Prioritising Feature Representations in Visual Working Memory

Jasper Hajonides van der Meulen

New College
University of Oxford

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

Approximate word count: 45.000

Trinity Term, 2021

Abstract

Representations of our external environment and of our internal memories together guide our awareness and actions. While large numbers of features and objects from the boundless external environment compete for sensory processing, working-memory capacity is highly limited. Though extracting and integrating features into relevant objects are major objectives of perception, feature integration and well separated object representations may be basic characteristics of working memory. As discussed in Chapter 1, attentional selection of both external and internal content is pivotal, but its workings may differ according to the distinct representational formats. This thesis investigates how attention prioritises feature information in working memory to guide future actions. To test my hypotheses, I design and conduct behavioural studies, analyse neurophysiological recordings, and apply computational modelling. My empirical findings describe two main ways through which feature representations in working memory can be modulated: through goal-related prioritisation and through interacting with lingering sensory history. Chapters 2 and 4 describe tasks in which internally maintained information can be prioritised based on task demands. They demonstrate prioritisation of items and feature dimensions during working-memory delays, as well as the ability to protect and update features during maintenance. Chapter 4 also outlines how stimulus history modulates sensory processing and interacts with goal-based influences. Sensory content from same-trial or previous-trial information exerts opposing systematic performance biases. Applying neural decoding methods that I develop in Chapter 3, I reveal in Chapter 4 that both same-trial and previous-trial sensory history generates a repulsive neural bias at encoding, modulated by task relevance of previous stimuli. The findings suggest that other biases at post-perceptual stages may also contribute to the final opposing biases in performance. Taken together, the studies in the thesis extend our understanding of how visual feature information in working memory can be modulated and prioritised to serve future behaviour adaptively.

Prioritising Feature Representations in Visual Working Memory





Jasper Hajonides van der Meulen

New College

University of Oxford

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

Trinity Term, 2021

This thesis is dedicated to the memory of
Dr Theo Hajonides van der Meulen (1928 - 2021)

Acknowledgements

Over my time in Oxford, I have had the privilege to spend time with such a high calibre of friends and colleagues. While it's difficult to find the words to fully express my thanks for those who made an impact on me over the past four years, I will make a worthy attempt to do so.

First and foremost, I would like to express my deepest gratitude to my supervisory team. **Kia**, I am ever grateful for your support and guidance over the years. You have been such an inspiration when it comes to scientific rigour and leadership – and stayed cool through it all. Thank you for providing me with so many opportunities and for keeping me on my toes. Your mentorship will continue to have an immeasurable impact on my future trajectory. Thanks, **Freek**, for keeping your foot on the gas pedal throughout the DPhil and for always being up for stimulating conversations, feedback, puns, or to bounce research ideas. I count myself lucky to have had you on board; my DPhil trajectory would not have been the same if it wasn't for your input. Keep nailing it. I would also like to thank **Mark** for his guidance, kindness insights, and for cultivating such a welcoming lab environment. I am grateful for you taking me on.

Additionally, I would like to thank Nick Myers. I am thankful for having been able to work with you over the past year. I have a deep appreciation for your rigour, kindness, and patience (especially when it came to CW and CCW labels).

I also owe my thanks to my fellow DPhil students Sage, Sammi, and Giedrė for being there with me through it all. And other friends and colleagues from the lab Dejan, Lauren, Andrew, Oana, Nora, Sarah, Dani, Jemma, Anna-Katharina, Marcus, Oliver, Irene, Priyanka, Frida, Michael, Michał, Emilia, and Sam, have had such a positive impact on my research and time in Oxford.

This doctoral degree would not have been possible without the generous funding from a Scatcherd European scholarship and a Grand Union DTP studentship award, with significant contributions from New College and the Medical Sciences Division at the University of Oxford. I am also grateful for all the departmental, college, and research staff that have played an important background role in supporting my work and time in Oxford.

On a different note, I would like to thank all the local triathlon, cycling, running, rowing, swimming athletes and coaches that have together played such a stimulating role over my time in Oxford. In particular, I want to thank Laura J. for spending 400+ hours swimming together in the Iffley pool over the past 2.5 years and for keeping me sharp. Without doubt, I also owe a word of thanks to my housemates Matt, Flavia, Flo, and Giorgia for making the lockdown and write-up much more enjoyable. A further thanks to my friends in the Netherlands for tolerating my erratic remote commentary, the entertaining conversations, and for your friendship throughout the years. Also, thank you Lisa, for your support and for shaping my time in Oxford and beyond – and Susan for your kindness and generosity throughout lockdown.

I would also like to express my gratitude to my grandparents for encouraging me to start this journey, and in particular my grandfather for being deeply interested and involved every step of the way. I also owe thanks to my parents and my brothers for supporting me over the past years.

Abstract

Representations of our external environment and of our internal memories together guide our awareness and actions. While large numbers of features and objects from the boundless external environment compete for sensory processing, working-memory capacity is highly limited. Though extracting and integrating features into relevant objects are major objectives of perception, feature integration and well separated object representations may be basic characteristics of working memory. As discussed in Chapter 1, attentional selection of both external and internal content is pivotal, but its workings may differ according to the distinct representational formats. This thesis investigates how attention prioritises feature information in working memory to guide future actions. To test my hypotheses, I design and conduct behavioural studies, analyse neurophysiological recordings, and apply computational modelling. My empirical findings describe two main ways through which feature representations in working memory can be modulated: through goal-related prioritisation and through interacting with lingering sensory history. Chapters 2 and 4 describe tasks in which internally maintained information can be prioritised based on task demands. They demonstrate prioritisation of items and feature dimensions during working-memory delays, as well as the ability to protect and update features during maintenance. Chapter 4 also outlines how stimulus history modulates sensory processing and interacts with goal-based influences. Sensory content from same-trial or previous-trial information exerts opposing systematic performance biases. Applying neural decoding methods that I develop in Chapter 3, I reveal in Chapter 4 that both same-trial and previous-trial sensory history generates a repulsive neural bias at encoding, modulated by task relevance of previous stimuli. The findings suggest that other biases at post-perceptual stages may also contribute to the final opposing biases in performance. Taken together, the studies in the thesis extend our understanding of how visual feature information in working memory can be modulated and prioritised to serve future behaviour adaptively.

General Abstract

When going about our daily tasks, our behaviour is based both on information out in the environment as well as in our thoughts and memories. To guide our actions, we shift attention dynamically between objects in our external environment and information maintained internally, in our working memory. The scientific literature suggests that representations in working memory are not simply a mirror image of the sensory environment. For example, visual features (e.g., shape and colour) are initially processed independently when viewing the item but are thought to be integrated into object representations (e.g., a piece of fruit) in working memory. This prompts the broad question regarding whether and how we can focus attention on features within working memory? To test how we shift attention towards features in working memory, I collected behavioural responses and measured brain activity. To address my experimental hypotheses, I applied statistical analyses and advanced computational modelling.

The findings in the study emphasise the flexibility of attention in working memory, showing that attention can operate at the level of features in multiple ways. My results show we are capable of directing our attention voluntarily not only to a bounded object (with all its features) in memory but also towards feature dimensions (e.g., all colours) distributed across objects when this is advantageous for task performance. This specific weighting of information in working memory in accordance with task goals is an efficient way of storing only the relevant information in our memory. In addition, some processes influence the content of working memory in an automatic fashion. Patterns of behavioural responses and neural markers show that feature information from previous stimuli systematically bias the sensory processing of features in the current stimuli. Interestingly, therefore, our perception and working memory are not simply based on what we see but also on previously processed information. The findings further show that biases may occur at different points along stimulus processing and depend on task goals.

The research in the thesis greatly enriches our understanding of the flexibility of working memory. Counter to our common-sense intuition, the brain does not build a faithful representation of the external world but instead picks up and stores useful information in its current context. Voluntary attention based on *task relevance* and automatic effects of stimulus *history* both shape representations of feature information in working memory to guide our behaviour adaptively.

Statement of Contribution

The work presented within this thesis has greatly benefitted from the guidance of colleagues and my DPhil supervisors. Chapter 2 of this thesis was designed in collaboration with Mark Stokes and Kia Nobre, I collected the data and analysed it, with guidance of Freek van Ede. The manuscript was written up under supervision of my DPhil mentors and can be found under https://doi.org/10.1167/jov.20.8.25. Early work on Chapter 3 started with code from Michael Wolff and discussions with Mark Stokes before I analysed and wrote up the manuscript. Freek van Ede and Kia Nobre have been instrumental in their feedback on the writing and review rounds and helped me to publish my work in NeuroImage (https://doi.org/10.1016/j.neuroimage.2021.118030). Finally, Chapter 4 builds on MEG data that has been collected in 2015 by Nick Myers, under supervision of Mark Stokes. I worked closely with Nick Myers throughout the preprocessing, analysis, and write-up phase. Kia Nobre and Freek van Ede also made important contributions to this work throughout the analysis and write-up phase. The work will be submitted to a peer-reviewed journal in the near future.

Contents

List of Figures **xix**

List of Tables **xxi**

List of Abbreviations **xxiii**

1 Attentional Prioritisation of Visual Representations in Working Memory **1**

1.1 Definitions 4

1.1.1 Attention 5

1.1.2 Working Memory 6

1.2 Frameworks of Attention 9

1.2.1 Sources of Attention 9

1.2.2 Types of Attention 11

1.3 Attention in Working Memory 16

1.3.1 Investigating Internal Orienting of Attention 16

1.3.2 Action-Oriented Representations 18

1.4 Comparing Internal and External Forms of Attention 21

1.4.1 Capacity of Internal Representations 21

1.4.2 Representational Differences 23

1.5 Sensory History for Adaptive Internal Representations 25

1.6 Aims and Overview of the Thesis 28

2 Comparing the Prioritisation of Items and Feature-dimensions in Visual Working Memory **31**

2.1 Abstract 32

2.2 Introduction 33

2.3 Methods . . . 37
2.3.1 Participants . . . 37
2.3.2 Experimental Set-up and Stimuli . . . 37
2.3.3 Task and Design . . . 38
2.3.4 Behavioural Analysis . . . 40
2.3.5 EEG Acquisision . . . 42
2.3.6 EEG Preprocessing . . . 43
2.3.7 Time-frequency Processing . . . 44
2.4 Results . . . 46
2.4.1 Behavioural Cueing Benefits . . . 46
2.4.2 Distractor Processing . . . 49
2.4.3 Mixture Modelling . . . 49
2.4.4 Alpha Attenuation Following Feature and Item Retro-cues . 50
2.5 Discussion . . . 54

3 Decoding Visual Colour From Scalp Electroencephalography Measurements 59

3.1 Abstract . . . 60
3.2 Introduction . . . 61
3.3 Methods . . . 65
3.3.1 Participants . . . 65
3.3.2 Visual Display . . . 65
3.3.3 Stimuli and Task Design . . . 66
3.3.4 EEG recording . . . 68
3.3.5 EEG preprocessing . . . 69
3.3.6 Classification on EEG data . . . 71
3.3.7 Statistical Analysis . . . 74
3.4 Results . . . 76
3.4.1 Colour Can Be Decoded From Scalp EEG, and Is Comparable to Orientation Decoding. . . . 76
3.4.2 Colour Decoding Is Driven by Visual Signals With a Contralateral-Posterior Topography. . . . 78

3.4.3 Colour Decoding Is Parametrically Organised 78
3.4.4 Decoding Is Sensitive to Luminance (L*) Differences but Primarily Driven by Difference in Colour (a*b*) 80
3.4.5 Dependence of Decoding on Trial and Participant Numbers . 82
3.5 Discussion 85

4 Behavioural and Neural Biases in Visual Working Memory 91
4.1 Abstract 93
4.2 Introduction 95
4.3 Methods 100
4.3.1 Participants 100
4.3.2 Experimental set-up 100
4.3.3 Task 100
4.3.4 Behavioural Analysis 102
4.3.5 Mixture Modelling 103
4.3.6 Performance Bias Calculation 104
4.3.7 MEG Acquisition 105
4.3.8 MEG Data Preprocessing 105
4.3.9 LDA Classification 106
4.3.10 Bias Computation 107
4.3.11 Statistical Testing 108
4.4 Results 110
4.4.1 Error Rates 110
4.4.2 Sources of Performance Errors Within Trials 111
4.4.3 Performance Biases Within Trials 112
4.4.4 Attractive Performance Bias Between Trials 113
4.4.5 Task Dependence of Attractive Performance Bias Between Trials 115
4.4.6 Neural Classification of Presented Orientations 117
4.4.7 Within-trial Classification Bias Away From Previous Stimulus 118
4.4.8 Repulsive Neural Biases Between Trials Away From Sensory History 118

4.4.9 Cross-Decoding Evidence for Previously Presented Stimuli . 121
4.5 Discussion 125

5 General Discussion 133
5.1 Introduction 134
5.1.1 Adapting Working-Memory Representations for Future Use . 136
5.1.2 Use It or Lose It 137
5.1.3 Top-Down Prioritisation of Relevant Features 138
5.1.4 Bottom-Up Modulation of Feature Representations 140
5.2 Future Directions 142
5.2.1 Bottom-Up Modulation of Feature Representations 142
5.2.2 Naturalistic and Dynamic Environments 144
5.3 Final Concusions 146

Appendices

A Supplementary Results for Chapter 2 149
A.1 Mixture Modelling for Orientation and Colour Dimensions 149
A.2 Modulation of Alpha Following a Retro-cue 151
A.3 Event-related Potentials Following a Retro-cue 152
A.4 Distractor Processing 153

B Supplementary Results for Chapter 3 155
B.1 Decoding Distractor Features 155
B.2 Classification Using Single Time-point Data 157
B.2.1 Decoding Time Course 157
B.2.2 Localisation of Feature Decoding 157
B.2.3 Individual Feature Decoding 158
B.2.4 Trial and Subject Numbers for Decoding 159
B.2.5 Feature Decoding Using Alpha Power 159

C Supplementary Results for Chapter 4 163
C.1 Decoding Distractor Features 163

D Disruption Statement Due to the COVID-19 Pandemic **165**

D.1 Visuo-Motor Selection: Simultaneous EEG-fMRI 166

D.1.1 State of Progress 169

Works Cited **171**

List of Figures

2.1 Experimental design with timings. 41
2.2 Performance benefits of item and feature-dimension retro cueing. . . 48
2.3 Cueing effects on mixture modelling parameters. 51
2.4 Induced neural EEG responses to the left and right item retro-cues and to colour and orientation dimension retro-cues. 53

3.1 Stimulus characteristics and presentation. 69
3.2 ERPs and decoding performance for colours and orientations. . . . 77
3.3 Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. 79
3.4 Decoding of individual features between 150 and 350 ms. 80
3.5 Decoding measured luminance and a*b* differences between stimuli. 83
3.6 Feature decoding as a function of trial and participant numbers. . . 84

4.1 Experimental task design. 103
4.2 Within-trial repulsive performance biases. 113
4.3 Attractive behavioural biases towards the relevant target orientation on the previous trial. 115
4.4 Orientation decoding of presented grating. 117
4.5 Shift in tuning curve relative to previous grating orientation. . . . 119
4.6 Bias on encoding imposed by previous trial. 122

A.1 Mixture modelling parameters for colour and orientation. 150
A.2 Mixture modelling parameters for colour and orientation. 151
A.3 Event-related potentials following informative and neutral retro-cues. 152

B.1 Decoding the colour and orientation of the visual distractors. 156

B.2 Decoding performance for colours and orientations with single time-point decoding. 157
B.3 Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. 158
B.4 Decoding of individual features between 150 and 350 ms. 159
B.5 Feature decoding as a function of trial and participant numbers. . . 160
B.6 LDA classification of colour and orientation using Alpha-band power in the stimulus encoding and distractor presentation epoch. 161

C.1 Cross-decoding Evidence for Orientation 1 Upon Presentation of Orientation 2. 164

D.1 Task design and data of interest. 167
D.2 Analysis relating variance in neural oscillations to BOLD response magnitude. 169

List of Tables

4.1 **Descriptive statistics of error for all experimental conditions.** 110

A.1 **Within-Subjects Effects for *Orientation Recall: Error*.** . . . 153
A.2 **Within-Subjects Effects for *Colour Recall: Error*.** 153
A.3 **Within-Subjects Effects for *Orientation Recall: RT*.** 154
A.4 **Within-Subjects Effects for *Colour Recall: RT*.** 154

List of Abbreviations

BOLD . . . . . blood oxygen level dependent

EEG . . . . . Electroencephalography

MEG . . . . . Magnetoencephalography

MRI . . . . . Magnetic Resonance Imaging

The true art of memory is the art of attention.

—Samuel Johnson (1709 - 1784)

1

Attentional Prioritisation of Visual Representations in Working Memory

Contents

1.1 Definitions 4
1.1.1 Attention 5
1.1.2 Working Memory 6
1.2 Frameworks of Attention 9
1.2.1 Sources of Attention 9
1.2.2 Types of Attention 11
1.3 Attention in Working Memory 16
1.3.1 Investigating Internal Orienting of Attention 16
1.3.2 Action-Oriented Representations 18
1.4 Comparing Internal and External Forms of Attention 21
1.4.1 Capacity of Internal Representations 21
1.4.2 Representational Differences 23
1.5 Sensory History for Adaptive Internal Representations 25
1.6 Aims and Overview of the Thesis 28

The internal representation of our environment is composed of both current visual information as well as previous experiences and serves to guide our future actions. Imagine you are buying groceries at the local supermarket. You set foot in the door and have memorised a small set of items you would like to purchase. As

you walk through the different sections of the store, items from your shopping list will sequentially gain priority as you approach the shelf containing the item. You search through a large set of similar-looking shopping products while maintaining the item of your shopping list in memory. The item that is internally represented in memory can have a distinctive colour from all other items on the shelf, which facilitates the search, and you can identify and pick out the product before moving on and prioritising the next item on your list. Altogether, attention links memory and perception to conceive a pragmatic internal representation of our environment.

It is of fundamental importance to understand how memory and attention interact to generate our representation of the world and guide our actions. This thesis describes novel experimental work that extends what we currently know about the internal and external representations of our environment. More specifically, I seek to investigate processing of feature-specific information of visual items and explore behavioural and neural mechanisms underlying feature representation as a function of task demand and context. This research builds on various methodologies and theories advanced by pioneers in neuroscience, psychology, and related fields. The experimental work relies on a range of experimental methods that are each sensitive to a different set of processes: 1) sophisticated analysis of behavioural responses in carefully designed laboratory tasks of attention and working memory, 2) characterisation of voltage differences to different experimental conditions measured with EEG, 3) analysis of time-resolved oscillatory neural activity, and 4) identification of distributed multivariate patterns of MEG/EEG sensors using supervised and unsupervised learning.

The present chapter will review key topics that lay the foundation for the subsequent empirical chapters. I start with reviewing the fundamental topics of attention and working memory. Subsequently, I discuss how attention interacts with external representations, what information is represented in working memory, how attention can be oriented towards these internal representations, relevant differences between internal and external representational formats, and how reconfigurations of working memory are an optimal way of representing information for future actions. I will conclude this chapter by summarising concrete aims of the thesis.

1.1 Definitions

The notion that mental representations are a reflection of our environment constructed through attentional filtering dates back centuries (James, 1890; Malebranche & Olscamp, 1997; von Helmholtz, 1867). Over the course of the past two centuries, discoveries in life and mind sciences have changed our perspective dramatically on human physiology and cognition. Leaps in the fields of electrophysiological recordings, animal models, microelectrodes, MEG measuring techniques, PET scans, functional and structural MRI sequences, have removed barriers to study structure and function of the human brain (Berlucchi, 2009; Nobre & van Ede, 2020; Savoy, 2001). These advances have proven instrumental in developing our understanding of cognitive processes like working memory and attention, which can otherwise only be measured at the output stage, through behavioural responses. Combining methods, analytical tools, and theories from neuroscience and cognitive psychology has redefined our understanding of cognition. In this thesis I will combine state-of-the-art methodologies in neuroscience and psychology to further our understanding working memory and attention.

It should be noted that the terms "attention" and "working memory" encompass sets of complex cognitive processes that are ill defined at best (Whissell, Abramson, & Barber, 2013). These are two central concepts in psychology and neuroscience, which have both been studied largely separately but are intertwined and share a range of functions (e.g., comparing orienting of attention and working memory encoding; sustained attention and working memory maintenance). Most commonly,

we find that attention is a construct rooted in modulating information in accordance with top-down and bottom-up signals to guide adaptive behaviour (Nobre & Kastner, 2014; Nobre & Stokes, 2020, p. 1204). In contrast, definitions of working memory often emphasise its representational nature, through which stimulus information is sustained over a maintenance period (Baddeley, 2003). It is useful to define these constructs explicitly, as they shape the scope of the current thesis.

1.1.1 Attention

We are presented with an abundance of sensory information from all our senses. Yet, we only process a fraction of this information at any moment in time. Amongst other factors, this feat could arise because of the limitations of cellular metabolism in the brain (Attwell & Laughlin, 2001; Lennie, 2003). This poses limitations on bioenergetically expensive action potentials and postsynaptic potentials that are associated with perceptual processing (Attwell & Laughlin, 2001; Bruckmaier, Tachtsidis, Phan, & Lavie, 2020). Yet, selecting the most important information from our environment is not merely a limitation. Through selective sampling, observers proactively select, prioritise, and process the sensory information that is deemed most relevant (Nobre & Van Ede, 2018).

When introducing the concept of attention, it is difficult not to refer back to William James (1890, p. 404) who encompassed the concept of attention as *"... taking possession of the mind, in clear and vivid form, of one out of what seems several simultaneously possible objects or trains of thought. Focalisation,*

concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others. . . " This description notes the critical ability to select internal or external information voluntarily. It also refers to the selective nature of attention and the need to prioritise one over competing information steams. In present day literature, the term attention is commonly used to refer to an ensemble of functions that prioritise relevant information according to goals and expectations that guide task performance and adaptive behaviour. These functions include the selection of information among a set of competing stimuli through unique visual features, as well as anticipatory processing guided by top-down signals that prepare the sensory system for information to come. This thesis adheres to the contemporary definition of attention, which defines attention as *"the selective prioritisation of neural representations that are most relevant to one's current behavioural goals"* (Buschman and Kastner, 2015, Nobre and Kastner, 2014, p. 1204). Attentional prioritisation acts on both sensory and working-memory representations (Nobre & Stokes, 2020). In this thesis, I will focus specifically on the selection and prioritisation of visual information in working memory.

1.1.2 Working Memory

Working memory is considered to be a core component of higher cognitive functions (Awh, Jonides, & Reuter-Lorenz, 1998; Baddeley, 1992; Engle, Tuholski, Laughlin, & Conway, 1999; Postle, 2006). It shares significant overlap with other cognitive functions and is proposed to be functionally involved in item maintenance (Rottschy

et al., 2012), planning (Miller, Galanter, & Pribram, 1960), manipulation (Cowan, 1988), attentional control (e.g., Engle, 2002), and guiding behaviour (Baddeley, 2003). Different measurements, tasks, and analyses emphasising different aspects of working memory have given rise to many definitions of this core cognitive domain (Cowan, 2017). While there are commonalities between each of the proposed definitions, the differences can generate discrepancies in the literature. In this thesis, I define working memory as *"the informational content maintained or manipulated to guide adaptive behaviour."*

Mounting evidence suggests working memory does not arise as a function of a dedicated system but instead as an emergent property drawing on processes such as attention, short-term memory or maintenance, and action-related functions (Postle, 2006). Jointly, these functions support the execution of goal-directed actions. The action-oriented nature of working memory (sometimes referred to as "premembering" Heuer, Ohl, & Rolfs, 2020; Myers, Stokes, & Nobre, 2017; Nobre & Stokes, 2019; van Ede, 2020) implies that representations emphasise task-relevant content. Rather than sustaining visual content passively until task completion, the emerging view is that working memory configures the available information for optimal perception-to-action mapping over extended delays - thereby increasing the flexibility of our behaviour. The nature and content of the task therefore has important implications for the representation of content in working memory.

It should be noted that interactions between attention and working memory are prevalent. Working memory enables attention through representing goal-relevant information, thereby guiding attentional control (Desimone & Duncan, 1995; Olivers,

Meijer, & Theeuwes, 2006; Soto, Hodsoll, Rotshtein, & Humphreys, 2008). In turn, attention also influences working memory as prioritisation and selection of sensory information is critical for determining what features are encoded into working memory (Gazzaley, 2011; Murray, Nobre, Clark, Cravo, & Stokes, 2013; Myers, Stokes, Walther, & Nobre, 2014). Furthermore, selective attention continues to operate upon information maintained in working memory to prioritise and select internalised information flexibly as goals and expectations change (Chun, Golomb, & Turk-Browne, 2010; Gazzaley & Nobre, 2012; van Ede, 2020). This thesis will predominantly study the account of flexible prioritisation of information after it has been encoded into working memory.

1.2 Frameworks of Attention

1.2.1 Sources of Attention

Studies of attention and working memory build on research investigating at many different levels of analysis. While empirical work in this thesis focusses on the study of macroscopic physiological recordings, studies across spatial and temporal scales inform each other bi-directionally.

Foundational work in the field of selective attention investigated activity from single neurons using single-unit recordings. Neurons across the visual system were observed to increase their firing rate when a preferred stimulus is presented within their receptive field (Hubel & Wiesel, 1959, 1962). During sensory processing, competition between stimuli arises when multiple stimuli are presented within a receptive field (Moran & Desimone, 1985; Reynolds, Chelazzi, & Desimone, 1999). Classically, studies of attention show that directing attention to a single stimulus resolves this conflict and blocks the response to the unattended stimulus (Chelazzi., Miller., Duncan., & Desimone., 1993; Luck & Vogel, 1997; Moran & Desimone, 1985). This finding supports the *biased competition* model of attention where attention biases neural responses through top-down signals in favour of the most relevant stimulus (Desimone & Duncan, 1995). Together, according to the standard model of biased competition, the different sources of bias prioritise most relevant information to provide an advantage during the natural, competitive process of perceptual analysis.

