back and deleted) on accuracy. The results indicated significant main effects for condition, F(1.91, 60.97) = 51.60, p < .001, η^2 = .621, and presentation speed, F(1.58, 50.57) = 49.16, p < .001, η^2 = .618. As well as this we see a significant interaction between the presentation speed and display condition, *F*(3.78,121.09) = 22.71, *p* < .001, ηp^2 = .429. Mean accuracies per condition are displayed in Table 3.1.

Table 3.1.

The mean accuracy for each condition separated by speed.

Condition and Presentation Speed	Mean (ms)	SD (ms)
33 Single	93.36	8.39
33 Front	88.28	11.89
33 Back	66.13	20.00
33 Deleted	78.91	14.64
50 Single	97.27	7.60
50 Front	94.64	7.20
50 Back	81.19	15.67
50 Deleted	90.07	9.04
100 Single	100.00	.000
100 Front	98.05	4.22
100 Back	96.93	2.77
100 Deleted	97.77	3.01

To further investigate the main effect of condition pairwise t tests of conditions showed significant differences between all pairs (*p* < .011). The lowest accuracy was found in the back condition (81.71 percent), followed by deleted (89.20 percent), front (93.66 percent) with the highest accuracy in the single condition (96.97 percent).

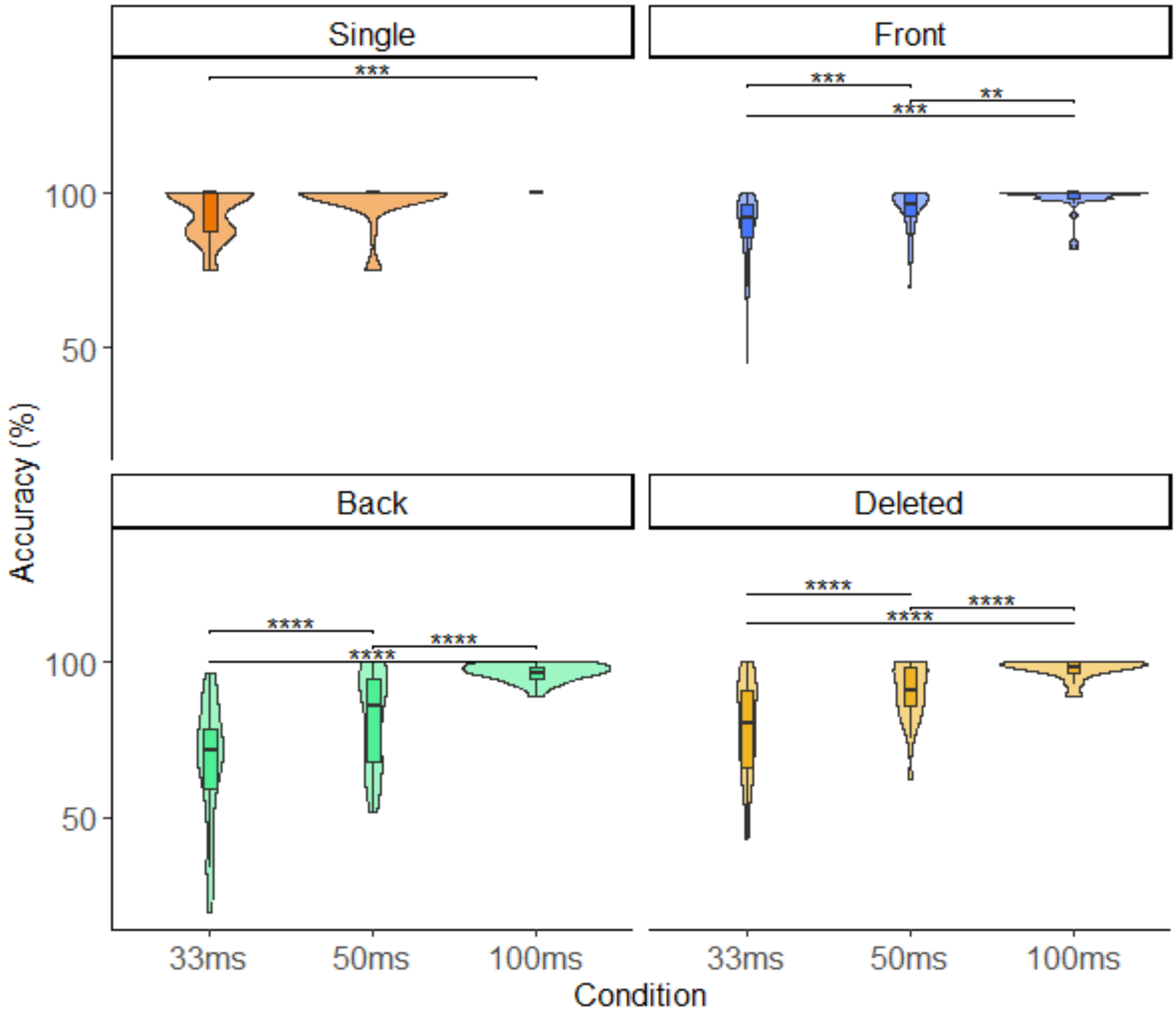
In presentation speeds, pairwise comparisons revealed that all pairs were significantly different (*p* < .001). With the pattern of highest accuracy in the 100ms presentation speed (98.15 percent) followed by 50ms (90.96 percent) and then 33ms (82.10 percent).

Post-hoc tests

Pairwise t tests were used to analyse the significant pairings within the conditions and presentation speeds (see Figure 3.5). These tests reveal a similar pattern for front, back and deleted conditions across speeds, but not single objects, where recognition is notably more accurate.

Figure 3.5.

Violin plots of accuracies for each condition and speed. Significant post hoc tests (pairwise t tests, fdr corrected) are represented by a connecting line, with significance denoted using asterisks.

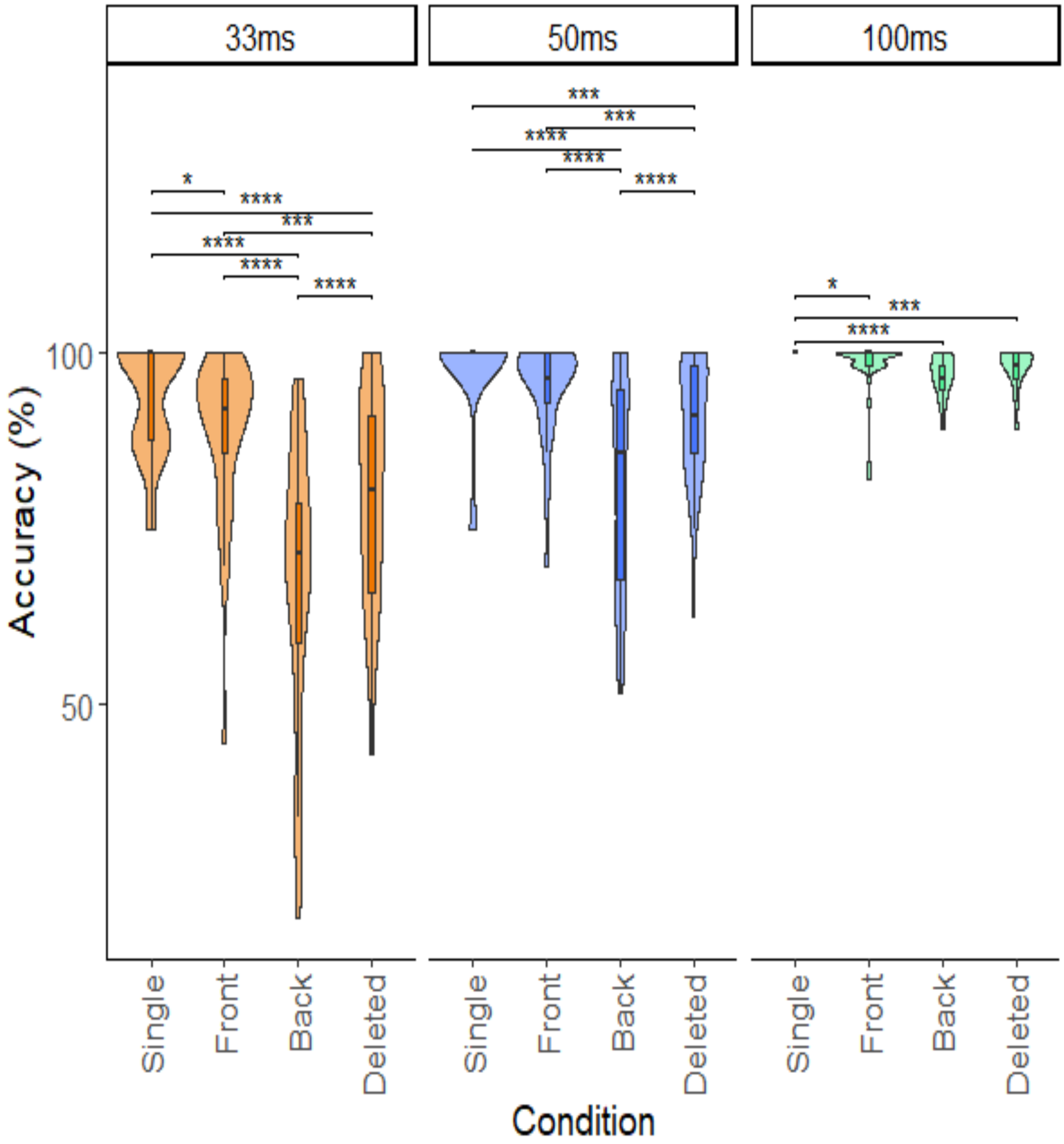


Note. * p<.05, ** p<.01, *** p<.001, ****p<.0001.

The results demonstrate that across speeds, there are significant variations (Figure 3.6). At 33ms speed, significant differences were observed between the single condition and all other conditions, suggesting that the rapid presentation time accentuates distinctions in task processing. Similarly, at 50ms and 100ms, a similar pattern emerged between single and other conditions. Though at 100ms, the deleted condition is not significantly different to the occluded pairing front condition, whereas at 50 and 33ms it is, which may be due to the easier recognition of the front object than the deleted object with its cut-outs until 100ms where recognition is easier to achieve robustly. Additionally, at all speeds there were significant differences found between front and back conditions, suggesting that there is a distinct difference in ability to recognise the occluded and unoccluded objects.

Figure 3.6.

Violin plot of speed and condition within that. Significant post hoc tests (pairwise t tests, fdr corrected) are represented by a connecting line, with significance denoted using asterisks.



Note. * p<.05, ** p<.01, *** p<.001, ****p<.0001.

3.3.2.2. Reaction time

ANOVA

A second ANOVA was conducted to look at the effect of presentation speed and display condition on mean reaction time and revealed significant main effects for condition, *F(*2.21, 70.59) = 6.937, $p = .001$, $\eta^2 = .178$, and presentation speed, *F*(1.96, 62.76) = 43.10, $p < .001$, $\eta^2 = .574$. Additionally, there was a significant interaction between condition and presentation speed, *F*(3.90, 124.81) = 5.001, $p = .001$, $\eta^2 = .135$. Mean speeds for each condition and presentation speed pairing can be seen in Table 3.2.

To examine the main effect of presentation speed further post-hoc pairwise t tests indicated significant differences between all three pairs of conditions (33ms-50ms, 50ms-100ms, 33ms-100ms; all *p*'s < .001). Overall, the slowest RTs were found in the 33ms condition (806.56ms) followed by 50ms (754.39ms) and 100ms (712.64) had the fastest RTs.

The effects of condition were further measured using post-hoc pairwise t tests, indicating a significant difference between reaction times for back (819.08ms) and front (757.61ms), ($p < .001$), back and single (750.81ms), ($p = .008$) as well as back and deleted (712.47ms), ($p < .001$).

Table 3.2.

The mean RTs for each condition separated by speed.

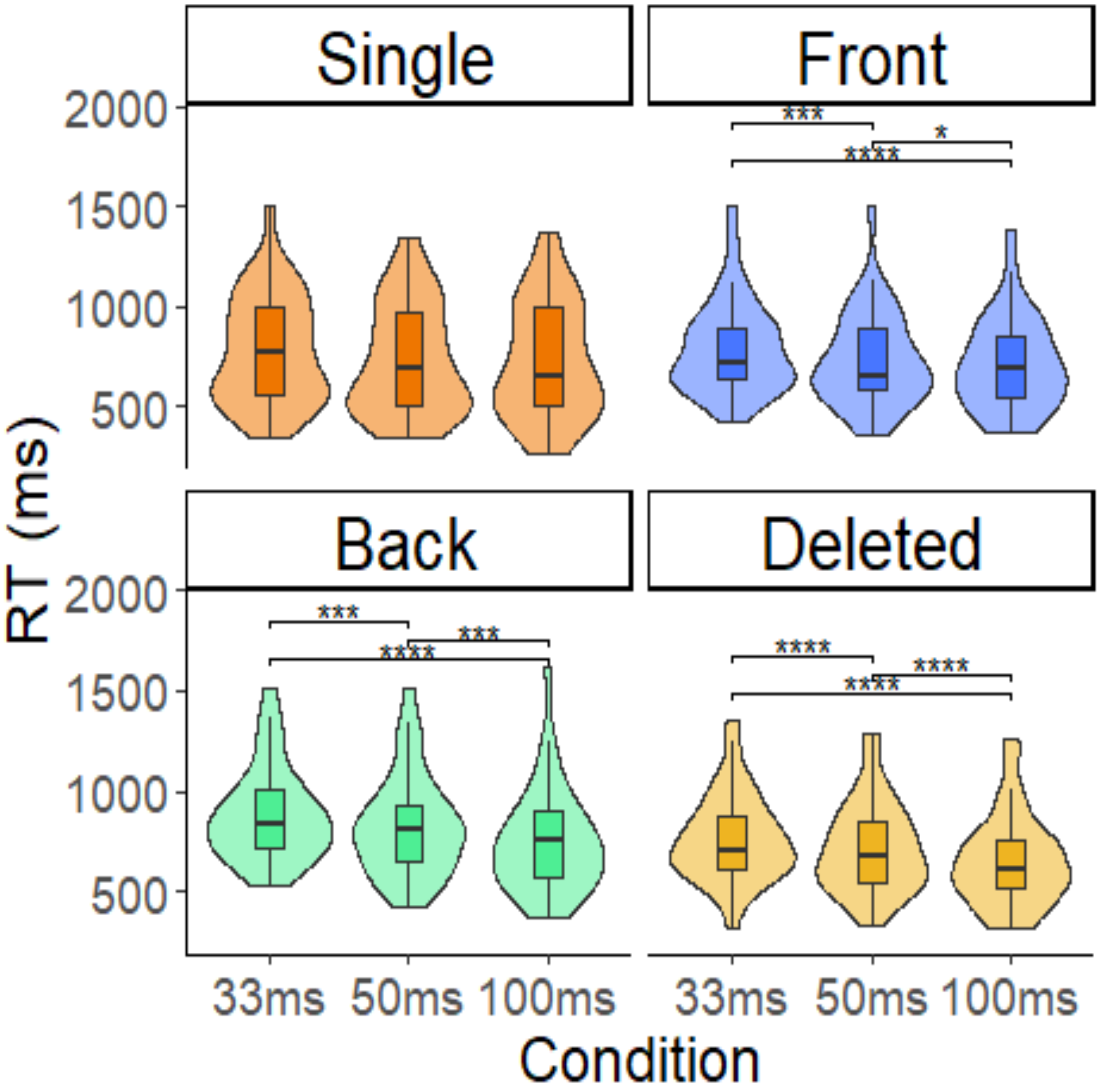
Condition and Presentation Speed (ms)	Mean (ms)	SD (ms)
33 Single	782.00	271.89
33 Front	795.13	235.91
33 Back	882.27	232.67
33 Deleted	766.15	220.41
50 Single	734.89	271.97
50 Front	737.13	237.55
50 Back	826.09	256.69
50 Deleted	717.02	225.52
100 Single	735.54	291.47
100 Front	713.43	227.72
100 Back	748.22	260.53
100 Deleted	653.36	226.61

Post-hoc tests

Pairwise t tests compared the conditions across speeds and vice versa to examine any significant differences for the interaction effect between speed and condition (see Figure 3.7). We demonstrate here that there is always a significant difference between speeds in the deleted, back and front conditions, whereas in the single condition there is no difference between any conditions. It is also worth noting that the significant differences between some of the front conditions are more marginal than those in the deleted and back conditions, perhaps suggesting more challenge in recognition for the latter two conditions.

Figure 3.7.

Reaction times for each condition and speed. Significant post hoc tests (pairwise t tests, fdr corrected) are represented by a connecting line.

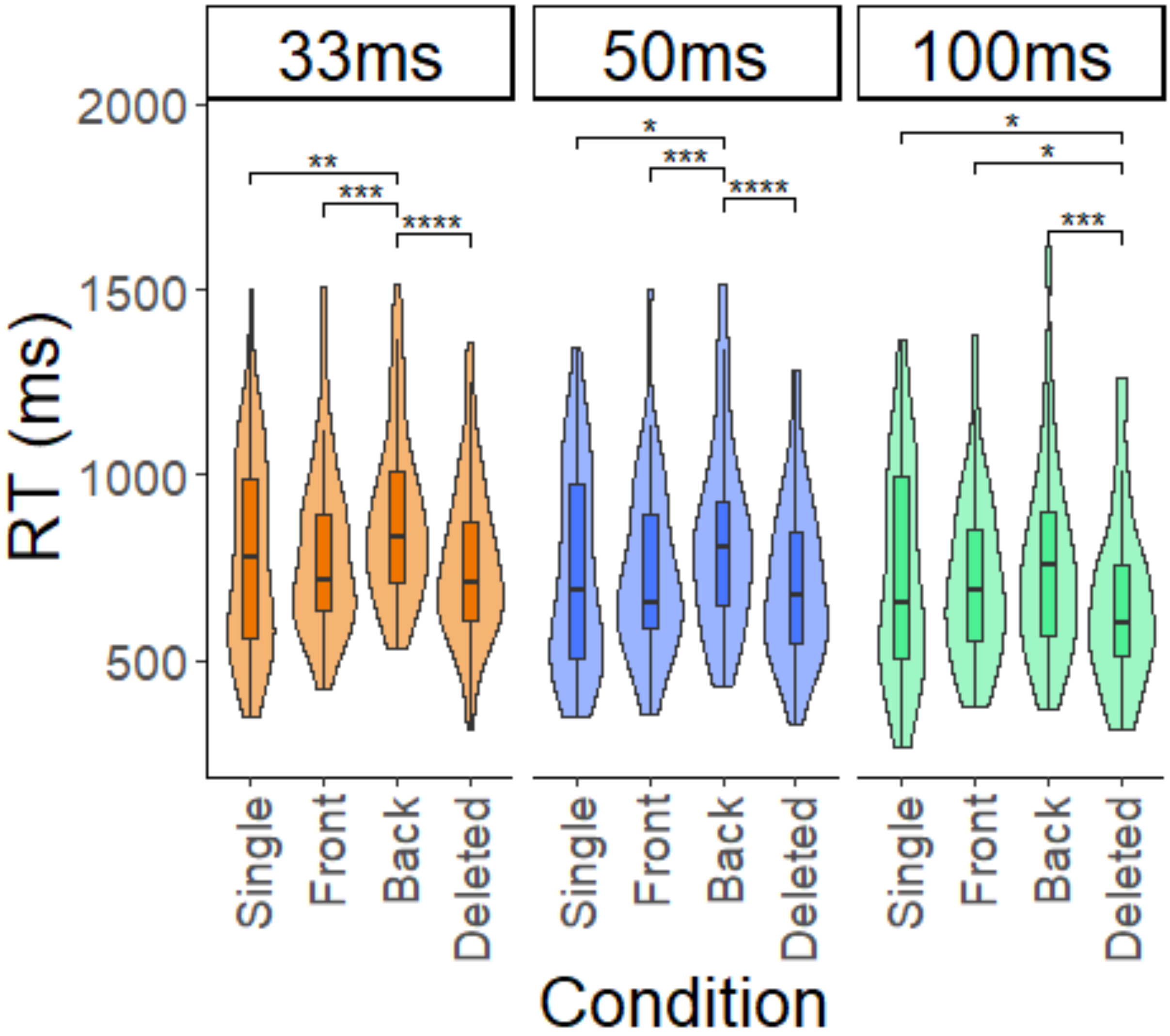


Note. * p<.05, ** p<.01, *** p<.001, ****p<.0001.

When looking at the pairwise comparisons, with condition split across speeds (Figure 3.8), we show that at every speed the back condition has the slowest reaction time, particularly when compared to the deleted condition. At 33ms and 50ms presentation times there were also significantly faster reaction times for the single condition than the back condition, which is in line with our expectations.

Figure 3.8.

Reaction times for each condition across presentation speeds. Significant post hoc tests (pairwise t tests, fdr corrected) are represented by a connecting line.



Note. * p<.05, ** p<.01, *** p<.001, ****p<.0001.

3.3.3. Discussion

The study conducted a comprehensive examination of the impact of presentation speed and display condition on both accuracy and RTs. We found that there are significant main effects for both display and presentation speed as well as interactions between the two factors.

Within accuracies, as predicted, the single condition was the most accurate, maintaining high levels of accuracy across all participants at all speeds of object presentation. The other conditions were revealed by post-hoc tests to be less accurate. The front condition, where participants identified the occluding front object of an object-pair stimuli, was still highly accurate, which is expected, as the object of interest was not occluded in any way. However, the back condition was significantly less accurate than all other conditions. All conditions were the most accurate at the 100ms presentation speed and were less accurate at 50ms and 33ms in turn.

When looking at post-hoc tests for accuracy, faster presentation speeds demonstrated large variations in accuracy, with a variation of 2.23 percent between conditions at 100ms, rising to 27.23 percent at 33ms. This may be due to participants having less time to properly process the correct object identity, especially when in the occluded pairing conditions (front and back) there was more visual data to contend with, whereas in the deleted condition there was missing detail from the central part of the image. In the 100ms presentation speed, we see the most significant post-hocs are between the single and other conditions, speaking to the ease of recognition for participants in the single condition at this presentation speed.

The data also showed that the accuracy in the deleted condition was significantly more accurate (7.49 percent higher) than the back condition overall. This does not align with the findings of Johnson & Olshausen (2005), however it can be argued that this is due to the differences between the real-world occluders we used compared to their meaningless occluders. These ovals did preserve the dimension of depth within the trials, which has proved important in providing visual cues necessary to facilitate recognition. The ovals providing the depth information to the object image were solidly coloured and lacked the complexity of the occlusion of objects in natural scenes that humans encounter every day, which speaks to this cost of processing two objects simultaneously. Much like in the digit clutter tasks employed by Spoerer et al. (2017) where recurrent mechanisms were implicated when digits also acted as occluders, we show that additional objects to represent increases error in recognition.

Examining RT results, we show a significant interaction between condition and presentation speed. Delving further, results revealed that 100ms presentation speeds led to significantly faster RTs then the other conditions, followed by 50ms, with 33ms presentation speed leading to the slowest RTs. Across conditions, the back condition yielded significantly slower RTs than all other conditions, indicating the most challenging object recognition. This may be explained by the required increase in processing needed to identify the objects in the back, occluded condition compared to the relatively simpler single stimuli. The front condition having slower RTs than the single condition, but faster than back aligns with this, speaking to the distraction to the object recognition that having a second object to represent plays, when that object is not relevant to the task for that condition. The two real-world objects in either of the conditions using occluded pairs may both have been identified before the target object, either front or back depending on specific condition, was able to be selected individually.

Post hoc tests found within RTs, the most significantly different conditions across presentation times were those in the back and deleted conditions, where all pairings were highly significant. This is as expected as the visual system would need to process more visual information – taking more time for the back and deleted conditions. These conditions were also significantly different across each presentation speed, with the deleted condition showing faster RTs at every presentation speed than the back condition, which was not the case across speeds for other conditions.