In addition to considering selective attentional prioritisation for specific locations

from the perspective of receptive fields, attention is also known to spread across objects (Duncan, 1984; Scolari, Edward, & Serences, 2014; Treisman & Gelade, 1980). As attention preferentially processes one item over its competitors, the features used to select the target item during visual search facilitate processing of other features belonging to the same item to drive feature integration (Desimone & Duncan, 1995; Duncan, 1984). According to the influential/foundational feature integration theory (Treisman & Schmidt, 1982), attention is necessary for binding together features into a single object representation. Without attention allocated to the group of features that comprise an object, the group of features will be unbound and floating freely and prone to swap errors between locations.

In addition to spatial and object-based attention, it is also possible to prioritise feature values (Egeth, Virzi, & Garbart, 1984; Liu & Hou, 2011; White & Carrasco, 2011). Feature-based attention has been found to act across locations and objects. Throughout the visual system, neurons are at least partially selective for stimulus features (e.g. orientation, shape, colour). These stimulus properties give rise to characteristic tuning curves for feature spaces, where firing rate is increased when the preferred feature is presented within the receptive field (Hubel & Wiesel, 1959). There is an ongoing debate on how attention modulates neuronal tuning curves. Researchers have argued attention acts through contrast gain (selective increase of contrast; Martínez-Trujillo & Treue, 2002; Reynolds & Pasternak, 2000; Treue, 2004), multiplicative response gain across all features (McAdams & Maunsell, 1999), or sharpening of tuning curves (Spitzer, Desimone, & Morant, 1988). It is likely that observed differences in response tuning as a function of attention can

be reconciled by taking task context and sensory conditions into consideration. For example, Reynolds and Heeger (2009) propose a normalisation model of attention, compatible with most empirical findings, in which the consequence of attention related modulation depends on the properties of attentional focus as well as the sensory stimulation and competition. In this thesis, neural responses and the attentional modulation thereof are measured at the macroscopic scale through extracranial recordings. Specifically, in **Chapters 3 and 4**, I employ computational methods that build on these selective tuning properties and the attentional modulation of items and features.

1.2.2 Types of Attention

Picture yourself waiting for your friend in a busy shopping street. While scanning the crowd to catch a glimpse of your friend, your brain is making numerous computations to detect the target of your search in an efficient manner. To find your visual target, you rely on your memories to guide your search; attention is oriented to locations in visual space where it is likely for your friend to appear at times when this is most efficient, and, importantly, we utilise information about what visual features that set your friend apart from the dazzling stream of visual input. This search continues until we discover sensory features that match those in our mental template, and we can execute the task at hand.

To this end, it has been suggested that attention is directed to the representational format that sets the target apart from competing stimuli, or stimulus features.

These attentional search templates serve the purpose of facilitating the search in our environment for task-relevant information (Boettcher, Gresch, Nobre, & van Ede, 2021). However, in a changing environment, where viewing angles change, partial occlusion is common, and diverse sets of distractors, maintaining a photographic target image in mind is of little use; flexible mental templates are required. The emerging view of the literature is that attentional templates are designed to increase representational dissimilarity between targets and non-targets, depending on task circumstances (Becker, Harris, Venini, & Retell, 2014; Navalpakkam & Itti, 2007; Yu & Geng, 2019). Thus, it is suggested that task-relevant information that is stored in working memory in an adaptable format where attention can be oriented towards different representational formats, some of which are discussed below.

Space- and Object-based Attention

Selective attention is highly versatile and can be deployed in pursuit of different goals (Nobre & Van Ede, 2018). In early studies, attentional selection is studied in the context of space (Posner, 1980), where one location is preferentially processed over other locations. Classically, the attention spotlight is defined by space and all information that falls outside this locus is ignored (Posner, 1980). In most common settings, we direct our attention to locations containing an item that consists of a complex set of features (e.g. a teapot consisting of a range of visual features like colours, size, orientations, textures, location, shape). In such settings, we preferentially encode all the attended features that comprise the object (Baylis & Driver, 1993; Duncan, Humphreys, & Ward, 1997; Egly, Driver, & Rafal, 1994;

Kahneman & Henik, 1981). Integrated into an object, is it more difficult to suppress features that are not task relevant when they belong to the same object (Kramer & Jacobson, 1991; Stroop, 1935), as it exemplified in the Stroop task where incongruency between word meaning and the colour of the word are involuntarily processed together and impede performance of the observer.

Feature- and Dimension-based Attention

Yet, in other scenarios it can be advantageous to direct attention to a limited set of features (e.g. in search for a red teapot among teapots of different colours), here, we can selectively attend features that differentiate the target item from the competing distractors or facilitate perception. Studies of feature-based attention demonstrate that observers can search for a target feature while restricting processing of irrelevant features (Egeth, Virzi, & Garbart, 1984; Treisman & Gelade, 1980; Wolfe, 1994). Feature-based attention is suggested to act globally, independent of location (Sàenz, Buraĉas, & Boynton, 2003; Treue & Martínez-Trujillo, 1999). For example, Saenz et al. (2003) demonstrated that when observers attended the same feature (colour or motion direction) across two items performance was increased. This was contrasted with trials where the items did not share a common feature (see also Liu & Hou, 2011; Rossi & Paradiso, 1995; Treue & Martínez-Trujillo, 1999; White & Carrasco, 2011). The location-independent account of feature-based attention has been enforced by neuroimaging studies (Liu, Larsson, & Carrasco, 2007; Sàenz, Buraĉas, & Boynton, 2003).

Depending on task circumstances, attention towards a single feature value or towards a feature dimension (e.g., size, orientation, colour) can be deployed to find a visual target more efficiently (Found & Müller, 1996; Kumada, 2001; Müller, Heller, & Ziegler, 1995; Treisman, 1988). Investigating benefits of feature-dimension based attention, Treisman (1988) conducted a study where observers searched for a target among distractors, defined by three feature-dimensions (colour, orientation, and size). Target search times were similarly low when participants were told the feature value within the dimension as when participants were made aware of just the feature-dimension of the target. When no information about the relevant feature dimension was provided, a significant increase in search times was observed. Thus, Treisman (1988) concluded dimension-weighting yields an attentional benefit.

In **Chapter 2**, I consider feature-dimension-based prioritisation in working memory, where I directly compare it to item-based prioritisation. To date, most work in feature-dimension prioritisation has come from studies of attentional selection during perception. Whether such prioritisation may also apply across objects maintained in working memory – which may have a different representational format for attention to interact with - remains largely unaddressed.

Temporal Attention

Beyond spatial attention, location that is defined by receptive fields or feature-dimension based attention defined by functionally specific regions, other types of information can also be used to guide attention. Through temporal attention, observers can modulate perceptual processing using the temporal structure of events

(Coull & Nobre, 1998; Ghose & Maunsell, 2002; Nobre & Van Ede, 2018). Previous studies have argued that early sensory visual modulations by temporal expectations may interact with spatial attention, and possibly other features (Doherty, Rao, Mesulam, & Nobre, 2005; Rohenkohl, Gould, Pessoa, & Nobre, 2014). Although this type of attentional prioritisation is not central to this thesis, temporal attention is inherently present in dynamic and flexible anticipation and mnemonic prioritisation of different feature dimensions across time.

1.3 Attention in Working Memory

Earlier studies have described primarily how external information and attentional modulation thereof changes behaviour and cognition. More recently, it has become clear that attention can also operate directly on objects maintained in working memory. The sections below, and the upcoming empirical chapters, will discuss and investigate through which cognitive and neural principles and mechanisms attention can prioritise informational content that is maintained in working memory.

1.3.1 Investigating Internal Orienting of Attention

A recent trend, accelerated by the development of neurophysiological methods, has emphasised how attention can be directed to internal representations (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Panichello & Buschman, 2021; van Ede & Nobre, 2021). Many properties of selective attention that have been discussed so far generalise between external and internal attention (Chun, Golomb, & Turk-Browne, 2010; Gazzaley & Nobre, 2012). However, as I will discuss later, there are also relevant differences between external and internal attention.

Experimentally, the priority of individual memory items can be controlled by presenting participants with an array of items succeeded by a cue signalling the relevance of a subset of memory items. In the seminal study by Griffin and Nobre (2003), participants were presented with an array of four crosses with distinct colours in each quadrant which they had to remember over a brief delay until they were probed for their memory of one of the colours. In line with previous external

attention studies, on some trials participants were shown a cue prior to the onset of the array that indicated which colour was most likely to be probed at the end of the trial. Critically, as a novel manipulation, on other trials, the cue appeared 1500 - 2500 ms after the array had been removed from the screen. These cues presented retrospectively have come to be referred to as *retro-cues*. While pre-cues prioritise item encoding through external attention; retro-cues act on internal representations. Importantly, bringing items stored in working memory into the focus of attention has been demonstrated to have behavioural significance and improve performance for the cued items stimuli relative to uncued items or trials where the cue did not carry information (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Makovski & Jiang, 2007; Sligte, Scholte, & Lamme, 2008). Importantly, prioritisation of internal representations brings about behavioural benefits beyond merely preventing the degradation of the memory content over time (Murray, Nobre, Clark, Cravo, & Stokes, 2013; Sligte, Scholte, & Lamme, 2008; Sprague, Ester, & Serences, 2016).

After sensory input is encoded for future use and held in a heightened state of availability, it competes with other items that are also maintained in working memory (Franconeri, Alvarez, & Cavanagh, 2013; Lewis-Peacock & Norman, 2014). Classically, studies have demonstrated that precision of recall decreases with time (Panichello, DePasquale, Pillow, & Buschman, 2019; Shin, Zou, & Ma, 2017) and as more items are stored in working memory (Bays, Catalao, & Husain, 2009) but also limited by the amount of items that can be maintained (Baddeley, 1978, 1992; Luck & Vogel, 1997), suggesting a decrease in the fidelity of the stored representation. It has become apparent that noise accumulation plays a major role in the decrease

of memory errors over time (Panichello, DePasquale, Pillow, & Buschman, 2019; Pertzov, Manohar, & Husain, 2017; Rademaker, Park, Sack, & Tong, 2018). In summary, both protecting working memory against decay as well as attentional orienting to a subset of items in working memory are two ways that are commonly considered to be involved in boosting the fidelity of the attended relative to the unattended items.

1.3.2 Action-Oriented Representations

Empirical work in this thesis discusses how working-memory representations encode and prioritise information to serve future behaviour. Working memory represents content optimised for upcoming actions (Nobre & Van Ede, 2018). Task-relevant information is encoded in working memory with a preferential status. As discussed in **Chapter 2**, presenting more information to the observer by means of internal or external cues can help represent a subset of information maintained in working memory more efficiently for the task at hand. In addition to selection of information, working memory reconfigures sensory content to adapted resource use for task demands (see **Chapter 4**; Myers, Stokes, & Nobre, 2017). For example, when encoding an orientation in memory, the purpose of the task is considered. The memory representation would be different depending on if the response required is a continuous, precision report or a categorical judgement against a template orientation. In the former task, high fidelity memories are necessary to perform the task accurately, however, in the latter case this is not always required. If the

item (say, and oriented stimulus) in memory ought to be compared to a known boundary (e.g. 90 degrees) and following a delay a button should be pressed if the perceived orientation is clockwise or counterclockwise from the boundary, working-memory representations are maintained uniquely different. In this extreme case, an action-oriented memory format may not represent the orientation but instead a categorical judgement or button to press at the end of the delay period.

Evidence shows that working memory is tailored towards task demands, which guide processing of feature values in a goal-directed way (Freedman, Riesenhuber, Poggio, & Miller, 2001; Rainer, Asaad, & Miller, 1998). Additionally, our memory content designed to guide behaviour and can bias its feature values for better performance (Boettcher, Nobre, & van Ede, 2020; Geng, Di Quattro, & Helm, 2017; Navalpakkam & Itti, 2007; Yu & Geng, 2019). More specifically, the stimulus dimensions that are most relevant to the task at hand are represented most strongly, while the other – equally available – dimensions are underrepresented (Boettcher, Nobre, & van Ede, 2020). This means that working memory can often contain non-veridical representations that overrepresent task-relevant features can be functionally adaptive to make subtle decisions in the task at hand.

In other cases, current and previous task-relevant information in working memory can interact with our environment when making perceptual decisions (Cicchini, Mikellidou, & Burr, 2018; Fischer & Whitney, 2014; Fritsche, Spaak, & de Lange, 2020). In **Chapter 4**, I investigate how current and previous internal representations can alter feature values during perceptual and working memory processing. Doing so, I look at how memory content is altered in – what in

naturalistic settings would be – an adaptive way through sensory history effects. These modulations of content in working memory are adaptive for future goals in the sense that they reduce noise from an imperfect visual representation and improve performance by referring to previously relevant mnemonic information (**Stocker2009**).

1.4 Comparing Internal and External Forms of Attention

Studies on attentional orienting have drawn many parallels between orienting attention to internal and external representations (Chun, Golomb, & Turk-Browne, 2010; Nobre, Coull, et al., 2004; Oberauer, 2019). Both internal and external attention act through voluntary and involuntary sources (van Ede, 2020). Evidence also suggests internal and external attention could act on many but also similar levels of representation. For example, internal attention can be oriented towards a subset of locations or items in memory (Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003) or to a feature dimension (Heuer & Schubö, 2016b; Lepsien, Thornton, & Nobre, 2011; Matsukura & Vecera, 2015; Niklaus, Nobre, & van Ede, 2017). Yet, it is likely that attentional processes may act differently on consolidated information maintained internally than on raw sensory information present externally – as the formats of information may differ.

1.4.1 Capacity of Internal Representations

While investigating the format of working memory, it is important to consider the richness of external compared to internal representations. While the external environment contains immeasurable detail, memory capacity is highly limited. Attentional selection in perception serves to filter sensory input and regulates most of the information that is encoded into working memory. In **Chapters 2 and 3**,

I describe the breadth of detail that is present at the initial stages of processing and is carried forward, subject to working-memory storage limitations.

Early studies on the capacity of visual working memory had estimated humans could retain around four items in memory during delayed recall tasks (so-called slot models; Cowan, 2000; Luck & Vogel, 1997; Pashler, 1988; Vogel & Machizawa, 2004; Zhang & Luck, 2008) with resolution limited by attentional resources (precision models; Bays, Catalao, & Husain, 2009; Ma, Husain, & Bays, 2014; Van Den Berg, Shin, Chou, George, & Ma, 2012). These capacity limits seem to be relatively consistent for items containing a single feature (e.g., a display with four colours) compared to multi-feature items (e.g., four lines that each comprise a colour and an orientation; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). This suggested that features of a single object are bound together into integrated objects for efficient processing (see also Xu & Chun, 2006). Yet, follow-up studies show feature binding does not increase information in working memory without limits. Maintaining objects with multiple features (e.g., two or more colours) leads to a reduction in working-memory capacity (Olson & Jiang, 2002; Wheeler & Treisman, 2002) and, therefore, for increasingly complex items the capacity has also been observed to be lower (Alvarez & Cavanagh, 2004; Eng, Chen, & Jiang, 2005; Song & Jiang, 2006). Hence, it is possible that, following perception, features belonging to the same item are processed and stored together. Although recent studies suggest that it is still possible to select single features of an integrated item (Niklaus, Nobre, & van Ede, 2017; Park, Sy, Hong, & Tong, 2017; Yu & Shim, 2017), it is unclear to what extent attention to features in working memory can decouple from the

item representations. **Chapter 2** directly addresses this question by comparing prioritisation of items and feature dimensions in working memory and contrasting behavioural benefits and neural correlates of internal attention between the two types of attention.

1.4.2 Representational Differences

Apart from disparities in capacity, the format of information in working memory differs from that in perception. Accepting that attention at the perceptual stage is instrumental in binding features together into integrated items, it is proposed that features in working memory that belong to the same item are more closely bound together after encoding (Luck & Vogel, 1997; Wheeler & Treisman, 2002; Xu & Chun, 2006). Furthermore, bound item representations may be more independent from each other in memory than in perception. Internal representations can be orthogonalised from perception suggesting a representational transformation takes place post-perception (Libby & Buschman, 2021; Panichello & Buschman, 2021; Wan, Menendez, & Postle, 2021). Orthogonalisation can reduce interference between simultaneously encoded stimuli in the same neural population.

In one way, the differences in representational format may result in orienting attention towards internal representations being more flexible than towards features in our environment. Shifting attention towards internal representations is more dynamic and is associated with fewer costs and is reversible (LaRocque, Riggall, Emrich, & Postle, 2017; Myers, Stokes, & Nobre, 2017; Rerko & Oberauer, 2013;

Van Moorselaar, Olivers, Theeuwes, Lamme, Victor, & Sligte, 2015; van Ede, Niklaus, & Nobre, 2017a). More specifically, item representations that have been encoded in working memory are not lost when attention is oriented away; working-memory representations may be restored when put back into the focus of attention shortly afterwards.

On the other hand, differences in representational format may also place constraints on how attention can be oriented within working memory. Not all types of information may be accessible to the same extent. If the primary unit of working memory are items, containing sets of features bound together at the encoding stage, then individual feature dimensions or values may be less accessible to attentional prioritisation. As alluded to above, if this is the case, orienting of attention to feature-dimensions in working memory would automatically prioritise the other features belonging to the same item, reducing overall cueing benefits. Interestingly, recent studies show that attentional selection can act on feature dimensions within working memory (Heuer & Schubö, 2017; Niklaus, Nobre, & van Ede, 2017; Park, Sy, Hong, & Tong, 2017; Pilling & Barrett, 2016; Yu & Shim, 2017). Still unclear, however, is how benefits of feature-dimension attention compare to those of object-based attention in working memory. Understanding the nature and comparative efficacy of these effects can in turn be highly informative about the representational format of working memory. In **Chapter 2**, I describe an experimental study where I directly contrast the prioritisation of integrated items or feature-dimensions maintained in working memory and observe their performance benefits and neural signatures.

1.5 Sensory History for Adaptive Internal Representations

Working memory is not modulated solely by current task demands and current sensory input. In **Chapter 4**, I will test how working-memory encoding leaves a lasting trace that affects processing of subsequent sensory content. This chapter will consider how stimuli presented on the current and previous trial modulate content encoded into working memory – and therefore thus also consider longer-term memory influences on working memory. Additionally, I test how attention is not only relevant in the current trial but how attention effects in one trial could interact with sensory history to jointly shape content in working memory at later times.

To construct a stable representation of our visual environment, the visual system relies in part on visual adaptation (**Stocker2009**; Clifford, 2014; Wainwright, 1999; Webster, 2015) and memory traces of sensory information (Bliss, Sun, & D'Esposito, 2017; Kiyonaga, Scimeca, Bliss, & Whitney, 2017). Together, these sources have been argued to support temporarily stable representations by exploiting the temporal redundancies in the sensory input (Cicchini, Mikellidou, & Burr, 2018; Srinivasan, Laughlin, & Dubs, 1982).

Working memory and perception have been argued not to be independent, but rather to share overlapping resources in the brain (Christophe, Hebart, & Haynes, 2012; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009). It is hypothesised that sensory regions are recruited when retaining precise feature information online for future use (termed the sensory recruitment hypothesis; Awh & Jonides, 2001; D'Esposito, 2007; Pasternak & Greenlee, 2005; Postle, 2006).

However, with present neuroimaging and neurophysiological methods, it will remain difficult to reconcile the evidence for sensory recruitment in visual areas during working memory maintenance. If working-memory representations share processing in overlapping neuronal populations in sensory regions, this raises the question of how new perceptual input interacts with lingering representations (Makovski & Jiang, 2008; Teng & Kravitz, 2019). Studying the relationship between perceptual and working-memory representations, neurophysiological and neuroimaging studies show that, while distractor presentation during working memory maintenance has little effect on behaviour, it diminishes the read-out of any sensory working-memory representation (Bettencourt & Xu, 2015; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; Miller, Erickson, & Desimone, 1996) but not in prefrontal and parietal regions (Bettencourt & Xu, 2015; Miller, Erickson, & Desimone, 1996). These findings draw into question whether sensory areas are necessary for visual working memory and how information is relayed between other areas of the brain. Bettencourt and Xu (2015) have argued that memory readout in visual areas is epiphenomenal and is not required to retain visual information in working memory. Instead, the authors suggest that parietal regions maintain distractor-resistant working-memory representations. Interestingly, neural stimulation of visual regions provide further causal evidence of the involvement of sensory areas in working-memory representations (Rademaker, Van De Ven, Tong, & Sack, 2017).

An recent line of work has theorised that working memory is further supported through coding strategies that do not solely rely on persistent firing during the maintenance period. One such proposed mechanism is that information is maintained

in activity-silent representations that rely on short-term synaptic plasticity (Mongillo, Barak, & Tsodyks, 2008; Stokes, 2015). This would allow for metabolically efficient storage that can simultaneously represent incoming sensory input and information in working memory (Sreenivasan, Curtis, & D'Esposito, 2014; Sreenivasan & D'Esposito, 2019) . Studies in support of activity-silent codes have shown that presenting strong visual stimulus (a so-called "ping") will send a wave of activity through visual areas that can revive recurrent activity through the short-term synaptic changes for readout (Wolff, Ding, Myers, & Stokes, 2015; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020; Wolff, Jochim, Akyürek, & Stokes, 2017). Moreover, memory information in short-term synaptic codes can interact with new perceptual information and thereby bias behaviour (Teng & Kravitz, 2019). Thus, it is suggested that synaptic traces related to previously relevant sensory information can have a lasting effect and interact with subsequent sensory processing. Even though the exact mechanisms remain elusive, the history effects in sensory processing are becoming increasingly appreciated. In **Chapter 4**, I will test how sensory history affects processing of task-relevant and irrelevant features in a working memory task.

1.6 Aims and Overview of the Thesis

In this doctoral thesis, I investigate feature information processing in working memory and its modulation by top-down and bottom up processes. Using a range of electrophysiological and behavioural methods, I investigate encoding of external information into working memory and how this process is modulated by attention through task demands. I started this thesis laying the foundation of fundamental principles on selective attention and representations in working memory (**Chapter 1**). This chapter is aimed to provide a mere summary of the status of a field that develops continuously. In **Chapter 2**, I investigate retro-cueing benefits for two features of the same item or two features within the same feature-dimension but spread out across the visual field. By collecting behavioural responses and EEG measurements, it is possible to test the cueing benefits as well as the neural dynamics following either type of cue. Comparing prioritisation of item- and dimension-based information in working memory is informative about the representational format internally represented content. Subsequently, **Chapter 3** tests the extent to which feature processing at the working memory encoding stage can be traced. I aim to extend the decoding methods beneficial describe neural dynamics of colour – and orientation – processing. Capitalising on these sensitive classification methods that I developed in **Chapter 3**, **Chapter 4** investigates systematic modulations in neural representations as a function of sensory history and the long-lasting effect of task-relevance. More specifically, I investigate the repulsive and attractive bias observed in working memory towards same and previous trial orientations,

respectively. I use MEG data to test the biases in feature representations in working memory guiding behaviour. In **Chapter 5**, the general discussion of the present work delineates the insights my work has provided into feature modulations through top-down task relevance and bottom-up sensory history. In this final chapter, I further outline the applications and future directions of this research field.

In summary, there are several main strands that define this thesis. I investigate *feature representations* and how feature information from different feature-dimensions is processed and prioritised according to task demands. In **Chapter 3**, I specifically analyse concurrent processing of information on multi-item, multi-feature displays. **Chapter 4** examines modulatory biases in working memory encoding and post-perceptual processes. **Chapter 2** describes neural and behavioural comparisons between prioritising feature-dimensions and prioritising single items that contain information from multiple feature dimensions. Furthermore, this thesis regards *representational dynamics*; how external features leave a lasting trace as internal representations. **Chapters 3 and 4** characterise representation of sensory stimuli using state-of-the-art decoding methods. Here, **Chapter 4** contributes by describing the dynamics of feature encoding and modulation by lingering information from past stimulus events and past trials. Furthermore, **Chapters 2 and 4** examine modulations of working-memory representations by assessing alpha desynchronization (**Chapter 2**) and multiple sensory-history biases across time (**Chapter 4**).

2

Comparing the Prioritisation of Items and Feature-dimensions in Visual Working Memory

Contents

2.1 Abstract 32
2.2 Introduction 33
2.3 Methods 37
2.3.1 Participants 37
2.3.2 Experimental Set-up and Stimuli 37
2.3.3 Task and Design 38
2.3.4 Behavioural Analysis 40
2.3.5 EEG Acquisition 42
2.3.6 EEG Preprocessing 43
2.3.7 Time-frequency Processing 44
2.4 Results 46
2.4.1 Behavioural Cueing Benefits 46
2.4.2 Distractor Processing 49
2.4.3 Mixture Modelling 49
2.4.4 Alpha Attenuation Following Feature and Item Retro-cues 50
2.5 Discussion 54

2.1 Abstract

Selective attention can be directed not only to external sensory inputs, but also to internal sensory representations held within visual working memory. To date, this has been studied predominantly following retrospective cues directing attention to particular items, or their locations in memory. In addition to item-level attentional prioritisation, recent studies have shown that selectively attending to feature dimensions in working memory can also improve memory recall performance. However, no study to date has directly compared item-based and dimension-based attention in working memory, nor their neural bases. Here, I compared the benefits of retrospective cues (retro-cues) that were directed either at a multi-feature item or at a feature-dimension that was shared between two spatially segregated items. Behavioural results revealed qualitatively similar attentional benefits in both recall accuracy and response time, but also showed that cueing benefits were larger following item cues. Concurrent EEG measurements further revealed a similar attenuation of posterior alpha oscillations following both item and dimension retro-cues when compared to non-informative, neutral retro-cues. I argue that attention can act flexibly to prioritise the most relevant information – at either the item or the dimension-level – to optimise ensuing memory-based task performance, and I discuss the implications of the observed commonalities and differences between item-level and dimension-level prioritisation in visual working memory.

2.2 Introduction

Visual working memory provides a means to maintain relevant information independently of continued visual input, to guide adaptive behaviour (Baddeley, 1992, 2003). Because working memory has limited capacity and/or resources (Bays & Husain, 2008; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008), it is essential to distribute memory processes efficiently to complete tasks at hand effectively. Over the past decade, it has become increasingly clear that working memory is more flexible than originally thought. Focused attention continues to prioritise and select contents maintained in working memory as goals and predictions about goals change (Griffin & Nobre, 2003; Kuo, Stokes, & Nobre, 2011; Landman, Spekreijse, & Lamme, 2003; Souza & Oberauer, 2016). To bring about behavioural benefits, attention-related modulatory signals must interact with mnemonic information that is available within working memory. Thus, by studying what forms of attention confer benefits to working memory also provides insight about the format of information held in working memory.

In perceptual attention, research has revealed a multitude of representational formats available for modulation. These include spatial locations, objects, features, semantic associations, time intervals, and likely more (for an overview, see e.g., Nobre & Van Ede, 2018). Whether the same diversity of representational formats is available in visual working memory is an important and informative question. Information in working memory results from attentional filtering of incoming sensory processing (Vogel, McCollough, & Machizawa, 2005). Thus, the representational

information might be kept in an altered, more compact format. For example, it has been suggested that the primary representational unit of working memory involves integrated (feature-bound) items (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). Under this framework, one might expect that the primary target for attentional selection in visual working memory should be at the level of individual items. Accordingly, most studies looking at the role of attention in working memory to date have used spatial retro-cues, linked to item-based representations, and have shown clear benefits (e.g., Griffin & Nobre, 2003; Landman, Spekreijse, & Lamme, 2003; Souza & Oberauer, 2016).

At the same time, recent studies have demonstrated that attention can also facilitate behaviour when directed to feature dimensions that are shared among multiple items in working memory (Heuer & Schubö, 2017; Niklaus, Nobre, & van Ede, 2017; Park, Sy, Hong, & Tong, 2017; Pilling & Barrett, 2016; Sahan, Sheldon, & Postle, 2019; Ye, Hu, Ristaniemi, Gendron, & Liu, 2016; Yu & Shim, 2017). It remains unclear, however, how such effects of dimension-based attention compare to item-based attention in working memory, as no study has directly compared these two forms of attentional facilitation in working memory. Here, I directly compare these two types of attentional facilitation.

In addition to comparing item and dimension-based attention at the level of behavioural performance, I examined their effects on an electrophysiological marker linked to attention in working memory: the attenuation of posterior alpha oscillations. Several studies have revealed that item-based prioritisation in working memory is associated with the attenuation of alpha oscillations in posterior brain areas,

suggesting modulation of visual areas involved in representing the mnemonic items (Myers, Walther, Wallis, Stokes, & Nobre, 2015; Poch, Capilla, Hinojosa, & Campo, 2017; van Ede, 2018; van Ede, Niklaus, & Nobre, 2017b; Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015; Wolff, Jochim, Akyürek, & Stokes, 2017). It remains unclear whether alpha attenuation also occurs during the attentional prioritisation of feature dimensions that are shared across multiple items held in working memory.

In the current study, I therefore compared and contrasted behavioural and neural effects of internal shifts of attention to multi-feature items and to single feature dimensions that were shared across multiple items. Through the behavioural data, the aim was to test whether there is a clear primacy of the object-level information in working memory. If the representational format in working memory organises items as integrated objects, this should also be the primary level at which attention can operate. Accordingly, benefits from item-directing retro-cues should be substantially larger. If, however, attention has similar access to multiple levels of information in visual working memory, then retro-cueing benefits for feature dimensions and individual items may be similar. By recording EEG and measuring alpha oscillations, I further tested whether a similar alpha modulation occurs when attention is directed to a cued item or to a visual feature dimension that is distributed across multiple items held in working memory.

To address these questions, I used a task in which participants were presented with two Gabor gratings, each of which contained both colour and orientation information. On half of the blocks, participants were presented with an item-directing retro-cue and on the other half with a feature-dimension-directing retro-

cue. Both blocks contained neutral (uninformative) retro-cues, against which I compared the effects of both types of informative retro-cues. I observed qualitatively comparable retro-cueing effects, though benefits following item cues were larger. Both effects were accompanied by similar alpha attenuation following the cues, and both item- and dimension-level benefits on behaviour were highly correlated across participants.

2.3 Methods

2.3.1 Participants

The study was approved by the Central University Research Ethics Committee of the University of Oxford and is conducted in accordance with the Declaration of Helsinki. Thirty-two healthy volunteers (19 female; mean age 28.3; range 18-35) took part. Participants had normal or corrected-to-normal vision and were not colour blind. Participants provided written informed consent before participating in the study and were paid £15 per hour. Data from two participants were excluded from analysis, one for terminating the experiment early and the other due to hardware failure.

2.3.2 Experimental Set-up and Stimuli

Participants were seated in front of a 23-inch monitor (1920 × 1080, 100 Hz). Stimuli were generated using Psychophysics Toolbox version 3.0.11 (Brainard, 1997) in MATLAB 2014b (MathWorks, Natick, MA). Head position was set at 90 cm from the monitor, and participants used a chinrest. The stimuli consisted of luminance-defined sinusoidal Gabor gratings generated in MATLAB 2014b. Fourty-eight evenly spaced colours were drawn from a circle in CIE L*a*b colour space (center at L = 54, a = 18, b = -8, radius = 59). Gratings were presented using one of 48 different orientations (3.75° to 180° in steps of 3.75°) and 48 different colours.

2.3.3 Task and Design

Participants performed 960 trials of a visual working memory task (**Figure 2.1**) in which they were asked to reproduce the colour or orientation of one out of two memory items at the end of a memory delay of 2.3 seconds. At the start of each trial, two Gabor stimuli with a radius of 2.2° positioned left and right from fixation (centred 3.1° of visual angle) were presented simultaneously for 300 ms. Participants were instructed to remember the colour and the orientation of both items. At the end of the trial, they were probed to report the orientation or the colour of one of the items. The to-be-reported feature dimension was indicated with the probe circle that was either a colour wheel (colour report) or a white wheel (orientation report), while the to-be-reported item was indicated by the location of the probe circle (left/right, corresponding to the original location of the probed memory item). Orientation and colour values varied independently between the two items, with the constraint that no two equal orientations or colours were presented on the same trial. Colours and orientations were counterbalanced so that each was presented equally often across trials.

Two events occurred during the memory delay: first a retro-cue appeared, which could provide information about the item or feature dimension that would be probed. Second, 700 ms after the retro-cue offset, an irrelevant 'distractor' stimulus was presented that contained either colour or orientation information (**Figure 2.1** for examples). I inserted the distractor with the intention to study the neural response to the distractor as a function of cueing condition. In the end, this aspect did not yield clear results and was therefore omitted for clarity. However, the single-feature

nature of the distractor did allow us to investigate whether the congruence of the distractor feature and the probed feature had any effect on recall.

Participants completed 20 blocks in total, each containing 48 trials. On even-numbered blocks, informative retro-cues indicated the location (left, 'L', or right, 'R') of the item that would be probed at the end of the trial, without giving information about what dimension would be probed. On odd-numbered blocks, informative retro-cues indicated the feature dimension to be probed (colour, 'C', or orientation, 'O'), without giving information about what item would be probed. I thus cued either a single item that contained two features, or a single feature-dimension that was shared between two items. I note that what I refer to as "feature-dimension" selection is thus distinct from using feature-based cues to orient attention to specific items (as in e.g., Heuer & Schubö, 2016a; Heuer, Schubö, & Crawford, 2016). When informative, the retro-cue was always valid. Both blocks also contained 33% non-informative neutral cues ('X') that provided no information about what item or dimension would be probed at the end of the trial. Participants were encouraged to use the informative retro-cues to select the relevant item or feature dimension.

Following another fixation period of 700 ms (after the retro-cue), a bilateral distractor was presented for 100 ms. The distractor consisted of either a single colour or a black-and-white oriented Gabor grating. Sixteen colours and 16 orientations were used (11.25° to 180° in steps of 11.25°). These varied randomly from the orientation/colour of the items in the memory array. However, I ensured that the three types of cues (C/O/X or L/R/X) all contained the same range of 16

distractor features. Out of these 16 distractor features, I randomly assigned 8 as colour features and 8 as orientation features.

After another fixation period of 700 ms (after the distractor), the colour-wheel or a white-circle probe appeared with the dial initialised at a random location. To keep the orientation and colour recall as similar as possible, I presented the colours at a fixed position on the colour wheel. Participants were instructed to respond as accurately as possible by using the 'J' and 'F' key to rotate the probe clockwise and counter-clockwise, respectively. Participants were instructed to use their left index finger to press the 'F' key and the right index finger to press the 'J' key. Although there was no explicit time limit for the response time, I logged reaction times as the time between the onset of the probe and the first button press that initiated the 'dial-up' report. Reaction time therefore serves as a proxy for the time it took participants to access the relevant memory information before commencing their reproduction report.

2.3.4 Behavioural Analysis

I computed the error for each trial for each participant by subtracting the target orientation or colour (in radians around the colour circle in CIE L*a*B space) from the probe response. All error scores were mapped onto a -½π to ½π space. All trials for which the reaction time was more than 4 standard deviations above a participant's mean decision time were discarded (0.9% $\pm$ 0.3%). To calculate the retro-cue benefit, I subtracted the absolute error on cued trials from the absolute error on neutral

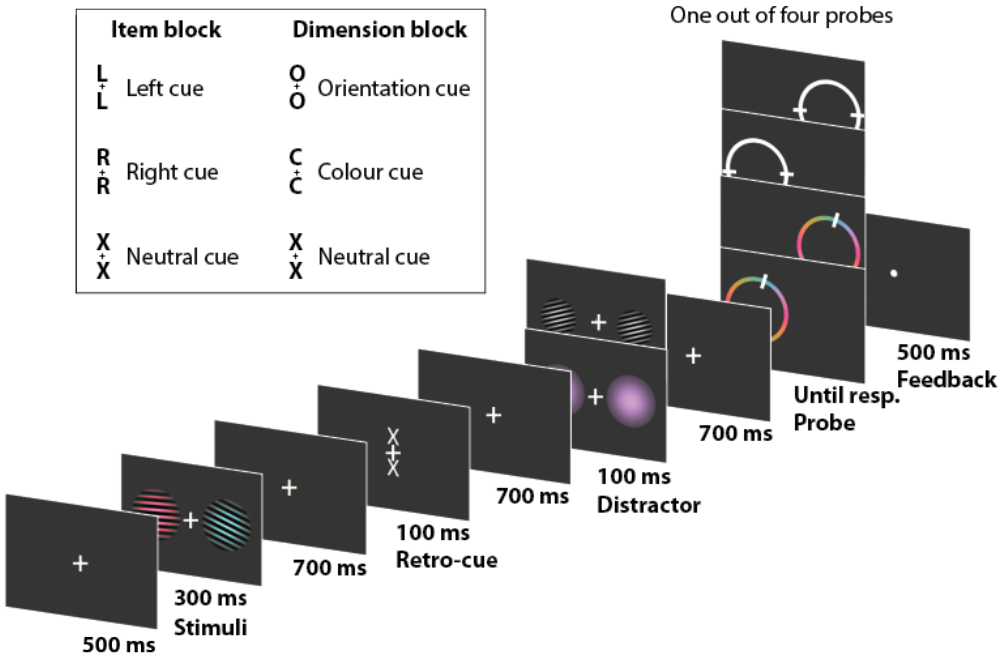


Figure 2.1: Experimental design with timings. Participants were presented with two coloured Gabor gratings to memorise. Subsequently, a retro-cue informed participant which item or feature dimension would be probed. After either a colour distractor or an orientation distractor with a semi-randomly drawn orientation or colour, the probe was presented. Participants adjusted the probe dial to match the feature in memory.

trials. In all my analyses, I only compared trials of one retro-cueing condition with neutral trials from the same block types (i.e. item blocks or dimension blocks).

A mixture model was fitted separately for each retro-cueing condition and respective neutral condition, modelling target response rate, guess rate, swap responses, and precision to the error data of each subject (Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008). I fitted the mixture model separately for every subject, colour or orientation recall, spatial or dimension retro-cues, and informative or neutral retro-cues. Estimating the mixture-model parameters allowed for estimation of different components that contribute the overall error; I estimated the fidelity of the representation independently of the guess and swap rate. I used the mixture model made available by Bays et al. (2009).

When comparing more than two conditions, I applied a repeated-measures analysis of variance (rmANOVA) and report η^2 as a measure of effect size. When evaluating retro-cueing benefits, I applied dependent samples t-test, comparing informative vs neutral cues, as well as the cueing effects between item and dimension retro-cues. When the assumption of normality was violated I instead applied a Wilcoxon signed-rank test. I report Cohen's d as a measure of effect size for parametric tests and matched rank biserial correlation for non-parametric effect size. For evaluation I used two-sided tests with a critical alpha value of 0.05.

2.3.5 EEG Acquisision

EEG data were collected using Synamps amplifiers and Neuroscan software (Compumedics). I used a 61 Ag/AgCl sintered electrodes (EasyCap, Herrsching, Germany), laid out according to the international 10-10 system, with mastoids behind the left and right ear. The left mastoid was used as an active reference during the recordings. Offline, an average-mastoids reference was derived using the left and right mastoids. The ground electrode was placed on the left arm above the elbow. Horizontal EOG was measured using lateral electrodes next to both eyes while vertical EOG was measured above and below the left eye. Data were sampled at 1000 Hz, and stored for subsequent analysis.

2.3.6 EEG Preprocessing

Data were imported into MATLAB 2017a using *pop_loadcurry()* and further analysed using Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) and the OHBA Software Library (OSL; https://ohba-analysis.github.io/). Analysis started by cutting out the epochs between 100 ms before and 2200 ms after retro-cue onset (*ft_redefinetrial*) followed by re-referencing the data to the average of the mastoids (*ft_preprocessing*). EEG data were down sampled to 200 Hz to reduce computational demands and storage space (*ft_resampledata*).

Next, EEG data were further de-noised using Independent Component Analysis (ICA; *ft_componentalanalysis*) applying the FastICA algorithm (Hyvärinen, 1999) to all EEG sensors. ICA separates the EEG signal into non-Gaussian subcomponents of the data that are statistically independent from one another. Spatial components strongly correlated ($r > 0.4$) with electrooculogram (EOG) channels were removed from the EEG data. I set-out to remove trials on which participants blinked during the window of 100 ms prior up to 200 ms post retro-cue presentation. After baselining the horizontal EOG signal at -300 to -100 ms trials on which horizontal EOG voltage surpassed 200 μV (approximately ½ of the maximum voltage evoked by a typical blink) were flagged and later removed from EEG and behavioural analyses (0.446% ± 1.23%; mean ± standard deviation). Subsequently, I removed epochs based on within-trial variance of the broadband signal at a 0.05 significance threshold using a generalised ESD test (Rosner, 1983, implemented in OSL) and discarded 2.48% ± 2.18% (mean ± standard deviation) of the trials.

2.3.7 Time-frequency Processing

Time-frequency decomposition of the EEG signal was done using *ft_freqanalysis*. Spectral power between 2 and 50 Hz was computed on Hanning-tapered data using a short-time Fourier Transform, with a 300-ms sliding time window that was advanced in steps of 15 ms. I zoom in on modulations in posterior alpha oscillations, by averaging the time-frequency plots for the left cues (P1, P3, P5, P7, PO3, PO7, O2) and right (P2, P4, P6, P8, PO4, PO8, O2) posterior electrodes and calculating the normalised differences in power following between informative and neutral retro-cues

$$\frac{\text{informative} - \text{neutral}}{\text{informative} + \text{neutral}} \times 100$$

I did this separately for left and right item retro-cues, and for colour and orientation dimension retro-cues. Using the same quantification, I also compared alpha power for contralateral vs. ipsilateral electrodes after item, and alpha for colour vs. orientation following a dimension cue. For statistical evaluation, I applied a two-sided cluster-based permutation analysis (Maris & Oostenveld, 2007) with 5000 permutations at an evaluation threshold of 0.05.

To characterise the onset of alpha attenuation after the retro-cue, I extracted the time course of 7-12 Hz power modulation (in the specified informative vs. neutral cue contrast) and focused on the 0-1000 ms period post retro-cue onset. On these data, I then identified the earliest time point in which the power modulation reached half of its minimal value for each condition. This latency was used as a measure to compare neural modulation by dimension retro-cues and item retro-cues.

To depict the topography of the power modulations analysed in the predefined set of posterior electrodes (depicted in **Figure 2.4**), I calculated the relevant contrast for each electrode and averaged over the time-frequency window of 400-800 ms and 7-12 Hz. In addition, to focus on alpha lateralisation, I contrasted activity in electrodes left posterior electrodes (O1, PO7, PO3, P7, P5, P3, P1) and right posterior electrodes (O2, PO8, PO4, P8, P6, P4, P2), contralateral vs ipsilateral to the cued item following informative item retro-cues. To visualise the topography of the difference between electrodes contra- and ipsilateral to the item cue, I mirrored the left and right electrodes onto one side. I then subtracted electrodes that would be ipsilateral relative to the cued item from the electrodes that would be contralateral relative to the cued item.

Topographies were intended solely to portray the nature of the modulation and were not subjected to further statistical testing.

2.4 Results

2.4.1 Behavioural Cueing Benefits

Figure 2.2A shows behavioural performance as a function of experimental condition (collapsed over distractor type, as this did not yield consistent results as discussed below). To analyse the effects of item and feature-dimension retro-cues, I quantified retro-cueing benefits as the difference between the trials with informative and neutral retro-cues (**Figure 2.2B**).

To formally quantify the effects of retro-cue informativeness (valid or neutral) and retro-cue block type (item retro-cue block or dimension retro-cue block) I used a 2 × 2 rmANOVA. I ran this separately for RT and response error, and separately for both colour and orientation recall reports. I observed a significant main effect of retro-cue informativeness, with better performance following informative vs. neutral retro-cues on all four dependent variables: orientation error ($F_{1,29} = 18.888, p < .001, \eta^2 = 0.394$), colour error ($F_{1,29} = 26.387, p < .001, \eta^2 = 0.476$), orientation RT ($F_{1,29} = 67.27, p < .001, \eta^2 = 0.699$), and colour RT ($F_{1,29} = 65.15, p < .001, \eta^2 = 0.692$). At the same time, I found that the behavioural benefits of retro-cue informativeness were larger in item retro-cue block than in dimension retro-cue block, yielding a significant interaction for colour error ($F_{1,29} = 9.065, p = .005, \eta^2 = 0.238$), orientation RT ($F_{1,29} = 19.64, p < .001, \eta^2 = 0.404$), and colour RT ($F_{1,29} = 21.00, p < .001, \eta^2 = 0.420$). Though I found the same trend, this did not reach significance for orientation error ($F_{1,29} = 3.750, p = .063, \eta^2 = 0.115$). Finally, in line with the larger benefit of item-retro-cues, I also found a significant

main effect of block type, constituted by better performance in item retro-cue blocks for error and reaction times (orientation error, $F_{1,29} = 39.634, p < .001, \eta^2 = 0.577$; colour error, $F_{1,29} = 8.343, p = .007, \eta^2 = 0.223$; orientation RT, $F_{1,29} = 50.51, p < .001, \eta^2 = 0.635$; and colour RT, $F_{1,29} = 32.08, p < .001, \eta^2 = 0.525$).

They had smaller errors ($t_{29} = 4.235, p < .001, d = 0.773$) and responded faster ($t_{29} = 7.854, p < .001, d = 1.434$), compared to trials with neutral retro-cues in the same block types. Similarly, orientation reports benefitted significantly from dimension cues in both reproduction error ($t_{29} = 3.748, p = .001, d = 0.684$) and response onset time ($t_{29} = 6.302, p < .001, d = 1.151$) compared to neutral trials within the dimension retro-cueing blocks. Item cues conferred numerically larger benefits than dimension cues. The difference was not statistically significant for error (0.023 rad., 48%; $t_{29} = 1.936, p = .063, d = .354$), but reached significance for reaction times (79 ms, 94%; $t_{29} = 4.431, p < .001, d = 0.809$).

The same pattern of results was found for the error and reaction times in the colour recall trials: colour reports benefitted from both item cues ($t_{29} = 5.060; p < .001; d = 0.924$) and dimension cues ($t_{29} = 4.069; p < .001; d = 0.743$) and responses were also faster for item cues ($t_{29} = 9.097; p < .001; d = 1.661$) and dimension cues ($t_{29} = 4.951; p < .001; d = 0.904$) compared to their respective neutral trials. For colour reports, I also found greater benefits of item retro-cues compared to dimension retro-cues for both error (0.039 rad., 86%; $t_{29} = 3.011; p = .005; d = 0.550$) and reaction time (63 ms, 91%; $t_{29} = 4.583; p < .001; d = 0.837$).

Benefits of item-based and dimension-based retro-cueing showed strong positive correlations across individuals for both colour and orientation reports (**Figure**

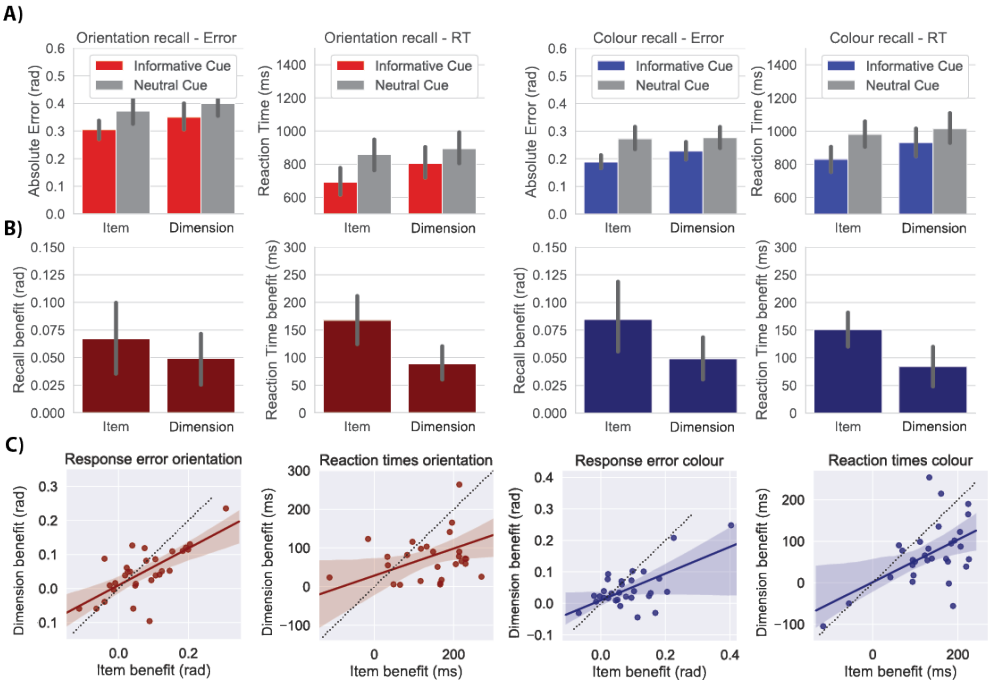


Figure 2.2: Performance benefits of item and feature-dimension retro cueing. A) The four columns show absolute error and reaction times for trials with an informative cue or a neutral cue. Trials in which orientation was probed are displayed in red while colour-probe trials are displayed in blue. Each panel shows the data separately for item retro-cue blocks and dimension retro-cue blocks. B) Behavioural benefit of retro-cues. Subtracting the mean absolute error on trials with an informative cue from the neutral trials gives the performance benefit of the retro-cue – here expressed as positive values. Orientation benefit is depicted in dark red and colour benefit in dark blue. C) Correlations across participants between the item and dimension retro-cue benefits in shown in B. Error bars show 95% confidence intervals.

2.2C). For orientation reports, I found significant correlations between retro-cueing benefits following item cues and dimension cues for both error ($r = .709; p < .001$) and reaction time ($r = .539; p = .002$). Likewise, for colour reports, I found significant correlations between retro-cueing benefits following item cues and dimension cues for both error ($r = .697; p < .001$) and reaction time ($r = .596; p < .001$). Thus, participants who benefitted most from item retro-cues also benefitted most from dimension retro-cues.

2.4.2 Distractor Processing

I also considered the third factor 'distractor congruence' (i.e., when the distractor contained the same or the other feature dimension as the to-be-recalled memory dimension), but found no systematic effects of distractor congruence across my four dependent variables, nor interactions with the factors of interest – see **appendix A**.

2.4.3 Mixture Modelling

In addition to the raw behavioural scores, I also modelled sources of error using a mixture model (**Figure 2.3**; Bays, Catalao, & Husain, 2009). I modelled four components 1) precision, characterised by width (STD^{-1}) of the target centred response distribution, 2) proportion of target responses modelled by the gaussian centred around the target, 3) proportion of random responses characterised by the height of the uniform response distribution, 4) proportion of responses to the non-cued feature of the same dimension as the cued dimension (non-target report or 'swap' errors). **Figure 2.3** show the retro-cueing effects (informative vs. neutral) on each of these four parameters, separately for item and dimension-cues (collapsed over colour and orientation reports, after fitting the model for each dimension separately; see **appendix A.1** for mixture model parameters separated for colour and orientation reports). As depicted in **Figure A.1** informative (vs. neutral) retro-cues significantly increased precision for item retro-cues (item: $t_{29} = 2.736, p = .011, d = 0.500$) though this did not reach significance for dimension retro-cues ($t_{29} = 0.578, p = .568, d = 0.105$). At the same time, both item and

dimension retro-cues increased target response rates (item, $t_{29} = 5.595, p < .001, d = 1.022$; dimension, $Z_{29} = 382, p = .001, rrb = 0.643$), and decreased guess rates (item, $t_{29} = -2.131, p = .042, d = -0.389$; dimension, $Z_{29} = 78, p < .001, rrb = -0.665$), and item retro-cues further decreased swap rate (item, $Z_{29} = 21, p < .001, rrb = -0.910$; dimension, $Z_{29} = 137, p = .080, rrb = -0.368$).