As previous research demonstrated, visual clutter causes interference in object processing (Spoerer et al., 2017), which may explain why the single condition had faster RTs than the front condition, where despite the lack of occlusion in the target object, there was still an additional object to represent. In the pattern of Johnson & Olshausen (2005), the deleted condition was predicted to have slower RTs than all other conditions, particularly the back condition, but this was not the case. The deleted condition has significantly faster RTs than the back condition, perhaps suggesting that the multiple object representations requiring processing in the occluded object pairing had a detrimental effect on RTs. Whereas in the deleted condition, participants only had the cut-out silhouette of the front object, which lacked the 'real' object detail and did not elicit this competition effect.

There is a growing body of literature suggesting that computational principles such as recurrence allow better understanding of the dynamics of the visual system (Ernst et al., 2019; Spoerer et al., 2019). Other researchers in the neurocomputing domain who have created object tracking and recognition models, have found them to be more successful when accounting for occlusion (Wu et al., 2020). They achieved this by programming their model to treat a single visual scene as a combination of objects and occlusion regions rather than simply ignoring anything other than the main focal object. This allows occluders, often dynamic parts of a visual scene, to be recognised and provide added contextual detail. This understanding aligns with the work on DNNs and RCNNs taking place suggesting recurrent connections are critical in achieving a more human-like ability to process occluded objects (Kar et al., 2019; Spoerer et al., 2020; Tang et al., 2018). Compared to models only using feedforward connections, recurrent models can identify objects in increasingly difficult visual search tasks with only a small loss of speed for a near-perfect degree of accuracy (Nicholson & Prinz, 2020). This demonstrates the need for multidisciplinary collaboration between neuroscience and computational science to be taken to further understanding of occlusion and object recognition.

Using the real-world objects for occlusion and deletion was important to maximise the ecological validity and best mirror the conditions the visual system contends with daily. There is known to be a highly diverse and dynamic computational process as information travels along the ventral stream as well as within the ventral stream regions themselves, emphasising intra-area computations (Kietzmann et al., 2019). This makes it clear that recurrent network connections and associated models are a key way to enable greater understanding of how occlusion is so robustly managed by the visual system in both humans and primates.

Adapting the present study paradigms to utilise computer models would perhaps provide more clarity on the differences between the occluded and deleted conditions. This would allow a much larger and more varied set of objects to be employed to train DNN models, across additional categories, to show whether an ability to account for occlusion can improve an artificial neural network. One way this could be approached is by training an RNN to recognise objects in conditions similar to these, incorporating recurrent connections to account for occlusion (Spoerer et al., 2019; Spoerer et al., 2017). The model could be designed to treat visual stimuli as a combination of objects and occlusion regions, mirroring the dynamic of real-world vision. By simulating occluded and deleted conditions, the RNN could provide insights into the computational principles underlying object recognition. Additionally, having human behavioural data to compare this to would allow a quantitative benchmark for assessing the recurrent computations in this process. Multidisciplinary approaches are crucial at present to contribute a more nuanced understanding of the visual system so this could be a key future approach.

Overall, this experiment has revealed that presentation speed and condition have important effects on the RTs and accuracy by which people can recognise objects. The single condition consistently exhibited high accuracy, while the occluded back condition, was considerably less accurate. Faster presentation speeds led to larger variations in accuracy, though the single condition exhibited perfect performance at the slower 100ms presentation speed. Interestingly, in the deleted condition, where details were missing, we found unexpectedly higher accuracy than the occluded back condition, suggesting that having multiple objects displayed at once makes recognition harder. Reaction times mirrored accuracy patterns, with the back condition exhibiting significantly slower responses than the other conditions. The pairings of real-world objects used allowed the difference between the back and deleted conditions to be displayed. This is because even when recognition of only the back object is required in the task, the nature of visual processing means that the irrelevant front object is processed in addition to the target back object in this condition. Future research should utilise objects as occluders to obtain more natural results, as well as embracing computational methods to further understand the processes involved in visual object recognition.

3.4. Experiment 2

Our previous fMRI study left us with the question of how neural representations of objects under occlusion relate to human perception of the same objects. We know that neural processing in IT typically maps to human perception, responding selectively to specific objects and categories, while EVC is often tied to lower level visual processes (Groen et al., 2017). Here we combine our previous fMRI data (Chapter 2) with novel MVPA analyses to improve our understanding of the mechanisms of object recognition under challenging visual conditions. Creating weight scores from the fMRI data collected for the same stimuli as Experiment 1, allows us to analyse the behavioural data regarding neural patterns. A weight was created for each occluded object to reveal how well the full object pattern was present from this. Based on prior research (MacEvoy & Epstein, 2009; Reddy et al., 2009) we are able to relate these weights to the accuracy achieved for each occluded object in our behavioural experiment.

As we know that performance has been found to improve with higher percentage of an object visible (Tang et al., 2018), we would also predict that as EVC is implicated in low level properties, the magnitude of occlusion should influence the ability to recognise the occluded back object. However, based on our prior work (Mansfield et al., 2023), we believe that this will show a more pronounced effect in early visual areas, which are more susceptible to changes related to multiple object representations and competition effects. In IT, the expectation would be that the weights would relate more to behaviour than low level features.

3.4.1. Methods

3.4.1.1. MRI data acquisition

The pre-processed fMRI data is the same as Chapter 2. The MVPA analysis pipeline differed in this study, with linear regression applied instead of decoding to better mirror the work of Reddy et al. (2009).

3.4.1.2. Analysis software

Data were analysed using SPSS statistics (29.0) as well as R (RStudio 2022.02.1), where the latter was also used to create figures for data visualisation. MATLAB (v.2020b) was used to complete the linear regression to calculate the beta weights.

3.4.1.3. Weights analysis

To expand on the findings from the decoding analyses in Chapter 2, where the object category of the occluded object was predicted, we looked at the activity patterns themselves. These beta weights were defined using standard linear regression in the manner of Reddy et al., (2009). Linear regression was used to predict each occluded trial activity pattern from the linear combination of the activity patterns of each of the constituent objects that made up that trial in isolation (i.e. the single object patterns that correspond to the front and back objects of that trial). We used leave one run out cross-validation to compute this from the single object patterns estimated from N-1 runs. The occluded trial activity was predicted on the left out run, which was then cycled and averaged. Accuracies and RTs (median) computed across participants per occluded item when recognising the back (occluded) object at the quickest presentation speed of 33ms were used. This was under the rationale that this speed is the most challenging and hence would offer more variance. It is important to note that in the following analyses we only looked at the beta weights for the back object. These weights were related to the accuracy and RT scores. This enabled us to keep the scope of the study focused on the most challenging visual object to recognise, the occluded back object.

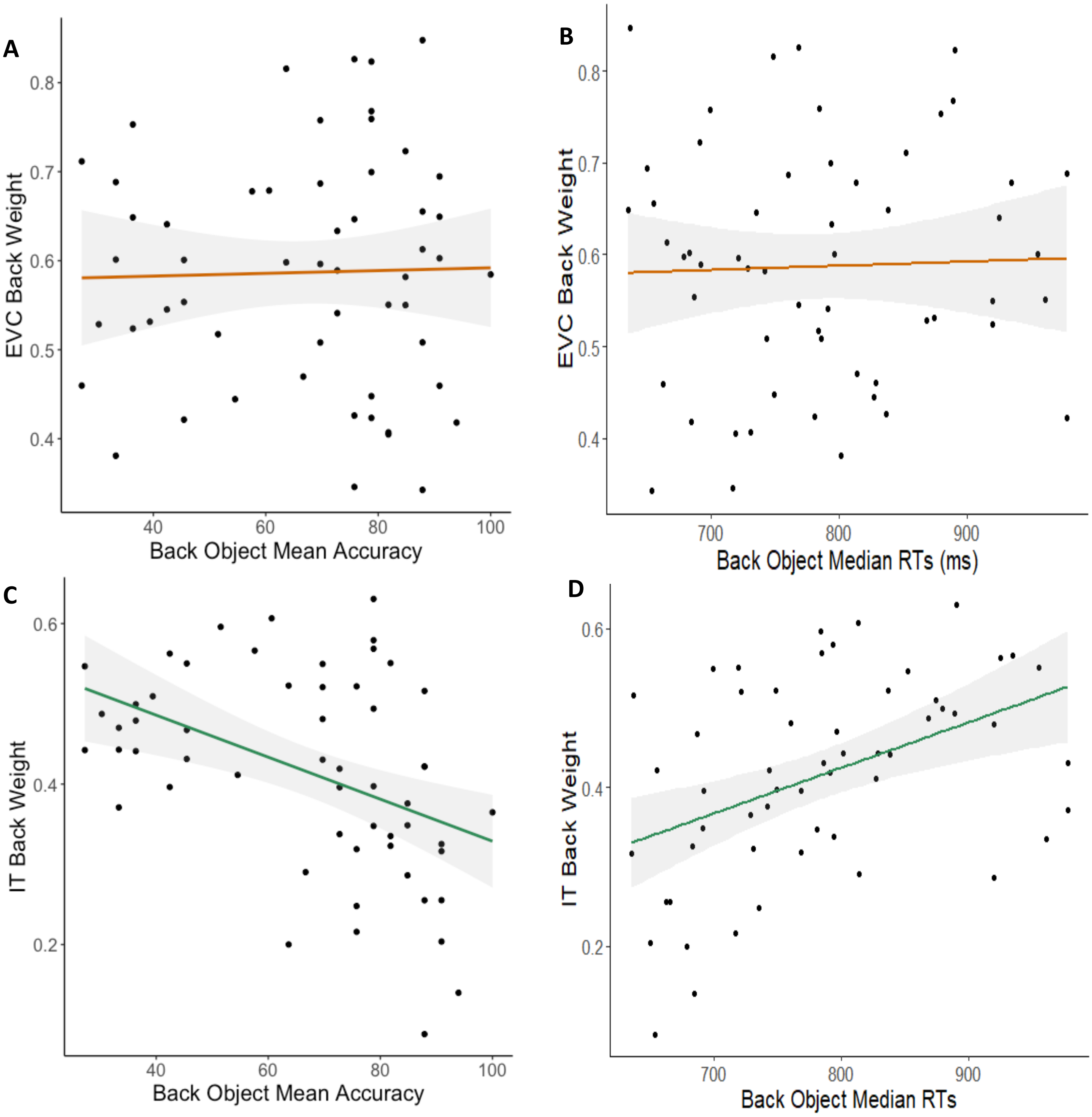
Magnitude of occlusion was computed by calculating the ratio of how many pixels were present in the back object of an occluded trial, as a percentage of how many were present on the whole trial. This measure was inverted to provide an index of the degree of occlusion.

3.4.2. Results

Spearman's correlations were run to determine any effects between the weights of the back object in the occluded object pair in either early visual areas or IT cortex with reaction times and accuracies respectively. These analyses show there to be non-significant results in EVC for both accuracy, *r* (54) = -.001, *p* = .992, and RT, *r* (54) = .020, *p* = .884. However, in IT there were significant correlations found, with a negative correlation in accuracy, *r* (54) = -.462, *p* < .001, as well as a positive correlation in RT, *r* (54) = .45, p < .001 (see Figure 3.9).

Figure 3.9.

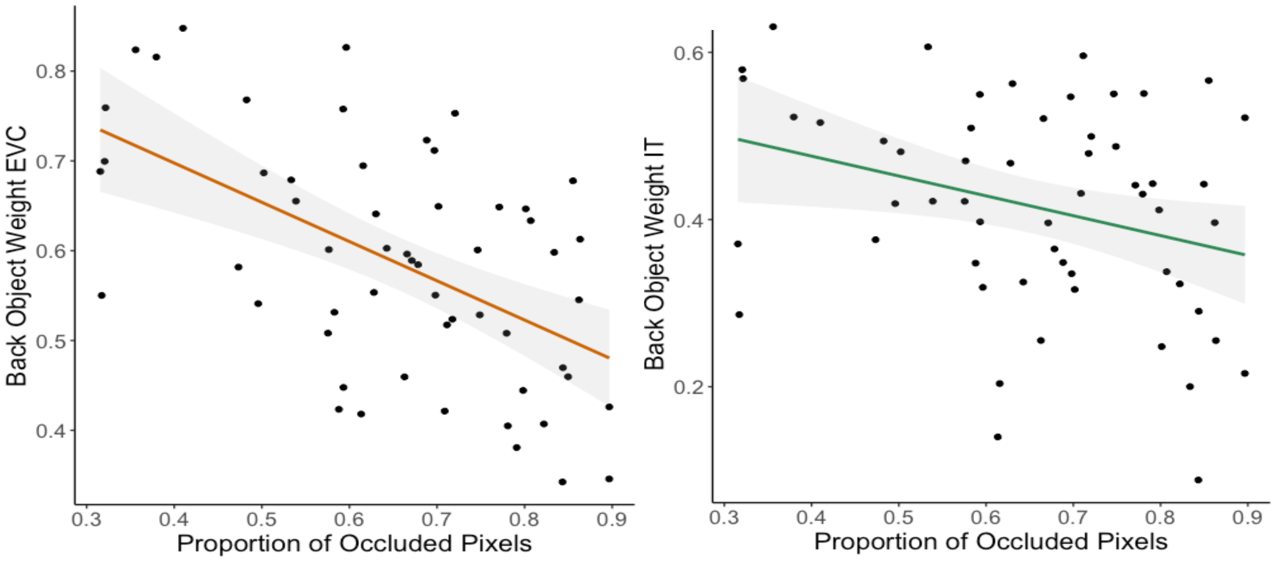
Correlations of the back object weight in EVC for accuracy (A) and RTs (B) as well as IT object weights with the mean accuracy (C) and median RTs (D) for the back object of the occluded pairings. Medians were used for RTs as they are more robust to outliers.



When correlating the proportion of occluded pixels across brain regions of interest we have found that there is a significant negative correlation between back object weight in EVC and the proportion of occluded pixels, *r* (54) = -.488, *p* < .001. As well as this, there is a marginally significant spearman's correlation between IT weight and the proportion of occluded pixels, *r* (54) = -.267, *p* = .047, (see Figure 3.10).

Figure 3.10.

Correlations of the back object weight for (A) EVC and (B) IT against the proportion of occluded pixels.



In order to control for magnitude of occlusion in IT back object weights, we ran a partial correlation on IT back object weights against back object mean accuracies and median RTs, controlling for proportion of occluded pixels a found that correlations were still significant for both accuracy (*r* (54) = .547, *p* < .001) and RT (*r* (54) = .461, *p* < .001). Thus, demonstrating statistically that the results are not simply due to the proportion of occluded pixels. The relationship between the back weight and IT is not dependent on how many pixels are present, suggesting higher level functioning is at play here.

3.4.3. Discussion

In this study we tested the effects of occluded pair back object weights against behavioural RT and accuracy scores to understand how the representation of the back object in IT and EVC relate to behaviour versus magnitude of occlusion. The correlations between the beta weights for the back objects in EVC and IT reveal differences between how these brain regions handle the presence of occluders. In EVC there were no significant correlations in either RT or accuracy. However, in IT there was evidence to suggest that when RTs are slower, and when accuracies are lower, there is a greater weight assigned to the back object. In other words, when recognition is more difficult, a higher weight, and thus more processing, is assigned in IT. This may be indicative of the recognition process taking more cycles of recognition to complete the pattern under more challenging visual conditions (Kar et al., 2019). These delays in recognition suggest the need for additional computations to interpret partially visible images. In line with the existing literature, it is clear that there is a robust ability within the visual system to successfully recognise objects even when they are highly occluded (Rajaei et al., 2019; Tang et al., 2018; Zhu et al., 2019).

Both EVC and IT have significant correlations with occluded pixel proportion, with EVC featuring a larger correlation. The correlation value in IT is around half of that of the EVC value. This demonstrates that the early visual areas are more attuned to the proportion of occluded pixels, with the weight of the back object decreasing dramatically when exposed to greater occlusion. This may be indicative of the focus on the front object (and thus less weight on the back object) of an occluded pairing instead of the back object when more pixels of the back object are occluded. Though an alternative view may be that EVC processing is increased when there is a higher proportion of the back object present. This is representative of the smaller scale of focus V1 receptive fields are able to represent (DiCarlo & Cox, 2007). Whereas in IT, there is greater ability to untangle more of the representation and disambiguate multiple objects represented simultaneously (DiCarlo & Cox, 2007), irrelevant to the proportion of occluded pixels. Thus, we argue that IT is better able to process these complex object stimuli, which are shown to scale with difficulty, suggestive of specific mechanisms well equipped for this challenging visual process, whereas this is not the case in EVC.

These results potentially provide insight into how visual recognition may be taking place in the visual world. The constant presence of multiple objects in our visual fields means that the visual system is highly adapted to cope with this, perhaps using amodal and pattern completion mechanisms (Ao et al., 2023; Tang et al., 2018). The challenge for researchers is understanding the process by which these complex visual conditions are recognised and understood.

3.5. General discussion

Overall, across both behavioural and neural methods we have explored how the presence of additional objects affects recognition. Neurally, we show that occlusion affects the neural processing assigned to each object in IT, and behaviourally we find that IT responses to occluded objects relate to the participant's recognition of the same object (Johnson & Olshausen, 2005; Smith & Muckli, 2010; Spoerer et al., 2017; Tang et al., 2018), where the magnitude of occlusion has larger effects on EVC. Behaviourally, multiple objects and a behavioural masking task enabled us to reveal a greater accuracy and faster RT of judgement for objects in unoccluded conditions. However, diverging from previous occlusion research (Johnson & Olshausen, 2005; Spoerer et al., 2017; Tang et al., 2018; Wyatte et al., 2012), we discovered that when objects act as both the occluding and occluded factors, there are differences in the recognition of the back, occluded object dependent on the presence of an object or cut-out. Specifically, that the deleted condition was more easily recognised than the back condition, despite the visible information of the object of interest being identical across these two conditions, the only difference being the presence of the second object representation in the occluded pairs and the cut-out in the deleted.

A cost of processing multiple representations was demonstrated in Spoerer et al. (2017) digit clutter paradigm where the error for more digits, and thus more visual clutter, was significantly higher than for fewer digits. This reflected particularly high errors in feedforward compared to recurrent models. In our study, where sections of multiple objects are displayed simultaneously, this split within visual resources may explain why the deleted condition leads to greater recognition than the occluded back object condition.

The observed increase in beta weight for the back object in IT, coupled with the slower reaction times and decreased accuracy, suggests a unique processing pattern under challenging recognition conditions. We discern that when accuracy is lower and RTs are slower, there is an effect on the beta weight derived for the back object of the pair where this weight is increased linearly with the difficulty of recognition. This phenomenon speaks to the adaptability and resource allocation within IT, allowing for a more comprehensive object recognition process (Conway, 2018). In contrast, the non-significant correlations in EVC suggest that there is more consideration for the visible features, with processing power being devoted as a function of this visibility. We posit here that this finding implies that IT invests more processing cycles in recurrent processing when faced with challenges, such as occlusion. The ability to combine behavioural and neuroimaging data allows an extra layer of depth, as even though the groups of participants were different, the stimuli remained the same throughout, allowing comprehensive comparisons of the effects of each stimulus pairing. The trends and interactions we find present a clear picture of how object recognition under occlusion could be occurring and how research has to adapt to more ecological methods to capture the nuance of vision.

While this study has successfully used both behavioural and neuroimaging data to investigate the underlying processes of object recognition under occlusion, there are still advances that could be made to further our understanding. The analysis of experiment 2 largely used the back object weight, which misses a layer of complexity regarding the front object. We see in the behavioural results that there is a still a cost associated with the presence of the back object when compared to the single, unoccluded object condition. Therefore, looking specifically at the results for the front object of the occluded pair may offer additional understanding into the result of there being increased visual input, even if the back object does not require recognition itself for the task. This may provide insight that highlights the roles of attention and prediction within recognition and would be a useful route for additional analysis.

In line with the pivotal role of recurrence in object recognition (Han et al., 2018; Jia et al., 2020; Kietzmann et al., 2019; Lamme & Roelfsema, 2000; Wyatte et al., 2014), leveraging computational methods becomes crucial for developing our understanding. The vast datasets accessible through these methods offer the opportunity to explore diverse objects and degrees of occlusion, shedding light on the intricate interplay of feed-forward, feedback, and recurrent mechanisms in recognition processes.

DNNs provide a promising avenue to grow our knowledge, especially when adopting a predictive coding objective during training where generally, these networks learn to predict future occurrences, with network layers making recurrent, local connections and only feeding forward the deviations, or prediction errors, to subsequent network layers (Lotter et al., 2017). This process has been successfully utilised, where through the use of an iterative 'predictive' training method, convolutional DNNs have even been able to solve some illusory contours whereas this was not the case prior to this advance (Pang et al., 2021). Ali et al. (2022) also emphasises that recurrent neural networks inherently engage in prediction as a consequence of the efficiency of recognition. Furthermore, in the future DNNs could be used to test what occurs during the processing of occluded objects in this experiment. One potential avenue would be to look into how many cycles a recurrent net takes to recognise an occluded object, compared to the results of a feedforward neural network and linking these findings back to behavioural results and neural results in IT specifically.

3.5.1. Conclusion

Overall, in this Chapter we first of all built on prior work, exploring the behavioural recognition of objects under conditions of occlusion, demonstrating that lower accuracy and slower RTs in occluded back conditions than deleted conditions. This suggests an effect of competition across multiple object representations. To take this further, we explored the link between this data and neural representations using linear regression. Here, we determined beta weights for each occluded pair of objects in the stimuli set and found a rich pattern of results suggesting that representations in IT relate to behavioural difficulty, while this is not the case in EVC. This is because there are strong correlations between the back object weight with accuracy and RT in IT but not in EVC. In EVC we find more reliance on the visibility of the objects, but after confirming that the IT differences were not due to a simple change in magnitude of occlusion, we were able to confirm that there are higher weights assigned in IT when recognition was more difficult. In summary, we determine that occlusion is a challenge that is better solved in the IT cortex where resource allocation can be more complex.

Chapter 4. Does predictive processing explain responses to occluded objects in primary visual cortex?

4.1. Abstract

Previous studies have determined that visual details can be found in early visual cortex even when sections of a scene are occluded. Expectation suppression has been used in prior studies of EVC to measure prediction, though this requires the combination of decoding and univariate analysis. The goal of the present study was to utilise expectation suppression to determine how these visual effects may be explained when multiple object representations are presented and occluded, and whether sharpening or dampening accounts of predictive processing better represent this. Participants (*N*=18) in this event-related fMRI study saw cue-target pairs of images of objects, first with an occluded bar and then a matching or mismatching bar representing the occluded area while performing a colour response task. Mapping checkerboards at the end of each run allowed us to define our ROIs in EVC as those areas which responded more to the occluded bar section (the target area) or the remaining area (known as the surround) respectively. Decoding analyses demonstrated an effect of expectation suppression where matching trials were represented more accurately than mismatching trials in both target and surround ROIs. Our decoding data showed support for the sharpening account of predictive processing. Whilst the univariate data demonstrated a more surprising picture, exhibiting no traditional expectation effect in line with either sharpening or dampening. Instead, the data show a large effect of the neutral (noise) condition, particularly in the surround region. Overall, our results reveal that expectation suppression may be occurring across cue-target pairings of objects even when a section of a stimulus is occluded.

4.2. Does predictive processing explain responses to occluded objects in primary visual cortex?

4.2.1. Predictive processing and expectation suppression

Sensory processing was thought to be a feedforward process, however more recent findings suggest that the brain constantly constructs an internal model of the world that incorporates prior knowledge alongside sensory input (Clark, 2013; de Lange et al., 2018). This prior knowledge is used within the theory of predictive processing, which proposes a unified account of mental functioning that is primarily focused on minimising surprise (Melloni et al., 2011; Ransom et al., 2020). Within the cortex, top-down connections are purported to convey predictions about lower-level activity while bottom-up processes are thought to transmit prediction error to the higher order areas (Boutin et al., 2021).

The overarching theory of predictive processing suggests that the brain works to optimise processing efficiency by actively minimising prediction errors. This is achieved through the transmission of only the unpredicted portion of a signal, filtering through predicted details. This process is facilitated by feedforward, feedback and recurrent connections (Mills et al., 2021; Tang et al., 2018; Williams, 2018). Predictive processing models suggest that sensory regions are not passive recipients of signals, but instead continuously engage in conveying predictions and associated errors throughout the visual stream (Kok & De Lange, 2015).

Motivated by the desire to unravel the roots of predictive processing in object recognition, our study builds on prior findings (Smith & Muckli, 2010) to explore the connection between observed effects and the theory of predictive processing. Smith and Muckli (2010) found that activity patterns of an occluded quadrant of a scene were significantly related to feed-forward stimulation and were driven largely by V1. Thus, providing motivation to further look into the effects of occlusion in early visual areas including V1 to shed light on the processes responsible for this. The recognition of objects under conditions of occlusion poses a computational challenge, with recurrent neural networks emerging as a fitting model to address the associated energy constraints (Ali et al., 2022). This computational knowledge aligns with the brains strategy of inhibiting predictable sensory input to conserve resources. Ali and colleagues (2022) found this proficiency was mirrored in computational models, which developed distinct error and prediction units when trained for efficiency.

Computational vision research aligns with the unified account of predictive processing (de Lange et al., 2018). Here, efficient connections are steered by the predictability of stimuli. It is clear that the challenge of recognising objects under conditions of occlusion is addressed more successfully with the addition of recurrent connections into visual models (Ernst et al., 2021; Han et al., 2018; Tang et al., 2018). Therefore, a logical next step in deciphering the complexities of predictive processing in object recognition is to introduce more intricate visual stimuli, especially those with occluded sections.

In parallel, expectation suppression serves as a measure to comprehend predictive processing. The effect is, in its simplest terms, explained by a reduction of neural response to an expected stimulus (Feuerriegel et al., 2021). To diminish stimulus activation, it stands to reason that there is a process to determine the expected features from the feedforward information. The rationale of expectation suppression is built here, asserting that prior knowledge enables the propagation of prediction errors instead of an entire object representation, thereby requiring fewer neural resources due to the expected nature of the representation (Alink & Blank, 2021; Feuerriegel et al., 2021; Kok & De Lange, 2015). This perspective aligns with what is commonly known as the sharpening account of expectation suppression (Kok et al., 2012). In this framework, suppressed neurons represent unpredictable stimuli, emphasising a selective mechanism that refines neural responses to expected features, creating a greater contrast between the expected and unexpected stimuli. Ongoing debate surrounding sharpening versus dampening effects in predictive processing add complexity to our understanding. Studies supporting both mechanisms across different temporal windows introduce nuances that challenge a straightforward classification (Xu et al., 2021).

Dampening is thought to predominantly suppress neurons that are attuned to an expected stimulus, resulting in a reduction of contrast in activity patterns, sharpening has more of an effect on the neurons not aligned with an expected stimulus, causing an increased activity pattern contrast (Alink & Blank, 2021; de Lange et al., 2018). There have been studies that supported both the sharpening (González-García & He, 2021; Jiang et al., 2013; Kok et al., 2012) and dampening (Walsh & McGovern, 2018) accounts. Thus, showcasing the richness of neural responses to expected stimuli. Further complexity arises in the consideration of visual state conditions, as demonstrated by (Rossel et al., 2022), who observed a sharpening effect in lower perceived blurriness of a predictable image compared to an unpredictable one. The interplay of sharpening and dampening effects becomes apparent when exploring decoding results, where the correlation between reduced activation and decoding accuracy reveals distinctive patterns for each account.

Decoding results in the sharpening account showcase an association in which the reduction of activation is inversely correlated with the accuracy of decoding, where a more predictable stimuli is represented by lower activation but a greater decoding accuracy (Kok & De Lange, 2015; Richter & de Lange, 2019; Summerfield & de Lange, 2014). In contrast, the dampening account anticipates a flatter overall picture in decoding results, with less contrast between conditions due to the selective suppression of neurons attuned to predictable stimuli (Kok & De Lange, 2015; Walsh & McGovern, 2018).

However, more complexity arises with studies suggesting that both mechanisms operate, just across differing time points. Xu et al. (2021) in an EEG study, presented evidence for sharpening effects during N1 and dampening effects during N2. This was achieved by employing a double-flash task involving an auditory cue preceding a pair of oncoming flashes of light with unpredictable stimulus onset asynchrony (SOA) values (1000/1500/2000ms) both between the cue and flashes and between the flashes themselves (400/600/900ms). The participants held a temporal template in their minds for each block, reflecting the SOA between the flashes, and subsequently pressed a response button after each double-flash to judge whether the experienced SOA matched or mismatched their temporal prediction. Their results showed an expectation suppression effect, where evoked EEG energy was lower for matching than mismatching predictions. Though the authors did not use MVPA as is the norm in neuroimaging expectation suppression analysis, in ERP analyses the N1 period showed suppression in the unexpected condition, in line with sharpening. However, the N2 was enhanced for the unexpected compared to the expected condition, which is indicative of the dampening account of expectation suppression. This temporal element shown through the amplitude changes in evoked potentials supports the suggestion that both accounts could be active but working in differing temporal windows.

The impact of expected responses is robust even across various experimental designs and methods. Aitken et al. (2020) used MEG, asking participants to observe moving dots following an auditory cue. Researchers found using decoding that the anticipated direction of the dots influenced the neural representation merely 150ms after their appearance, highlighting the incredible predictive abilities of the human brain and the influence recurrent processing may have in modulating this prediction. An fMRI study using face and house stimuli to analyse the Fusiform Face Area (FFA) presented a colourful frame that briefly preceded the stimuli throughout a task that required participants to respond to the presence of an inverted stimulus using a speeded button press (Egner et al., 2010). The prediction was driven by the colour of the frame (blue, yellow or green) and strongly supported a predictive processing model of visual cognition. Specifically, Egner and colleagues (2010) found that using fMRI and computational simulations showed that the predictive processing model incorporated the expectation of the faces with the surprise responses of houses more effectively than a feature detection model was able to. Feature detection acts as a low level processing step engaged in the identification of features such as points, lines and curves among others (Li et al., 2015). Thus, demonstrating that predictive processing does possess an important effect in visual understanding and recognition. The array of methods showing specific effects of expectation highlight the importance of neural resource allocation and predictive ability in the visual cortex. The ability to represent objects is therefore key where novel occlusion stimuli are used to study these prediction effects.

In previous expectation suppression studies using grating stimuli, EVC demonstrated expectation suppression effects related to attention and task-relevance (John-Saaltink et al., 2015; Kok, Jehee, et al., 2012). John-Saaltink et al. (2015) used auditory tones to predict gratings, where tasks were either to predict the spatial frequency of the grating stimuli, to perform a 1-back letter repetition task with added noise increasing difficulty, or a working memory task where targets were 2-back colour repetitions with no noise but higher difficulty for the working memory system. The authors suggest that expectation suppression is not an automatic phenomenon, instead dependent on attentional state and cognitive resources, where an irrelevant task led to greater expectation suppression, particularly when compared with a working memory task. Therefore, future designs to capture the most information on expectation suppression abilities should use irrelevant task designs as to not overload perceptual resources. Additionally, in Kok et al. (2012), their findings of expectation suppression suggest that prediction and attention interact in EVC. When predicted stimuli are unattended and task-irrelevant, reduced neural activity in these stimuli are representative of expectation suppression. Though they show that when stimuli are attended and task-relevant, the findings are reversed, with attention then modulating the effects of expectation. This aligns with the idea that attention can counteract or modify suppression effects, allowing for a more detailed and precise processing of predicted stimuli. Hence, it is clear from these studies that to sufficiently capture the nuances of expectation suppression, care must be taken with the task design to ensure that the task is irrelevant from the stimuli so as not to draw too much attention.

4.2.2. Visual occlusion

Occlusion, where an object is blocked by another, is a process which requires the ability to 'fill in' missing information in order to comprehend a visual scene. This amodal pattern completion is thought to again be a response of recurrent computations within the cortex passing efficiently via feedback and lateral connections (Kietzmann et al., 2019; Wyatte et al., 2012). Amodal contour completion is well-suited to deal with occluded regions of objects or scenes, perceptually representing occluded regions based on previous experience (Scherzer & Ekroll, 2015). This ability to infer the missing object information, relying on previous visual experience aligns well with predictive processing where the whole object is predicted and only missing information is coded as the prediction error. In this manner, rich occlusion related responses have been determined from occlusion research. It has been determined through multiple studies that rich visual responses are present in the earlier visual areas (e.g., V1-V4) even when a visual scene has been occluded, which is argued to align with prediction (Morgan et al., 2019; Muckli et al., 2015; Smith & Muckli, 2010).

In occlusion research, investigations have revealed the remarkable capability of the IT cortex to fill in occluded information, enabling the recognition of intricate stimuli despite substantial obstruction (Mansfield et al., 2023; Tang et al., 2018). Contrary to traditional perspectives that attribute only low-level features to EVC, studies, such as those by Kok et al. (2016) and Lee (2003) have unveiled the surprising proficiency of the EVC in processing rich and challenging stimuli. Notably, in the context of illusory contours, where simple bottom-up information is lacking, the EVC demonstrated remarkable processing capabilities. The extension of these findings to 7T scanning, particularly in response to illusory contours like the Kanizsa illusion (Kanizsa, 1976), reveals a laminar profile of the BOLD response in EVC. This profile highlights distinctions between bottom-up stimulation and top-down activity, illustrating that in V1, bottom-up stimulation activates all cortical layers, whereas feedback induced from illusory figures selectively activates deeper cortical layers (Kok et al., 2016). These insights challenge conventional views of EVC function and provide a nuanced understanding of its involvement in processing complex visual stimuli, necessitating further research.

Subsequent research delved deeper, illustrating V1 activity patterns in fMRI effectively filling in occluded regions, aligning even with line drawings of the absent sections (Morgan et al., 2019). They found that this elucidated the internal models of V1, revealing the extraction of scene category information in the early brain regions. The interplay between behavioural and neural insights underscores the importance of comprehending internal models that drive recognition in the presence of occluded sections within the visual environment.

Interestingly, there has been a surge of research into how occlusion is represented in movement. One area where this has been largely popularised is in self-driving cars, where safety features would require cars to be able to recognise obstacles and people in the path of the car even when they may be partially occluded (Cheng et al., 2020; Wu et al., 2020). Automated tracking in sports has also benefitted from this work. For instance, Video Assistant Referees (VAR) have been used in football to help referees make correct decisions, but this requires the ability to recognise the different players, who may be occluded, as well as the ball itself, often at speed (de Oliveira et al., 2023). This also is addressed in the scene understanding literature where work into recognising many people all with partial occlusions in circumstances such as crowds have also been analysed through use of models refined by scene layout and temporal reasoning (Tang et al., 2014). Therefore, real-world implications of a more comprehensive understanding of occlusion are vast.

These additional fields seeking to comprehend and account for occlusion again suggest the necessity for multidisciplinary work that allows these advances in knowledge to be collated to a greater overall understanding of how occlusion works in more naturalistic settings. As spatial and temporal integration are known to plan an important role in pattern completion mechanisms (Tang et al., 2018; Wyatte et al., 2014) it makes sense that movement and temporal features of stimuli are a key aspect to investigate. Though the present study does not involve directly moving stimuli, there is a temporal and spatial dimension demonstrated through showing cue-target pairs that have different spatial constraints across different time points. This may enable opportunity to understand how the early visual areas respond to managing expectations across temporal and spatial factors.

4.2.3. Our motivation

The overarching goal of the project is to advance from previous research that established the EVC’s reception of contextual, predictive information even in the absence of visual stimulation (Morgan et al., 2019; Smith & Muckli, 2010). In doing this, our focus shifts towards investigating whether predictive processing mechanisms are at play in response to occluded visual objects within V1. Departing from the paradigm employed in previous chapters, we maintain the use of eight single-object stimuli to delve into occlusion and assess the potential occurrence of expectation suppression and predictive processing. Unlike chapters 2 and 3 where multiple objects were presented across space in pairs, this study introduces a temporal dimension, revealing entire objects across time through cue-target combinations.

While BOLD response amplitudes alone may not provide insights into sharpening and dampening effects, the pattern of activation gained through decoding analyses should be enlightening. For instance, when a cue image featuring a face with an occluding bar is presented, predictive theories posit that the visual system should complete the image by filling in the missing section based on prior visual experience. Subsequent presentation of the target condition, revealing the missing section, allows us to validate or invalidate prior knowledge through a matching or mismatching image respectively (i.e., a section of a cup).

Decoding outcomes will unveil whether expectation suppression changes the representation in EVC by examining the representations in the target section. According to sharpening studies, successful prediction is characterised by lower activation but higher decoding accuracy. Specifically, greater pattern discrimination for matching cue-target combinations compared to neutral and mismatching conditions would be demonstrative of a sharpening effect. Conversely, evidence of diminished pattern discrimination for matching cue-target combinations compared to neutral and mismatching conditions would be indicative of a dampening effect (Kok & De Lange, 2015). This study aims to determine whether predictive processing underlies the processing of visual information under occlusion, and, if so, which account of expectation suppression more accurately represents these effects.

4.3. Methods

4.3.1. Participants

Self-reported right-handed healthy participants (*N* = 18, 7 Male, 2 Non-Binary, 9 Female, mean age = 25.17, SD = 5.74) participated in this fMRI experiment. All participants reported normal or corrected to normal vision and were deemed eligible after meeting MRI screening criteria. Informed consent was obtained in accordance with approval from the Research Ethics Committee of the University of East Anglia School of Psychology. Participants were reimbursed for their time at a rate of £12 an hour.

4.3.2. Stimuli and design

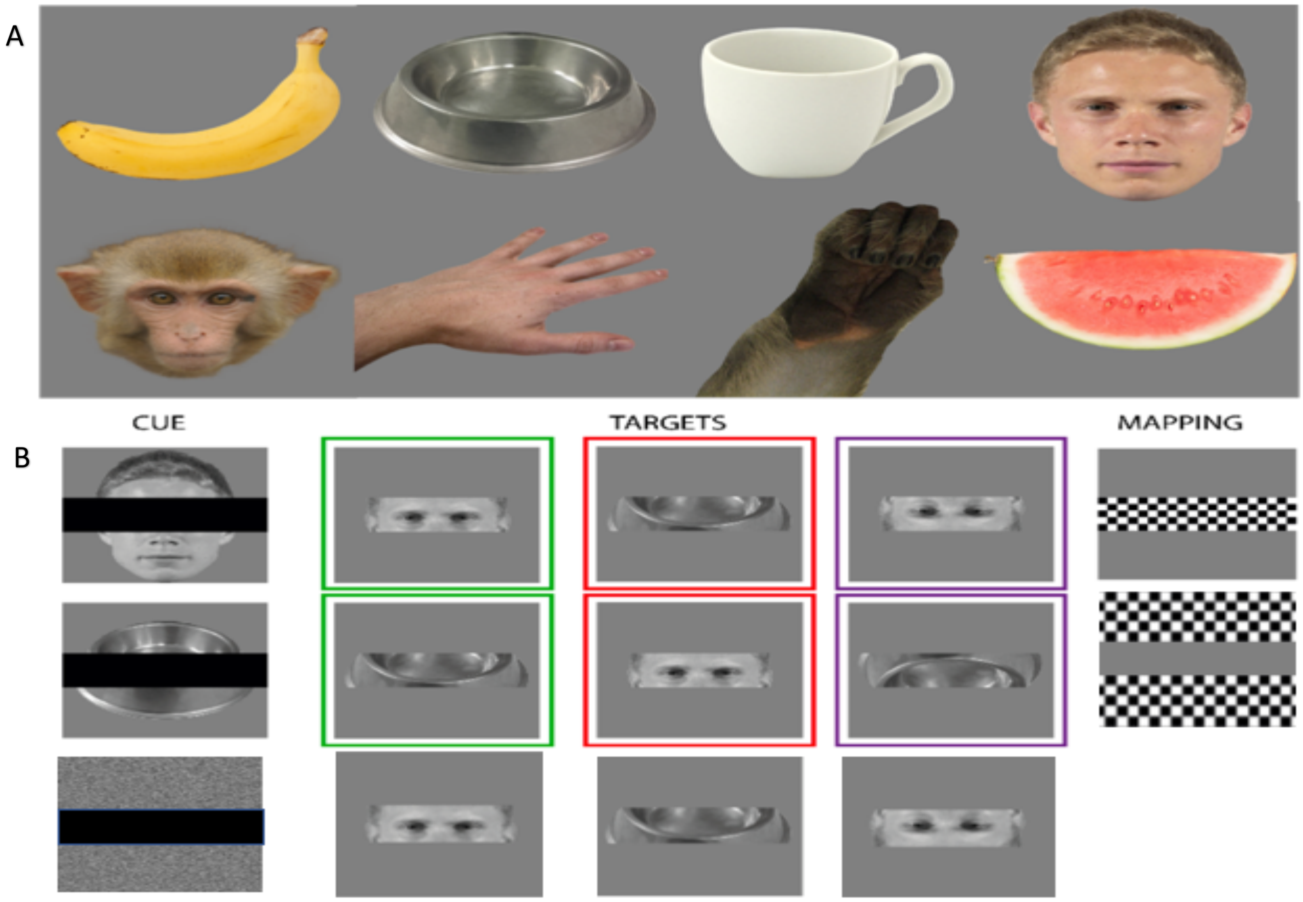
The study utilised a rapid event-related fMRI method where participants were presented with a cue object followed by a matching or mismatching target stimulus. Participants performed a recognition task requiring a button press when any cue object was shown in red. This occurred once for each of the objects and once for the neutral condition (9 total). These response trials were not included in subsequent analysis.

The stimuli were made up of eight objects of roughly the same real-world size: banana, dog bowl, mug, human face, monkey face, human hand, monkey hand and watermelon. These objects spanned various salient semantic categories (e.g., animate/inanimate, human/non-human, natural/artificial), which are easily recognisable in human vision (Rosch et al., 1976). PNG images were presented in greyscale on a grey background at 799x799 pixels, the visual angle was presented at 11 degrees high, with the occluded section measuring 2.75 degrees high.

Cue images featured an opaque black bar occluding the central third of the image, such that only the top and bottom thirds of the image were visible. The target stimuli were comprised of the central third of the objects. The target stimuli related to the cue by either matching (e.g., the cue showed the banana with a missing block and the target was the central upright strip of the banana image) or mismatching. The mismatches were either caused by orientation differences (the same image as the cue but inverted) or by object (a different target image to the cue, either upright or inverted) (see Figure 4.1 for examples). A neutral condition was created where the cue was not an object, but instead made up of noise, which was followed by a target that would be either upright or inverted.

Figure 4.1.

A) The eight whole stimuli. **B)** Examples of trial stimuli. Cue stimuli have an occluding block across them, green boxes represent match trials. Red and purple outlines represent mismatch trials, either by object or orientation. Neutral trials have noise occluded using a bar as a cue, followed by a target as normal. Mapping stimuli allows specific analysis of the occluded section as well as the surround section.



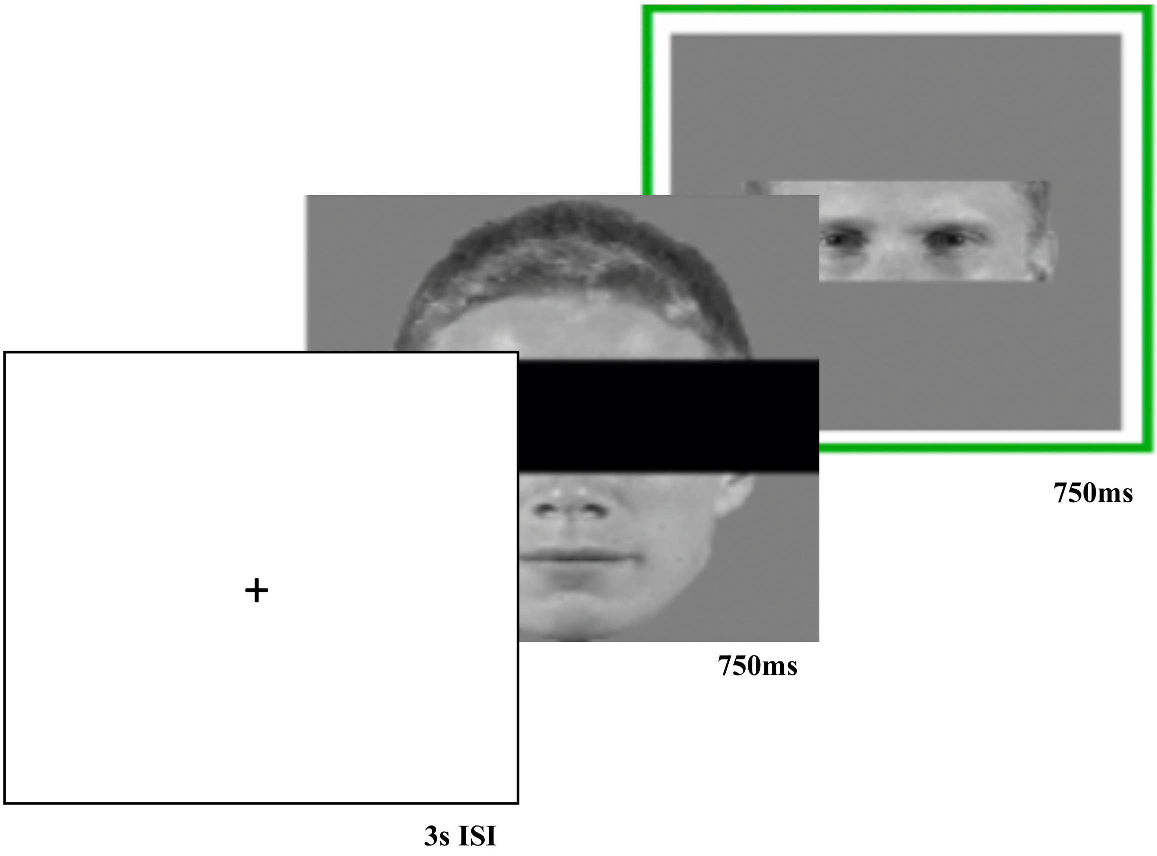
4.3.3. Procedure

Before the study commenced, participants saw an information sheet and gave informed consent to take part in the study. They each completed an MRI eligibility checklist to ensure their safety in the scanner. If eligible and happy to proceed, they were taken to the scanning room to begin the task. Prior to the task, the eye tracker was calibrated for the participant using calibration and validation settings on the EyeLink 1000 Plus. Participants saw a fixation cross which they were asked to look at continuously. The use of eye tracking allows us to ensure that any results are not due to different looking behaviour during certain trials or conditions. We can ensure sure that participants were staying fixated on the cross so the visual experience would be identical across participants.

For each trial, participants saw a cue comprised of one of the whole objects or a neutral stimulus with an occluding bar for 750ms, which was then followed by a target stimulus which was always one of the 8 cut out regions for 750ms. These cue-target pairs constituted either matching, mismatching or neutral pairings, see Figure 4.2. An attention-checking task involved participants pressing a response button when a red-hued version of the cue stimulus was presented, these trials were not included in subsequent analysis. There were nine red stimuli in total, one for each object and one for the neutral condition. Participants took part in 4 runs of the main experiment, with each run lasting approximately 9.5 minutes. There were 89 total trials per run: 9 red trials (excluded from further analysis), match 32 trials (each object four times); mismatch 16 trials (each object twice); and neutral 32 trials (each object four times). Half of each condition's trials were upright and the other half were inverted. Each trial took 4.5 seconds. Mapping checkerboards were displayed at the end of each block, with 12s of checkerboard in the location of the target stimuli and 12s in the surround area with 12s fixation between each (see Figure 4.1). Using this method allowed us to functionally define the areas in EVC that were more active when viewing the target compared to the surround, which helped us to define our ROIs.