Direct comparisons between item and dimension retro-cue benefits showed a significantly greater reduction in the rate of swap errors by item retro-cues relative to dimension retro-cues (Wilcoxon signed-rank test; $Z_{29} = 95, p = 0.002, rrb = 0.591$; see **Figure 2.3**). Effects for the other three parameters were not statistically different between item and dimension retro-cues (all p > .10).

2.4.4 Alpha Attenuation Following Feature and Item Retro-cues

Figure 2.4 shows the time- and frequency-resolved modulations in spectral EEG power in posterior electrodes following item and dimension retro-cues, expressed as a difference from the neutral retro-cueing condition (neutral retro-cue minus informative retro-cue). After both item retro-cues and dimension retro-cues, I observe an attenuation of alpha power starting at around 400 ms after presentation of the retro-cue (clusters all conditions p < .001). The alpha attenuation in trials with informative retro-cues re-emerges after the distractor onset, in the window just prior to the probe. To reveal the spatial layout of the significant clusters, I visualised the EEG topographies of the alpha-band power at 400 to 800 ms after the retro-cue. Similar topographies were associated with the later alpha

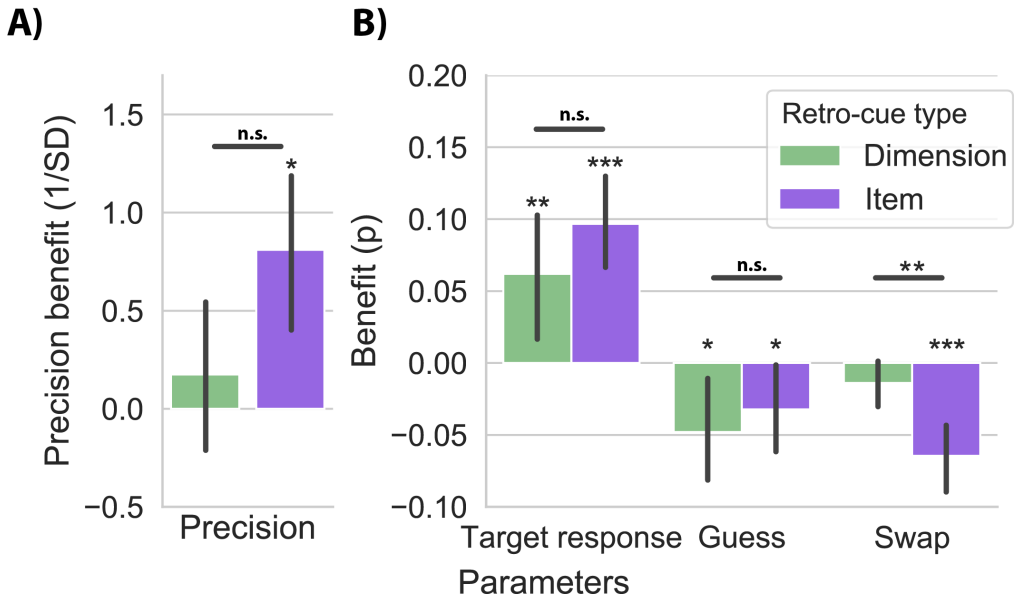


Figure 2.3: Cueing effects on mixture modelling parameters. A) Mixture-model estimates for the cueing effects on A) precision, B) target response, guess rate, and swap rate, for trials where item retro-cues or dimension retro-cues were presented compared to neutral trials. Hence, valid retro-cues positively influenced target response proportions and negatively influenced guess rate and swap rate. The asterisks indicate significant differences of, respectively, item- or dimension benefits from zero (i.e., benefits following informative vs. neutral cues). Asterisks above horizonal lines indicate significant differences between item and dimension retro-cueing benefits. Error bars indicate 95% confidence intervals. n.s: $p > .05$, * $p < .05$, ** $p < .01$, *** $p < .001$.

modulation after the distractor and with the early modulation in the higher 13-30 Hz band (topographies not depicted).

In addition to this 'global effect' when comparing informative to neutral retro-cues, I also evaluated the specific difference between left and right item cues (from the perspective of electrodes contra- and ipsilateral to the cued item), and between colour and orientation (**Figure 2.4**, bottom row). In line with several prior studies (Myers, Walther, Wallis, Stokes, & Nobre, 2015; Poch, Capilla, Hinojosa, & Campo, 2017; van Ede, Niklaus, & Nobre, 2017a; Wallis, Stokes, Cousijn, Woolrich, & Nobre, 2015; Wolff, Jochim, Akyürek, & Stokes, 2017), following item cues, alpha attenuation was most pronounced contralateral to the memorised location of the

cued item. In contrast, following dimension retro-cues no clear differences were observed between colour and orientation cues, which directed attention to a single feature-dimension that was shared between the left and right items. Finally, I found that the alpha attenuation had very similar latencies following item-directing and dimension-directing retro-cues ($t_{29} = 1.273, p = .213$; **Figure A.2**). Further exploring the data, I also analysed event-related potentials (ERP) following retro-cues to investigate ERP components linked to attentional selection in working memory. I found clear selection-related potentials following retro-cues that afforded item-level selection (**Figure A.3**). At the same time, however, I found no clear lateralisation of ERPs after item cues, nor significant ERP modulations associated with attentional selection following dimension cues.

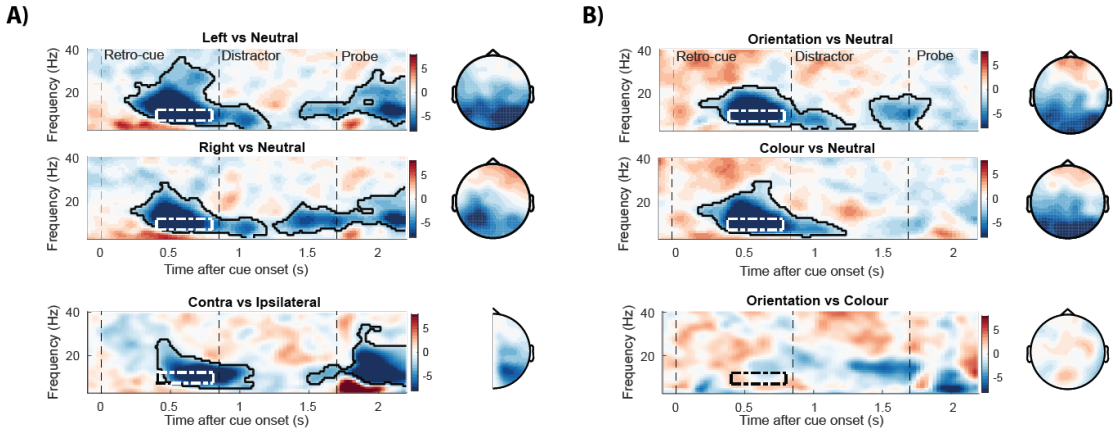


Figure 2.4: Induced neural EEG responses to the left and right item retro-cues and to colour and orientation dimension retro-cues. A) Time-frequency representation of the difference between left/right cued trials vs. neutral trials in item-cue blocks in left (P1, P3, P5, P7, PO3, PO7, O2) and right (P2, P4, P6, P8, PO4, PO8, O2) electrodes. The bottom panel depicts the time-frequency map of the difference between contralateral and ipsilateral responses calculated across left and right posterior electrodes, as described in the methods section. Topographies show the associated differences in alpha power (7-12 Hz) in the 400-800 ms window (indicated in the dotted-line boxes in all time-frequency maps). For the contralateral-vs-ipsilateral topography (bottom), I projected the relevant contrast values into the right electrode of each electrode pair. B) Same representations as outlined above but here I compare colour and orientation with their respective neutral trials or with one another, in left and right electrodes. Highlighted areas with the black solid outline indicate significant clusters (permutation test, n = 30, cluster-forming threshold $p < .05$, corrected significance threshold $p < .05$).

2.5 Discussion

I demonstrate that both item-based and dimension-based attentional prioritisation during working memory maintenance decreases recall error and speeds response initiation times following the probe. Hence, I support the finding that selective attention can retrospectively prioritise not only items (Griffin & Nobre, 2003; Kuo, Stokes, & Nobre, 2011; Landman, Spekreijse, & Lamme, 2003; Souza & Oberauer, 2016), but also feature dimensions maintained in working memory (Heuer & Schubö, 2017; Niklaus, Nobre, & van Ede, 2017; Park, Sy, Hong, & Tong, 2017; Sahan, Sheldon, & Postle, 2019; Ye, Hu, Ristaniemi, Gendron, & Liu, 2016; Yu & Shim, 2017). I thereby replicated a feature prioritisation effect in working memory for more than one location (unlike; Makovski and Jiang, 2007; but see also: Heuer and Schubö, 2016b; Lepsien, Thornton, and Nobre, 2011; Matsukura and Vecera, 2015). Building on this work, my experimental design uniquely allowed us to compare the magnitudes of both types of behavioural retro-cue benefits within a single experiment, and to correlate their strengths across participants. While the item benefit was larger than dimension benefit, both were highly robust. They were each evident across both colour and orientation reports and in both recall accuracy and response initiation times. Moreover, I found strong correlations between the benefits that followed item and dimension cues, and qualitatively similar neural modulations, which suggest that the two types of retro-cueing benefits may share at least overlapping (though not necessarily equivalent) attentional mechanisms.

The notion that both retro-cueing types yield behavioural benefits that are

qualitatively similar was further supported by the similar retro-cueing effects on guess-rate, and target-response rate parameters estimated by the mixture model. At the same time, I observed that only item cues significantly enhanced precision and reduced the probability of swaps (non-target responses) – the latter being the only parameter that also differed significantly between item and dimension retro-cue benefits. This difference is likely explained by the fact that swaps are calculated between items (within a single dimension). Provided that feature-cues always concerned one dimension, shared across both items, they may have helped up-regulate the relevant feature-dimension, but not to separate the two spatially-segregated items and thereby to reduce swap rates (in contrast to item cues that directly targeted the relevant item from the two memorised items). The mixture modelling therefore enables the separation of error into its different sources, such as swap rate and guess rate. Overall error should reflect the aggregate of all mixture modelling components.

In a strict account in which the primary unit of working memory is integrated items (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001), one may predict that attention in working memory will primarily operate at the level of items, leaving little room for attentional facilitation of specific features that are shared among items. Alternatively, if working memory consists of a hierarchy of representations, with both item-level and dimension-level representations (Bays, Wu, & Husain, 2011; Brady, Konkle, & Alvarez, 2011; Fougnie & Alvarez, 2011; Töllner, Conci, Müller, & Mazza, 2016); then one may expect that attention can operate similarly at distinct levels, depending on the nature of the task at hand. My data are in line

with a mixture of both scenarios – showing that attention can operate qualitatively similarly at both item and dimension levels, while also revealing an additional benefit when attention is directed at two dimensions of a single item (following item cues), compared to a single feature dimension across two items (following dimension cues).

At the same time, I note that attentional benefits in behavioural performance in working memory tasks need not only reflect changes in the quality of representational information. Factors related to prospective task preparation may also contribute (González-García, Formica, Liefooghe, & Brass, 2020; Myers, Stokes, & Nobre, 2017; van Ede, Chekroud, Stokes, & Nobre, 2019). Therefore, while my data provide clear evidence for the benefit of dimension retro-cues – which is qualitatively similar to, and correlated with, the benefit following item cues – it remains possible that at least part of these benefits are due to factors other than a change in the underlying mnemonic representation (and this holds for both item and dimension retro-cueing benefits).

In addition to the behavioural performance data, I also observed commonalities in the neural modulation following item and dimension cues; both cases showing robust alpha attenuation over posterior electrodes, arising around the same time, with a similar magnitude. The neural responses therefore provide important relevant complementary data to my behavioural performance data. I observe more direct evidence for an early modulation in posterior (putatively visual) brain areas following both types of retro-cues; compatible with a modulation at the level of the memorised visual representations. This may reflect re-weighting of sensory input in order to optimise upcoming memory-guided behaviour (Myers, Stokes, &

Nobre, 2017). However, despite the similar appearance of this modulation following item- and feature-dimension retro-cues, it remains possible that the neural signals I measure reflect the engagement of distinct, but overlapping, brain areas and neural computations (Heuer, Schubö, and Crawford, 2016; just like the recruitment of different neural substrates for item-selection based on spatial vs. feature cues; Heuer and Schubö, 2016a). However, because I used visual retro-cues, I cannot fully rule out the possibility that at least part of this modulation may be driven by differential visual processing of informative versus a neutral retro-cues per se – though I note how the neutral retro-cues were designed to be similar to the informative retro-cues, ruling out more obvious differences due to bottom up visual features such as retro-cue size and saliency. One important difference between the two types of selection is that spatial orienting is inherently linked to item selection and not to dimension selection. Items occupied distinct locations and each dimension was always shared between two segregated items. In future studies it will also be informative to equate spatial factors when comparing item and dimension selection, such as by using multi-feature items that occupy the same spatial location (e.g. overlapping coloured moving dot clouds).

In conclusion, retro-cueing studies have typically shown that internally directed attention can prioritise a subset of mnemonic representations (Griffin & Nobre, 2003; Rerko, Souza, & Oberauer, 2014; Van Moorselaar, Olivers, Theeuwes, Lamme, Victor, & Sligte, 2015). These representations are typically thought of as integrated item of features bound together into a discrete mnemonic item (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). My results show that attention can also

effectively be directed to specific visual dimensions that are shared across multiple items in memory – and for the first time reveal that such dimension cues yield qualitatively similar (albeit weaker) behavioural benefits and neural modulations or latency, as do item cues, and that item and dimension cueing benefits are correlated across individuals. I argue that retro-cues help place memorised visual stimuli into a goal-oriented format, such that relevant information at both the item and the dimension-level can be optimised for upcoming task performance.

3

Decoding Visual Colour From Scalp Electroencephalography Measurements

Contents

3.1 Abstract 60
3.2 Introduction 61
3.3 Methods 65
3.3.1 Participants 65
3.3.2 Visual Display 65
3.3.3 Stimuli and Task Design 66
3.3.4 EEG recording 68
3.3.5 EEG preprocessing 69
3.3.6 Classification on EEG data 71
3.3.7 Statistical Analysis 74
3.4 Results 76
3.4.1 Colour Can Be Decoded From Scalp EEG, and Is Comparable to Orientation Decoding. 76
3.4.2 Colour Decoding Is Driven by Visual Signals With a Contralateral-Posterior Topography. 78
3.4.3 Colour Decoding Is Parametrically Organised 78
3.4.4 Decoding Is Sensitive to Luminance (L*) Differences but Primarily Driven by Difference in Colour (a*b*) 80
3.4.5 Dependence of Decoding on Trial and Participant Numbers 82
3.5 Discussion 85

3.1 Abstract

Recent advances have made it possible to decode visual feature values from various feature-dimensions from presented items from patterns of scalp EEG measurements during perceptual processing. As of recently, such multivariate methods have been commonly used to decode visual-spatial features such as location, orientation, or spatial frequency. In the current chapter, I investigate working-memory encoding of colour and orientation feature-dimensions. I show that it is also possible to track visual colour processing by using Linear Discriminant Analysis on patterns of EEG activity. Building on other recent demonstrations, I show that colour decoding: (1) reflects sensory qualities (as opposed to, for example, verbal labelling) with a prominent contribution from posterior electrodes contralateral to the stimulus, (2) conforms to a parametric coding space, (3) is possible in multi-item displays, and (4) is comparable in magnitude to the decoding of visual stimulus orientation. Through subsampling my data, I also provide an estimate of the approximate number of trials and participants required for robust decoding. Finally, I show that while colour decoding can be sensitive to subtle differences in luminance, my colour decoding results are primarily driven by measured colour differences between stimuli. Colour decoding opens a relevant new dimension in which to track working-memory encoding and visual processing using scalp EEG measurements, while bypassing potential confounds associated with decoding approaches that focus on spatial features.

3.2 Introduction

Our external environment presents us with a large amount of visual information. The mnemonic representations that are formed rely on the principles and limitations of perceptual processing. Working-memory encoding of visual features therefore determines our internal representations on which we draw to guide our future actions. Recent advances in electrophysiological recordings have accelerated our ability to reliably record neural activity during working-memory encoding and ultimately decode feature representations during this process, thereby advancing our understanding of visual processing. EEG and MEG have enabled researchers to measure synchronous activity in large neural populations non-invasively, thus providing a macroscopic readout of brain activity. Whereas extracranial measures of brain activity carry the problem of source ambiguity (Koles, 1998), it is nevertheless possible to compare scalp topography of brain activity to differentiate spatial patterns generated by subtly different configurations of neuronal populations (Stokes, 2015). Recently, studies have shown that the information across an array of electrodes can be used in multivariate analyses to decode perceived visual feature-dimensions such as stimulus orientation, spatial location, spatial frequency, and motion direction (Bae and Luck, 2019; Cichy, Ramirez, and Pantazis, 2015; Foster, Bsales, Jaffe, and Awh, 2017; King and Dehaene, 2014; Myers, Walther, Wallis, Stokes, and Nobre, 2015; e.g., Ramkumar, Jas, Pannasch, Hari, and Parkkonen, 2013; Sandhaeger, von Nicolai, Miller, and Siegel, 2019; Stokes, 2015).

Fewer studies have successfully applied multivariate approaches to decode non-spatial features, such as colour. Decoding of visual-spatial features, which have typically been the focus in EEG decoding studies to date (such as orientation, location, and spatial frequency) is susceptible to contributions of spatial attention and/or eye movements, such as spatial biases in micro-saccades (Engbert & Kliegl, 2003; Hafed & Clark, 2002; Hollingworth, Matsukura, & Luck, 2013; Mostert et al., 2018; Quax, Dijkstra, van Staveren, Bosch, & van Gerven, 2019; Thielen, Bosch, van Leeuwen, van Gerven, & van Lier, 2019; van Ede, Chekroud, Stokes, & Nobre, 2019). Therefore, the extension of decoding methods to features like colour, which are not defined by spatial parameters, provides an important validation of the approach by ensuring that decoding is based on the neural processing of feature-specific content. The demonstration of colour decoding from scalp EEG should therefore have important implications – affording novel ways to track perceptual feature representations in various contexts.

To date, a number of MEG (Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020; Rosenthal, Singh, Hermann, Pantazis, & Conway, 2021; Sandhaeger, von Nicolai, Miller, & Siegel, 2019; Teichmann, Grootswagers, Carlson, & Rich, 2019; Teichmann, Quek, et al., 2020) and, to a lesser extent, EEG studies (Bocincova & Johnson, 2018; Sandhaeger, von Nicolai, Miller, & Siegel, 2019; Sutterer, Coia, Sun, Shevell, & Awh, 2021) have applied decoding approaches to distinguish among colour values upon visual presentation. I have built on these initial studies to further establish and characterise the properties of colour decoding from scalp EEG measurements. For example, it remains to be determined whether the relevant

signals represent the sensory processing of colours and are independent from other associated factors, such as verbal labels or decision-making in tasks, or from subtle differences in luminance that can occur between rendered colours.

Like orientation, colour also follows an organisational structure in visual cortices (Engel, Zhang, & Wandell, 1997; Kleinschmidt, Lee, Requardt, & Frahm, 1996), and research has shown that it is possible to decode colour with functional MRI measurement with millimetre resolution (Brouwer & Heeger, 2009) across a range of cortical visual areas (Brouwer & Heeger, 2009; Persichetti, Thompson-Schill, Butt, Brainard, & Aguirre, 2015). Therefore, finding that colour-related decoding follows parametric variations in stimuli (as opposed to binary or categorical boundaries) in a manner akin to decoding of stimulus orientation is an important step toward ensuring that decoding can reveal colour-specific neural information (in line with Rosenthal, Singh, Hermann, Pantazis, & Conway, 2021). Further reassurance would come from the involvement of electrodes sensitive to visual processing to the decoding, at posterior sites posterior sites contralateral to the decoded stimulus (akin to recent MEG research; Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020). Finally, the ability to decode the colours of multiple stimuli presented concurrently (as opposed to a single central stimulus in isolation, as in most prior colour decoding studies) would greatly add to the flexibility and analytical power of colour decoding as a tool to study dynamic modulation of feature representations in future studies.

In the present study, I therefore aimed to track parametric variations in colour processing in scalp EEG, from multi-item displays, while also qualitatively comparing 'colour decoding' to the more established decoding of stimulus orientation (in term

of robustness, tuning profile, timing, spatial topography, and so on). To do so, I presented 30 participants with two sinusoidal Gabor gratings (one left, one right) that each contained a unique colour and orientation, and I required participants to attend to both features of both items (for a later memory test, which is beyond the scope of the current study). Using Linear Discriminant Analysis, I obtained reliable (and comparable) decoding of the colour and orientation of both stimuli. My data also provide a clear case for a "visual" account of this decoding, by showing that colour decoding has a contralateral-posterior topography, and follows the parametric colour-coding space.

3.3 Methods

3.3.1 Participants

Ethical approval for this study was obtained from the Central University Research Ethics Committee of the University of Oxford. Thirty healthy adults (28.3 years old on average; ranging from 18 - 35 years old; 17 females) were included in this study after excluding two participants due to technical difficulties and incomplete data collection. All participants had normal or corrected-to-normal vision and reported no deficiencies in their colour vision. All participants were reimbursed £15 per hour for their time.

3.3.2 Visual Display

Stimuli were generated using the Psychophysics Toolbox version 3.0.11 (Brainard, 1997) in MATLAB 2014b (MathWorks, Natick, MA). The stimuli were presented on a 27-inch monitor (1920 × 1080, 144Hz; default gamma correction of 2.2). The monitor was integrated in a noise-proof booth with a glass barrier right in front of the display. Viewing distance was 90 cm from the monitor where participants were stabilised at a fixed position using a chin rest. Using a photodiode we established a 24 ms delay between the stimulus triggers received on the acquisition software and the visual presentation of the stimulus on screen. I corrected for this in my analyses accordingly.

3.3.3 Stimuli and Task Design

The critical stimuli for this experiment were presented within the context of a visual working-memory task that ensured attentive viewing of the visual stimuli at encoding. Task details are presented here for clarity, but the analyses only focus on decoding the memory stimuli at encoding (for a full description of the task and participants' behavioural performance, see: **Chapter 2**). At the start of each trial, two Gabor stimuli with a diameter of 4.3° of visual angle (one to the left, the other to the right of fixation, centred at ±4.3° each) were presented simultaneously for 300 ms (**Figure 3.1C**) on a grey background (defined RGB: 0.196 0.196 0.196; measured CIE L*a*b*: 15.18, 3.146, -3.822). The stimuli consisted of bilateral luminance-defined sinusoidal Gabor gratings and each had a unique colour and orientation. Colours were drawn from a CIELAB colour space, which is often used to define perceptually uniform stimulus sets. Forty-eight evenly-spaced colours were drawn from a circle in CIELAB colour space with a fixed lightness of L = 54 (**Figure 3.1A**; centre is at a = 18, b = -8, radius =59).

To get a precise measurement of the obtained luminance of my stimuli on the monitor, forty-eight rendered colours were measured post-hoc using a CRS ColorCal KMII Colorimeter (Cambridge Research Systems, Kent, UK). Luminance and CIE chromaticity (in CIE xyY coordinates) were measured on a uniform colour patch containing only the average colour of the sinusoidal grating as an approximation of the grating colour with varying intensity. CIE xyY coordinates were converted to CIELAB coordinates using a d65 white point (**Figure 3.1B**). Deviations of the rendered colours from the CIELAB colours can be caused through various ways. For

example, CIELAB colours were converted using a linear colour conversion algorithm (Image Processing Toolbox, MATLAB, The MathWorks, Inc., Natick, MA) with an industry standard d65 white point. Without monitor calibration, the sampled RGB colours could deviate from originally defined CIELAB colours. Also, not all colours can be rendered by computer monitors. This is likely to impose small non-systematic imprecisions on the generation of individual colours. Finally, as previously described, there was a glass barrier between the monitor and the participant. The glass barrier could be introducing reflections, resulting in light loss. I added these deviations to my analysis to assess the relative contributions of stimulus rendering offsets and to map out (and rule out) potential alternative explanations for my colour decoding findings (such as decoding of luminance differences, rather than colour differences, between my stimuli).

In total, 48 different colours and 48 different orientations were presented, with feature values ranging from 3.75° to 180° in evenly spaced steps. Orientation and colour values were initially allocated randomly within each block. However, to minimise trials in which the two items had the same or very similar colour or orientation values, I redrew the colour and orientation values of all trials in a block when more than 2 trials were allocated two items with colour or orientation values within the same bin (out of 12 bins). To increase the amount of samples for feature, I binned the 48 colour and orientation features into 12 - equally spaced – bins (**Figure 3.1A, B**). To match the orientation space more closely, I opted for a symmetrical colour probe that included all colours in the $0 – \pi$ range of the probe.

The colour wheel was not rotated from trial-to-trial but was always presented in the same orientation as displayed in **Figure 3.1C**.

At the time of stimulus presentation, participants were instructed to attend to and remember both the colour and the orientation of both gratings. At the end of each trial, participants were probed to report a single feature from one of the two items. Which feature and item would be probed was unpredictable at encoding. During the 2400-ms memory delay, two events occurred: (1) 1000 ms after encoding onset a cue could provide information about the item or feature dimension that would be probed and (2) 1800 ms after encoding onset, a bilateral colour or orientation distractor was presented to distort visual processing. These trial details are not critical to the current investigation, which focuses almost exclusively on the encoding phase of the experiment. More detailed description of the experiment from which I here re-analysed the EEG data can be found online (see published manuscript of **Chapter 2**; Hajonides, van Ede, Stokes, & Nobre, 2020).