Figure 4.2.

An example of a matching trial and associated timings.



Participants were also asked to lie still while an anatomical scan was run to allow clearer analysis of the areas of the cortex of interest. Additionally, they completed a localiser scan run, which used an N-back task in images from the categories of faces, houses, bodies and scrambled requiring them to press a response button when they saw the same stimuli presented in trial N and N-1. A debrief was completed at the end of the scanning session, with the entire scanning session lasting no longer than 90 minutes.

4.3.4. MRI data acquisition

Structural and functional MRI data was collected using a high-field 3-Tesla MR scanner (Siemens Prisma). High resolution T1 weighted anatomical images of the brain were obtained with a three-dimensional magnetisation-prepared rapid-acquisition gradient echo (3D MPRAGE) sequence (34 Volumes, 1mm isotropic). BOLD signals were recorded using a multiband echo-planar imaging (EPI) sequence: (444 volumes, TR = 1268ms; TE = 30ms; flip angle 74; 34 slices, matrix 78 x 78; voxel size = 2X2X2; slice thickness 2mm; no interslice gap; field of view 192; multiband factor 2, Partial Fourier = 7/8, no Grappa). Slices were positioned to cover occipital and temporal lobes. The visual display was rear projected onto a screen behind the participant via an LCD projector, participants observed the screen through a mirror attached to the head coil. Eye movements were recorded using an EyeLink 1000 Plus to ensure fixation. This eye tracker was mounted onto the display screen, using the mirror to observe eye movements throughout, calibration and validation for each participant was completed at the start of the experimental runs.

A short 5-volume posterior-anterior opposite phase encoding direction scan was acquired before main functional scans, to allow for EPI distortion correction (Fritz et al., 2014; Jezzard & Balaban, 1995). An independent functional localiser (Faces, Places, Objects and Scrambled – see (Charest et al., 2014) was run, which utilised a block design where a one-back task kept participant attention. Two runs of this localiser were run per participant, taking approximately 15 minutes.

4.3.5. MRI data pre-processing

Functional data for each experimental run, in addition to localiser runs was pre-processed in Brain Voyager 20.4 (Brain Innovation, Maastricht, The Netherlands; Goebel et al., 2006), using defaults for slice scan time correction, 3D body motion correction and temporal filtering. Functional data were intra-session aligned to the pre-processed functional run closest to the anatomical scan of each participant.

Each participant's T1 weighted anatomical image was pre-processed to extract the brain from the head-volume. Functional data were then coregistered to the participant's ACPC anatomical scan. Note no Talairach transformations were applied, since such a transformation would remove valuable fine-grained pattern information from the data that may be useful for MVPA analysis (Argall et al., 2006; Dale et al., 1999; Fischl et al., 1999; Goebel et al., 2006; Kriegeskorte & Bandettini, 2007). For the main MVPA analyses (described further below) we conducted a GLM analysis independently per run per participant, with a different predictor coding stimulus onset for each stimulus presentation convolved with a standard double gamma model of the haemodynamic response function (Greening et al., 2018; Smith & Muckli, 2010). The resulting beta-weight estimates are the input to the pattern classification algorithm described below (see multivariate pattern analysis). A GLM with one predictor per unique image (89) with separated GLMs by run was used for decoding.

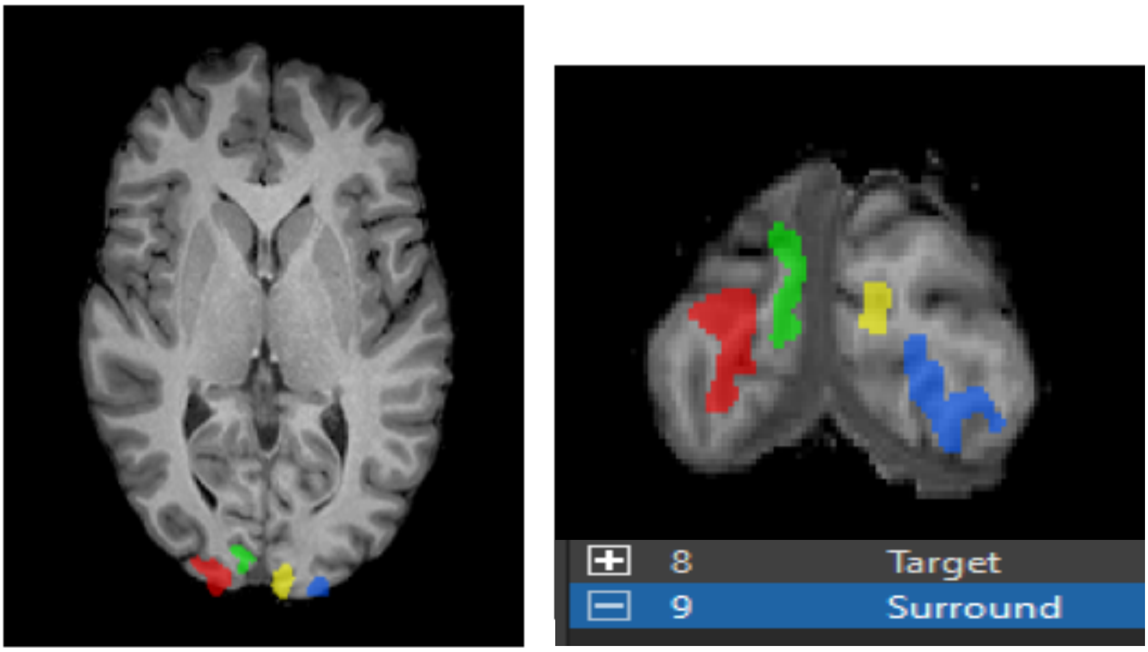
Deconvolution analysis was used for univariate analysis due to the increased ability to model the BOLD response in event-related designs, representing the hemodynamic response function effectively (HRF; Chen et al., 2023). This was computed first of all within participants, then the specific time points of interest and tested across participants for significant differences. The time point analysed was between 3-6 volumes for each of the six conditions (match-upright, match-inverted, neutral-upright, neutral-inverted, mismatch-upright, mismatch-inverted). ANOVAs were run through R studio to analyse the effects of the conditions against potential effects of hemisphere (left or right), orientation (upright or inverted) and ROI (target or surround). Note that if orientation was used as a separate variable, that condition was collapsed into match, mismatch and neutral.

4.3.6. Anatomical regions of interest

These ROIs were created in each hemisphere by using the mapping checkerboards to define which area of EVC had higher activation to the target area (here the occluded bar) and the surround area (the unoccluded area). A contrast of target minus surround was applied and the resulting areas were defined as the target ROI (where there was higher activation for the target over surround areas) and surround ROI (where there was less activation for the target than the surround), which then created our ROIs (see Figure 4.3).

Figure 4.3.

An example of the ROI allocation for one participant from BrainVoyager. Red and blue indicate surround and green and yellow represent the target. The contrast of target minus surround was applied across averaged runs for each participant to ascertain the areas in EVC that had more activation for the target area than the surround area, as defined using mapping checkerboards.



4.3.7. Multivariate pattern analysis

Linear SVM decoding (leave one run out cross-validation) used to decode object identity of target patches separate for each condition (match-upright, match-inverted, neutral-upright, neutral-inverted, mismatch-upright and mismatch-inverted). Trials were subsampled in the match and neutral condition (16 upright, 16 inverted) to equal the number of trials in the mismatch condition (8 upright, 8 inverted) for decoding, which was then iterated 10 times. The LIBSVM toolbox (Chang & Lin, 2011) was used to implement the linear SVM algorithm, using default parameters (C = 1), which uses the 1vs1 method for multiclass classification. The activity pattern estimates (beta weights) within each voxel in the training data were normalised between -1 to 1, before being used in the SVM (Bailey et al., 2023; Greening et al., 2018; Knights et al., 2021; Muckli et al., 2015).

Test data were normalised using the same parameters as the training set, to optimise classification performance. To test whether group level decoding accuracy was significantly above chance, we performed non-parametric Wilcoxon signed-rank tests using exact method on all MVPA analyses, against the computed empirical chance level (Formisano et al., 2008; Greening et al., 2018), with all significance values reported two-tailed. We used a permutation approach – randomly permuting the mapping between each condition and each label, independently per run, to calculate the empirical chance level for each participant and each decoding analysis separately (Bailey et al., 2023).

4.4. Results

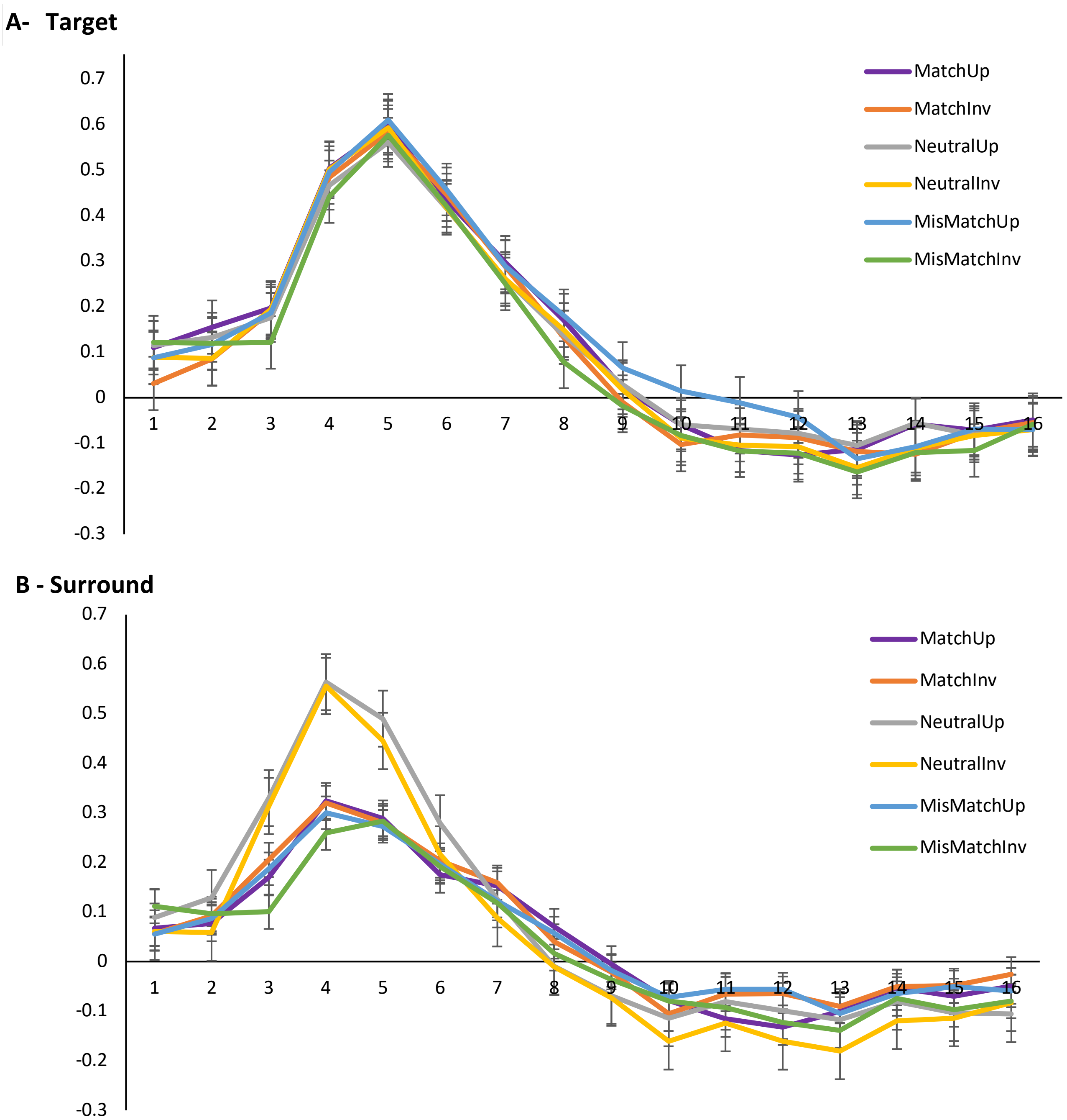
4.4.1. Univariate deconvolution analysis

4.4.1.1. Time course

Deconvolution analysis run through BrainVoyager allowed a time course to be plotted for each ROI (pooled across hemispheres) averaged across participants. These informed our decision to use time points three to six for our analyses as this time frame contained the peak amplitude across conditions and regions, see Figure 4.4.

Figure 4.4.

HRF time course plots for all participants split across Conditions for the **A**) Target and **B**) Surround region of interest. Error bars represent standard error.



A- Target

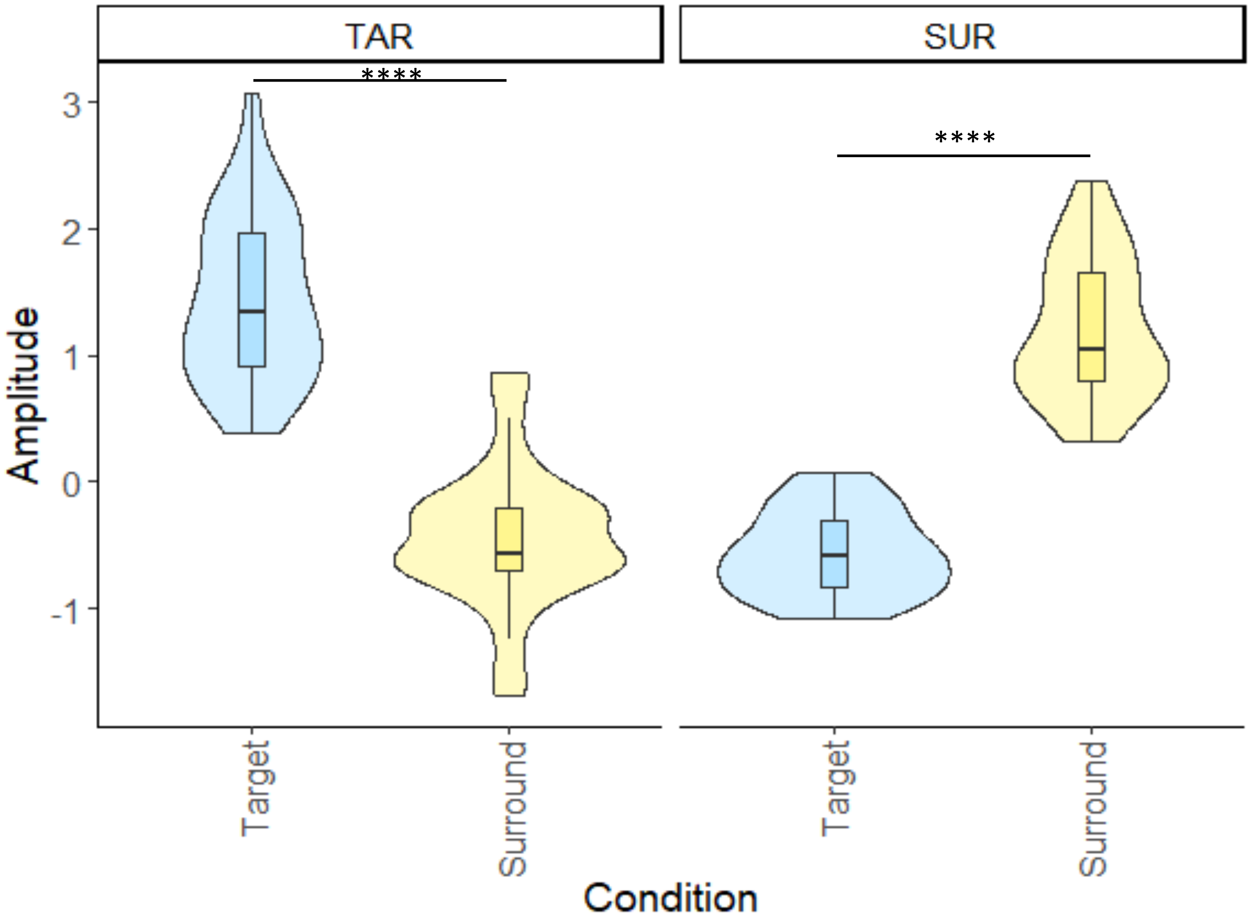
B - Surround

4.4.1.2. ROI validation

We completed a simple circular analysis of the target and surround ROIs to check they had been correctly selected. To do this, we analysed the differences between the mapping conditions using pairwise t tests, FDR corrected. As expected, the target ROI boasted a significantly higher amplitude for the target over surround condition and vice versa in the surround ROI (*W* = 595, *p* < .001)., see Figure 4.5. the target and surround conditions were not analysed in our main ANOVAs.

Figure 4.5.

Violin plot of the differences between the target and surround conditions in each ROI.



Note. * *p*<.05, ** *p*<.01, *** p<.001, ****p<.0001. FDR corrections applied.

4.4.1.3. ANOVA

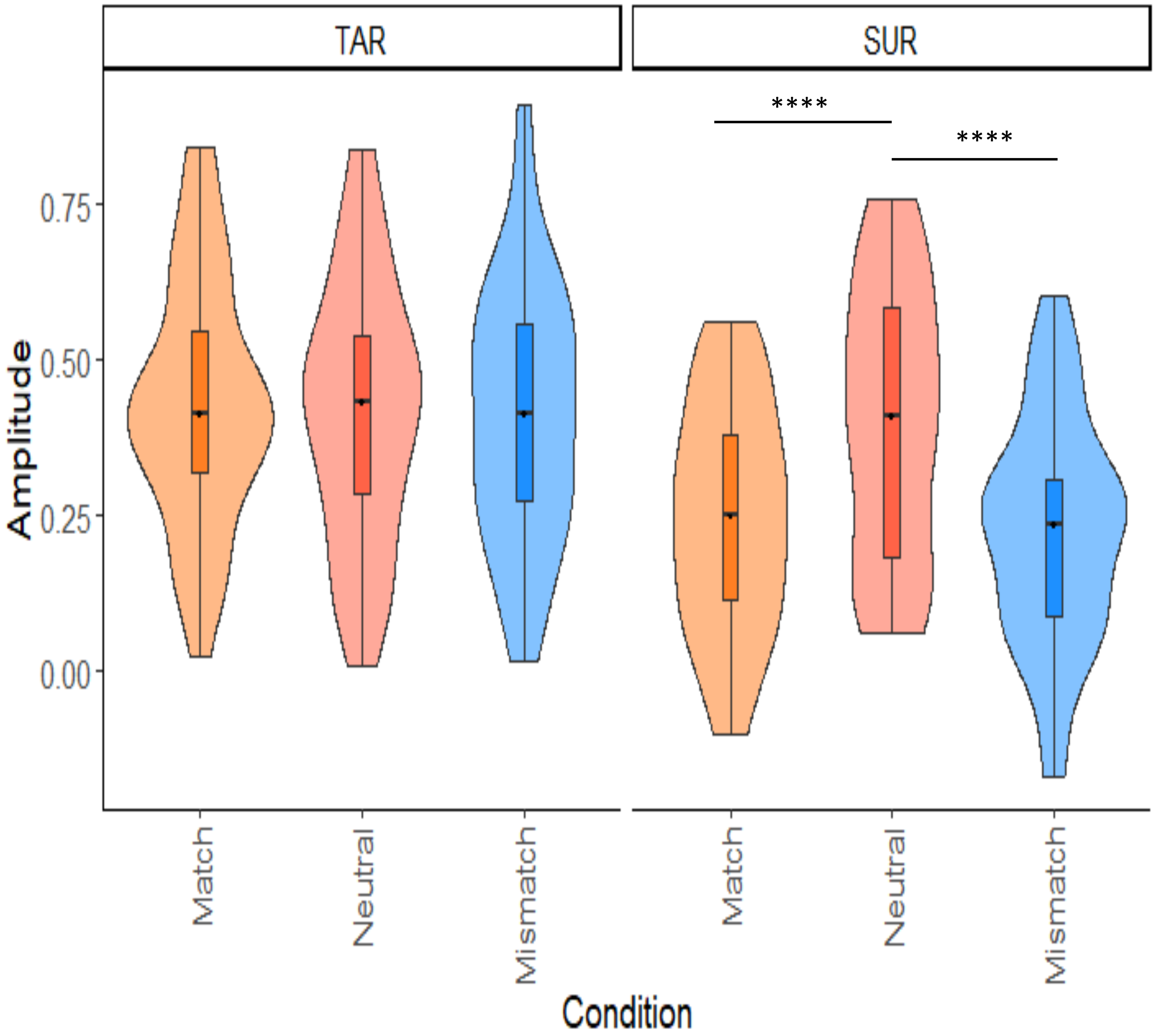
ANOVA testing of condition (matching, mismatching and neutral), orientation (upright and inverted), hemisphere (left and right) and ROI (target and surround) found a significant main effect of condition, *F*(2, 32) = 14.00, $p < .001$ as well as ROI *F*(1,16) = 21.08, $p < .001$. There was also a significant interaction between condition and ROI, *F*(1.38, 22.15) = 41.93, $p < .001$. As well as this, a significant interaction emerged between condition, hemisphere and ROI, *F*(2, 32) = 4.40, $p = .02$.

The interaction between condition, ROI and orientation, while marginally non-significant, is still interesting to consider, *F*(1.35, 21.55) = 3.31, $p = .072$. Thus, post-hoc tests have also been utilised here to analyse these effects for exploratory purposes.

First of all, we explored where the interaction between condition and ROI arose from. Collapsing across inversion, we sought to explore these factors. We found that when pooled across hemispheres, an ANOVA on the target ROI had no significant findings, whereas the Surround ROI showed a significant main effect of condition, *F*(2, 32) = 31.00, $p < .001$. Post-hoc tests revealed significance between conditions across the surround area (see Figure 4.6) where the neutral condition had significantly higher amplitudes than both the match and mismatch conditions. There were also differences between ROIs when split across condition (see Figure 4.7), where match and mismatch conditions both had significantly higher amplitudes in target than in surround ROIs.

Figure 4.6.

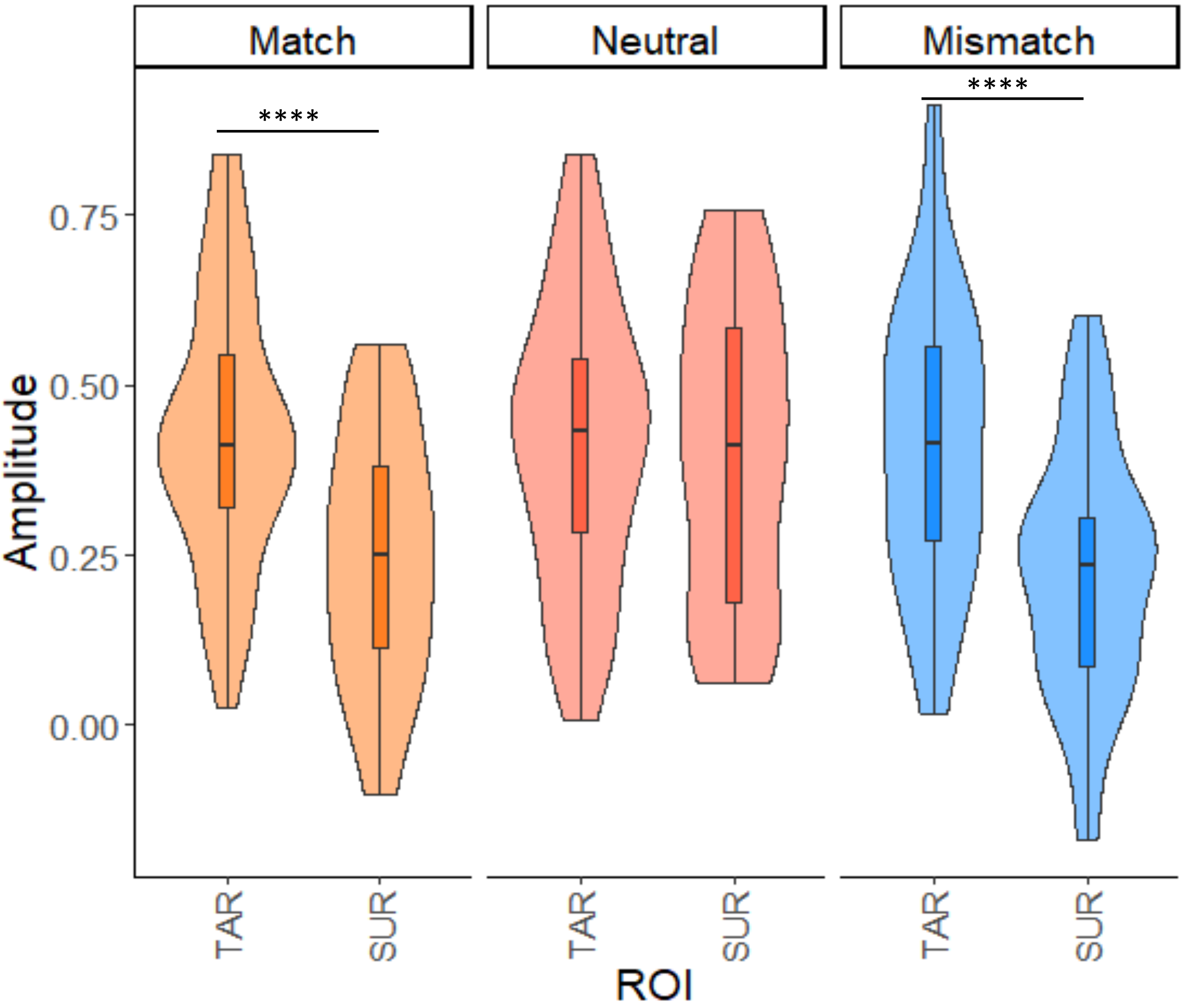
A violin plot of the deconvolution amplitude of each condition, split across ROI (target and surround). Significant paired t tests indicated by asterisks.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001. FDR corrected.

Figure 4.7.

A violin plot of the deconvolution amplitude of each ROI, split across condition. Significant paired t tests indicated by asterisks.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001. FDR corrected.

To follow up the condition, ROI and hemisphere interaction, we ran an ANOVA of these three variables, collapsed across orientation and found effects of condition and ROI as before and a main effect of condition and ROI. To try to unpick why the hemisphere interaction had occurred in the large ANOVA, we ran additional ANOVAs of condition and ROI separately for each hemisphere. It was discovered that the left hemisphere had an effect of condition, $F(2,32) = 15.24, p < .001$, as well as ROI, $F(1,16) = 7.69, p = .014$, and an interaction between condition and ROI, $F(2,32) = 17.23, p < .001$. In the right hemisphere this pattern was mirrored, with a main effect of condition, $F(2,32) = 9.92, p < .001$, ROI, $F(1,16) = 12.96, p = .002$, and an interaction between condition and ROI, $F(2,32) = 47.89$, $p < .001$. In FDR corrected post-hoc t test analyses, the left hemisphere showed significant pairings in the surround ROI split across conditions: match – neutral ($p < .001$) and neutral – mismatch ($p < .001$). In the right hemisphere surround ROI there were also significant pairings: match - neutral ($p < .001$) and neutral – mismatch ($p < .001$), see Appendix J for visual representation. Though this pattern is reflected in both hemispheres, the more significant effects in the right hemisphere may be responsible for the initial interaction of hemisphere, condition and ROI.

4.4.2. Decoding accuracies

4.4.2.1. Wilcoxon

Before completing further analysis, we first tested whether group-level decoding accuracy was significantly above chance we used Wilcoxon signed-rank tests against empirically derived chance levels (Bailey et al., 2023; Formisano et al., 2008; Greening et al., 2018), all significance levels two-tailed. This revealed decoding of object identity as significantly different from chance (0.125) in 19 of 24 conditions ($p < .025$) (pairings comprised of each of the four regions of interest - target and surround for left and right - across the six conditions), with the five non-significant pairings all being within the mismatch condition ($p > .177$). Thus, we used non-parametric Wilcoxon methods and did not analyse the mismatch condition separately.

4.4.2.2. ANOVA

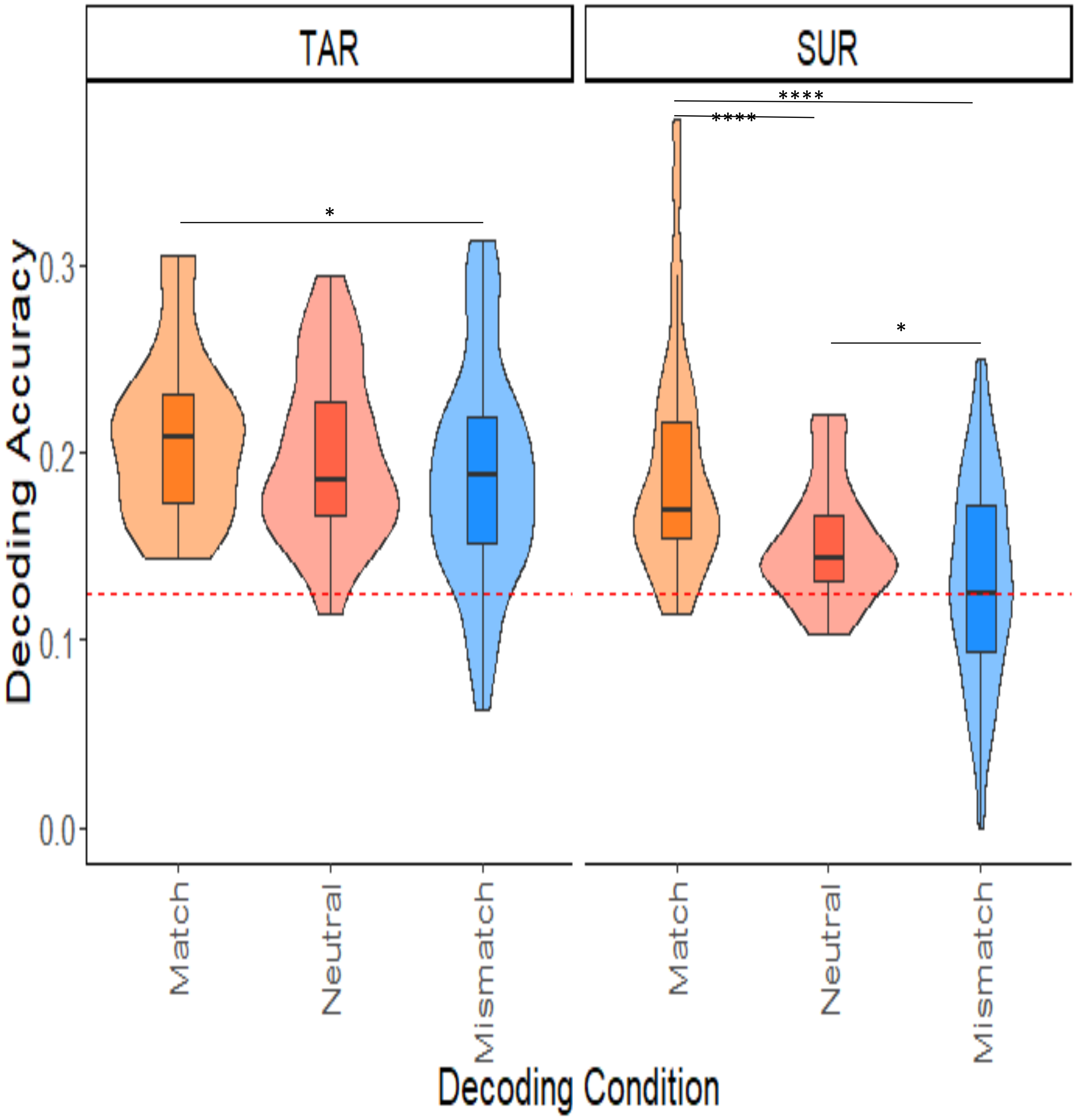
A permutation ANOVA test with 10000 permutations was computed on variables of condition (match, mismatch and neutral), orientation (upright and inverted) and ROIs (target and surround) on decoding. There was a significant main effect of condition $F(2,34) = 11.15$, $p < .001$, as well as ROI $F(1,17) = 30.85$, $p < .001$. There was a significant interaction between condition and ROI, $F(2,34) = 4.40$, $p = .02$.

4.4.2.3. Post-hoc tests

Following this, we collapsed the results across orientation and one way-ANOVAs were carried out to determine the effects of condition on each region of interest. Testing the effect of condition in each region, we revealed target to be marginally non-significant, $F(2,34) = 2.5$, $p = .09$, while surround was significant, $F(2,34) = 13.07$, $p < .001$. Subsequently, Wilcoxon signed rank tests were computed between each decoding condition, and the significant pairings, as seen in Figure 4.8, were corrected for pairwise errors using FDR. The significant differences between the match and mismatching conditions in each ROI suggest an expectation suppression effect, where expected – or predictable – stimuli are better represented in EVC than unexpected stimuli. Comparing between the ROIs when split across conditions (see Figure 4.9) we reveal that decoding in the target ROI is greater than the surround ROI for all conditions.

Figure 4.8.

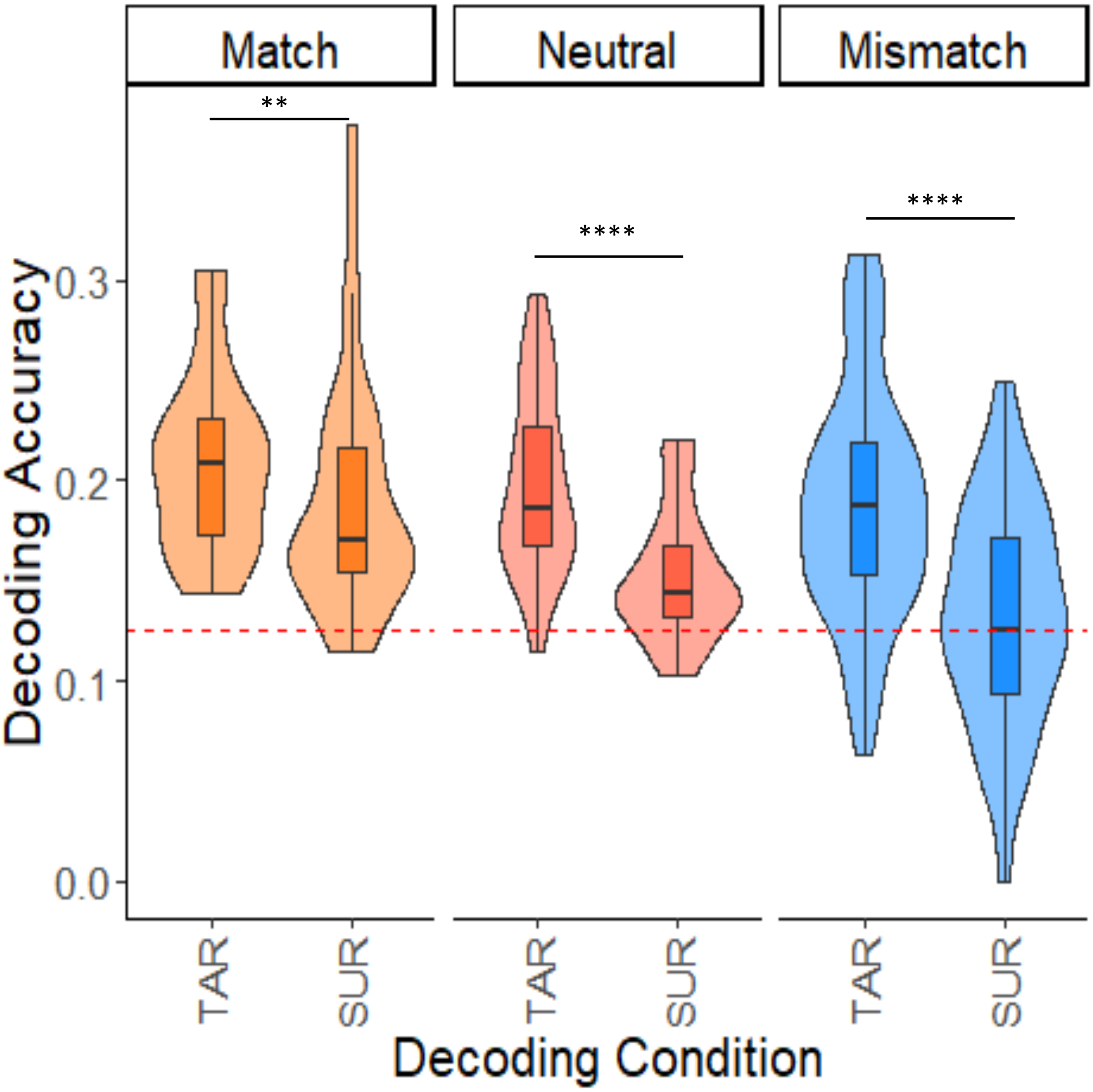
Plot of the decoding accuracy of the decoding conditions split across target (TAR) and surround (SUR) regions of interest.



Note. * *p*<.05, ** *p*<.01, *** p<.001, ****p*<.0001. FDR corrections applied.

Figure 4.9.

Plot of the decoding accuracy of the target (TAR) and surround (SUR) regions of interest split across conditions (match, mismatch and neutral).



Note. * *p*<.05, ** *p*<.01, *** p<.001, ****p*<.0001. FDR corrections applied.

4.5. Discussion

In this study, our aim was to unravel predictive processing in occlusion, specifically through the lens of expectation suppression. Our interpretation of MVPA decoding results leans towards expectation suppression, particularly the sharpening effect. The consistently higher decoding accuracy in the matching, expected condition compared to the mismatching, unexpected condition, in both target and surround regions, supports this view. However, the univariate analyses paint a more surprising picture, showing no effect of match versus mismatch as would be expected in a traditional picture of expectation suppression. We found the neutral condition had significantly higher amplitudes than the match or mismatch condition in the surround region particularly.

More specifically, our decoding analysis revealed an effect consistent with expectation suppression, exhibited through the higher decoding accuracy in the match compared to the mismatch condition across both ROIs. This is as expected, supporting the view that an expected stimulus would be more accurately decoded due to the increased predictability (González-García & He, 2021; Kok et al., 2012; Summerfield & de Lange, 2014; Walsh & McGovern, 2018). Therefore, we can argue that there is evidence of predictive processing occurring throughout this experiment, where the propagation of object representation information and prediction errors cause this effect. This dynamic updating of predictions within trials to explain away expected changes even aligns with results in eye-tracking that discovered V1 activity for updating apparent motion predictions (Edwards et al., 2017), making it clear how robust the mechanisms for prediction are across multiple sensory modalities and paradigms. In the present study, as the representation of the expected match condition is decoded more accurately than the mismatch condition, this aligns with the sharpening account, where neurons attuned for the unexpected stimuli have their activity suppressed (Kok et al., 2012), creating this heightened contrast. The dampening account would have demonstrated a lower contrast between conditions in decoding analyses, where the response for the expected, matching condition would be suppressed relative to the unexpected mismatching condition (Kok & De Lange, 2015).

Our results regarding the ROIs were somewhat unexpected, as we found huge effects in the surround ROI across all analyses, where we primarily expected to find effects in the target ROI. However, these differences may suggest that the surround ROI is more affected by the spatial context shaping the visual processing in early visual areas. It has been determined that temporal and spatial context do play a role in object recognition, specifically when undertaking pattern completion, thus we suggest that these surround findings may be a result of this (de Haas & Schwarzkopf, 2018; Tang et al., 2018; Wyatte et al., 2014). The surround region initially receives full stimulation with high contrast, which is not simply ignored as the trial continues from the cue occluded image to the target bar image.

Additionally, these somewhat unexpected results between the target and surround regions may also be the result of the higher complexity of our stimuli and paradigm. This is because where previous expectation suppression research tends to only use one complex visual scene, image or motion stimuli per trial paired with a colour or auditory cue to investigate the ability to investigate effects of expectation (Edwards et al., 2017; Egner et al., 2010; Smith et al., 2018; Smith & Muckli, 2010; Walsh & McGovern, 2018), our trials contain a cue-target pair subject to occlusion that may contain two object identities the visual system is required to parse. This may be the cause of the match versus mismatch effects as there are two different object identities in the mismatch but not the match trials. We are aware that in higher visual areas, there are improved abilities to recognise multiple objects simultaneously, while this is a more taxing process in early visual areas (Mansfield et al., 2023). Thus, this added complexity may offer an alternative explanation for the effects we found in the univariate analysis. We may have unlocked some insight here into how prior context shapes the surround region, even when there is not a stimulus being processed in the target section of the cue-target pair.

Deconvolution analysis first ensured that our target and surround regions were correctly representative. Our analyses then moved towards investigating the specifics of the interactions between condition and ROI, as well as the significant condition, ROI and hemisphere interaction and the marginally non-significant interaction of condition, ROI and orientation to determine what was driving these effects. The results of the deconvolution analysis did not yield the expected trend of mismatching conditions showing higher amplitudes. However, as the decoding is much more statistically powerful than the univariate analyses it may be that the complexity of the paradigm, with multiple ways to mismatch in addition to the neutral condition and potential multiple object representations across trials caused the effects only to occur statistically when more powerful analyses were used.

The findings regarding the time course analyses for the univariate analyses paint an intriguing picture of the data, where in the target ROI the amplitudes peak at around 0.6 for the task conditions, whereas in the surround ROI the amplitudes for the main conditions peak around 0.3, whilst the amplitudes for only the neutral condition peak around 0.6. Findings suggest a differing effect between how the neutral stimulus condition was processed and demands additional thought to try to unpick the implications of this in the surround versus target regions. It may be that the neutral stimulus pattern of the prime acted as a highly mismatched and unpredictable cue-target pair, with the expectation never able to be fulfilled. It may be due to the difference in low level statistics where the very visually different neutral stimuli could be driving the effects. But these speculative arguments would need additional research to compare them to the match and mismatch conditions, as there was no control for the neutral condition. Perhaps future research could use neutral patterns as cue and target so that the expectation could be fulfilled on occasion and the neutral condition is more controlled for. Or we could utilise blurred target patches to increase the chances of finding an effect due to increased difficulty in potential prediction.

The interaction between condition, ROI and hemisphere introduces additional complexity, suggesting nuanced hemispheric involvement in processing these conditions within the target and surround regions. Intriguingly, post-hoc tests conducted separately for each hemisphere revealed a consistent pattern of results, indicating that the observed nuances may be due to the larger effect in the right hemisphere in the initial interaction.

To address the unexpected findings within the univariate analysis, there has been some consideration of limitations of the paradigm. While this is a novel project for looking at this topic, this has meant that several of the methodological choices were made from best guesses based on previous, similar but not matched, research. Looking at previous fMRI studies utilising expectation suppression methods, they often train using arbitrary pairings for predictions, for example, a certain auditory beep signals a grating at a 45 degree angle whereas a different one signals a grating at 145 degrees (Kok et al., 2012). In the present study we used recognisable objects and relying on the implicit associations that participants already had to these objects so did not teach a new association. This difference may go some way in explaining why a pattern of results in the univariate analyses is different to the previous expectation suppression results.

We could argue that the differences between the potential expectation suppression results in the decoding analyses and the lack thereof in the univariate could come down to the high contrast stimuli of the target object. As participants saw this high contrast stimuli in all conditions, it may be that the responses are all equal in the target ROI because of this. This could be due to the high contrast saturating the neurons within the visual system to a higher extent than had been anticipated. The pattern of results in the decoding does align with expectation suppression, which is statistically more powerful than univariate analysis, however without the combination of the decoding and univariate analyses it is not possible to determine for definite that expectation suppression is responsible for these effects.

However, results from Smith and Muckli's (2010) occlusion study determined interesting early visual effects even without being able to test for expectation suppression effects as their results only used MVPA and not univariate analysis. Showing that robust effects of spatial context still enabled advancement in understanding of the early visual cortex, specifically V1, even without being able to test for expectation suppression. In future research, we may be better placed to do some computational modelling to try to grasp the patterns found here, particularly using recurrent neural networks as these connections are thought to be effective in challenging recognition scenarios (Nayebi et al., 2018).

Moreover, the bar section in the cue currently aligns with the whole target image which is only comprised of the bar. This may be the reason that the surround region has rich effects. So, to keep a clear focus on the target section alone, the surround area of the target stimuli of the cue-target pair could be the same throughout the cue-target pair, with just the target bar area changing, rather than being blank. This would perhaps cause neurons to suppress the expected surround section, as there would always be a validated expectation there, causing less of an effect in the surround region.

Considering the findings within IT from previous studies (Mansfield et al., 2023; Tang et al., 2018), it would be a beneficial avenue to pursue to analyse the results here in the higher visual areas to see if there is a particular point in the ventral stream that enables expectation suppression to occur with this occlusion stimuli. Maybe the additional visual complexity of representing two objects is more robustly represented in the higher order areas which are more equipped to allow overall pattern completion from both inferred and visible object information. This could explain to us whether the surround region of the stimulus is particularly affecting the EVC specifically or whether the paradigm itself just did not account for the spatial representation of the surround region across time to cause such large effects in the neutral condition in the univariate results.

It may be the case in the current study that the surround area, with its high contrast, clear visual information, was exhibiting its own type of predictive effect, albeit in a different order to the cue-target pairing we has designed. It may be that the participants were drawing across the spatial representation of the surround sections of the objects across the time of the trial and predicting what should be carried across to complete the object. For example, if in one cue-target pair, the cue was the face with the occluded bar, the representation of the face may be filling in the occluded bar, but across time the representation of the cue stimulus would still shape the representation in the surround area even during the target bar. This is similar to the face paradigm employed by Smith et al. (2018) that used sections of faces to look for predictive processes occurring. Therefore, it would be an interesting avenue to approach this paradigm with EEG, which has a better ability to gain fine-grained temporal data. This may allow us a more thorough understanding of how the representations within prediction and amodal completion take place across time in the early visual areas. There is currently a clearer view of how this is achieved in higher order areas such as lateral occipital cortex and fusiform face area than early visual cortex which is generally regarded as unclear (Thielen et al., 2019).

4.5.1. Conclusion

In combination, it appears as if there is tentative evidence for a sharpening account of predictive processing due to the superior representation for matching versus mismatching in both ROIs across decoding, with the matching condition even having higher decoding accuracy than neutral in the surround ROI, where decoding is statistically more powerful than the univariate analysis. However, the lack of expectation suppression-like patterns appearing in the univariate analysis demonstrates a need to improve the paradigm to better capture the intricacies of expectation suppression across object recognition. This could perhaps be addressed by the surround section of the cue image being showed again in the target section of the cue-target pair so the occluding bar section would be the only part of the stimulus pair that changed.

Chapter 5. General discussion

5.1. Thesis overview

The overarching goal of this thesis was to advance our knowledge of the processes underlying object recognition, particularly in challenging visual scenarios such as occlusion. To achieve this, we employed both neural and behavioural measures, delving into two key areas. Firstly, we explored the impact of utilising multiple objects as occlusion stimuli, employing real object images to serve as both the occluding and occluded objects (discussed in Chapters 2 and 3). Secondly, we investigated whether predictive processing mechanisms play a pivotal role in facilitating recognition across objects with occluded sections (explored in Chapter 4).

This concluding chapter serves as a comprehensive synthesis of the prior chapters, analysing the outcomes of the three experimental chapters. It not only examines the practical implications of our findings but also delves into the theoretical underpinnings. Furthermore, this chapter will address potential limitations, discuss methodological considerations, and propose potential avenues for future research. Through this, we aim to contribute to a deeper understanding of the mechanisms involved in object recognition and representation under challenging visual conditions.

5.2. Summary of findings

5.2.1. Summary of Chapter 2 results

In this study on visual object recognition under conditions of occlusion and deletion, we investigated the differential impact on EVC and IT across different decoding methods. Our results revealed significant differences between EVC and IT in terms of their responses to multiple objects. Notably, IT exhibited greater tolerance to the presence of multiple objects compared to EVC.