3.3.4 EEG recording

The EEG signal was acquired using a 10-20 system using 61 Ag/AgCl sintered electrodes (EasyCap, Herrsching, Germany). The data were recorded using a Synamps amplifier and Curry 8 acquisition software (Compumedics NeuroScan, Charlotte, NC). An electrode placed behind the right mastoid was used as the active reference during acquisition. Offline, data were re-referenced to the average of the left and right mastoids. The ground was placed above the left elbow. Bipolar

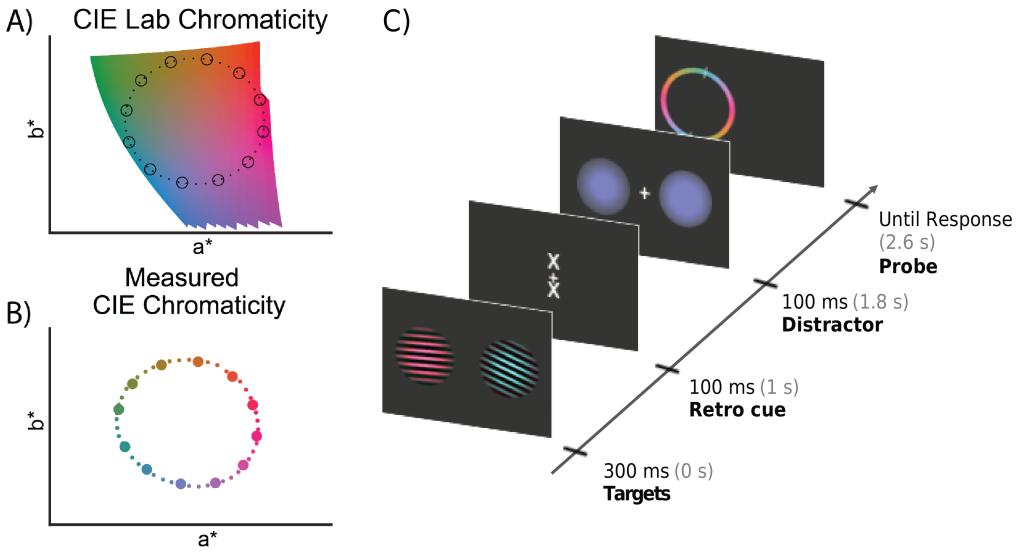


Figure 3.1: Stimulus characteristics and presentation. A) Forty-eight colours - depicted in small dots - were sampled from CIELAB colour space with a fixed lightness ("L") of 54. During analysis, colours were binned into 12 evenly spaced bins, depicted in circles. B) Measured CIE chromaticity a* and b* coordinates on the monitor using the Colorimeter. C) Participants were presented with two Gabor gratings for 300 ms with both colour and orientation information. Colour and orientation were varied independently, using circular spaces with 48 possible feature-values (indicated in panel A and B). Participants were required to reproduce the exact colour or orientation of either stimulus after a brief memory delay, this ensured attentive encoding of the stimuli. The memory delay contained two events: a retro cue, and a visual distractor that consisted of a bilateral colour circle or two identical oriented Gabor gratings. I focus my analyses on the 1000 ms after onset of the gratings. The schematic shows an example trial prompting a report for the colour of the left grating. Stimulus durations are shown next to the tick marks along with in brackets the onset time after the start of the trial. See **Chapter 2** for the full details of the task design.

electrooculography (EOG) was recorded from electrodes above and below the left eye and lateral of both eyes. EEG data were digitised at 1000 Hz with an anti-aliasing filter with a cut-off frequency of 400 Hz. Impedances were kept below 7kΩ.

3.3.5 EEG preprocessing

EEG data were analysed in MATLAB 2017a using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) in conjunction with the OHBA Software Library (OSL;

https://ohba-analysis.github.io/). After importing, the data were epoched from -300 ms to 1000 ms around presentation the presentation of the Gabor gratings (*ft_redefinetrial*) and re-referenced to the average of both mastoids (*ft_preprocessing*). Raw epoched data were downsampled to 100 Hz, and stored for subsequent analyses (analysis scripts and data is made available online at OSF https://osf.io/j289e/?view_only=b13407009b4245f7950960c34a5474a6).

Subsequently, I used the EOG to identify trials in which participants blinked and might have missed the visually presented stimuli. The vertical EOG was baselined between 300 ms to 100 ms before stimulus presentation. Trials that exceeded half of the maximum voltage elicited by blinks (half of 400 μV) within the timeframe of 100 ms prior to and 100 ms post stimulus presentation were marked and later discarded. Additional eye-related artefacts were identified using Independent Component Analysis (ICA; *ft_componentalanalysis*) by applying the FastICA algorithm (Hyvärinen, 1999) to the full EEG layout. Components were rejected if they had a correlation of $r > 0.40$ with the EOG electrodes (on average 2.7 components per participant, and a maximum of 4 components). In addition to trials where participants blinked at the moment of stimulus presentation, trials with high within-trial variance of the EEG broadband signal were removed using a generalised ESD test at a 0.05 significance threshold (Rosner, 1983, implemented in OSL), removing 2.48% $\pm$ 2.18% (mean $\pm$ standard deviation) of all trials.

3.3.6 Classification on EEG data

To focus my decoding on visual processing, I restricted my main analyses on the 17 most posterior electrodes (P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz and O2; Wolff, Jochim, Akyürek, & Stokes, 2017).

To make class predictions, I employed Linear Discriminant Analysis (LDA) using the Scikit-learn toolbox (Pedregosa et al., 2011) and estimated the likelihood for all classes. LDA is a multi-class decoding algorithm that performs well for neuroscientific time-course data (for a comparison of classifiers see Grootswagers, Wardle, & Carlson, 2017).

To increase the sensitivity of my decoding analysis, I capitalised on both spatial and temporal patterns for decoding (Grootswagers, Wardle, & Carlson, 2017; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). To do so, I used a sliding-window approach and concatenated topographical patterns from t_0 up to t_{-19} (20 steps of 10 ms, ranging over 200 ms) into a single vector. This increased the amount of features in the LDA classifier from 17 (17 posterior electrodes) to 340 (17 electrodes $\times$ 20 time steps). To avoid baselining issues and to utilise dynamic activity rather than stable brain states, I subtracted the mean activity across the 20 time points within each time window for each channel and trial separately. This analysis method thus exploits embedded informative temporal variability, reduces noise by combining data over time, and circumvents baselining issues (Grootswagers, Wardle, & Carlson, 2017; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). In a supplementary analysis, I confirmed that similar (be it less robust) colour

and orientation decoding could be obtained when decoding based on purely spatial patterns, on a timepoint-by-timepoint basis.

Principal component analysis was applied to the data to capture 95% of the variance in the data (*sklearn.decomposition.PCA*). Trial feature labels were binned into 12 classes (each subtending $12/\pi$ for orientation and $6/\pi$ for colour; numpy.digitize). These data were split into train-and-test sets using 10-fold stratified cross-validation (*sklearn.model_selection.RepeatedStratifiedKFold*). Next, training data were standardised (*sklearn.preprocessing.StandardScaler*) and after the same standardisation was applied to the test data, I fit the LDA to the training data and labels (*sklearn.discriminant_analysis.LinearDisciminantAnalysis*; using singular value decomposition and a default threshold for rank estimation of 0.001). Likelihoods for each class were estimated after training and testing the LDA classifier separately for each time point, features (colour and orientation from left and right sides), and participant. The resulting likelihoods were centred so that the predicted class was always in centre surrounded by neighbouring feature classes. These predicted-class likelihoods were convolved with a cosine ([- 1/2 π, 1/2 π] for orientation and [- π, π] for colour) to provide a single measure of evidence. This serves to convert two-dimensional tuning curves and obtain a single measure of evidence of each time point. For additional analyses, where I looked at dissimilarity matrices, no cosine convolution was applied. Training and testing were performed separately for left and right features and only at the final stage I averaged across both sides.

To confirm the posterior locus of decodability, I also ran a searchlight decoding analysis (Kriegeskorte, Goebel, & Bandettini, 2006), iteratively considering a small

group of electrodes. To this end, I applied the same temporal classification on the voltages between 150 and 350 ms, but this time only considered the data from a given electrode together with the data from its immediately adjacent (left/right/above/below) neighbours.

Building on RSA (representational similarity analysis; Kriegeskorte, Mur, & Bandettini, 2008), I additionally investigated how the measured differences in luminance (L*) and colour (a*b*) each contributed to the identified parametric decoding of colour. RSA allowed us to examine relative contributions of different neural codes. The preceding analyses assumed a perfectly circular organisation of colours (**Figure 3.1A**) with identical tuning curves for each colour. In practice, rendered colours slightly violate this hypothesised organisation. To test the relative contribution of luminance and colour in my data, I therefore constructed 'dissimilarity models' for both features using the measured Colorimeter data. For measured luminance (L*), a one-dimensional variable, I characterised the difference in luminance of each of the 12 colour bins as the absolute difference of one colour bin with all other colour bins. Doing so for each colour, I obtained a 12-by-12 matrix containing differences of measured luminance of each colour with all other colours with a value of zero across the diagonal. Similarly, I generated a model characterising differences in measured a*b*. Here, I used Euclidean distance between measured a*- and b*-coordinates of each colour bin and all other colour bins as a proxy of colour dissimilarity. A third model acted to compare these two models to the circular model (that used the desired CIELAB colour space as its basis) described above. All models were z-scored. Subsequently, all three models were multiplied

- and subsequently summed - with LDA evidence of every colour bin for each of the 12 possible classifier predictions for every subject, using the spatial-temporal decoding method described above. Higher scores indicate higher similarity between the patterns in the data and those in the model.

3.3.7 Statistical Analysis

For the analysis of the classifier predictions, cosine-convolved evidence for each decoded feature-dimension (colour, orientation) was compared to mean evidence of 40 permutations where the labels were randomly shuffled prior to training the classifier for each time point using a cluster permutation test (mne.stats.permutation_cluster_test; Gramfort, Luessi, Larson, Engemann, Strohmeier, Brodbeck, Parkkonen, et al., 2014) with 10000 iterations and a F-statistic threshold of 2.045 ($p < .05$ with 29 degrees of freedom). Cluster-based permutation tests were also applied to the similarity measure obtained in the RSA analysis. Here, I compared similarity of the measured a*b* model and that of the defined CIELAB model against the luminance model to ask whether I could decode CIE chromaticity over and above measured differences in luminance. For completeness, I also tested whether measured luminance could be significantly decoded, for this I tested similarity scores of the luminance model against similarity scores of a 12-by-12 matrix of z-scored random values multiplied with the data.

For further characterisation of the data, I also ran several analyses with data over the 150-350 ms post-stimulus time-window in which both colour and orientation

decoding were particularly prominent. I first used this window to assess the reliability of the LDA likelihoods for each of the 12 colours and orientations separately. For each bin, I tested decoding against zero using a one-tailed one-sample t-tests, and report both the uncorrected results and the results after applying a Bonferroni correction for multiple comparisons. I also used the data from the same window to investigate the parametric coding spaces for colour and orientation by applying multi-dimensional scaling (MDS; *sklearn.manifold.MDS*; Kruskal, 1964) to a matrix with rows for decodable colour or orientation and columns for classifier predictions. This matrix was inverted so that the diagonal had the smallest likelihoods. I ran MDS with 1000 initialisations and a maximum of 1000 iterations to find the configuration with the lowest stress score for a two-dimensional embedding of the 12 colours and orientations. Finally, I varied the number of participants and trials/participant I included in the analysis. Doing so, I randomly sampled a subset of participants (5-30) and the first n trials/participants (where n ranged from 96 to 912 in steps of 48), I took this approach to simulate having run a shorter experiment. Subsequently, I evaluated the cosine-convolved classifier evidence resulting from spatial-temporal decoding between 150 ms and 350 ms for colours and orientations, averaged over left and right stimulus presentations. To ensure random sampling I drew each subset of participants over 100 permutations and averaged over the results. In contrast to other analyses, a 5-fold cross validation was used to deal with the full range of trial numbers. The resulting matrix was slightly smoothed using a gaussian kernel with a sigma of 0.5 (*scipy.ndimage.gaussian_filter*), where edge artefacts were avoided by replicating the nearest data point.

3.4 Results

3.4.1 Colour Can Be Decoded From Scalp EEG, and Is Comparable to Orientation Decoding.

I extracted EEG data from the 17 posterior electrodes ranging from 300 ms prior to the presentation of the two gratings to 1000 ms after stimulus presentation onset. Average ERP traces revealed similar voltage profiles across the range of tested colours (**Figure 3.2A**, left) and orientations (**Figure 3.2A**, right). An LDA classifier was able to uncover a clear reconstruction of the presented colours and as well orientations (**Figure 3.2B**; averaged over left and right stimulus presentations). To increase sensitivity and to ease visualisation, I reduced these tuning-profiles from **Figure 3.2B**, to a single decodability value per time point, by convolving the tuning profiles with a cosine function (van Ede, Chekroud, Stokes, & Nobre, 2018; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2019; Wolff, Jochim, Akyürek, & Stokes, 2017). Decoding time-courses are depicted in **Figure 3.2C**. I found highly significant decodability for both visual features. Colour decoding first reached significance after 115 ms and remained significant until 875 ms. Orientation decoding first reached significance after 135 ms and remained significant until 945 ms. Colour decoding could also be demonstrated for the processing of the visual distractors, which were not attended and did not require a behavioural report (**Figure B.1B**)

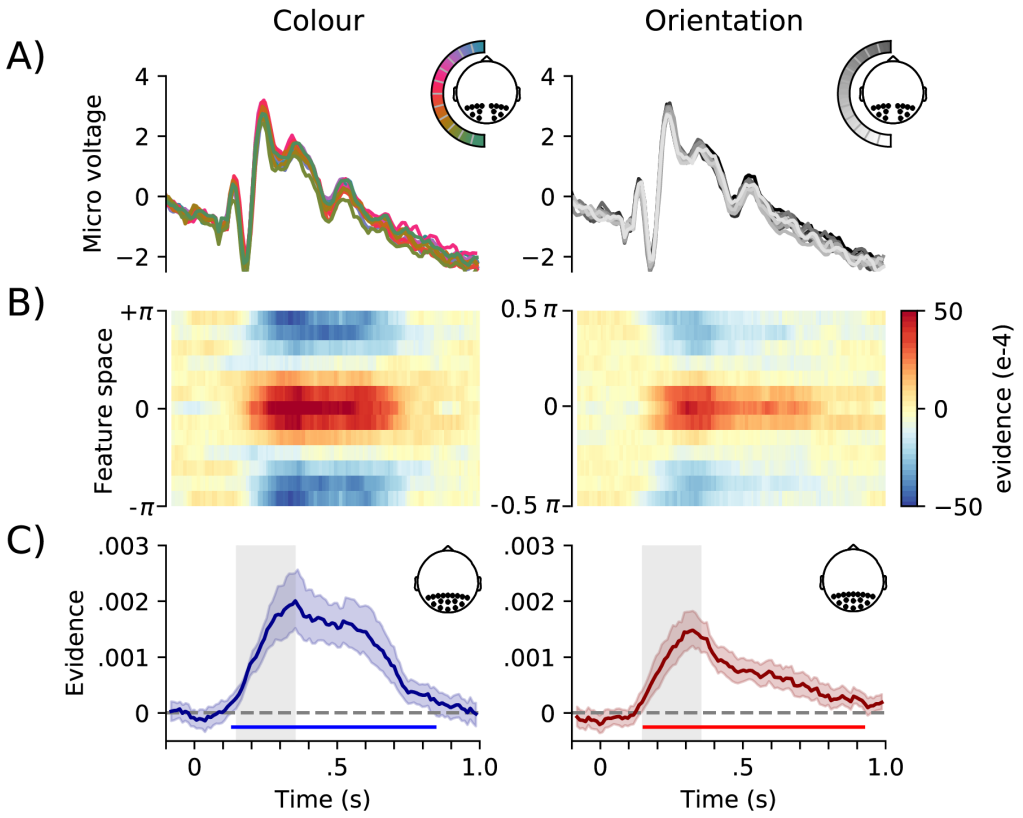


Figure 3.2: ERPs and decoding performance for colours and orientations. A) Event-related potentials for the 7 posterior electrodes contralateral to the presented feature (see inset), averaged over left and right, locked to stimulus presentation. Displaying micro voltage for all twelve different orientations or colours. B) Two-dimensional, decoding tuning curves with the likelihood for target colour or orientation (value "0") and all neighbouring values of theta. For visualisation purposes, a 13^{th} bin was added at the bottom on the diagram mirroring the top bin. These multivariate analyses were done using only the 17 posterior electrodes (see inset panel C). C) Mean cosine-convolved evidence with red lines for colour, blue for orientation. Cluster-permutation corrected significant time points are indicated with horizontal lines. Grey-shaded area between 150 to 350 ms is used for subsequent analyses. These multivariate analyses were done using only the 17 posterior electrodes (see inset panel C). Error bars show 95% confidence intervals, calculated across participants (n = 30).

3.4.2 Colour Decoding Is Driven by Visual Signals With a Contralateral-Posterior Topography.

To understand what contributed to my ability to decode colour, I next ran a searchlight decoding analysis (Kriegeskorte, Goebel, & Bandettini, 2006) across all sensors (van Ede, Chekroud, Stokes, & Nobre, 2019; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). This demonstrated that classifier evidence (in the 150 to 350 ms post-stimulus period in which I found robust decoding; grey-shaded windows in **Figure 3.2C**) peaked in posterior electrodes, with primary contributions from electrodes contralateral to the decoded stimuli. This was the case both for decoded colour (**Figure 3.3A**), and orientation (**Figure 3.3B**). This provides compelling evidence for the "visual" nature of this decoding (as opposed to, for example, decoding distinct "verbal labels" associated with distinct colours).

3.4.3 Colour Decoding Is Parametrically Organised

To test how well each of the feature-value bins could be decoded (relative to all other feature-value bins within the colour and orientation spaces), I analysed mean cosine-convolved LDA likelihoods in the data pooled between 150-ms and 350-ms window that showed robust decoding of both features (indicated in grey in **Figure 3.2C**). Decoding, as a function of the decoded feature-values, is shown in **Figure 3.4A**. For colour bins, all 12/12 colours could be decoded significantly against a $p < .05$ threshold (mean $t_{29} = 4.523, \min = 2.440, \max = 6.175$), and for orientation 11/12 orientations could be decoded significantly against an uncorrected single-tailed threshold (mean $t_{29} = 4.096, \min = 1.585, \max = 7.390$). After applying Bonferroni

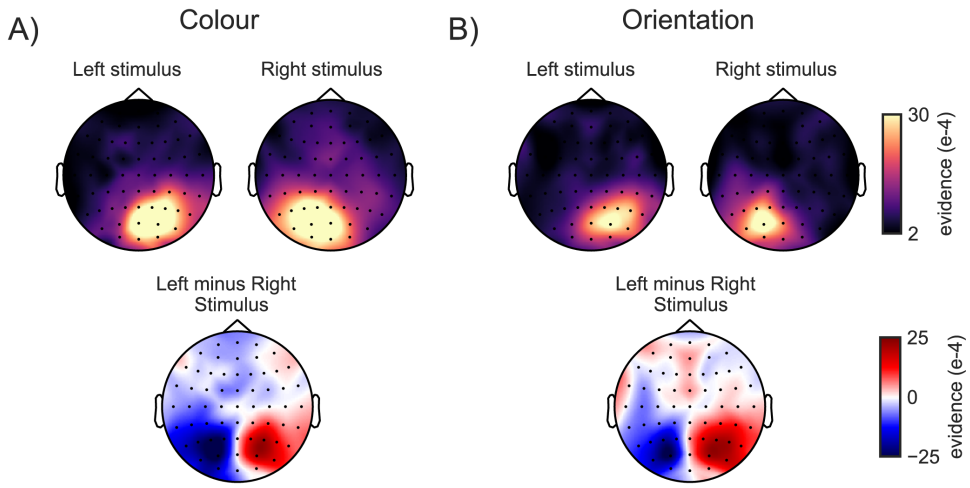


Figure 3.3: Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. Top row topographies show which electrodes show highest mean cosine-convolved evidence for A) colours and B) orientations, of stimuli that were presented on the left or on the right of fixation. Posterior contralateral electrodes show highest evidence. Bottom panels show the difference between left-stimulus and right-stimulus decoding topographies, highlighting the lateralisation of the evidence, and the comparable lateralisation obtained for colour and orientation decoding.

correction, 9/12 colour bins were significant and 8/12 orientations. Using multi-dimensional scaling I revealed the representational space in two dimensions, showing a circular, or oval, configuration for both colour and orientation bins (**Figure 3.4B**). For orientation, the cardinal orientations (0°, 90°, and 180°) showed higher decoding than the off-cardinal decoding. Interestingly, due to some non-linearities in the circular space, I observe some colours (pink, red, orange) in one part of the representational space and other colours (purple, blue, green) on the other (see also Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020)

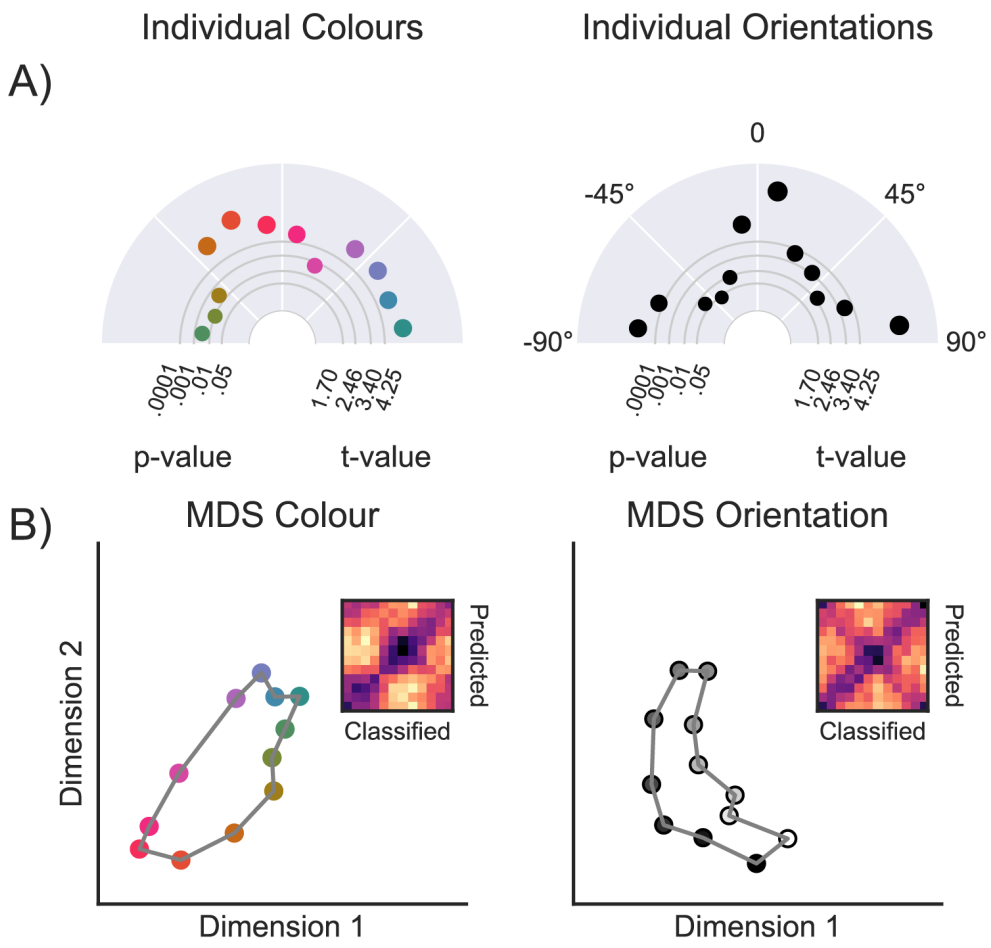


Figure 3.4: Decoding of individual features between 150 and 350 ms. A) This polar plot illustrates the decoding evidence of 12 individual colours and orientations. The distance from the centre depicts the t-value of testing the cosine-convolved tuning curve for a single feature, left and right combined, across participants, relative to zero. The size of the dots mirrors the magnitude of the evidence for that feature B) Two-dimensional visualisation of the similarity matrix in which I observed a circular configuration for colour features. Dissimilarity matrices show feature (colour or orientation) on the y-axis and distance between bins of the tuning curve on the x-axis.

3.4.4 Decoding Is Sensitive to Luminance (L*) Differences but Primarily Driven by Difference in Colour (a*b*)

Colours generated for this experiment were drawn from a plane in CIELAB with identical (desired) lightness, and with equal distances between the colours in the circular colour landscape (**Figure 3.1A**). Nevertheless, because of slight

luminance and colour differences generated during the conversion to RGB colours and rendering process on the monitor, measured colour properties were slightly different. In the next step, I therefore used the physically observed measured luminance (L)* and measured colour (a*b*) to dissect their relative contributions to the identified colour decoding.

To this end, I created three "models" (**Figure 3.1A**) and used RSA to ask how well each of them independently fit the empirical EEG data. I observed that measured differences in luminance (L*; Luminance model in **Figure 3.1A**) between the 12 colour-bins could be decoded from the EEG signal yielding a significant cluster between 115 - 765 ms (**Figure 3.1B**). Critically, however, in addition to the luminance model, I also compared the data to the dissimilarity matrix of the measured colour values (a*b*; **Figure 3.1A**), as well as the (highly similar) CIELAB model used in the previous analyses. This revealed that the colour model and the CIELAB model each fit the observed data significantly better than the luminance model, with significant clusters between 155 - 855 ms and 165 - 735 ms, respectively (indicated by the increased line width in the horizontal cluster-significance lines in **Figure 3.1B**). This is also appreciated by comparing the three models to the empirical dissimilarity matrices, at several time slices, in the bottom of **Figure 3.1B**. These data confirm that the empirical dissimilarity matrix in the EEG data is more similar to the colour and CIELAB models, than to the model constructed from measured luminance values, and that this is so consistently across time. Thus, although unintended luminance differences between my stimuli may have made some contribution, the vast share of colour decoding appears to be driven

by colour(a*b*) differences between my stimuli.