Cross-decoding analyses illuminated the distinctions in processing the visible and inferred features of object pairs in IT, demonstrating its ability to decode the identity of multiple objects simultaneously with great efficiency, even if features were only inferred rather than visible. This differed in EVC where a higher reliance on visible features was observed. Comparisons with Johnson and Olshausen's (2005) work underscored the effects of multiple objects when recognising occluded objects, showing potential competition effects in recognising multiple object identities.

Synthetic decoding, motivated by multiple object presentations being decodable from the combination of activity patterns to each stimulus in isolation, was extended to work here with occlusion rather than two simultaneously presented (but separate) objects (MacEvoy & Epstein, 2009). Our findings in EVC again indicated a reliance on the visible visual features of objects, while IT demonstrated similar decoding accuracies for both visible and inferred features. This suggests that IT employs a higher-level model involving hidden visual features for predicting occluded object identity. The study supports the concept of pattern completion through the visual stream (Kanizsa, 1976; Tang et al., 2018; Zhu et al., 2019), with IT processing completed object representations well.

In summary, our study contributes valuable insights into the neural mechanisms underlying object recognition in challenging visual conditions, emphasising the importance of including additional visual information as occluders and highlights the differing processing capabilities of EVC and IT.

5.2.2. Summary of Chapter 3 results

This chapter provided an exploration into the impact of occlusion and multiple objects on both behavioural and neural aspects of recognition. The behavioural and neural results combined highlight the differential effects between IT and EVC, with EVC in particular being more affected by the magnitude of occlusion and the associated challenges of representation of occluded objects, while this was not the case in IT. In IT the correlations reflected a linear increase in beta weights with recognition difficulty, which is potentially representative of more processing being allocated, when the recognition was more difficult. Specifically, when accuracy was lower or RTs were slower.

Behaviourally, the study revealed enhanced accuracy and faster RTs for unoccluded conditions. When objects served as both the occluding and occluded objects, recognition differences appeared for the occluded back object, with the deleted condition being more successfully recognised. This departure from previous research (Johnson & Olshausen, 2005) suggested a unique aspect of recognition when multiple objects are being perceived at once.

The integration of the behavioural and neuroimaging data provided a more comprehensive understanding of how object recognition takes place under occlusion. This study emphasises the importance of adapting research methods to capture the nuances of vision in more ecological contexts. It also makes it clear that approaching the same question from a number of different methods or analysis types offers greater insight and speaks to a multidisciplinary approach in neuroscience and vision being worthwhile to pursue.

Overall, this chapter has gone some way to extend prior research by examining both behavioural and neural representations of objects under challenging visual conditions of occlusion. The observed behavioural difficulties in the occluded back condition, coupled with distinct neural patterns in IT provide valuable insights into the intricate interplay of factors influencing object recognition when objects are occluded or deleted. These include the ability to of IT to adjust resource allocation and processing in the face of increased difficulty, as reflected in lower accuracy and slower reaction times. Furthermore, the behavioural differences in recognising the occluded back object condition in the presence of both occluding and occluded objects provide novel perspectives. The finding that the deleted condition is more accurately identified than the occluded back condition highlights the complexity of object recognition when perceiving multiple objects simultaneously.

5.2.3. Summary of Chapter 4 results

Chapter 4 delved into the theory of predictive processing, specifically looking at this with regards to occlusion, using expectation suppression in EVC. Decoding results from MVPA indicate an expectation suppression effect, which aligns particularly with the sharpening account of predictive processing. Consistently higher decoding accuracy in the expected, matching, condition compared to the mismatch condition, in both target and surround ROI supports the anticipated predictive processing pattern (González-García & He, 2021; Kok, Jehee, et al., 2012; Summerfield & de Lange, 2014; Walsh & McGovern, 2018).

Univariate analyses did not reveal results in line with expectation suppression, instead showing a substantial influence of the surround ROI on observed effects. These findings suggest a potential impact of spatial context on visual processing in the early visual areas. This was a novel task, using complex stimuli involving recognisable objects which hold implicit associations and having the potential to show multiple object representations across time and space in each trial. Thus, even as more questions arise about exactly how these effects occur, there have still been valuable insights garnered into how context shapes the surround region, even without a stimulus present in the target section of the image.

Overall, this study offers tentative evidence for a sharpening account of predictive processing during occlusion in EVC. The fact that univariate results did not differ highlights the need for paradigm improvements to capture the intricate dynamics of expectation suppression during object recognition under conditions of occlusion.

5.3. Integrating findings and relation to the broader literature

Our ability to recognise highly occluded objects in busy visual scenes is striking. Relying on the interplay of areas of the visual cortex provides detail from the broad spatial and feature-based facets of objects to the specific contextual and categorical details informed by prior experience (Kok & De Lange, 2015; Schyns et al., 1998; Tang et al., 2014). Using multiple objects appearing together in paired occluded configurations was our attempt at increasing the ecological validity of object recognition, where we very rarely – if ever – see an isolated object. This complexity within the novel stimulus set created for use in Chapters 2 and 3 added some additional detail to the knowledge of occlusion, particularly when multiple objects are being represented simultaneously. This gave us valuable information regarding the brain regions that were particularly implicated in this process, specifically allowing the assessment of how higher order and early visual areas represented occluded compared to deleted objects. Though previous studies had looked at occlusion and visual clutter (Johnson & Olshausen, 2005; Reddy & Kanwisher, 2007; Spoerer et al., 2017; Tang et al., 2018), there was little research that had looked at how the impact of another object as the occluder was mitigated visually.

Grounded in ecological validity, Chapters 2 and 3 underscore the necessity of adapting research methods to capture the intricacies of vision in real-world scenarios. This methodological shift helps to both refine our experimental approach as well as contributing to a broader trend in neuroscience toward more ecologically valid research. In the realm of vision where variations in size, shape, luminance, contrast and visual occlusion abound (Carlson et al., 2011), viewing one stimulus at a time falls short of capturing these nuances appropriately. The contrast between the findings of the present study in Chapter 2 and Johnson and Olshausen's (2005) work further emphasises the significance of ecological context. The inclusion of an ecologically valid multiple object pairing revealed a cost for simultaneously processing multiple objects, even when the focus was only one object in the task itself. This finding diverged from Johnson and Olshausen (2005) where the occluded condition was better recognised than the deleted condition, potentially influenced by their use of two-dimensional ovals for occluding and deleting the target object which lacked additional visual information.

It became clear when examining the literature that areas within the ventral visual stream were key for visual object recognition and classification (DiCarlo & Cox, 2007; Fyall et al., 2017; Gazzaniga et al., 2018; Goodale & Milner, 1992; Sorooshyari et al., 2020; Wyatte et al., 2012). A specific area of focus was V1, the primary visual cortex is implicated in studies where grating stimuli can be predicted based on learned associations (Kok et al., 2012), as well as in more complex occlusion paradigms (Morgan et al., 2019; Smith & Muckli, 2010). This research necessitated the exploration of EVC under more challenging visual conditions to determine the complexity of the object representation and categorisation that could be achieved here. Additionally, higher order visual areas such as IT have been determined to represent highly complex objects incredibly accurately (Kreiman, 2008; Mur et al., 2013). Thus, our comparisons of visual occlusion across these two areas of the ventral visual stream have allowed us to determine benchmarks of processing. Chapters 2 and 3 allowed comparison of EVC with IT and mid-visual regions to understand how the ability to process occluded objects changed along the visual stream, whereas Chapter 4 had a focus specifically on EVC to better understand potentially impacts of predictive processing in early vision. Consistent with prior findings we discovered that representations of occluded objects were still created in EVC, but IT was more tolerant to multiple object representations.

The concept of predictive processing has emerged as a guiding principle throughout our exploration of object recognition. The idea that the brain better explains sensory input by minimising any error in the predictions propagated along the visual stream provides an efficient theory for challenging conditions of object recognition among other concepts (Clark, 2013; Mills et al., 2021). Whilst we can argue that the effects in Chapters 2 and 3 may be due to predictive effects, it was not possible to quantify directly. Thus, Chapter 4 was dedicated to specifically investigating predictive effects during occlusion in order to reinforce and expand upon the consistent previous findings across EVC (Morgan et al., 2019; Smith & Muckli, 2010). Using expectation suppression methods allowed this, with the understanding that a reduction in the measure of neural activation following predicted stimuli represents predictive processing (Feuerriegel et al., 2021), though our univariate results did not show this. The sharpening account aligns with the heightened decoding accuracy in the expected condition, contributing empirical weight to the theoretical framework introduced in the literature review of Chapter 1 (González-García & He, 2021; Kok, et al., 2012; Kok & De Lange, 2015).

Throughout the thesis it has been clear that there is much to be gained by utilising an interdisciplinary approach. The combination of both behavioural and neural measures has not only improved our understanding of the intricacies of object recognition under challenging visual conditions but has also emphasised the importance of employing varied methodologies (Kriegeskorte, 2015). Chapter 3 in particular displayed the importance of integrating neuroimaging and behavioural data. By examining both aspects, we not only unravel the differential effects of activation in IT and EVC, but also glean insights into the cognitive processes involved in recognising occluded objects. This holistic perspective surpasses the confines of our specific investigations, offering a broader understanding of the interplay between neural mechanisms and behavioural outcomes in the realm of object recognition. Taking a step further, the integration of computational methods like CNNs using complex stimuli like those presented in our studies could enhance the analysis of visual cognition processes (Bracci & Op de Beeck, 2023; Nayebi et al., 2018). This combination may empower us to simulate the brain's processing of occluded objects, facilitating hypothesis testing and refinement of theoretical frameworks, Such advancements build upon foundational insights derived from human neural and behavioural studies (Kriegeskorte, 2015). In essence, the multidisciplinary approach allows us to grasp the complexities of vision more comprehensively, leading to richer insights that go beyond the confines of individual research methodologies alone.

Amodal completion was a focus within Chapter 2 and beyond, showing how mechanisms unfold during the representation of occluded objects (Tang et al., 2018). The nuanced exploration of visible and inferred features in IT hints at a higher-level model involved in predicting occluded object identity, where IT better accessed the whole object representation from the parts available. This contributes not only to our specific study but integrates with broader discussions on how the visual stream engages in pattern completion (Ao et al., 2023; Nanay, 2018b; Thielen et al., 2019; Weigelt et al., 2007). The ability to link missing parts of an object and still robustly recognise this within a short time frame has been well-researched as an explanation for object recognition under occlusion. In the current work, we have found that objects may be ‘completed’, that is, they are still able to be accurately represented even when another object is occluding them.

These results suggesting EVC lacks significant ability to complete objects contrasts with the findings of Smith and Muckli (2010), who observed more prominent EVC effects. It’s key to note that the difference in results may be attributed to methodological distinctions. Smith and Muckli (2010) employed natural scenes with rich details, occluding the lower right quadrant with a white box. In contrast, our study introduced a layer of additional complexity with occluding front objects, potentially influencing neural processing in a different way. The increase in low-level visual properties increased for the paired object representations, meaning there was a potential increase in attentional demand due to the additional visual components requiring understanding to process the identity of the target object. It is also worth noting that Smith and Muckli (2010) found distinctions between V1 and V2 effects in this study, with V1 acting as the driving force for context effects, with V2 showing less evidence for this. As our EVC region encompassed V1-3, it may be that the overall EVC effects are somewhat averaged across regions. Their findings alongside the present results so make it clear that the context surrounding the target object and the attention allocated to that is a key topic of future study.

The cost of processing multiple objects has been well-observed in Chapters 2 and 3, and potentially Chapter 4. Competition has been found to be an integral part of visual recognition, as there is no way to look at everything in a scene simultaneously (Trapp & Bar, 2015), which aligns well with an associated cost of processing multiple objects at once in the case of occluded or cluttered objects (Spoerer et al., 2017). In Chapter 2 this can be observed in the first cross decoding analysis, where the presence of more than one object in the occluded pairing causes significantly reduced decoding accuracy when compared to the single object showing in the EVC during the deleted condition. This differed from the research by Johnson and Olshausen (2005) where deleted objects were more poorly recognised than occluded objects due to the relative lack of meaningful object information and thus competition that the two dimensional ovals introduced in the occluded condition compared to our meaningful object occluders. In early visual areas we revealed a competition effect that was not seen in IT, which is known to be more tolerant to the presence of multiple object representations simultaneously.

Additional decoding enhanced our understanding of the potential mechanisms of neural elicitation of object representations across these regions of interest, with the finding that EVC responses to occluded objects were better determined by the visible visual features, while in synthetic decoding IT the visible and inferred features were equally successful at predicting the identity of the occluded back object. Though in the second cross decoding condition, there were further differences in the pattern of decoding between IT and EVC, where there were significant differences in IT as well as EVC, though these findings were in opposing directions. In EVC, again the lower level model explanation where visible visual information was shown presented a higher decoding accuracy. But in IT, there was a significantly higher decoding accuracy for inferred over visible features. This may demonstrate that the ability to predict an object using conceptual prior experience and top-down predictions is incredibly robust within IT.

In Chapter 3 this was compounded by the combination of neural and behavioural results. In IT, the increase in beta weight for the back occluded object couples with slower reaction times and decreased accuracy to suggest that there is more room to adapt resources in this area when difficult visual conditions occur (Jozwik et al., 2023; Spoerer et al., 2019).

Whereas in EVC this was not the case, which may explain why there is more of an effect of the magnitude of occlusion. Here we instead discovered that when more visible visual information was present, the ability to correctly represent the object was better, while this effect was not as prevalent in IT.

Chapter 4 underscores the impact of spatial context, demonstrating particular effects in the surround region of stimuli. This exploration introduced layers of additional complexity to visual perception, challenging conventional approaches that often create simple arbitrary associations across trials in expectation suppression research (Aitken et al., 2020; Egner et al., 2010; Kok et al., 2012). Previous paradigms often lacked the nuance of using existing associations for objects that could be carried across time and space during trials, despite the integration of spatial and temporal features often playing an important role in pattern completion mechanisms (Tang et al., 2018; Wyatte et al., 2014). Thus, this study prompts a re-evaluation of existing models, highlighting the pivotal role of contextual information in shaping early visual representations, impacting predictive processing and object recognition.

In light of the need for paradigm improvements, particularly in the context of predictive processing, Chapter 4 urges us to refine our experimental approaches. To capture intricate dynamics of expectation suppression during object recognition under occlusion, we acknowledge the evolving nature of experimental design. Whilst creating novel tasks generates challenges, being able to add new understanding regarding the spatial influence of the target and surround areas in object recognition and expectation suppression is important to keep developing the field of vision. Here we added extra understanding to previous work that showed V1 activation in occluded sections using MVPA decoding (Smith and Muckli, 2010), and were able to measure expectation suppression through our method using MVPA, though not univariately. Additionally, the stimuli from Chapter 2 and 3, utilising pairs of occluded stimuli and the deleted counterpart tells us a lot about how more naturalistic research can be achieved and what we stand to gain from this. That is, more ecologically valid stimuli will potentially be able to inform us about how the human brain processes real-world objects much more efficiently than the same study not using real-world object images would have done. While simple stimuli like gratings are informative in exploring basic processes, it is important that we can conceptualise this and take it further where it may have real-world applications in fields like computer vision. Our findings would then hold broader implications for existing cognitive models of object recognition. This thesis has demonstrated that to improve our cognitive understanding of object recognition we should be utilising more life-like visual scenarios to prompt more life-like visual responses, which in turn could improve understanding at human levels as well as beyond.

5.4. Limitations and future directions

In striving for high quality research, we acknowledge areas where improvements could be made, particularly when using novel stimuli and paradigms. Though each study in this thesis was developed with prior research firmly in mind, there is a novel aspect to each. In Chapters 2 and 3, this provided an additional avenue to measure neural mechanisms of occlusion compared to deletion in a way similarly employed by other researchers (Johnson & Olshausen, 2005; Tang et al., 2018), while Chapter 4 took a slightly different approach. When measuring expectation suppression, previous research often did not have an occlusion basis (Egner et al., 2010), used very different stimuli such as gratings or scenes (Aitken et al., 2020; Kok et al., 2012), and often created arbitrary expectation pairs (Xu et al., 2021). This meant that it was hard to develop the task and stimuli to ensure the best data collection, knowing that task relevance plays a part in the attentional basis of expectation (Kok et al., 2016; Kok, Rahnev, et al., 2012). Our novel data decoding results still paint a picture that suggests an effect of prediction, specifically presenting a sharpening account; however, the lack of a univariate result in line with expectation suppression suggest that there could be improvements to the paradigm and method to better capture the nuances of recognition.

There was a substantial difference of contrast in the target and surround area of the cue stimuli, with the black bar and the clearly presented object which may have caused a saturation or even supersaturation of the EVC neuron receptive fields due to the high contrast. Nonlinear in nature, the response of some neurons to being faced with high contrast can cause the neural response to plateau or even decrease, though research on supersaturation is not common and is largely based in non-human samples (Peirce, 2007). This pattern may suggest that areas of high contrast like the occluding bar compared to the surround region may have incredibly distinctive patterns of activation, which is the case in our Chapter 4 results where the surround region presents large effects. To combat this, future studies could benefit from visually degrading stimuli, potentially through techniques like blurring, as demonstrated by Rossel et al. (2022). Another way to do this could be through visual degradation of the target region while leaving the cue intact, in a way similar to Blank & Davis (2016), who in a sound study used degraded words to measure how prior expectations affected perception of degraded speech. In vision, this adjustment of including visually degraded stimuli would aim to prevent the saturation of visual processing resources within EVC, allowing greater power to detect differences in univariate analysis.

Temporal dynamics are pivotal in understanding recognition and expectation processes (Rohenkohl & Nobre, 2011). As Xu et al. (2020) demonstrated through their temporally distinct measures of predictive processing accounts, being able to understand the time points at which these processes of recognition and expectation occur is beneficial, particularly when noting the short time frame complex recognition occurs in (Tang et al., 2014). To offer a more illuminating picture of how the temporal effects of trials in Chapter 4 may have affected the expectation suppression we could see in fMRI, it would be important to run an MEG study as a counterpoint using the same stimuli structure. Measuring the influence of these occluded stimuli, with particular interest paid to early visual areas, would allow us to better understand the way that prediction or expectation could be manifesting across time when implicit expectation associations were being called upon. Previous research by Doherty et al. (2005) demonstrated effects of spatial and temporal attention working in combination to modulate perceptual attention in a 2D occluded motion task where a 2D ball image seemingly moved behind a grey occluding strip and appeared on the other side either at the expected trajectory and speed of motion or not. Therefore, while their 2D movement occlusion task is not the same paradigm as Chapter 4, it is noteworthy that their effects of spatial and temporal orientation increase performance accuracy and change ERP modulations. This could provide some insight into the effects that we observed in the surround region for the neutral condition, where there were significantly higher amplitudes here over the match and mismatch conditions. In addition to improvements within the stimulus presentation as mentioned above, this combination of methods to create a broader and more comprehensive view of the process would enhance our understanding of how predictive processing could be affecting object recognition under occlusion.

Using an online study to conduct the behavioural study for Chapter 3 was a necessity of the COVID-19 pandemic and allowed us to collect a useful behavioural counterpart to the fMRI study of Chapter 2. This meant greater opportunity for distraction with less ability to control how the task was carried out. To attempt to mitigate this, we did include catch trials and those who did not pass these were disregarded from further analysis. We also narrowed down the devices and browsers that participants could use to ensure that the timing was as accurate as possible. This required participants to use Firefox as a browser on a desktop or laptop, not a mobile device. Gorilla, the platform itself, was determined to have the highest degree of timing accuracy under these conditions (Anwyl-Irvine et al., 2021). We did find reliable comparisons between the behavioural data and independently measured fMRI signals in IT, which suggests we were successful in capturing the attention and effort of our participants. Though if we were to run this as a lab-based study in the future, an advantage would be that we would be able to use incredibly fast masking which links highly to interrupting recurrent processing which has been implicated in the ability to represent object identities (Tang et al., 2018).

Additionally, while Chapter 4 has revealed much about the early visual areas and how they may process expected versus unexpected cue-target pairings, it would be remiss to not suggest a future plan to analyse the activation in IT. In light of our findings from earlier Chapters, IT has an incredible ability to represent occluded object pairs, often in a different way to EVC. Thus, being able to measure and directly compare this potential expectation suppression and prediction effect in terms of early and higher visual areas we could learn how the spatial and temporal context and implicit associations affect IT. While expectation suppression has been measured largely in the EVC (Aitken et al., 2020; de Lange et al., 2018; Egner et al., 2010; Kok et al., 2013), the ability to understand what is occurring in the higher visual areas would improve our knowledge of the overall picture of processing challenging visual conditions like occlusion. Based on our previous research and the associated literature, it would be expected that using this paradigm, we would see an effect of prediction in the IT cortex, where the expected results had lower amplitude and higher decoding accuracy. This would be due to the prediction of the cue target pair being easily inferred in IT from the visual context available in the cue stimulus.

In emphasising the move towards naturalistic and ecological validity in science, it is crucial to acknowledge the intricate nature of visual processes influenced by factors like contrast, size and luminance (Pinto et al., 2008). While understanding the fundamental aspects of these processes is essential, the integration of dynamic stimuli, such as videos in fMRI studies (Lahner, 2022) or interactive experience in virtual reality and mobile EEG setups could offer significant benefits in measuring neural and behavioural responses to occlusion. Therefore, by introducing elements like moving objects and occlusion, researchers gain new insights into the complex dynamics of visual experience providing more of a bridge between controlled experiments and the intricacies of everyday visual encounters, which have up until now not been considered or measured in this way. This approach not only enhances our comprehension of neural and behavioural responses, but also aligns with the capabilities of comprehensive artificial neural networks and classifiers, where it is easier than it has ever been to test and refine theories.

It has become abundantly clear that the use of computational modelling can account well for findings regarding visual processes. The ability to train and test theories and models on huge datasets provide results that can unlock more knowledge about how the human visual system is working. Being able to model computational visual systems on the brain overcomes challenges like occlusion can help to overcome computational hurdles. It has been determined that recurrence within convolutional neural networks has improved recognition capabilities even when objects are occluded (Spoerer et al., 2017; Tang et al., 2018). This process has been achieved by combining expertise from neuroscience and computer science and adopting a predictive processing objective during training where networks learn to predict future occurrences efficiently (Ali et al., 2022; Lotter et al., 2017; Pang et al., 2021). However, there are still improvements to be made in this way.

A potential avenue from the work of this thesis involves examining how many cycles an RNN needs to recognise an occluded object, specifically in later layers which correspond to the IT region of the human visual system. The goal would be to correlate patterns of representations in this layer and the weights assigned in IT, as measured by fMRI. This approach could thus extend our previous findings linking recognition difficulty to higher beta weights in IT, with the overarching goal to connect occluded object recognition cycles in the model with the corresponding neural activity in IT, emphasising the mapping of these cycles to fMRI-measured weights. When observing our results of more ecologically valid occlusion, where objects act as both the occluded and occluding objects, it became clear that further research would benefit from using computational models to improve this knowledge. The incorporation of ecologically valid stimuli in our study is particularly advantageous. By using stimuli that closely mirror real-world scenarios, we can gain insights into how object representations are attended in natural environments. The ability of ANNs to use large datasets incredibly quickly for testing and training phases in computational methods allows hypotheses to be tested rapidly and with the opportunity to refine paradigms with much more ease than when testing participants. This resulting increase in knowledge and ecological validity could lead to refinements across predictive models, creating more efficient models utilising predictive effects (Lotter et al., 2017) based on the observed neural mechanisms implicated in challenging visual conditions.