3.4.5 Dependence of Decoding on Trial and Participant Numbers

The current study included 30 participants who each completed over 900 trials. To provide guidance on "how many" participants and trials would typically be required to achieve robust decoding, I ran the following analysis. I randomly sampled x participants (in 100 permutations) and, for each included participant, I incrementally increased the number of trials used for the decoding analysis, always including the first y trials of the session (as if I had only collected data up to this trial). For each combination of # participant by # trials per participant, I assessed feature decoding between 150 to 350 ms after stimulus onset, and plotted the resulting matrix (**Figure 3.6**). Both the number of trials and the number of participants have a strong effect on the reliability of feature decoding. It is evident that both sources of statistical power influence my decoding scores.

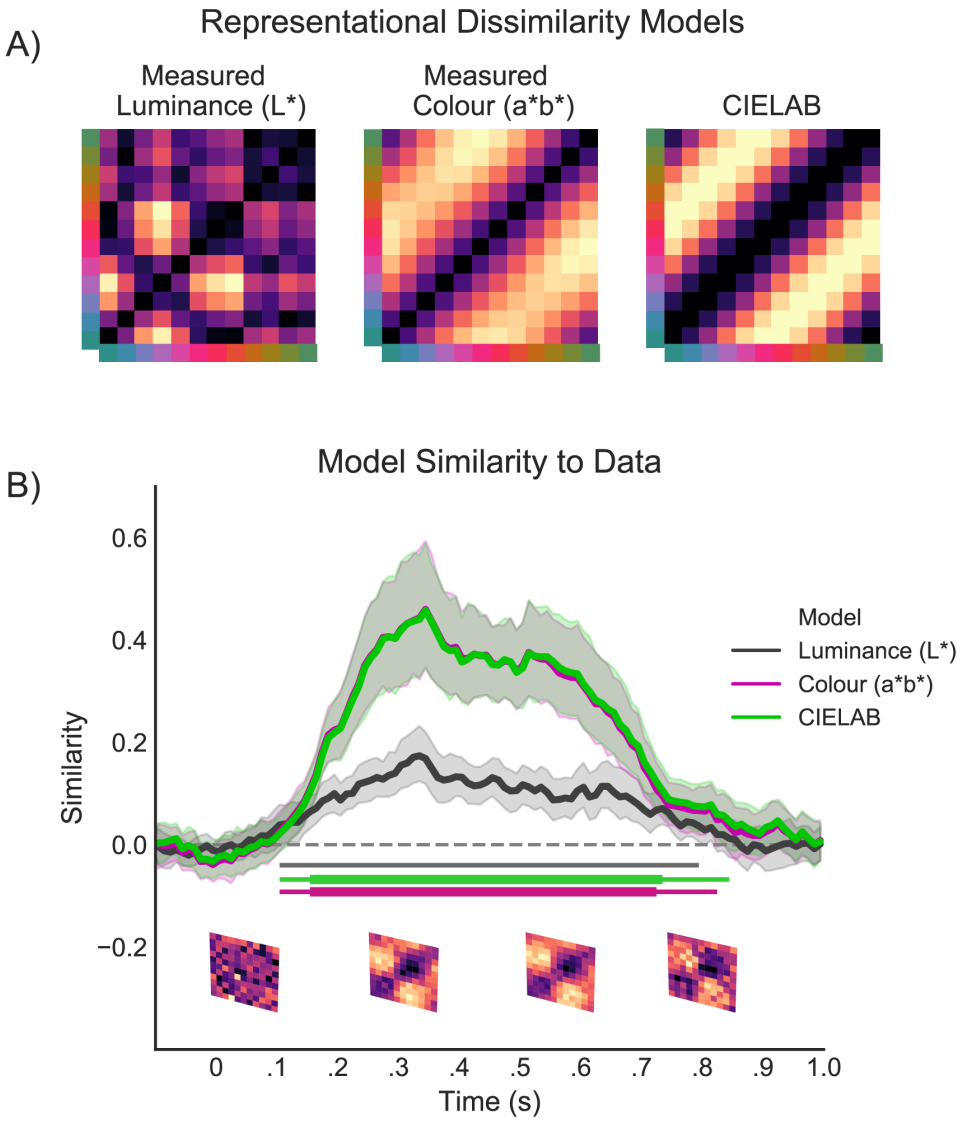


Figure 3.5: Decoding measured luminance and a*b* differences between stimuli. A) Z-scored dissimilarity models representing the measured differences is luminance (L*), Euclidean distances between 12 pairs of measured a*b* values, and a CIELAB model that assumes equal dissimilarity between all neighbouring colours (and that was used in all other analyses). B) Similarity between dissimilarity models and the LDA evidence in the EEG data for each time point between all 12 colour bins. Going from left to right, the matrices below the x-axis show the average LDA evidence between -100 to 150 ms, 150 to 400 ms, 400 to 650 ms, and 650 to 900 ms post stimulus onset. Clusters of significant differences between a random model and measured a*b* (purple), measured luminance model (black) and CIELAB (green) are indicated with the coloured horizontal lines below the similarity time courses. Larger line width of the horizontal green and purple lines indicate cluster-corrected differences with luminance evidence. Error bars show 95% confidence intervals, calculated across participants (n = 30).

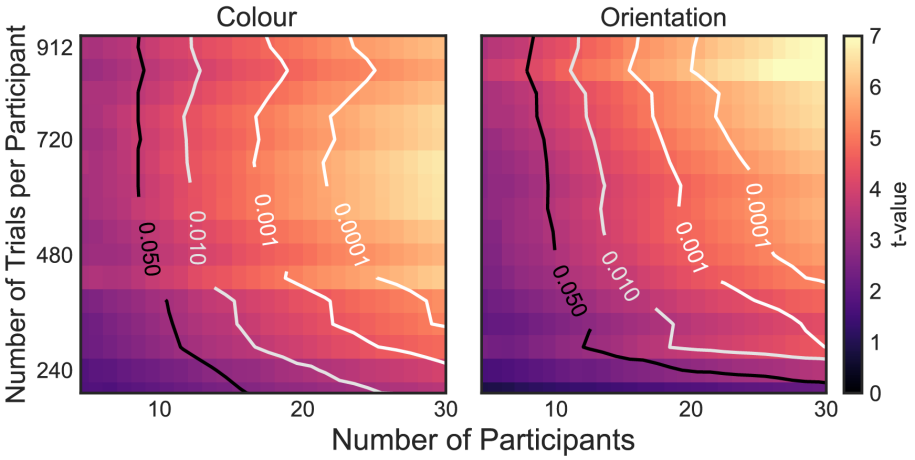


Figure 3.6: Feature decoding as a function of trial and participant numbers. Average cosine-convolved evidence of data between 150 and 350 ms for the average of left and right feature decoding. Contours mark the .05, .01, .001, and .0001 p-value thresholds for n-1 degrees of freedom.

3.5 Discussion

I investigated the ability to decode visual colours from scalp EEG measurements during working memory encoding. I built on other recent studies that have employed colour decoding in scalp EEG (Bocincova & Johnson, 2018; Sandhaeger, von Nicolai, Miller, & Siegel, 2019; Sutterer, Coia, Sun, Shevell, & Awh, 2021) and MEG (Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020; Rosenthal, Singh, Hermann, Pantazis, & Conway, 2021; Sandhaeger, von Nicolai, Miller, & Siegel, 2019; Teichmann, Grootswagers, Carlson, & Rich, 2019; Teichmann, Quek, et al., 2020) and extended these in several ways. I show that I can reliably classify four simultaneously presented visual features – two colours and two orientations – from posterior EEG recordings using multi-class LDA. Follow-up analyses revealed that activity in posterior electrodes contralateral to the decoded stimulus were the primary contributors to the decoding of both features, suggesting that visual sensory processing was the main source of decodable signals, ruling out alternative explanations of colour decoding, such as verbal labelling. Furthermore, colour decoding followed a colour circle in representational space, speaking to the parametric nature of visual colour decoding. Finally, by having both orientation and colour information present in my stimuli, I was able to decode both. This revealed how the decoding of colour is comparable in magnitude, and if anything even clearer, to the decoding of orientation in a visual stimulus.

The current results are in line with, and nicely build on, several other recent non-invasive human electrophysiological studies showing an increase in classifier

evidence following visual presentation of coloured stimuli (Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020; Rosenthal, Singh, Hermann, Pantazis, & Conway, 2021; Sandhaeger, von Nicolai, Miller, & Siegel, 2019; Teichmann, Grootswagers, Carlson, & Rich, 2019; Teichmann, Quek, et al., 2020). One critical difference between the current work and these previously described studies is that I here used a 61-electrode EEG setup rather than a MEG system which typically contains a much larger number of sensors. Though MEG and EEG are sensitive to similar types of activity (but see Lopes da Silva, 2013; Malmivuo, 2012), demonstrating that it is also feasible to decode colour in scalp EEG is an important advance. This is because EEG has lower spatial resolution than MEG and because EEG is more accessible as a technique to labs around the world. Adding to this particular advance, my data further reveal that it is possible to decode parametric colour values in multi-item displays with lateralised stimuli - which allowed us to demonstrate a predominant contralateral topography - and that decoding of colour is not necessarily inferior (at least in my data) to the decoding of stimulus orientation.

To increase sensitivity, in the current work I capitalised on both spatial and temporal activity patterns for classification (see Grootswagers, Wardle, & Carlson, 2017; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). However, to ensure that my results did not critically depend on this choice, I also ran all analysis using a more basic decoding approach whereby I relied only on spatial patterns for decoding; running the decoding analysis separately for each time point after stimulus onset. While this analysis was, as expected, less sensitive, I could nevertheless demonstrate that it was still possible to significantly decode visual colour from purely spatial

patterns in the EEG data; that colour and orientation decoding remained highly comparable in magnitude; and that colour and orientation decoding each had a contralateral-posterior topography and conformed to a parametric coding space (see results in **Appendix B)**.

One limitation of the current study is that the reported colour wheel was fixed, instead of being rotated from trial-to-trial. As a consequence, colour positions on the colour wheel could potentially be predicted by participants ahead of time. However, there are several reasons why I believe my colour-decoding results are not driven by processes related to spatial and/or response anticipation related to the upcoming colour wheel. Foremost, colour decoding was not limited to task-relevant memory content but could also be demonstrated following task-irrelevant distractor stimuli that did not require any behavioural report **Figure B.1**. If colour decoding indeed reflected anticipation of the colour wheel, this should have applied only to the task-relevant memory items, not the distractor. Furthermore, the decoding occurred very early, during the timing of initial sensory coding, and was prominent over electrodes most sensitive to sensory processing (countering potential interpretations related to response preparation). Also, while colour could be decoded well from the broadband EEG signal, I could not establish reliable colour decoding when using alpha power instead **Figure B.6**. This provides additional support against a spatial attention account, provided that alpha is often considered a sensitive marker of spatial attention (Bae & Luck, 2018, 2019; Foster, Bsales, Jaffe, & Awh, 2017; Sutterer, Foster, Adam, Vogel, & Awh, 2018; Thut, Nietzel, Brandt, & Pascual-Leone, 2006; Worden, Foxe, Wang, & Simpson, 2000).

Finally, in my task, colour decoding could be demonstrated when participants were simultaneously encoding four independent features into working memory. Even when considering only the task-relevant memory items, the possibility that the brain would project two orientations and two colours onto four equally probable future response scenarios and do so within 150 ms of stimulus onset in contralateral visual electrodes, appears unlikely.

Perceived colour and luminance depend on many factors, including monitor setup, local stimulus contrasts, and individual biology (Webster, Miyahara, Malkoc, & Raker, 2000). While it is therefore difficult to disentangle fully the respective contributions of each of these factors to my ability to decode colour, my primary aim was to demonstrate its primary source and usefulness in cognitive neuroscientific research. Furthermore, to characterise the relative contributions of subtle difference in luminance from the desired differences in colour empirically, I performed post-hoc measurements of each and used RSA to estimate their relative contribution to my neural colour decoding results. Though I observed significant evidence of luminance differences between the rendered colour stimuli, I also found significantly stronger evidence for the decodability of both the intended (CIELAB) and the physically measured colour differences between rendered stimuli. I further observed a clear parametric coding space that conformed to the circular colour space (that was defined for colour, not for luminance). Furthermore, recent neurophysiological studies described how luminance and hue may modulate brain activity independently (Hermann, Rosenthal, Singh, Pantazis, & Conway, 2020; Sutterer, Coia, Sun, Shevell,

& Awh, 2021), providing further validation that is it possible to decode colour, independently from luminance.

The ability to decode colour, as a reliable proxy of detailed visual processing, is important not only because it opens a new dimension through which to track detailed visual processing on the basis of scalp EEG measurements. It may also prove instrumental to by-pass a fundamental limitation faced by other decodable features such as spatial location, orientation, or spatial frequency. Colour, being a non-spatial parametric feature, should be much less susceptible to confounds associated with visual-spatial processing, such micro-saccades (Engbert & Kliegl, 2003; Hafed & Clark, 2002; Hollingworth, Matsukura, & Luck, 2013; Mostert et al., 2018; Quax, Dijkstra, van Staveren, Bosch, & van Gerven, 2019; Thielen, Bosch, van Leeuwen, van Gerven, & van Lier, 2019; van Ede, Chekroud, Stokes, & Nobre, 2019).

In conclusion, like several other recent studies, I show that colour decoding is possible from scalp EEG measurements. Building on this related recent work, I have now shown that this colour decoding reflects visual processing with a clear posterior-contralateral topography; that it conforms to a parametric colour-coding space; that it is possible in multi-item display; and that it is comparable to the decoding of stimulus orientation. This opens a relevant new dimension in which to track visual processing using scalp EEG measurements.

4

Behavioural and Neural Biases in Visual Working Memory

Contents

4.1 Abstract 93
4.2 Introduction 95
4.3 Methods 100
4.3.1 Participants 100
4.3.2 Experimental set-up 100
4.3.3 Task 100
4.3.4 Behavioural Analysis 102
4.3.5 Mixture Modelling 103
4.3.6 Performance Bias Calculation 104
4.3.7 MEG Acquisition 105
4.3.8 MEG Data Preprocessing 105
4.3.9 LDA Classification 106
4.3.10 Bias Computation 107
4.3.11 Statistical Testing 108
4.4 Results 110
4.4.1 Error Rates 110
4.4.2 Sources of Performance Errors Within Trials 111
4.4.3 Performance Biases Within Trials 112
4.4.4 Attractive Performance Bias Between Trials 113
4.4.5 Task Dependence of Attractive Performance Bias Between Trials 115
4.4.6 Neural Classification of Presented Orientations 117
4.4.7 Within-trial Classification Bias Away From Previous Stimulus 118

4.4.8 Repulsive Neural Biases Between Trials Away From Sensory History 118
4.4.9 Cross-Decoding Evidence for Previously Presented Stimuli 121
4.5 Discussion 125

4.1 Abstract

Our cognitive experience is jointly guided by external sensory signals and internal memory traces. A striking case is how sensory history has a pronounced effect on perception and behaviour inadvertently and even when the past stimuli are no longer relevant for the task at hand. The nature and direction of such serial-dependence biases in performance can differ between experimental settings – both attractive and repulsive biases towards previous stimuli have been observed. How and when these biases arise in the human brain – and whether opposite behavioural biases are associated with opposite neural biases – remains largely unexplored. They could occur either via a change in sensory processing itself, a change in post-perceptual decision-making processes, or both. Here, I analysed behavioural and MEG data from a working-memory task in which participants were sequentially presented with two Gabor gratings with random orientations. A cue, presented together with the second orientation, indicated which of the two orientations had to be reported at the end of the trial. Behavioural responses showed evidence for two distinct biases: 1) a within-trial repulsive bias away from the previously encoded orientation on the same trial and 2) a between-trial attractive bias towards task-relevant orientations on the previous trial. Multivariate classification of stimulus orientation on the MEG data revealed that neural representations during memory encoding of the stimulus were always biased away from the previous grating orientation, regardless of whether I considered the within-trial or the between-trial prior grating orientation – despite opposite effects observed in behaviour. These results suggest that repulsive

biases always occur at the level of sensory processing and can be overturned at post-perceptual stages to result in attractive biases in behavioural performance.

4.2 Introduction

Relevant sensory information from our environment is represented in working memory to guide future actions. Additionally, behavioural performance is modulated by features perceived in the recent past, which are no longer relevant for the task at hand. The incorporation of sensory history is a beneficial strategy within a world that is largely stable over short time scales. This reliance on temporal correlations is deeply engrained in the visual system (Singer & Gray, 1995). Leveraging past sensory evidence can be an effective prior to extract signals from the noisy sensory stream and to help maintain a stable representation to bridge blinks, eye movements, or visual occlusions.

Recent studies have revealed robust short-term influences of prior stimulation in delayed-response and working-memory tasks (Bae & Luck, 2017; Cicchini, Anobile, & Burr, 2014; Cicchini, Mikellidou, & Burr, 2018; Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019; Czoschke, Peters, Rahm, Kaiser, & Bledowski, 2020; Fischer & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017). In these studies, perceptual decision making is systematically influenced by stimulus features presented earlier in the trial (Bae & Luck, 2017; Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019; Czoschke, Peters, Rahm, Kaiser, & Bledowski, 2020; Fritsche & de Lange, 2019; Fritsche, Mostert, & de Lange, 2017), in the previous trial (Cicchini, Mikellidou, & Burr, 2017; Cicchini, Mikellidou, & Burr, 2018; Fritsche, Mostert, & de Lange, 2017; Makovski & Jiang, 2008), or multiple trials back (Fritsche, Solomon, & De Lange, 2021; Fritsche, Spaak, & de Lange, 2020; Gekas, McDermott, & Mamassian, 2019;

Suárez-Pinilla, Seth, & Roseboom, 2018). More specifically, items are reported as more *similar* to the task-relevant stimulus on a previous trial. This attractive bias, sometimes referred to as the *serial-dependency bias* (Cicchini, Anobile, & Burr, 2014; Fischer & Whitney, 2014), has been suggested to induce temporal stability in our perceptual image by acting as a prior that systematically biases present perceptual information towards recent stimuli (Fritsche, Spaak, & de Lange, 2020; Kiyonaga, Scimeca, Bliss, & Whitney, 2017). Conversely, item features presented in sequence within trials are often judged more *dissimilar* to each other (Born & Tootell, 1992; Fritsche, Mostert, & de Lange, 2017; Störmer & Alvarez, 2014). The magnitude of this repulsive bias varies widely with experimental design, contrast, presentation duration, and stimulus size, lasting tens of milliseconds up to multiple seconds (Fritsche, Solomon, & De Lange, 2021; Fritsche, Spaak, & de Lange, 2020; Patterson, Wissig, & Kohn, 2013; Priebe & Lisberger, 2002; Suárez-Pinilla, Seth, & Roseboom, 2018). Within this context, the observed biases could be adaptive for exploiting temporal redundancy and optimising perceptual decision making (Cicchini, Mikellidou, & Burr, 2018; Kiyonaga, Scimeca, Bliss, & Whitney, 2017; van Bergen & Jehee, 2019).

Despite their opposite directions, attractive and repulsive performance biases have been shown jointly to influence task processing (Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019; Fritsche, Mostert, & de Lange, 2017; Fritsche, Spaak, & de Lange, 2020; Sadil, Cowell, & Huber, 2021). Clever experimental designs combined with modelling approaches can reveal both biases within the same setting. Czoschke et al. (2019) investigated both types of performance biases by

sequentially presenting participants with two dot-motion stimuli and subsequently cueing participants to report one of the two directions before moving on to the next trial. A repulsive bias occurred between the two motion directions within trials, with direction recalled further away from the item that was not probed. In contrast, an attractive bias occurred between trials, with responses biased towards the task-relevant motion direction on the previous trial. The use of continuous responses in such working-memory tasks provides an overt means of measuring the feature representation maintained in memory with great sensitivity and can detect systematic response offsets. Yet, little is known about the neural mechanisms of performance biases. From behavioural studies alone, it can be difficult to infer what stage(s) of processing are modulated (since performance measures only provide evidence about the final state of the system, but see Kim et al. 2020).

The current study combined a precision working-memory task for orientation recall and MEG recordings. In each trial, participants viewed two orientations and were cued to reproduce one of them using a continuous response. I investigated the neural dynamics of performance biases resulting from stimulus sequences between and within trials. In addition, I manipulated the task-relevance of grating orientations within trials in order to probe the degree of automaticity of performance biases.

I hypothesised that, in line with Czoschke et al. (2019), I would observe both within-trial and between-trial performance biases. To identify the neural correlates of both types of bias I trained a classifier on MEG sensor data and estimated if the evidence for the presented feature was modulated by stimuli presented on the

same trial or on the previous trial. If within-trial repulsive performance biases originate from modulation of sensory stages of processing, this would be reflected in the classification results. Furthermore, if this repulsive bias is an automatic modulation and not under top-down control, the shift in the representation would be the same regardless of the task relevance of the stimulus feature for the subsequent recall task. Similarly, if the attractive between-trial bias originates at the sensory processing stage, a shift in the neural representation would be expected to occur *towards* the target orientation on the previous trial. Based on previous research (Bae & Luck, 2020; Fischer, Czoschke, et al., 2020), I expected that between-trial biases would be modulated by the task-relevance of the item on the previous trial, such that shifts in representation would be restricted to or significantly larger for previous task-relevant items.

Previewing the results, I confirmed that behavioural responses were systematically repulsed away from previous orientations on the same trial and attracted towards orientations recalled on the previous trial. Mirroring the behavioural repulsion results, the neural representation of a subsequent stimulus was shifted away from the orientation that was first presented on a given trial. This representational bias was only present when the second item was task relevant. However, I found no neural evidence of an attractive between-trial bias during orientation encoding. Instead, there was a repulsive bias away from the target item on the previous trial. This could be caused by lingering sensory information from the previous trial. I speculate that the behavioural attractive bias between trials may

represent modulation at post-perceptual levels of processing, such as during memory maintenance, decision-making, or response preparation.

4.3 Methods

4.3.1 Participants

Twenty healthy volunteers with normal or corrected-to-normal vision participated in the study. All participants were between 20 and 36 years old (mean 25.4 years old; eleven females). Prior to taking part in the study, volunteers provided their informed consent according to the procedures approved by the Central University Research Ethics Committee of the University of Oxford. Participants received £15 per hour compensation for taking part in this study.

4.3.2 Experimental set-up

Participants sat in the MEG scanner, which was situated in a dimly lit, sound-proof, and magnetically shielded room. A projection screen was placed at a viewing distance of 90 cm. Visual stimuli were projected at the back of the screen at a spatial resolution of 1024 × 768 pixels using a refresh rate of 60 Hz using a Panasonic DLP projector (PT-D7700E).

The task was programmed and presented using Matlab (Mathworks, Nantick, WA) in conjunction with the Psychophysics Toolbox (Brainard, 1997). Participants indicated their responses on an optic-fibre response box.

4.3.3 Task

Participants performed a precision working-memory task in which they reproduced the orientation of one of two grating stimuli presented sequentially with independent

orientations (**Figure 4.1**). Simultaneously with the presentation of the second grating, participants were cued as to which grating orientation to report when probed at the end of the trial. On half of the trials only the first or second grating was presented and used for reporting.

Each trial started with a central fixation cross (0.2° visual angle) on screen for 800 ms with a grey background (RGB: 127, 127, 127; **Figure 4.1**). Subsequently, a sinusoidal Gabor stimulus at a random angle was centrally presented on the visual display for 200 ms (diameter of 6° visual angle, 2 cycles per degree of visual angle, 50% contrast, tapered by a Gaussian envelope with a 1.5° standard deviation). On trials where the first grating was not presented the fixation dot changed colour from black to grey (RGB: 192, 192, 192) to signal the omission. After a delay of 1700 – 1900 ms, the second grating was presented. This second grating had the same properties as the first grating except for the angle, which was again randomly drawn independently from that of the first grating orientation. At the centre of the grating, the fixation dot changed colour (orange: (255, 161, 0); cyan: (0, 236, 255)) to signal that either the first orientation (*protect* working memory content) or that the second orientation would be probed (*update* working memory content). Cue-colour contingencies of the experiment were counterbalanced across participants and changed halfway through the experiment. After the colour contingencies switched, participants practised the new contingencies over the course of one block before continuing with the second half of the session. The colour cue was always valid in indicating which grating orientation was relevant for the reporting stage at the

end of the trial. On trials where the second grating was absent the cue would therefore signal the relevance of the first array.

After another delay of 1700 – 1900 ms, a probe grating was presented, which participants were asked to adjust to match the cued grating in memory using the response box. Adjustments were made by pressing buttons with the right hand to rotate the grating either clockwise (middle finger) or anticlockwise (index finger). Responses were confirmed by pressing a button with their left index finger. A 200-ms fixation period followed, after which participants were presented with 50 ms of feedback in the form of a grating indicating the correct orientation.

In total, participants completed 400 trials; 200 trials where two items were presented (100 with update cue and 100 with protect cue) and 200 trials where one item was presented (100 with only grating 1 and 100 only grating 2). The resulting factorial design included four conditions: protect with two items presented (protect, 2 items), protect with one item presented (protect, 1-item control), update with two items presented (update, 2 items), and update with one item presented (update, 1-item control). These were randomly mixed and presented in blocks of 50 trials, with each block lasting approximately 10 minutes.

4.3.4 Behavioural Analysis

Response error was quantified by calculating the circular distance of the recalled orientation to the cued orientation. All responses were mapped onto a -180° to 180° space to compute circular error. In all between-trial analyses, I excluded the first

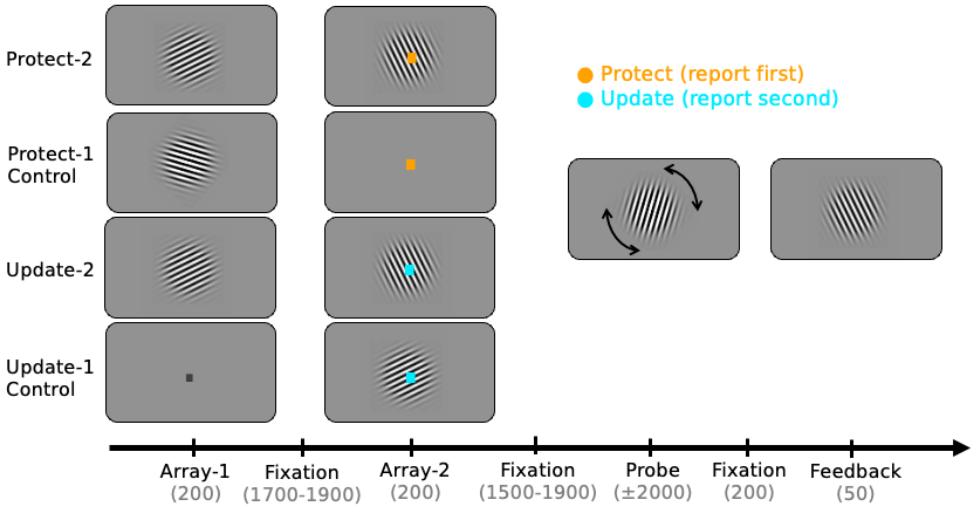


Figure 4.1: Experimental task design. Participants were presented with two arrays that contained an oriented bar grating or a place holder fixation dot. On any given trial, either the first (25% of trials), second (25%), or both (50%) orientations were presented. A colour cue that was presented on array 2 indicated the array that was relevant for the reporting stage later in the trial. The colour cue could indicate as relevant either the first array ("protect working memory content") or the second array ("update working memory content"). On trials where only one orientation was presented the cue was redundant but presented. Note that the colour cue contingencies are displayed for visualisation purposes; mappings between colour and cue meaning were counterbalanced. After matching the orientation in the probe display to the orientation in memory, participants were presented with feedback in the form of a grating with the correct orientation. Timings in brackets are in milliseconds. Cue size in array-2 is enlarged for visualisation purposes.

trial of each block.

4.3.5 Mixture Modelling

To investigate the relative rate of responses to targets, guesses, and erroneous responses to the wrong target ('swap errors'), I fit a classical mixture model frequently used in working memory literature (Bays, Catalao, & Husain, 2009; Schneegans & Bays, 2017; Zhang & Luck, 2008). The model was fit separately for each condition (protect, 2 items; protect, 1-item control; update, 2 items; update, 1-item control). The mixture model estimated the precision of the von Mises

distribution, target response rate, guess rate, and swap rate to the item that was presented on the same trial but not cued (only load-2 trials). The model provided single trial estimates of each of the three parameters.

4.3.6 Performance Bias Calculation

I calculated the performance bias (the circular difference between the response and the target orientation) as a function of the circular difference between the target orientation and the orientation that induced the bias (the previous target orientation on the preceding trial or the task-irrelevant grating on the same trial). I computed the difference between the target orientation and inducer through subtraction, with all angular differences mapped between -90° and 90°. These distances were binned into 64 equally-sized bins. I calculated the average signed error (bias) across all trials within each bin. To reduce noise, I applied a sliding window, computing the average response error over 25% of the bins (16 bins). Edge artifacts were avoided by wrapping the angular differences around, to ensure the average was always computed over 16 bins. Subsequently, I sign-flipped the bias in bins with a negative distance and averaged over negative (-90° to 0°) and positive distances (0° to 90°) between target orientation and inducing orientation, resulting in 32 bins. The integral over absolute distances (0° to 90°) calculated over the 32 bins for each condition and participant served as a measure of bias.

4.3.7 MEG Acquisition

Participants were seated in the MEG scanner after being instructed about the task specifics. They completed one practice block while seated in the scanner prior to MEG recording onset. Participants were instructed to maintain their gaze at the central fixation cross and to minimise blinking throughout the trial.

Neuromagnetic data were acquired using a whole-head VectorView system including 204 planar gradiometers and 102 magnetometers (Elekta Neuromag Oy, Helsinki, Finland) in a magnetically shielded room. Throughout the experiment, participants' head position was monitored continuously using index coils placed at four points on the head. Magnetic field strength was sampled at a rate of 1000 Hz and band-pass filtered on-line between 0.03 Hz and 300 Hz. In addition, vertical and horizontal electro-oculograms were measured using electrodes placed above, below, and adjacent to the eyes. Eye movements were monitored using an EyeLink 1000 (SR Research) eye tracker at a frequency of 1000 Hz.

4.3.8 MEG Data Preprocessing

The data were pre-processed offline using Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), OHBA software library (OSL) drawing on SPM8 (http://www.fil.ion.ucl.ac.uk/spm), and Elekta software. Prior to any preprocessing, the MEG data were visually inspected to remove and interpolate any sensors that displayed excessive levels of noise and were subsequently de-noised and motion corrected using Maxfilter Signal Space Separation (Taulu, Kajola, & Simola, 2004)

before removing independent components related to cardiac and eye-blink artefacts. Data were epoched around the first grating and second grating (from 400 ms prior to grating onset to 900 ms after onset) and downsampled to 200 Hz. Trials with high variance in either gradiometers or magnetometers were identified and excluded using a generalised ESD (extreme studentised deviate; Rosner, 1983) test at a 0.05 significance threshold. This resulted in 7.49% ± 11.55% (mean ± standard deviation) being excluded during preprocessing.

4.3.9 LDA Classification

Data were further pre-processed, where magnitudes of magnetometers were approximately matched to gradiometers by multiplication (factor 20) and subjected to spatio-temporal decoding as described in (Hajonides, Nobre, van Ede, & Stokes, 2021; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020; Wolff, Jochim, Akyürek, & Stokes, 2017). Data from all 306 MEG sensors across a sliding window of 30 time points (150 ms) were concatenated into a vector. Pre-stimulus baselining was not applied to maintain stable information from previously presented stimuli. Dimensionality was reduced for each time point through a principal component analysis, maintaining 90% of the variance (between 250 to 600 ms this was around 209 ± 39 components, mean ± std.). To train an LDA classifier, the data were split in train and test sets using 10-fold stratified cross-validation. Grating angles were binned into 10 equally spaced orientation bins (0° to 18°, 18° to 36°, 36° degree to 126°, 126° to 144°, 144° to 162°, 162° to 180°). For each trial and time point, I thus

obtained 10 LDA distances estimating the likelihood for each of the bins. Tuning curves were constructed by aligning evidence across trials around the same category bin. In cross-decoding analyses, LDA classifiers were trained on orientation bins of one event (e.g., presented grating) but classifier evidence aligned around bins of another orientation (e.g., target orientation on the previous trial). The resulting tuning curves were convolved with a cosine.

4.3.10 Bias Computation

For within-trial biases, I assessed processing of the second grating and only considered load-2 trials. The classifier was trained on all presentations of the second grating and bin likelihoods were generated for each trial. For between-trial analyses, I analysed orientation processing of both the first and second grating. For this reason, I trained the classifier on all trials and generated bin predictions for all trials.

Subsequently, based on the results from the performance-bias analyses, I selected trials where the angular distance between the inducer and the grating orientation on the display led to a significant behavioural bias at the group level. In the case of the within-trial repulsive bias, the inducer was the orientation of the first grating on the same trial; for the between-trial analyses the inducer was the target orientation on the previous trial. As a dependent variable, I considered likelihood estimations for each orientation bin, where I expected the highest likelihood for the angular bin that has zero offset to the presented orientation and decreasing likelihoods for bins with larger angular distances to the presented orientation. I separately assessed

likelihood estimations on trials where the inducer orientation was clockwise (CW) or counterclockwise (CCW) with respect to the current orientation. For both CW and CCW trials I separately averaged the evidence from the orientation bins CW (-72° to -18°) and evidence from the CCW bins (18° to 72°). Symmetry scores were computed by obtaining the difference between the two groups of angular bins (CW minus CCW). Finally, I calculated an overall neural bias score by subtracting symmetry scores on trials with CW vs. CCW inducers. Attractive biases resulted in a positive score (i.e., trials with CW angular distances resulted in more CW evidence, CCW angular distances resulted in more CCW evidence), whereas repulsive biases resulted in a negative score (i.e., CW angular distances resulted in more less CW evidence than CCW trials, and vice versa).

4.3.11 Statistical Testing

Statistical tests were computed using both JASP (Team, 2020) and Scipy (Virtanen et al., 2020).

I tested the time series of cosine-convolved classifier evidence against zero using a cluster-based permutation test (using MNE; Gramfort, Luessi, Larson, Engemann, Strohmeier, Brodbeck, Goj, et al., 2013). I ran 100,000 iterations and a F-statistic threshold of 2.093 ($p < .05$ with 19 degrees of freedom). The clusters where groups of time points were significantly different from zero are indicated in the relevant figures using horizontal lines. Cluster-based permutation testing was also applied to performance bias across angular distance between the presented

orientation and the inducer orientation.

Estimating the shifts in the tuning curves relied on the presence of a tuning curve. Hence, I only used the average of all timepoints where cosine-convolved classifier evidence for the presented grating orientation was significantly above zero in our comparison analyses (in all reported time averages, time points between 250 and 600 ms were used).

All tests were two-sided unless stated otherwise.

4.4 Results

4.4.1 Error Rates

Participants were accurate in reproducing the target orientation (mean response error 11.73° ± 0.70° S.E.M.; mean standard deviation 17.61° ± 1.07° S.E.M., see Table 1 for condition-wise performance). A 2-by-2 repeated-measures ANOVA showed main effects of cue type ($F_{1,19} = 16.49, p < .001, \eta^2 = 0.374$) and stimulus load ($F_{1,19} = 29.78, p < .001, \eta^2 = 0.075$). Cue type was significant for both load conditions, with absolute error higher on protect (1-item control) trials than on update (1-item control; Table **??**; $t_{19} = 3.948$, Bonferroni-corrected $p < .001, d = 0.883$) and higher on protect (2 items) than on update (2 items) trials ($t_{19} = 3.972, p < .001, d = 0.888$). By contrast, load primarily affected the protect conditions. Error was higher on protect (2 items) than on protect (1-item control) trials ($t_{19} = 5.665; p < .001, d = 1.267$) but did not significantly differ between update (2 items) and update (1-item control) trials ($t_{19} = 1.885, p = .075, d = 0.421$), leading to a significant interaction between the two factors ($F_{1,19} = 10.90, p = .004, \eta^2 = 0.026$).

Table 4.1: Descriptive statistics of error for all experimental conditions.

	Protect	**Update**
1 item	11.91° (±4.61°)	8.96° (±2.52°)
2 items	14.69° (±5.38°)	9.79° (±2.89°)

Note. Mean Absolute Error (± Std. Dev.). All units are in degrees.

4.4.2 Sources of Performance Errors Within Trials

To identify the main sources of performance error, I fit a mixture model to the response-error distribution (Bays, Catalao, & Husain, 2009) that identifies three sources of error – variability around the correct orientation, random guesses, and occasional swap errors where the orientation of an incorrect item is recalled. Critically, this model allowed us to estimate the proportion of swap errors made towards alternative targets that would normally be merged into response error and could potentially mask subtle attractive or repulsive biases (Huang, 2020). The data were modelled twice, to investigate swap-error contributions from task-irrelevant orientations on the same trial and target orientations from the previous trial.

Firstly, in load-2 trials, I modelled the contributions of task-relevant (target) orientations and task-irrelevant (swap) orientations. This revealed a high rate of responses based on the target orientation (.903 ± .019, mean ± S.E.M.), low guess rates (.065 ± .01), and low swap rates (.033 ± .01) towards the task-irrelevant grating. Low swap rates between items presented on the same trial indicate participants were unlikely to report the wrong item erroneously. In load-1 trials, the target-response rate was .934 ± .016 and the guess rate was .066 ± .017. Secondly, I modelled contributions of erroneous responses towards the task-relevant orientation on the previous trial. Performance was high, with responses based primarily on the correct grating (.916 ± .017, mean ± S.E.M), with only a low rate of guessing (.057 ± .012) or responding based on the wrong item (.027 ± .007 swap rate). All-in-all, errors originated predominantly from varying precision in responses to the orientation of

the correct target grating.

4.4.3 Performance Biases Within Trials

I analysed within-trial biases in behavioural performance by assessing whether the reported orientation was systematically reported as closer to or further away from the non-target orientation on the same trial. **Figure 4.2A** shows the performance bias for all absolute angular distances between the first and second grating orientation for update and protect trials. After encoding the first grating, participants either protected their memory content from interference of the second grating or updated their working memory to represent the second grating orientation. On trials with protect cues, there was no significant bias towards or away from the interfering second grating orientation that was not relevant to the task at hand ($t_{19} = 0.74, p = .467$). In contrast, trials with update cues revealed significant biases away from the initially encoded first grating orientation ($t_{19} = -2.33, p = .031$; illustrated in **Figure 4.2B**). The repulsive bias on update trials was confirmed using a cluster-based permutation test, showing a significant (cluster-corrected $p < 0.05$) cluster when the angular distance between the two orientations was between 8° and 48° (**Figure 4.2A**). Note that I did not consider load-1 control trials for this within-trial bias analysis, as these trials only contained a single item.

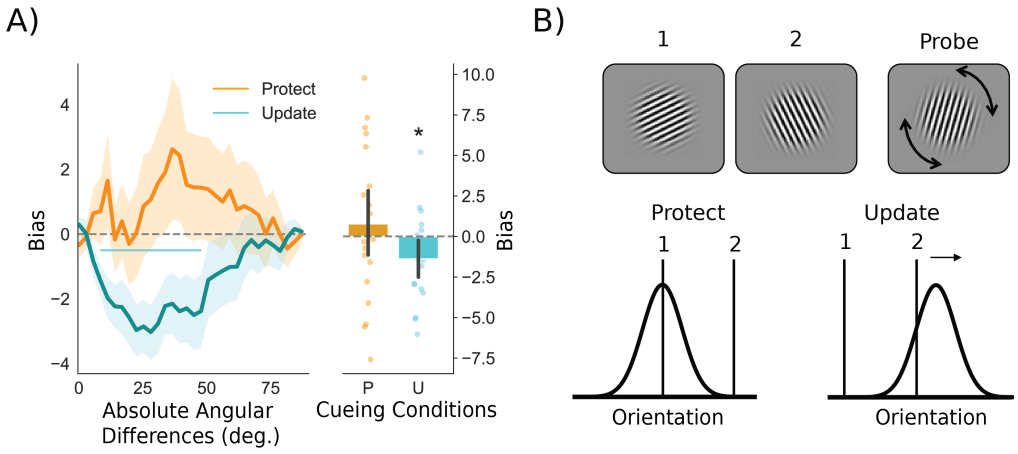


Figure 4.2: Within-trial repulsive performance biases. A) Within-trial bias as a function of the absolute angular distance between the two presented orientations. Trials were split for protect and update cues on load-2 trials. Shading indicates standard error of the mean. The horizontal line indicates a cluster of significant performance biases for a range of angular difference. The right panel shows the sum of integrals per subject. Error bars show 95% confidence intervals. P = protect trial; U = update trial. B) Schematic of the within-trial performance bias. A repulsive performance bias is observed for update trials only.

4.4.4 Attractive Performance Bias Between Trials

I next evaluated the between-trial bias on responses of the current trial towards the orientation that was cued on the previous trial (**Figure 4.3**). Again, I assessed the performance bias as a function of angular differences between the target grating on the current and on the previous trial. The analysis also considered the position of the target grating on the current trial (1^{st} or 2^{nd}) and the stimulus load on the current trial (load 1 or 2). For consistency, I term all trials where participants report the first grating orientation *protect trials* and trials where participants report the second orientation *update trials*, regardless of the number of gratings presented. Again, I quantified the integral of the bias across angular distances between targets on the current and previous trial **Figure 4.3A, B**. In contrast to the repulsive bias described in the previous section, here I found

that all conditions showed an attractive performance bias (all p < .05 in two-sided statistical tests; illustrated in **Figure 4.3E**). A repeated-measures ANOVA on the integrals indicated an effect of cue type, with larger biases occurring in protect trials ($F_{1,19} = 5.706, p = .027, \eta^2 = .172$), but not of the number of gratings presented in a trial ($F_{1,19} = .980, p = .335, \eta^2 = .007$). The two factors did not interact ($F_{1,19} = .377, p = .547, \eta^2 = .002$). This shows that it is not the updating or protecting perse that matters for this bias, but rather the time from encoding of the relevant item on the current trial, since the previous trial.

The magnitude of the performance bias was dependent on the absolute angular difference between the current target orientation and that of the previous trial (**Figure 4.3A**). On protect (2 items) trials, in which participants reported the first of two presented gratings, a cluster of attractive behavioural bias occurred across angular distances (0° – 56°, cluster-corrected $p < .001$). On protect (1-item control) trials I also observed an attractive bias cluster, across 0° – 65° ($p = .001$). Update trials also yielded attractive biases toward the target orientation in the previous trial. When two gratings were presented (update, 2 items), the bias cluster occurred for angular distances between 3° – 56° ($p < .001$). Finally, when the second grating was presented alone (update, 1-item control), a significant bias cluster occurred between 28° – 48° ($p = .045$).

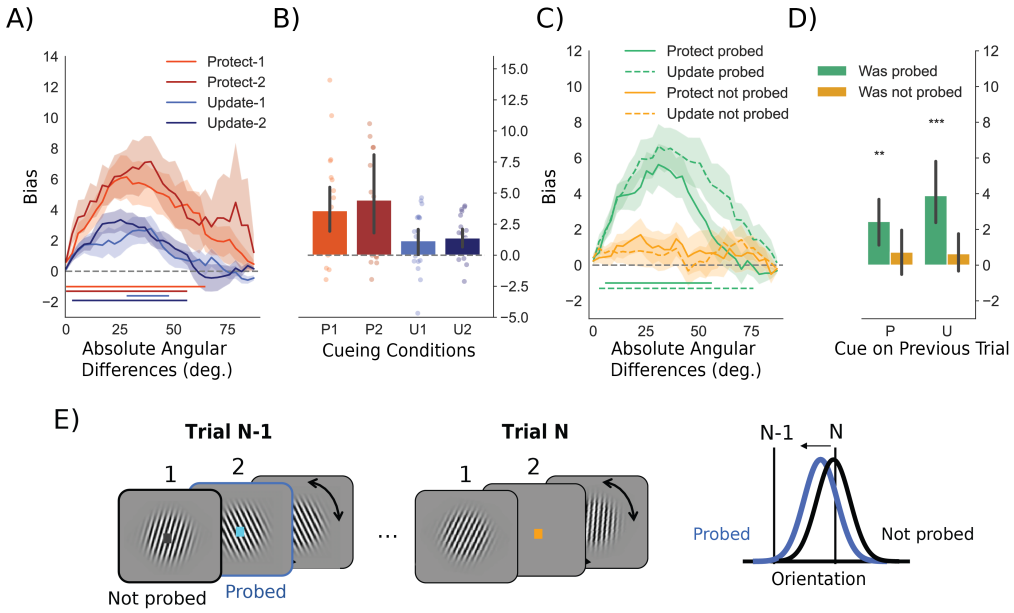


Figure 4.3: Attractive behavioural biases towards the relevant target orientation on the previous trial. A) Line plots for the attractive between-trial bias as a function of the absolute angular distance between the target orientation on the previous trial and the presented grating. Shading indicates standard error of the mean. Horizontal lines indicate clusters of significant performance biases for a range of angular difference. B) Bar plots indicate the average sum of integrals per subject and per condition as seen in panel 4.3A. C) Line plots show the attractive between-trial bias as a function of the absolute angular distance between the probed angle on the previous trial (green) or the presented but not probed angle in the previous trial (orange). The horizontal lines indicate cluster-corrected angular distances for which the bias is significant. D) Integrals of the response data shown in panel 4.3C where positive values indicate an attractive bias. P = Protect cue on previous trial; U = Update cue on previous trial. E) Schematic illustrating the attractive behavioural performance bias between trials. Behavioural responses on trial N were biased towards the orientation that was tested for recall (depicted in blue) and not towards the orientation that was not tested (black) on trial N-1. All error bars indicate 95% confidence intervals.

4.4.5 Task Dependence of Attractive Performance Bias Between Trials

Next, I repeated the same between-trial analyses but investigated the role of task relevance and cue type on the previous trial. This allowed us to test for and compare behavioural biases elicited by the probed (and reported) orientation and by the unreported orientation on the previous trial. I also tested whether the cue type

on the previous trial affected bias on the current trial. I only looked at trials in which two orientations were presented on the previous trial.

A repeated-measures ANOVA on the integral of the performance bias for task relevance and cue type confirmed an effect of task relevance ($F_{1,19} = 14.684, p = .001$; **Figure 4.3D**) but showed no effect of cue type ($F_{1,19} = 1.423, p = .248$) or interaction ($F_{1,19} = 1.633, p = .216$). Task-relevant orientations in trials of either cue type led to a significant bias (previous trial protect (2 items): $t_{19} = 3.524, p = .002$; previous trial update (2 items): $t_{19} = 4.476, p < .001$). Unprobed orientations did not lead to a significant bias (both $p > .2$). No reliable difference was observed between the strength of the bias between update (2 items) or protect (2 items) conditions on previous trials ($t_{19} = 1.691, p = .107$). Following up, I assessed the performance bias as a function of the angular distance between the current target orientation and previously presented orientations. Cluster-based permutation testing showed an attractive bias towards the task-relevant orientation on the previous trial when a protect cue ($p = .001$; 6° – 56°; **Figure 4.3C**) or an update cue ($p < .001$; 3° – 76°) was presented on the previous trial. For unreported orientations no bias was observed (no candidate clusters for protect or update cue). Together, this pattern of results is summarised in **Figure 4.3E**, showing an attractive between-trial bias, but only with regard to items that were relevant in the previous trial.

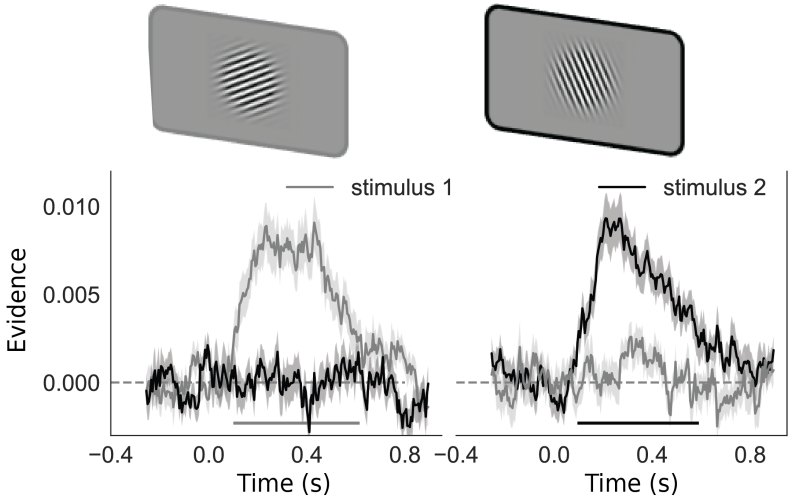


Figure 4.4: Orientation decoding of presented grating. Training an independent classifier to decode the first or second grating orientation in either the first- or second-time interval. The cue in the second display is removed for visualisation purposes. Error bars indicate standard error of the mean. Horizontal lines indicate periods of significant grating orientation decoding after cluster-based permutation tests against zero, p = .05.

4.4.6 Neural Classification of Presented Orientations

I extracted data from all 306 sensors ranging from 400 ms prior to stimulus onset up to 900 ms post stimulus onset. LDA classification reflected significant evidence for the presented orientation after visual onset of the grating **Figure 4.4**. This revealed both significant decoding of the first grating orientation (100 – 615ms; $p <$ .001) as well as the second grating orientation (95 – 590ms; $p <$ .001). A non-significant trend was observed for the evidence of the first grating orientation following the presentation of the second grating (250 – 600ms; $t_{19} = 1.971$; $p =$ .063; see also cross-decoding analyses below). There was no difference in classifier evidence after an update or protect cue (250 – 600ms; $t_{19} = 0.798$; $p = .435$).

4.4.7 Within-trial Classification Bias Away From Previous Stimulus

To investigate how information from the second grating was modulated by information from the first grating, I separately assessed trials with protect and update cues. Doing so, I evaluated the LDA evidence tuning curves elicited by the second grating. For these analyses I again only selected trials where both gratings were presented. I separated trials where the first grating was clockwise vs. counterclockwise relative to the second grating (angular distance of 10° - 50°, based on behavioural results, see **Figure 4.2A, B**). I considered the average of time points between 250 – 600 ms for all future analyses, since in this time window stimulus orientation could be decoded with reliable accuracy (see **Figure 4.4** and grey-shaded area in **Figure 4.5**). Echoing the performance biases, no significant bias occurred on protect trials (**Figure 4.5A, B**; $t_{19} = -0.723; p = .478$). However, on update trials, I observed a repulsive effect, away from the previously encoded grating orientation (**Figure 4.5C, D**; $t_{19} = -3.52; p = .002$).

Notably, there was no correlation, across participants, between the magnitude of the bias in the behavioural and neural data on protect trials ($r = .010; p = .968$) or update trials ($r = -.328; p = .158$).