It is worth considering that there are individual differences in how predictions from prior experiences are utilised. These may be a result of expertise, where an expert in a topic is less likely to have large errors in prediction than those without the expertise (Richler et al., 2019). There are also hypothesised differences in the precision of encoding in those with autism spectrum disorder (ASD) and psychosis (Takahashi et al., 2021; Utzerath et al., 2018). In relatively unambiguous situations, those with ASD can successfully learn and apply contingencies, yet when the predictive value is altered, for example in a new and volatile environment, the possibility for optimal processing becomes limited (Van de Cruys et al., 2014). It may be such that weaker top-down integration of prior and current information, results in stronger local, but weaker global processing (de Lange, Heilbron & Kok, 2018). Furthermore, those with psychosis may have difficulty distinguishing reality from delusion because of a decreased precision in encoding prior beliefs relative to sensory data (Sterzer et al., 2018; Kok & de Lange, 2015). It would be interesting to complete the studies from this thesis again, particularly the task from Chapter 4, with clinical groups compared to neurotypical populations. Providing a measure of how the results differed, as understanding how different populations navigate challenging visual conditions can have implications for tailored interventions as well as computational improvements.

Though the studies in this thesis were exploratory and novel in nature, they establish connections between our research objectives and broader literature concerning the potential implications and applications of object recognition and processing. In specific AI domains, such as object or person tracking, addressing occlusion challenges for ANNs holds immense value. This proficiency could yield widespread benefits, ranging from enhancing safety features in self-driving cars (Cheng et al., 2020; Wu et al., 2020), to refining sports event tracking technologies like VAR (de Oliveira et al., 2023). Moreover, it could contribute to bolstering security and surveillance systems, allowing for better adaptation to crowds and occlusions (Tang et al., 2014). A comprehensive understanding of how the visual system adeptly handles partial occlusion empowers us to apply this knowledge to computational methods and technologies, mitigating potential pitfalls. Employing more ecologically valid stimuli represents a crucial step towards elucidating these effects. Collaborations across disciplines in academia and beyond, particularly with industries engaged in augmented reality and object recognition technology, could yield tangible real-world implications for the handling and comprehension of object recognition.

5.5. General conclusion

The research presented in this thesis provides both a behavioural and neuroimaging perspective of object perception under conditions of occlusion. Following the use of novel stimuli pairs to compare occluded and deleted pairings where real-object images acted as occluded and occluding variables, we demonstrated robust occlusion effects. These differed across EVC and IT, showing more tolerance for multiple object representations in higher order visual areas compared to EVC. In these early visual areas, visible visual information was key in recognition while in IT the inferred features predicted the object identity well. When combining neuroimaging and behavioural study data we determined that when recognition is more difficult, IT assigns higher weights to allow for recognition, whereas in EVC this is not the case, perhaps explaining the lessened ability to tolerate this complexity brought by multiple objects presented simultaneously. When looking specifically at predictive processing via expectation suppression, we found some evidence from decoding results regarding an expectation suppression effect that aligned with the sharpening account of prediction, though univariate results of this novel stimulus set were less clear and demand further consideration and refinement.

This thesis has addressed the lack of realistic occlusion scenarios and whether predictive processing may be responsible for these effects. This was achieved by utilising object images of occluding and occluded objects, with a deleted condition to compare directly with how multiple object representations affect neural representations and recognition. We demonstrated differences in EVC and IT patterns that differed from previous research without meaningful occluders (Johnson & Olshausen, 2005; Spoerer et al., 2017). Additionally, to provide more insight into whether predictive processing is a suitable account for visual processing we tested these effects on recognition under occlusion in EVC. We found support for predictive processing effects in the sharpening account. Overall, due to the prevalence of occlusion in everyday life and our robust ability to process even heavily obscured objects, understanding these processes enables better ability to teach us about how we learn and adapt to our surroundings as well as to build better AI systems. It enforces the belief that more ecologically valid stimuli can reveal more to us about the dynamics of human visual processes.

References

Agam, Y., Liu, H., Papanastassiou, A., Buia, C., Golby, A. J., Madsen, J. R., & Kreiman, G. (2010). Robust selectivity to two-object images in human visual cortex. *Current Biology*, *20*(9), 872–879. https://doi.org/10.1016/j.cub.2010.03.050

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Aitken, F., Turner, G., & Kok, P. (2020). Prior expectations of motion direction modulate early sensory processing. *The Journal of Neuroscience*, *40*(33), 6389–6397. https://doi.org/10.1523/jneurosci.0537-20.2020

Ali, A., Ahmad, N., de Groot, E., Johannes van Gerven, M. A., & Kietzmann, T. C. (2022). Predictive coding is a consequence of energy efficiency in recurrent neural networks. *Patterns*, *3*(12), 100639. https://doi.org/10.1016/j.patter.2022.100639

Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage*, *231*. https://doi.org/10.1016/j.neuroimage.2021.117824

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience*, *30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*(4), 1407–1425. https://doi.org/10.3758/s13428-020-01501-5

Ao, J., Ke, Q., & Ehinger, K. A. (2023). Image amodal completion: A survey. *Computer Vision and Image Understanding*, *229*, 103661. https://doi.org/10.1016/j.cviu.2023.103661

Argall, B. D., Saad, Z. S., & Beauchamp, M. S. (2006). Simplified intersubject averaging on the cortical surface using SUMA. *Human Brain Mapping*, *27*(1), 14–27. https://doi.org/10.1002/hbm.20158

Avery, J. A., Lui, A. G., Ingeholm, J. E., Gotts, S. J., & Martin, A. (2021). Viewing images of foods evokes taste quality-specific activity in gustatory insular cortex. https://doi.org/10.1073/pnas.2010932118

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NeuroImage*, *70*, 37–47. https://doi.org/10.1016/j.neuroimage.2012.12.023

Bailey, K. M., Giordano, B. L., Kaas, A. L., & Smith, F. W. (2023). Decoding sounds depicting hand–object interactions in primary somatosensory cortex. *Cerebral Cortex*, *33*(7), 3621–3635. https://doi.org/10.1093/cercor/bhac296

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04126-7

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biology*, *14*(11). https://doi.org/10.1371/journal.pbio.1002577

Boutin, V., Franciosini, A., Chavane, F., Ruffier, F., & Perrinet, L. (2021). Sparse deep predictive coding captures contour integration capabilities of the early visual system. *PLoS Computational Biology*, *17*(1). https://doi.org/10.1371/JOURNAL.PCBI.1008629

Bracci, S., & Op de Beeck, H. P. (2023). Understanding human object vision: A picture is worth a thousand representations. *Annual Review of Psychology*, *74*(1), 113–135. https://doi.org/10.1146/annurev-psych-032720-041031

Breman, H., Mulders, J., Fritz, L., Peters, J., Pyles, J., Eck, J., Bastiani, M., Roebroeck, A., Ashburner, J., & Goebel, R. (2020). An image registration-based method for epi distortion correction based on opposite phase encoding (COPE) (pp. 122–130). https://doi.org/10.1007/978-3-030-50120-4_12

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., Majaj, N. J., & DiCarlo, J. J. (2014). Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLOS Computational Biology*, *10*(12), e1003963. https://doi.org/10.1371/journal.pcbi.1003963

Cao, R. (2020). New labels for old ideas: Predictive Processing and the interpretation of neural signals. *Review of Philosophy and Psychology*, *11*(3), 517–546. https://doi.org/10.1007/s13164-020-00481-x

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(10), 9. https://doi.org/10.1167/11.10.9

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 27:1-27:27. https://doi.org/10.1145/1961189.1961199

Charest, I., Kievit, R. A., Schmitz, T. W., Deca, D., Kriegeskorte, N., & Ungerleider, L. G. (2014). Unique semantic space in the brain of each beholder predicts perceived similarity. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(40), 14565–14570. https://doi.org/10.1073/pnas.1402594111

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. https://doi.org/10.1152/jn.1998.80.6.2918

Chen, G., Taylor, P. A., Reynolds, R. C., Leibenluft, E., Pine, D. S., Brotman, M. A., Pagliaccio, D., & Haller, S. P. (2023). BOLD response is more than just magnitude: Improving detection sensitivity through capturing hemodynamic profiles. *NeuroImage*, *277*, 120224. https://doi.org/10.1016/j.neuroimage.2023.120224

Cheng, Y., Yang, B., Wang, B., & Tan, R. T. (2020). 3D human pose estimation using spatio-temporal networks with explicit occlusion training. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(07), Article 07. https://doi.org/10.1609/aaai.v34i07.6689

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1), 27755. https://www.nature.com/articles/srep27755

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clarke, A. M., Herzog, M. H., & Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01193

Conway, B. R. (2018). The organization and operation of inferior temporal cortex. *Annual Review of Vision Science*, *4*(1), 381–402. https://doi.org/10.1146/annurev-vision-091517-034202

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), Article 3. https://doi.org/10.1038/nrn755

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194. https://www.sciencedirect.com/science/article/pii/S1053811998903950

de Haas, B., & Schwarzkopf, D. S. (2018). Spatially selective responses to Kanizsa and occlusion stimuli in human visual cortex. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-017-19121-z

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

de Oliveira, M. S., Steffen, V., & Trojan, F. (2023). A systematic review of the literature on video assistant referees in soccer: Challenges and opportunities in sports analytics. *Decision Analytics Journal*, *7*, 100232. https://doi.org/10.1016/j.dajour.2023.100232

Den Ouden, H. E., Kok, P., & De Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*, 548. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00548/full

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

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–341. https://doi.org/10.1016/j.tics.2007.06.010

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *Journal of Neuroscience*, *25*(36), 8259–8266. https://doi.org/10.1523/JNEUROSCI.1821-05.2005

Doostani, N., Hossein-Zadeh, G.-A., & Vaziri-Pashkam, M. (2023). The normalization model predicts responses in the human visual cortex during object-based attention. *eLife*, *12*, e75726. https://doi.org/10.7554/eLife.75726

Edwards, G., Vetter, P., McGruer, F., Petro, L. S., & Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-16093-y

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, *30*(49), 16601–16608. https://doi.org/10.1523/JNEUROSCI.2770-10.2010

Eickenberg, M., Gramfort, A., Varoquaux, G., & Thirion, B. (2017). Seeing it all: Convolutional network layers map the function of the human visual system. *NeuroImage*, *152*, 184–194. https://doi.org/10.1016/j.neuroimage.2016.10.001

Elias, P. (1955). Predictive coding–I. *IRE Transactions on Information Theory*, *1*(1), 16–24. https://doi.org/10.1109/TIT.1955.1055126

Emadi, N., & Esteky, H. (2013). Neural representation of ambiguous visual objects in the inferior temporal cortex | *PLOS ONE*. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0076856

Ernst, M. R., Triesch, J., & Burwick, T. (2019). Recurrent connections aid occluded object recognition by discounting occluders. *Artificial Neural Networks and Machine Learning – ICANN 2019: Image Processing* (pp. 294–305). Springer International Publishing. https://doi.org/10.1007/978-3-030-30508-6_24

Ernst, M. R., Triesch, J., & Burwick, T. (2021). Recurrent feedback improves recognition of partially occluded objects (arXiv:2104.10615). arXiv. http://arxiv.org/abs/2104.10615

Fallah, M., Stoner, G. R., & Reynolds, J. H. (2007). Stimulus-specific competitive selection in macaque extrastriate visual area V4. https://doi.org/10.1073/pnas.0611722104

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex (New York, N.Y.*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Feuerriegel, D., Vogels, R., & Kovács, G. (2021). Evaluating the evidence for expectation suppression in the visual system. *Neuroscience and Biobehavioral Reviews*, *126*, 368–381. https://doi.org/10.1016/j.neubiorev.2021.04.002

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207. https://doi.org/10.1006/nimg.1998.0396

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). 'Who' is saying 'what'? Brain-based decoding of human voice and speech. https://doi.org/10.1126/science.1164318

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Fritz, L., Mulders, J., Breman, H., Peters, J., Bastiani, M., Roebroeck, A., Andersson, J., Ashburner, J., Weiskopf, N., & Goebel, R. (2014). Comparison of EPI distortion correction methods at 3T and 7T. In *OHBM Annual Meeting, Hamburg, Germany*.

Fyall, A. M., El-Shamayleh, Y., Choi, H., Shea-Brown, E., & Pasupathy, A. (2017). Dynamic representation of partially occluded objects in primate prefrontal and visual cortex. *eLife*, *6*, e25784. https://doi.org/10.7554/eLife.25784

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2018). *Cognitive Neuroscience: Fifth International Student Edition*. https://books.google.co.uk/books?hl=en&lr=&id=1mSbDwAAQBAJ&oi=fnd&pg=PP3&dq=gazzaniga+ivry+mangun+2018&ots=JvlFToU74L&sig=Kj6vUxickouu-GNdLS48XcmRkqU#v=onepage&q=gazzaniga%20ivry%20mangun%202018&f=false

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews. Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, *27*(5), 392–401. https://doi.org/10.1002/hbm.20249

González-García, C., & He, B. J. (2021). A Gradient of sharpening effects by perceptual prior across the human cortical hierarchy. *Journal of Neuroscience*, *41*(1), 167–178. https://doi.org/10.1523/JNEUROSCI.2023-20.2020

Goodale, M., & Milner, A. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Greening, S. G., Mitchell, D. G. V., & Smith, F. W. (2018). Spatially generalizable representations of facial expressions: Decoding across partial face samples. *Cortex*, *101*, 31–43. https://doi.org/10.1016/j.cortex.2017.11.016

Grinter, E. J., Maybery, M. T., & Badcock, D. R. (2010). Vision in developmental disorders: Is there a dorsal stream deficit? *Brain Research Bulletin*, *82*(3–4), 147–160. https://doi.org/10.1016/j.brainresbull.2010.02.016

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160102. https://doi.org/10.1098/rstb.2016.0102

Guo, Y., Sohel, F., Bennamoun, M., Wan, J., & Lu, M. (2015). A novel local surface feature for 3D object recognition under clutter and occlusion. *Information Sciences*, *293*, 196–213. https://www.sciencedirect.com/science/article/pii/S0020025514009219

Gurney, K. (2003). *An Introduction to Neural Networks*. CRC Press.

Han, K., Wen, H., Zhang, Y., Fu, D., Culurciello, E., & Liu, Z. (2018). Deep Predictive coding network with local recurrent processing for object recognition. *Advances in Neural Information Processing Systems*, *31*. https://proceedings.neurips.cc/paper/2018/hash/1c63926ebcabda26b5cdb31b5cc91efb-Abstract.html

Haushofer, J., Livingstone, M. S., & Kanwisher, N. (2008). Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. *PLOS Biology*, *6*(7), e187. https://doi.org/10.1371/journal.pbio.0060187

Haynes, J. D. (2015). A primer on pattern-based approaches to fMRI: Principles, Pitfalls, and perspectives. *Neuron*, *87*(2), 257–270. https://doi.org/10.1016/j.neuron.2015.05.025

Helenius, P., Laasonen, M., Hokkanen, L., Paetau, R., & Niemivirta, M. (2011). Impaired engagement of the ventral attentional pathway in ADHD. *Neuropsychologia*, *49*(7), 1889–1896. https://doi.org/10.1016/j.neuropsychologia.2011.03.014

Horwitz, G. D., & Hass, C. A. (2012). Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing. *Nature Neuroscience*, *15*(6), Article 6. https://doi.org/10.1038/nn.3105

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *The Journal of Physiology*, *148*(3), 574–591. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1363130/

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science (New York, N.Y.)*, *310*(5749), 863–866. https://doi.org/10.1126/science.1117593

Janssen, P., Verhoef, B.-E., & Premereur, E. (2018). Functional interactions between the macaque dorsal and ventral visual pathways during three-dimensional object vision. *Cortex, 98*, 218–227. https://doi.org/10.1016/j.cortex.2017.01.021

Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, *34*(1), 65–73. https://doi.org/10.1002/mrm.1910340111

Jia, K., Goebel, R., & Kourtzi, Z. (2023). Ultra-high field imaging of human visual cognition. *Annual Review of Vision Science*, *9*(1), 479–500. https://doi.org/10.1146/annurev-vision-111022-123830

Jia, K., Zamboni, E., Kemper, V., Rua, C., Goncalves, N. R., Ng, A. K. T., Rodgers, C. T., Williams, G., Goebel, R., & Kourtzi, Z. (2020). Recurrent processing drives perceptual plasticity. *Current Biology*, *30*(21), 4177-4187.e4. https://doi.org/10.1016/j.cub.2020.08.016

Jia, X., Hong, H., & DiCarlo, J. J. (2021). Unsupervised changes in core object recognition behavior are predicted by neural plasticity in inferior temporal cortex. *eLife*, *10*, e60830. https://doi.org/10.7554/eLife.60830

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *Journal of Neuroscience*, *33*(47), 18438–18447. https://doi.org/10.1523/JNEUROSCI.3308-13.2013

John-Saaltink, E. S., Utzerath, C., Kok, P., Lau, H. C., & De Lange, F. P. (2015). Expectation suppression in early visual cortex depends on task set. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0131172

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), 3262–3276. https://doi.org/10.1016/j.visres.2005.06.007

Jozwik, K. M., Kietzmann, T. C., Cichy, R. M., Kriegeskorte, N., & Mur, M. (2023). Deep neural networks and visuo-semantic models explain complementary components of human ventral-stream representational dynamics. *The Journal of Neuroscience*, *43*(10), 1731–1741. https://doi.org/10.1523/JNEUROSCI.1424-22.2022

Kafaligonul, H., Breitmeyer, B. G., & Öğmen, H. (2015). Feedforward and feedback processes in vision. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00279

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*, 334–341. https://doi.org/10.1016/j.neuroimage.2017.12.065

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kanan, C., Tong, M., Zhang, L., & Cottrell, G. (2009, August). SUN: Top-down saliency using natural statistics. Routledge. https://doi.org/10.1080/13506280902771138

Kanizsa, G. (1976). Subjective contours. Scientific American. https://doi.org/10.1038/scientificamerican0476-48

Kar, K., & DiCarlo, J. J. (2021). Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *Neuron*, *109*, 164–176. https://doi.org/10.1016/j.neuron.2020.09.035

Kar, K., Kubilius, J., Schmidt, K., Issa, E. B., & DiCarlo, J. J. (2019). Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, *22*(6), Article 6. https://doi.org/10.1038/s41593-019-0392-5

Karapetian, A., Boyanova, A., Pandaram, M., Obermayer, K., Kietzmann, T. C., & Cichy, R. M. (2023). Empirically identifying and computationally modeling the brain–behavior relationship for human scene categorization. *Journal of Cognitive Neuroscience*, 1–19. https://doi.org/10.1162/jocn_a_02043

Karimi-Rouzbahani, H., Bagheri, N., & Ebrahimpour, R. (2017). Invariant object recognition is a personalized selection of invariant features in humans, not simply explained by hierarchical feed-forward vision models. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-13756-8

Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, *16*(3), 39. https://doi.org/10.1167/16.3.39

Kherad-Pajouh, S., & Renaud, O. (2015). A general permutation approach for analyzing repeated measures ANOVA and mixed-model designs. *Statistical Papers*, *56*(4), 947–967. https://doi.org/10.1007/s00362-014-0617-3

Khorsand, P., Moore, T., & Soltani, A. (2015). Combined contributions of feedforward and feedback inputs to bottom-up attention. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00155

Kietzmann, T. C., Spoerer, C. J., Sörensen, L. K. A., Cichy, R. M., Hauk, O., & Kriegeskorte, N. (2019). Recurrence is required to capture the representational dynamics of the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(43), 21854–21863. https://doi.org/10.1073/pnas.1905544116

Knights, E., Mansfield, C., Tonin, D., Saada, J., Smith, F. W., & Rossit, S. (2021). Hand-selective visual regions represent how to grasp 3d tools: brain decoding during real actions. *Journal of Neuroscience*, *41*(24), 5263–5273. https://doi.org/10.1523/JNEUROSCI.0083-21.2021

Knill, D. C., & Richards, W. (1996). *Perception as bayesian inference*. Cambridge University Press.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371–376. https://www.cell.com/current-biology/pdf/S0960-9822(15)01569-9.pdf

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of Neuroscience*, *33*(41), 16275–16284. https://www.jneurosci.org/content/33/41/16275.short

Kok, P., & De Lange, F. P. (2015). Predictive coding in sensory cortex. In *An Introduction to Model-Based Cognitive Neuroscience*. Springer New York. https://doi.org/10.1007/978-1-4939-2236-9_11

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex*, *22*(9), 2197–2206. https://academic.oup.com/cercor/article-abstract/22/9/2197/422638

Konkle, T., & Alvarez, G. A. (2022). A self-supervised domain-general learning framework for human ventral stream representation. *Nature Communications*, *13*(1), 491. https://www.nature.com/articles/s41467-022-28091-4

Kourtzi, Z., & Kanwisher, N. G. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506–1509.

Kreiman, G. (2008). Biological object recognition. *Scholarpedia*, *3*(6), 2667. https://doi.org/10.4249/scholarpedia.2667

Kriegeskorte, N. (2015). Deep Neural Networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*(1), 417–446. https://doi.org/10.1146/annurev-vision-082114-035447

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649–662. https://doi.org/10.1016/j.neuroimage.2007.02.022

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*. https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Kubilius, J., Bracci, S., & Beeck, H. P. O. de. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLOS Computational Biology*, *12*(4), e1004896. https://doi.org/10.1371/journal.pcbi.1004896

Lahner, B. (2022). An fMRI dataset of 1,102 natural videos for visual event understanding [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/144631

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Larsson, J., & Smith, A. T. (2012). FMRI repetition suppression: Neuronal adaptation or stimulus expectation? *Cerebral Cortex*, *22*(3), 567–576. https://doi.org/10.1093/cercor/bhr119

Lawrence, S. J. D., Formisano, E., Muckli, L., & de Lange, F. P. (2019). Laminar fMRI: Applications for cognitive neuroscience. *NeuroImage*, *197*, 785–791. https://doi.org/10.1016/j.neuroimage.2017.07.004

Layher, G., Schrodt, F., Butz, M. V., & Neumann, H. (2014). Adaptive learning in a compartmental model of visual cortex—How feedback enables stable category learning and refinement. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01287

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://www.nature.com/articles/nature14539

Lee, K. M., Ferreira-Santos, F., & Satpute, A. B. (2021). Predictive processing models and affective neuroscience. *Neuroscience and Biobehavioral Reviews*, *131*, 211–228. https://doi.org/10.1016/j.neubiorev.2021.09.009

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology-Paris*, *97*(2), 121–139. https://doi.org/10.1016/j.jphysparis.2003.09.015

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1907–1911. https://doi.org/10.1073/pnas.98.4.1907

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929. https://doi.org/10.1152/jn.1993.69.6.1918

Li, S., Zeng, X., Shao, Z., & Yu, Q. (2023). Neural representations in visual and parietal cortex differentiate between imagined, perceived, and illusory experiences. *Journal of Neuroscience*, *43*(38), 6508–6524. https://doi.org/10.1523/JNEUROSCI.0592-23.2023

Li, Y., Wang, S., Tian, Q., & Ding, X. (2015). A survey of recent advances in visual feature detection. *Neurocomputing*, *149*, 736–751. https://doi.org/10.1016/j.neucom.2014.08.003

Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. 3367–3375. https://openaccess.thecvf.com/content_cvpr_2015/html/Liang_Recurrent_Convolutional_Neural_2015_CVPR_paper.html

Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Frontiers in Computational Neuroscience*, *14*. https://www.frontiersin.org/articles/10.3389/fncom.2020.00029

Lotter, W., Kreiman, G., & Cox, D. (2017). Deep predictive coding networks for video prediction and unsupervised learning (arXiv:1605.08104). arXiv. https://doi.org/10.48550/arXiv.1605.08104

Lu, Z., Doerig, A., Bosch, V., Krahmer, B., Kaiser, D., Cichy, R. M., & Kietzmann, T. C. (2023). End-to-end topographic networks as models of cortical map formation and human visual behaviour: Moving beyond convolutions (arXiv:2308.09431). arXiv. https://doi.org/10.48550/arXiv.2308.09431

MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, *19*(11), 943–947. https://doi.org/10.1016/j.cub.2009.04.020

Mansfield, C., Kietzmann, T., van den Bosch, J., Charest, I., Mur, M., Kriegeskorte, N., & Smith, F. (2023). Neural representation of occluded objects in visual cortex. *Journal of Vision*, *23*(9), 4594–4594. https://doi.org/10.1167/jov.23.9.4594

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences*, *108*(24), 9739–9746. https://doi.org/10.1073/pnas.1105855108

Mehrer, J., Kietzmann, T. C., & Kriegeskorte, N. (2017). Deep neural networks trained on ecologically relevant categories better explain human IT. *Conference on Cognitive Computational Neuroscience. New York, NY, USA*. https://www2.securecms.com/CCNeuro/docs-0/5927d79368ed3feb338a2577.pdf

Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. https://doi.org/10.1523/JNEUROSCI.4570-10.2011

Mills, C., Zamani, A., White, R., & Christoff, K. (2021). Out of the blue: Understanding abrupt and wayward transitions in thought using probability and predictive processing: Probability in spontaneous thought. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817). https://doi.org/10.1098/rstb.2019.0692

Morgan, A. T., Petro, L. S., & Muckli, L. (2019). Scene representations conveyed by cortical feedback to early visual cortex can be described by line drawings. *The Journal of Neuroscience*, *39*(47), 9410–9423. https://doi.org/10.1523/JNEUROSCI.0852-19.2019

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. https://doi.org/10.1016/j.cub.2015.08.057

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00128

Nanay, B. (2018a). Multimodal mental imagery. *Cortex*, *105*, 125–134. https://doi.org/10.1016/j.cortex.2017.07.006

Nanay, B. (2018b). The importance of amodal completion in everyday perception. *I-Perception*, *9*(4), 2041669518788887. https://doi.org/10.1177/2041669518788887

Nayebi, A., Bear, D., Kubilius, J., Kar, K., Ganguli, S., Sussillo, D., Dicarlo, J. J., Yamins, D. L. K., & Program, N. P. (2018). Task-driven convolutional recurrent models of the visual system. https://github.com/neuroailab/tnn.

Ng, J., Bharath, A. A., & Zhaoping, L. (2006). A survey of architecture and function of the primary visual cortex (V1). *EURASIP Journal on Advances in Signal Processing*, *2007*(1), 097961. https://doi.org/10.1155/2007/97961

Ni, A. M., Ray, S., & Maunsell, J. H. R. (2012). Tuned normalization explains the size of attention modulations. *Neuron*, *73*(4), 803–813. https://doi.org/10.1016/j.neuron.2012.01.006

Nicholson, D. A., & Prinz, A. A. (2020). Deep neural network models of object recognition exhibit human-like limitations when performing visual search tasks [Preprint]. Neuroscience. https://doi.org/10.1101/2020.10.26.354258

O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. *Frontiers in Psychology*, *4*(APR). https://doi.org/10.3389/fpsyg.2013.00124

Pang, Z., O'May, C. B., Choksi, B., & VanRullen, R. (2021). Predictive coding feedback results in perceived illusory contours in a recurrent neural network. *Neural Networks*, *144*, 164–175. https://doi.org/10.1016/j.neunet.2021.08.024

Peelen, M., & Downing, P. (2022). Testing cognitive theories using multivariate pattern analysis of neuroimaging data. PsyArXiv. https://doi.org/10.31234/osf.io/rhzt9

Peirce, J. W. (2007). The potential importance of saturating and supersaturating contrast response functions in visual cortex. *Journal of Vision*, *7*(6), 13. https://doi.org/10.1167/7.6.13

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLOS Computational Biology*, *4*(1), e27. https://doi.org/10.1371/journal.pcbi.0040027

Pollicina, G., Dalton, P., & Vetter, P. (2022). Early visual cortex represents human sounds more distinctly than non-human sounds. *Journal of Vision*, *22*(14), 3705. https://doi.org/10.1167/jov.22.14.3705

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Proulx, M. J., & Egeth, H. E. (2008). Biased competition and visual search: The role of luminance and size contrast. *Psychological Research*, *72*(1), 106–113. https://doi.org/10.1007/s00426-006-0077-z

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on DSM-III-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. https://doi.org/10.1093/schbul/17.4.555

Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., & Khaligh-Razavi, S. M. (2019). Beyond core object recognition: Recurrent processes account for object recognition under occlusion. *PLoS Computational Biology*, *15*(5). https://doi.org/10.1371/journal.pcbi.1007001

Ramachandran, V. S. (2002). *Encyclopaedia of the Human Brain Set.* Academic Press-Elsevier Science USA.

Ramezanpour, H., & Fallah, M. (2022). The role of temporal cortex in the control of attention. *Current Research in Neurobiology*, *3*, 100038. https://doi.org/10.1016/j.crneur.2022.100038

Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T., & Todd, R. M. (2020). Affect-biased attention and predictive processing. *Cognition*, *203*. https://doi.org/10.1016/j.cognition.2020.104370

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

Rashal, E., & Wagemans, J. (2022). Depth from blur and grouping under inattention. *Attention, Perception, & Psychophysics*, *84*(3), 878–898. https://doi.org/10.3758/s13414-021-02402-1

Rauschenberger, R., Liu, T., Slotnick, S. D., & Yantis, S. (2006). Temporally unfolding neural representation of pictorial occlusion. *Psychological Science*, *17*(4), 358–364. https://doi.org/10.1111/j.1467-9280.2006.01711.x

Reddy, L., & Kanwisher, N. (2007). Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. *Journal of Vision*, *7*(9), 130. https://doi.org/10.1167/7.9.130

Reddy, L., Kanwisher, N. G., & Vanrullen, R. (2009). Attention and biased competition in multi-voxel object representations. In *PNAS December* (Vol. 15, pp. 21447–21452).

Richler, J. J., Tomarken, A. J., Sunday, M. A., Vickery, T. J., Ryan, K. F., Floyd, R. J., Sheinberg, D., Wong, A. C.-N., & Gauthier, I. (2019). Individual differences in object recognition. *Psychological Review*, *126*(2), 226–251. https://doi.org/10.1037/rev0000129

Richter, D., & de Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *eLife*, *8*, e47869. https://doi.org/10.7554/eLife.47869

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *Journal of Neuroscience*, *38*(34), 7452–7461. https://www.jneurosci.org/content/38/34/7452.short

Richter, D., Kietzmann, T., & de Lange, F. (2023). High-level prediction errors in low-level visual cortex. https://doi.org/10.1101/2023.08.21.554095

Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *Journal of Neuroscience*, *31*(40), 14076–14084. https://www.jneurosci.org/content/31/40/14076?utm_source=TrendMD&utm_medium=cpc&utm_campaign=JNeurosci_TrendMD_0

Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*(2), 713–726.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-X

Rossel, P., Peyrin, C., Roux-Sibilon, A., & Kauffmann, L. (2022). It makes sense, so I see it better! Contextual information about the visual environment increases its perceived sharpness. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(4), 331–350. https://doi.org/10.1037/xhp0000993

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Sakreida, K., Effnert, I., Thill, S., Menz, M. M., Jirak, D., Eickhoff, C. R., Ziemke, T., Eickhoff, S. B., Borghi, A. M., & Binkofski, F. (2016). Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience and Biobehavioral Reviews*, *69*, 89–112. https://doi.org/10.1016/j.neubiorev.2016.07.032

Sayres, R., & Grill-Spector, K. (2006). Object-selective cortex exhibits performance-independent repetition suppression. *Journal of Neurophysiology*, *95*(2), 995–1007. https://doi.org/10.1152/jn.00500.2005

Scherzer, T. R., & Ekroll, V. (2015). Partial modal completion under occlusion: What do modal and amodal percepts represent? *Journal of Vision*, *15*(1), 22. https://doi.org/10.1167/15.1.22

Schräder, J., Habel, U., Jo, H.-G., Walter, F., & Wagels, L. (2023). Identifying the duration of emotional stimulus presentation for conscious versus subconscious perception via hierarchical drift diffusion models. *Consciousness and Cognition*, *110*, 103493. https://doi.org/10.1016/j.concog.2023.103493

Schröger, E., Marzecová, A., & SanMiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, *41*(5), 641–664. https://doi.org/10.1111/ejn.12816

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*(1), 1–17. https://doi.org/10.1017/S0140525X98000107

Serre, T., Wolf, L., & Poggio, T. (2005). Object recognition with features inspired by visual cortex. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 994–1000 vol. 2. https://doi.org/10.1109/CVPR.2005.254

Smith, F., Petro, L., Muckli, L., & Adams, V. (2018). Decoding facial expressions across non-overlapping face features in early visual cortex. *Journal of Vision*, *18*(10), 913. https://doi.org/10.1167/18.10.913

Smith, F. W., & Goodale, M. A. (2015). Decoding visual object categories in early somatosensory cortex. *Cerebral Cortex.* https://academic.oup.com/cercor/article/25/4/1020/338635

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

Sobel, K. V., Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: The roles of top-down and bottom-up processing. *Psychonomic Bulletin & Review*, *14*(5), 840–845. https://doi.org/10.3758/bf03194109

Sorooshyari, S. K., Sheng, H., & Poor, H. V. (2020). Object recognition at higher regions of the ventral visual stream via dynamic inference. *Frontiers in Computational Neuroscience*, *14*, 46. https://doi.org/10.3389/fncom.2020.00046

Sovrano, V. A., & Bisazza, A. (2008). Recognition of partly occluded objects by fish. *Animal Cognition*, *11*(1), 161–166. https://doi.org/10.1007/s10071-007-0100-9

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLOS Computational Biology*, *16*(10), e1008215. https://doi.org/10.1371/journal.pcbi.1008215

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. *Frontiers in Psychology*, *8*(SEP). https://doi.org/10.3389/fpsyg.2017.01551

Spoerer, C., Kietzmann, T. C., & Kriegeskorte, N. (2019). Recurrent networks can recycle neural resources to flexibly trade speed for accuracy in visual recognition. Conference on Cognitive Computational Neuroscience, Berlin, Germany. https://doi.org/10.32470/CCN.2019.1068-0

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, *2*(2), 145–162. https://doi.org/10.1385/NI:2:2:145

Spratling, M. W. (2016). Predictive coding as a model of cognition. *Cognitive Processing*, *17*(3), 279–305. https://doi.org/10.1007/s10339-016-0765-6

Sterzer, P., Adams, R. A., Fletcher, P., Frith, C., Lawrie, S. M., Muckli, L., Petrovic, P., Uhlhaas, P., Voss, M., & Corlett, P. R. (2018). The predictive coding account of psychosis. *Biological Psychiatry*, *84*(9), 634–643. https://doi.org/10.1016/j.biopsych.2018.05.015

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), Article 11. https://doi.org/10.1038/nrn3838

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), Article 9. https://doi.org/10.1038/nn.2163

Takahashi, Y., Murata, S., Idei, H., Tomita, H., & Yamashita, Y. (2021). Neural network modeling of altered facial expression recognition in autism spectrum disorders based on predictive processing framework. *Scientific reports*, *11*(1), 14684.

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. https://doi.org/10.1016/j.neuron.2014.06.017

Tang, H., & Kreiman, G. (2017). Recognition of occluded objects. *Cognitive Science and Technology*, 41–58. Springer International Publishing. https://doi.org/10.1007/978-981-10-0213-7_3

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., Hardesty, W., Cox, D., & Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(35), 8835–8840. https://doi.org/10.1073/pnas.1719397115

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Ortega Caro, J., Hardesty, W., Cox, D., & Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, *115*(35), 8835–8840. https://doi.org/10.1073/pnas.1719397115

Tang, S., Andriluka, M., & Schiele, B. (2014). Detection and tracking of occluded people. *International Journal of Computer Vision*, *110*(1), 58–69. https://doi.org/10.1007/s11263-013-0664-6

Teichmann, L., Moerel, D., Rich, A. N., & Baker, C. I. (2022). The nature of neural object representations during dynamic occlusion. *Cortex*, *153*, 66–86. https://doi.org/10.1016/j.cortex.2022.04.009

Thiele, A., & Bellgrove, M. A. (2018). Neuromodulation of attention. *Neuron*, *97*(4), 769–785. https://www.cell.com/neuron/pdf/S0896-6273(18)30011-4.pdf

Thielen, J., Bosch, S. E., van Leeuwen, T. M., van Gerven, M. A. J., & van Lier, R. (2019). Neuroimaging findings on amodal completion: A review. *I-Perception*, *10*(2), 2041669519840047. https://doi.org/10.1177/2041669519840047

Thorat, S., Doerig, A., & Kietzmann, T. C. (2023). Characterising representation dynamics in recurrent neural networks for object recognition (arXiv:2308.12435). arXiv. http://arxiv.org/abs/2308.12435

Todorovic, A., & Lange, F. P. de. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Tootell, R. B. H., & Hadjikhani, N. (2001). Where is 'dorsal V4' in human visual cortex? Retinotopic, topographic and functional evidence. *Cerebral Cortex*, *11*(4), 298–311. https://doi.org/10.1093/cercor/11.4.298

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190–198. https://doi.org/10.1111/nyas.12680

Utzerath, C., Schmits, I. C., Buitelaar, J., & de Lange, F. P. (2018). Adolescents with autism show typical fMRI repetition suppression, but atypical surprise response. *Cortex, 109*, 25–34. https://doi.org/10.1016/j.cortex.2018.08.019

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

van Gerven, M., & Bohte, S. (2017). Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience*, *11*. https://www.frontiersin.org/articles/10.3389/fncom.2017.00114

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding Sound and imagery content in early visual cortex. *Current Biology*, *24*(11), 1256–1262. https://doi.org/10.1016/j.cub.2014.04.020

Võ, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205–210. https://doi.org/10.1016/j.copsyc.2019.03.009

Walsh, K. S., & McGovern, D. P. (2018). Expectation suppression dampens sensory representations of predicted stimuli. *Journal of Neuroscience*, *38*(50), 10592–10594. https://doi.org/10.1523/JNEUROSCI.2133-18.2018

Walsh, K. S., McGovern, D. P., Clark, A., & O’Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Wang, J., Xie, C., Zhang, Z., Zhu, J., Xie, L., & Yuille, A. (2017). Detecting semantic parts on partially occluded objects. http://arxiv.org/abs/1707.07819

Weigelt, S., Singer, W., & Muckli, L. (2007). Separate cortical stages in amodal completion revealed by functional magnetic resonance adaptation. *BMC Neuroscience*, *8*. https://doi.org/10.1186/1471-2202-8-70

Williams, D. (2018). Predictive processing and the representation wars. *Minds and Machines*, *28*(1), 141–172. https://doi.org/10.1007/s11023-017-9441-6

Wischnewski, M., & Peelen, M. V. (2021). Causal neural mechanisms of context-based object recognition. *eLife*, *10*, e69736. https://doi.org/10.7554/eLife.69736

Wu, F., Vong, C. M., & Liu, Q. (2020). Tracking objects with partial occlusion by background alignment. *Neurocomputing*, *402*, 1–13. https://doi.org/10.1016/j.neucom.2020.03.026

Wu, J., Zhou, C., Yang, M., Zhang, Q., Li, Y., & Yuan, J. (2020). Temporal-context enhanced detection of heavily occluded pedestrians. 13430–13439. https://openaccess.thecvf.com/content_CVPR_2020/html/Wu_Temporal-Context_Enhanced_Detection_of_Heavily_Occluded_Pedestrians_CVPR_2020

Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248-2261.

Xu, M., Meng, J., Yu, H., Jung, T. P., & Ming, D. (2020). Dynamic brain responses modulated by precise timing prediction in an opposing process. *Neuroscience Bulletin*. https://doi.org/10.1007/s12264-020-00527-1

Yan, C., Lange, F. P. de, & Richter, D. (2023). Conceptual associations generate sensory predictions. *Journal of Neuroscience*, *43*(20), 3733–3742. https://doi.org/10.1523/JNEUROSCI.1874-22.2023

Zhu, H., Tang, P., Park, J., Park, S., & Yuille, A. (2019). Robustness of object recognition under extreme occlusion in humans and computational models (arXiv:1905.04598). arXiv. https://doi.org/10.48550/arXiv.1905.04598

Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, *25*(36), 8150–8164. https://doi.org/10.1523/JNEUROSCI.2058-05.2005

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-Off between object selectivity and tolerance in monkey inferotemporal cortex. *Journal of Neuroscience*, *27*(45), 12292–12307. https://doi.org/10.1523/JNEUROSCI.1897-07.2007

Appendices

Chapter 2

Appendix A.

Significant pairings from wilcoxon signed ranks tests split across each brain region, adjusted for pairwise comparisons using FDR corrections.

Brain Region	Decoding Condition Pairing	W value
EVC	Single - Front	71*
	Single - Back	77**
	Single - Deleted	6**
	Front - Back	76**
	Front - Deleted	0**
	Back - Deleted	0**
MID	Single - Back	77*
	Front - Back	70*
	Back - Deleted	2**
IT	Single - Back	71*
	Front - Back	63*
	Back - Deleted	2**

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Appendix B.

Significant pairings from wilcoxon tests for the first cross decoding condition split across each brain region, FDR corrected.

Brain Region	Decoding Condition Pairing	W value
EVC	Single to Front – Single to Deleted	0***
	Single to Back – Single to Deleted	0***
IT	Single to Front – Single to Back	74*

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Appendix C.

Significant pairings from pairwise wilcoxon tests for the second cross decoding condition split across each brain region, FDR corrected.

Brain Region	Decoding Condition Pairing	W value
EVC	Single to Occluded – Deleted to Occluded	2**
IT	Single to Occluded – Deleted to Occluded	69*

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Appendix D.

Significant pairings from wilcoxon tests for the synthetic decoding condition split across each brain region, FDR corrected.

Brain Region	Decoding Condition Pairing	W value
EVC	Front Plus Deleted - Two Single	78***
MID	Front Plus Deleted - Two Single	74**

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Chapter 3

Appendix E.

Significant pairings from paired t tests for the accuracy condition, split across condition, adjusted using FDR corrections.

Condition	Speed pair	t value
Single	100 - 30	4.44***
Front	33 - 50	-3.91***
	100 - 33	4.81***
	100 - 50	3.16**
Back	33 - 50	-5.89****
	100 - 33	8.59****
	100 - 50	5.69****
Deleted	33 - 50	-5.63****
	100 - 33	7.05****
	100 - 50	4.71****

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Appendix F.

Significant pairings from paired t tests for the accuracy condition, split across speed, adjusted using FDR corrections.

Speed (ms)	Condition	t value
33	Front – Single	-2.44*
	Back – Single	-7.84****
	Deleted – Single	-5.36****
	Back – Front	-10.15****
	Deleted – Front	-4.22***
	Back – Deleted	-5.80****
50	Back – Single	-6.05****
	Deleted – Single	-4.40***
	Back – Front	-6.50****
	Deleted – Front	-3.90***
	Back - Deleted	-5.11****
100	Front – Single	-2.61*
	Deleted – Single	-4.30***
	Back – Single	-6.52****

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001.

Appendix G.

Significant pairings from paired t tests for the RT presentation speeds split across conditions, adjusted for pairwise comparisons using FDR corrections.

Condition	Speed pair	t value
Front	33 - 50	4.48***
	100 - 33	-6.64****
	100 - 50	-2.23*
Back	33 - 50	3.82***
	100 - 33	-8.22****
	100 - 50	-4.35***
Deleted	33 - 50	4.75****
	100 - 33	-8.43****
	100 - 50	-5.70****

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001

Appendix H.

Significant pairings from paired t tests for the RT conditions split across speeds, adjusted for pairwise comparisons using FDR corrections.

Speed	Condition Pair	t value
33	Back – Single	3.162**
	Back – Front	4.55***
	Back – Deleted	5.32****
50	Back – Single	2.64*
	Back – Deleted	5.38****
	Back – Front	4.25***
100	Deleted – Single	-2.74*
	Deleted – Front	-2.96*
	Back - Deleted	5.01***

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001

Chapter 4

Appendix I.

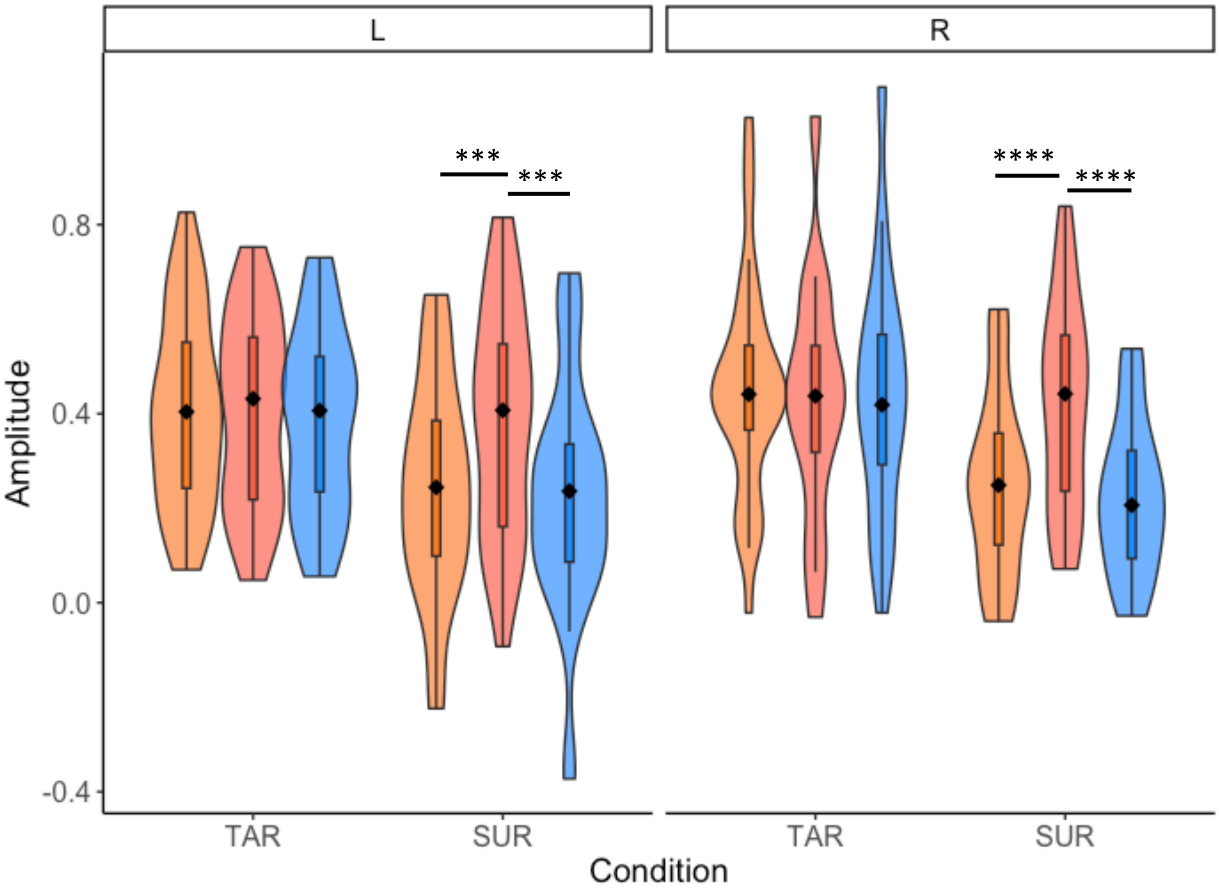
Significant pairings from Wilcoxon signed rank tests on decoding data.

Region of Interest	Decoding Condition Pairing	W value
Target	Match - Mismatch	3244*
Surround	Match – Mismatch	3920****
	Match – Neutral	3744****
	Neutral – Mismatch	3188*

Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001. FDR corrections applied to avoid pairwise error.

Appendix J.

Amplitudes of conditions in target and surround ROIs (orange is match, red shows neutral and blue represents mismatch) split across hemisphere.



Note. * *p*<.05, ** *p*<.01, *** p<.001, *****p*<.0001. FDR corrected.