4.4.8 Repulsive Neural Biases Between Trials Away From Sensory History

To probe for neural between-trial biases, I used the same approach as for within-trial neural biases. I tested the LDA-evidence derived during the stimulus encoding

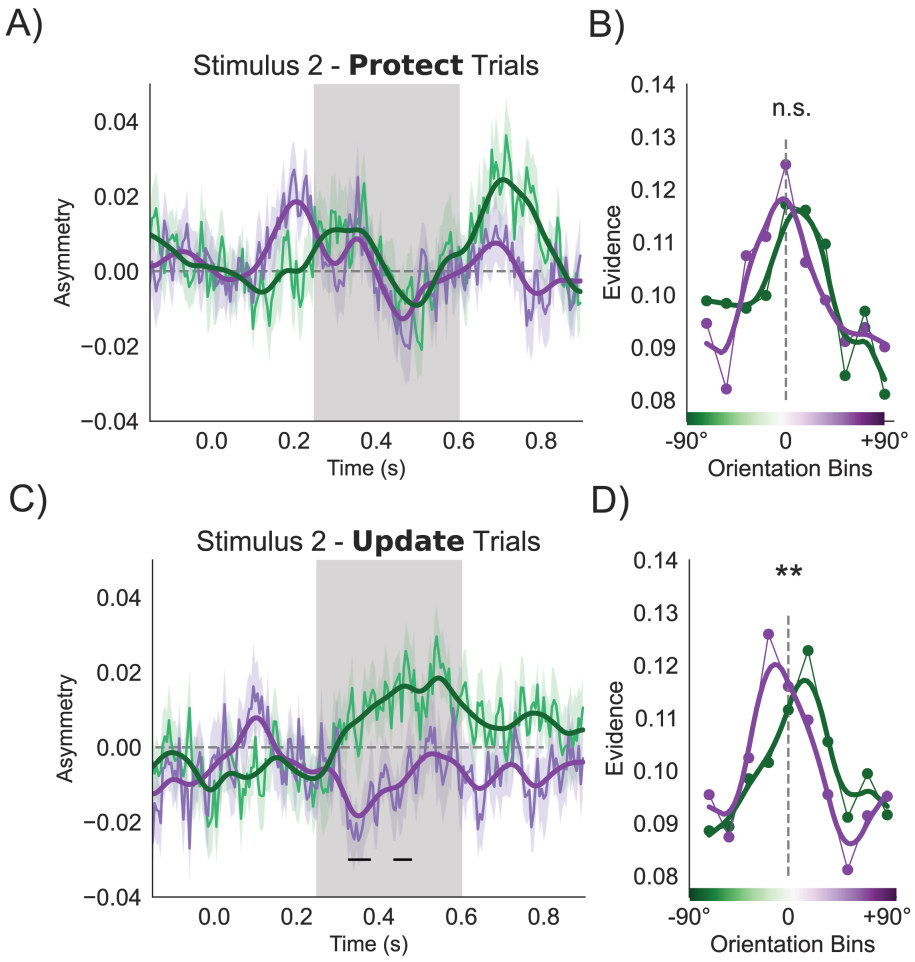


Figure 4.5: Shift in tuning curve relative to previous grating orientation. A) Asymmetry scores following the presentation of the second grating on protect (2 items) trials. The first grating orientation had a CCW (purple) or CW (green) angular distance relative to the current grating orientation. The shaded area indicates the time interval (250 - 600 ms) used for statistical analysis and to generate the tuning curves in the right panel. Smoothing (thicker lines in darker colours) was applied for visualisation purposes. B) shows the tuning curves with evidence for the presented grating orientation for trials with CW or CCW angular distances relative to the first rating. For visualisation purposes, data points were interpolated (from 10 to 50 data points) and fitted using a Savitzky-Golay filter (with a window length of 9 data points and polynomial of order 1). C) Asymmetry scores for update (2 items) trials, where participants encoded the grating orientation on screen into working memory with a bias away from the grating orientation that was presented earlier on the same trial. Shading indicates standard error of the mean. D) shows data from the grey-shaded region in panel 4.5C. Horizontal lines indicate periods of significant bias on unsmoothed data after cluster-based permutation tests against p = .05.

period for systematic deviations in likelihood estimations as a function of the angular distance between the current orientation and the probed orientation on the previous trial. If the behavioural between-trial bias reflects neural modulation during encoding of sensory features, I would expect to see an increase in likelihood estimations for orientations presented on the previous trial in line with the attractive performance bias. I trained the classifier on data following presentation of both the first and second grating orientation combined and included all cue types and loads. Informed by our behavioural analyses on between-trial performance biases, for the test set I selected trials where the previous probe angle had a relative difference of 0° – 60° (**Figure 4.2C, D**; derived from significant angular differences) CW or CCW from the presented grating orientation. The results were qualitatively the same and remained significant when other angular ranges were selected.

Contrary to our expectations, classifier evidence was significantly shifted away from the target orientation on the previous trial (250 – 600 ms post grating onset; $t_{19} = -2.83, p = .011$; **Figure 4.6A, B**) rather than mirroring the attractive behavioural bias. In practice, this would mean that the cued orientation on the previous trial was CW, classifier evidence for CCW bins was increased, and vice versa. The repulsive bias away from the target on the previous trial trended towards significance for the first ($t_{19} = -1.78, p = .090$) and was significant for the second grating ($t_{19} = -2.43, p = .025$) in the current trial when considered separately (**Figure 4.6C**). The between-trial repulsive bias during stimulus-two processing was present if no orientation was presented in the first interval ($t_{19} = -2.69, p = .014$) but not when the first grating was also presented ($t_{19} = -1.68, p = .110$).

Next, I tested whether this repulsive neural bias was affected by the task relevance of the grating and by the cueing condition of the previous trial. I quantified the bias for task-relevant and task-irrelevant grating orientations on the previous trial. Only trials where the previous trial was load-2 were included in this analysis. When averaging the neural bias over 250 – 600 ms, the task-relevant orientation showed a repulsive neural bias (t_{19} = -2.78, p = .012), but I found no neural bias for task-irrelevant orientations ($t_{19} = -0.33, p = .743$), though this difference did not reach significance ($t_{19} = 1.80, p = .088$). Yet, cluster-based permutation testing indicated a significant cluster where task-relevant orientations had a significantly stronger repulsive bias than task-irrelevant orientations (500 – 550ms; $p = .039$).

4.4.9 Cross-Decoding Evidence for Previously Presented Stimuli

If information about the previously presented orientation is still partially present in the visual system during and after the presentation of the current grating orientation, it could interact with encoding. One possibility is that the lingering representation is in an orthogonal representational format, which is different from sensory coding of features (Libby & Buschman, 2021). This means that there is little to no overlap between the activation pattern elicited during sensory input and the pattern related to the lingering representation of that same grating orientation. A classifier trained to separate perceptual information would not cross-generalise if tested on the memory code if this code is orthogonal. Alternatively, the lingering code could be present in a stable representation that shares similarity with incoming sensory information.

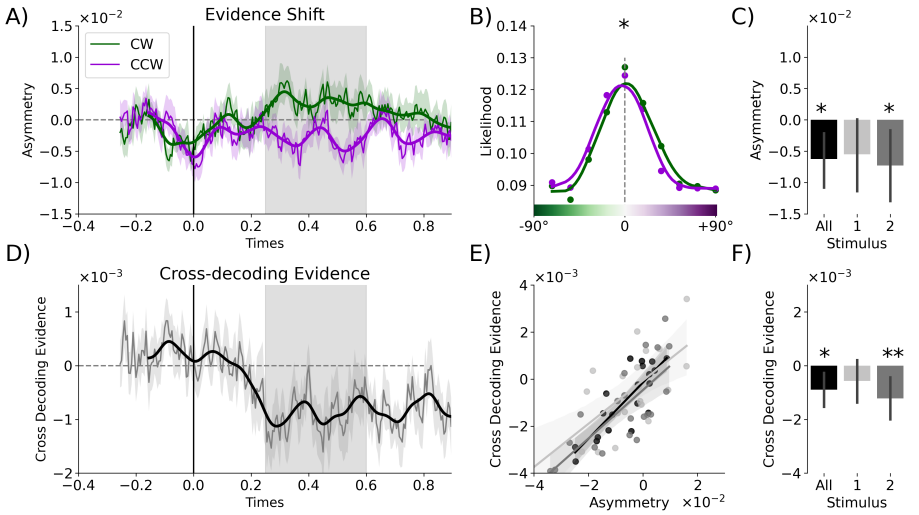


Figure 4.6: Bias on encoding imposed by previous trial. A) Asymmetry index for trials with a CW or CCW angular distance (0° – 60°) between the cued item on the previous trial and the presented grating (collapsed across the first and second grating). Asymmetry index time course is shown relative to the onset of grating presentation. A one-dimensional Gaussian filter with a kernel of 15 ms was applied to the data for visualisation purposes. The grey shaded area indicates which time points (250-600 ms) are used for analyses described in text and in panels 4.6B, C, and E. B) Average LDA orientation-likelihoods for trials with CW (green) and CCW (purple) angular distances with respect to the previous trial. Evidence is lower for CW orientation bins when in the previous trial a CW orientation was cued, and vice versa for CCW orientations. The same interpolation was applied as described in Figure Figure 4.5. C) Average bias relative to the previously cued orientation across all stimuli and for the first and second orientation separately. D) Cross-decoding evidence for the previously cued orientation (classifier trained on current orientation), locked to grating presentation. Same conventions as in panel A. E) Correlations between asymmetry index panel 4.6A and cross-decoding evidence from panel D for grating orientation processing of all stimuli combined (black), first grating (light grey) and second grating (dark grey). A least-squares linear regression model was applied for each of the conditions, with shaded 95% confidence intervals estimated using bootstrapping. F) Cross-decoding evidence from panel 4.6D summarised for all stimuli and separately for the first and second grating, with 95% confidence intervals. *: $p < .05$, **: $p < .01$

If this were the case, the classifier, trained on the data from the second grating orientation, would cross-generalise and identify information about the first grating.

I adapted cross-decoding to detect lingering orientation-selective activity from the previous trial (also see Wan, Cai, Samaha, & Postle, 2020). After training the classifier for the presented orientation and testing for evidence of the previous trial's target orientation, I observed significant negative classifier evidence in the period of 250 – 600 after grating onset (**Figure 4.6F**; concatenating over grating one and grating two: $t_{19} = -2.820, p = .011$; grating one alone, $t_{19} = -1.376, p = .185$; grating two, $t_{19} = -2.951, p = .008$).

Negative classifier evidence indicates that, while information is still present about the previous orientation, orientation-selective patterns may be sign-reversed relative to stimulus encoding. This suppression of evidence for the previous trial's orientation could have been the cause of the apparent repulsive bias in the decoding of the current trial's orientation. If this was the case, I would expect the two measures to be negatively correlated: stronger suppression of the previous orientation (negative classifier evidence) should lead to a stronger repulsive bias for the current orientation. I tested whether the magnitude of the between-trial neural bias and cross-decoding of previous trial were significantly correlated using Pearson correlations across participants. I found a significant correlation when assessing all grating presentations together (**Figure 4.6E**; $r = .832, p < .001$; grating one, $r = .685, p < .001$; grating two, $r = .751, p < .001$).

No correlation was observed between the attractive behavioural performance bias and the magnitude of the repulsive shift in the neural data ($r = .186, p = .432$)

nor with the magnitude of negative decoding ($r = .144, p = .544$).

The cross-decoding analysis and the correlation between cross-decoding and repulsive bias were repeated for the within-trial analyses in the previous section (see **Appendix C**). These results conceptually confirmed the results reported here, although not all analyses reached significance.

4.5 Discussion

The present study investigated biases from previously perceived and memorised information on neural coding and behavioural responses. I observed both neural and behavioural biases that demonstrate interactions between past and present sensory processing. Orientations presented on the same trial exerted repulsive biases on currently perceived orientations. In contrast, task-relevant orientations on the previous trial exerted an opposite, attractive bias, altogether providing evidence for two counteracting biasing processes. Multivariate decoding of neural data indicated that the two types of performance biases may arise from modulations acting on different stages of stimulus processing. Repulsive biases reflect at least in part modulation of sensory processing during item encoding, possibly reflecting mechanisms akin to visual adaptation that could promote visual discriminability. In contrast, the attractive performance bias observed across trials had no equivalent modulation at the sensory level. Interestingly, a neural repulsion occurred instead, suggesting the operation of post-perceptual modulatory mechanisms that can override any early sensory modulation and lead to attractive performance biases.

The present behavioural results provide evidence of two types of performance bias. Firstly, I identified a repulsive performance bias away from the first orientation in a trial. This bias was observed after participants updated their working-memory content to represent a second orientation. There was no retrospective repulsive bias when participants ignored the second grating and reported the first orientation, even though the second grating was presented closer in time to the probe. Secondly,

participants' responses on the current trial were biased towards task-relevant orientations on the previous trial (c.f., Bae & Luck, 2020; Fischer, Czoschke, et al., 2020). The attractive bias appeared strongest for items that were presented closest in time to the previous trial (first grating presentation) but was also present for the second grating, albeit with a smaller magnitude. Together, these results confirm a repulsive within-trial and an attractive between-trial performance bias.

The combination of our task design (grating orientations were sampled from a continuous feature space) and neuroimaging method (MEG) allowed us to decode the neural representation of encoded stimuli both with high fidelity and high temporal resolution, permitting insights about the origins of the observed performance biases. I could reliably decode the orientation of the presented grating and observed a significant shift in the decoded orientation as a function of sensory history. Previously perceived stimuli generated a repulsive shift in encoding of new sensory information, observed shortly after onset of grating presentation. This repulsive bias was only present when the presented orientation was task-relevant and therefore encoded into working memory. By contrast, task-irrelevant gratings that were not encoded into working memory could be decoded with similar precision but did not exhibit a significant bias. Gratings from previous trials that were associated with an attractive performance bias also led to repulsive neural biases during working memory encoding. Only task-relevant orientations led to a repulsive bias. These results suggest a consistent bias that is carried forward from previous sensory processing. This repulsive bias may arise from sensory adaptation, but is not a

purely low-level, automatic process. Instead, the modulation by task-relevance suggests the bias is context-dependent.

Visual adaptation has been proposed as the cause of repulsive neural biases between sequentially presented stimuli that arise automatically during the early perceptual stage (Jazayeri & Movshon, 2006, 2007; Kohn, 2007; Stocker & Simoncelli, 2008; Webster, 2015) . Visual adaptation can be interpreted in terms of down-weighting of neurons sensoritive to the inducer orientation (Clifford, Wenderoth, & Spehar, 2000; Wainwright, 1999). Visual adaptation to previously perceived stimuli can be beneficial for efficient use of resources when discriminating a high number of visual stimuli (Stocker & Simoncelli, 2008; Webster, 2015). This is because neurons can code for a larger range of stimuli when their responses are not saturated. By adjusting the sensitivity levels of neurons to their context and perceptual history, saturation can be reduced, and more information can be transmitted (Webster, 2015). While sensitivity adaptations have predominantly been described in early visual areas for long-duration exposure to the inducer orientation, adaptation has been observed even after very brief exposures (see introduction) and could therefore in part explain the current results. However, visual adaptation is generally thought of as an automatic phenomenon that is not modulated by top-down contextual factors (Kohn, 2007; Webster, 2015). Therefore, it cannot explain fully why I observed a repulsive neural bias only for task-relevant stimuli.

I observed a significant repulsive bias only on update trials, when the presented orientation was task-relevant and encoded into working memory, but not on protect trials, when it was task-irrelevant. This task-dependent modulation was unlikely

the result of reduced processing of task-irrelevant stimuli, as overall orientation decoding was not affected by task relevance. In line with recent studies, it is possible that a context-sensitive repulsive bias, possibly occurring at a post-perceptual stage (Fritsche & de Lange, 2019; Zamboni, Ledgeway, McGraw, & Schluppeck, 2016), exists alongside an early perceptual bias based on visual adaptation (Fritsche, Mostert, & de Lange, 2017). The interaction with task relevance suggests that stimulus processing is only biased when it is primed for later use in upcoming behaviour. Such an interplay of sensory and action systems has been shown to affect the encoding of visual features (Boettcher, Gresch, Nobre, & van Ede, 2021; Myers, Stokes, & Nobre, 2017; van Ede, Chekroud, Stokes, & Nobre, 2019). Circuits involved in the previous response could interact with motor encoding of the subsequent trial. Multiplexed encoding of task-relevant orientations in sensory and motor areas could make feature information less susceptible to degradation and loss of precision, but it may also lead to systematic biases. An alternative cause of the task-dependent neural bias could be that participants compare the relative orientation of the two gratings involuntarily when they are presented in sequence, generating a repulsive bias (Bae & Luck, 2017; Chunharas, Rademaker, Brady, & Serences, 2019; Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019; Czoschke, Peters, Rahm, Kaiser, & Bledowski, 2020). The repulsion could segregate the orientations for better storage of individualised items (e.g. Wei, Wang, & Wang, 2012). Under this explanation, only attended and encoded features would be subject to interactions with previously or presently stored features.

I observed a marked contrast between the attractive performance bias towards the orientation of the previous trial and the repulsive neural bias during encoding. This result appears to be at odds with previous studies that have assigned an early perceptual origin to attractive between-trial biases observed in behaviour (e.g., Cicchini, Mikellidou, & Burr, 2017; Cicchini, Mikellidou, & Burr, 2018; Fischer & Whitney, 2014). By contrast, I failed to find evidence of an attractive modulation of sensory representations towards the orientation of the previous trial at the working memory encoding stage. There have been few studies that have studied brain activity directly and have investigated shifts in neural - and behavioural - representations as a function of sensory history. EEG studies have shown that previous trial information can be decoded during the encoding phase of the current trial (Bae & Luck, 2019) or immediately prior to the current trial (Barbosa & Compte, 2020), and visually evoked ERPs in numerosity judgements are modulated by stimulus history (Fornaciai & Park, 2018, 2019). These results showing lingering effects of stimulus history have been interpreted as the neural basis of the attractive performance bias. But the mere presence of prior stimulus information does not indicate how it will influence processing of the current stimulus. Going beyond mere decoding of prior stimuli, St. John-Saaltink et al. 2016 observed an attractive behavioural and neural bias in early visual areas using fMRI. However, relating this study to the present findings is complicated by the differences in the tasks. The poor temporal resolution of the BOLD response can make it difficult to disambiguate early visual responses during encoding from re-entrant post-perceptual processes influencing responses in sensory areas. Finally, single-unit recordings

from frontal eye fields (FEF) have demonstrated a repulsive neural bias, paired with an attractive behavioural bias (Papadimitriou, White, & Snyder, 2017). Our results therefore suggest that neural repulsive biases observed previously in frontal cortex may generalise to parts of the brain involved in working memory encoding. Overall, systematic investigations of neural biases related to sensory history have been limited and relating this study to the present findings is complicated by the differences in the tasks and experimental methods.

Critically, the present study investigated sensory processing to infer the directionality of the resulting bias (Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020). I tested this using cross-decoding analyses - that train on the presented orientation and predict previous orientations – to show negative or inverted evidence for the stimuli from the previous trial. Congruent with a suppressive bias, patterns related to the orientation presented on the previous trial are downweighted, which implies a relative increase in the pattern strength for all other orientations. In turn, this may shift the tuning curve related to the current orientation away from the inducer, generating a repulsive bias. This relationship was confirmed by the robust correlation between cross-decoding and repulsive bias magnitude.

Altogether, through the neural bias analyses in this study, I demonstrate a consistent repulsive shift in neural evidence during working memory encoding. Our results imply that perceptual adaptation, along with context-sensitive factors, contributes to feature-selective downweighting to exert a repulsive bias away from recent stimulus features. Interestingly, no evidence of an attractive neural bias acting directly on sensory aspects of encoding was observed. Neural data thereby

provide indirect evidence for the post-perceptual account of attractive between-trial biases, rather than modulating encoding stages (Bae & Luck, 2020; Bliss, Sun, & D'Esposito, 2017; Fritsche, Mostert, & de Lange, 2017; Kim, Burr, Cicchini, & Alais, 2020; Pascucci et al., 2019). I speculate that the attractive between-trial bias instead arises through post-perceptual processing stages involving memory (Bliss, Sun, & D'Esposito, 2017; Fritsche, Mostert, & de Lange, 2017), perceptual decision-making, or motor planning (Boettcher, Gresch, Nobre, & van Ede, 2021; Machado, Neto, Bartels, Israelita, & Einstein, 2021; Sadil, Cowell, & Huber, 2021) . The source of the attractive between-trial bias, whatever its neural mechanism, may be strong enough to overcome the repulsive bias that I see during the perceptual/encoding stage.

In sum, I find that the brain integrates stimulus history during encoding and perceptual decision-making. Stimulus history manifests through attractive and repulsive biases. Our neurophysiological recordings highlighted the prevalence of repulsive sensory-processing biases, which are likely to combine with other sources of biases at other processing stages and over different time scales. Together, these co-existing biases may help guide efficient coding for nuanced perceptual discriminations and visual stability across our environment.

5
General Discussion

Contents

5.1 Introduction 134
5.1.1 Adapting Working-Memory Representations for Future Use 136
5.1.2 Use It or Lose It 137
5.1.3 Top-Down Prioritisation of Relevant Features 138
5.1.4 Bottom-Up Modulation of Feature Representations . . . 140
5.2 Future Directions 142
5.2.1 Bottom-Up Modulation of Feature Representations . . . 142
5.2.2 Naturalistic and Dynamic Environments 144
5.3 Final Concusions 146

5.1 Introduction

To interact with our environment in an efficient way, external and internal representations jointly determine goal-directed behaviour. During my doctoral studies, I investigated how feature information in working memory is modulated dynamically through selective attention to guide future actions. The empirical work suggests that working memory is more adaptable than previously thought. More specifically, the chapters in this thesis describe how individual visual features can be tracked during and after working-memory encoding and how this process is modulated by sensory history and task-relevance. Taken together, the body of research in this thesis emphasises the adaptability of working-memory representations, advocating for less of a veridical snapshot representation that would be maintained in memory. Instead, feature representations in working memory are highly adaptable and can be prioritised, dropped, biased, and reconfigured for future use (Nobre & Stokes, 2019). After recapping the previous chapters, I will discuss the relevance of the current findings more broadly and will subsequently highlight developments that I believe are relevant for the future in the field.

In **Chapter 2**, I conducted an EEG study in which I investigated and compared the behavioural and neural correlates of item-based and dimension-based prioritisation within visual working memory. Retro-cue benefits in response times and accuracy, compared to uninformative cues, were similar when observers could selectively attend a multi-feature item (consisting of colour and orientation information) or a feature dimension that was shared across two memory

items. Additionally, desynchronization of alpha oscillations, described as a marker of mnemonic modulation and selective attention, displayed qualitatively similar characteristics between item- and feature-dimension-based selection in working memory. We therefore show that item representations in working memory can be adapted for future tasks – and is not limited to the level of 'integrated objects'. Thus, attentional prioritisation acts similarly on different representational formats stored in working memory.

Expanding on the processing dynamics of multi-feature items, **Chapter 3** analysed EEG data at the working-memory encoding stage. Using novel supervised-learning methods, I demonstrate that colours (and also orientations) of both items can reliably be decoded from visual areas. Predicting 12 classes in a LDA decoding approach (with robust results for nearly all classes in both colour and orientation feature-dimensions), these methods allow for the fine granularity to study subtle shifts in feature representations (as seen in **Chapter 4**).

Chapter 4 further extends on feature processing in working memory by considering factors in a broader context that modulate working-memory representations to serve future tasks. In this empirical study, MEG data were collected while participants remembered an orientation and either *updated* their memory with new sensory input or maintained and *protected* memory content before reporting the relevant orientation. While working-memory content reflected precise feature information to perform the task at hand, participants' responses and neural representations were markedly affected by previously encoded features. In this chapter, I argue that adaptations to feature information from previously

presented stimuli systematically bias working-memory encoding of novel sensory input. Additionally, responses on the current trial were biased to be more similar to responses on the previous trial, I suggest that this second bias is imposed at a post-perceptual stage, overriding earlier sensory repulsive biases. While these biases are maladaptive in laboratory settings with static and isolated stimulus presentations, they are believed to be adaptive in naturalistic settings, where they can guide perception in a relatively stable flux of experience, reduce redundancies, and enable more accurate visual discriminations. Together, these biases shift working-memory representations to adapt to changing circumstances and are modulated by task-relevance.

5.1.1 Adapting Working-Memory Representations for Future Use

Throughout this thesis, I have provided evidence that furthers our understanding of feature representations in working memory and how attention can interact with these working-memory representations. Traditionally, visual working memory is conceptualised as a storage system that maintains veridical visual representations until this information can be used for the task at hand (D'Esposito & Postle, 2015; Luck & Vogel, 1997). This would mean that perceptual information is consolidated and prolonged over an extended period of time until used for the task at hand. Instead, it is suggested that working-memory representations are not bound by truthful depictions of our external environment and are served better by adaptive action-oriented content (Myers, Stokes, & Nobre, 2017; Nobre & Stokes, 2019; van Ede, 2020). These adaptive working-memory representations are subject to a

set of modulatory factors that is largely different from those that act on perception (Myers, Stokes, & Nobre, 2017).

This image of working memory is evolving to include more of the functional purpose and format (Myers, Stokes, & Nobre, 2017; Nobre & Stokes, 2019; van Ede, 2020); working-memory representations are designed to contain task-relevant information that can guide future behaviour (Boettcher, Gresch, Nobre, & van Ede, 2021; van Ede, 2020).

5.1.2 Use It or Lose It

There is an intuitive understanding that the physics and organisational structure of the external environment also applies to internally maintained information. In this thesis, the collection of empirical studies demonstrate how, guided by attentional selection, feature representations evolve to guide task performance and not necessarily organised in veridical fashion to represent the sensory memory stimulus. As described in previous chapters, relevant features are weighted and feature representations are shifted in working memory to suit the task at hand. Thus, working memory representations represent information in a functional way.

This is further illustrated by the capacity limits of internal representations, which require efficient storage of task-relevant features (see **Chapter 1, 2, and 4**). Furthermore, in **Chapter 3 and 4** I observe that, although visual information decoded from neural recordings is initially rich in detail, internal content is rapidly reconfigured to more sustainable neural code and serial information is incorporated

to adapt task performance. These findings highlight the prospective nature of memory (Nobre & Stokes, 2020).

Rather than describing working memory as a collection of sensory representations from the past, it can be better described as a functional state that anticipates and guides future behaviour (Nobre & Stokes, 2019). In line with this perspective, prioritising internally maintained information following a retro-cue can initiate a state that maps better onto behaviour required for the task at hand. In the appendix to this chapter, I describe the plans for an EEG-fMRI study designed to investigate the neural mechanisms associated with the prioritisation of relevant sensory information and concurrent anticipatory response coding. Due to the COVID-19 pandemic, this project was halted and could not be included in this thesis.

When investigating the ways through which sensory working-memory representations are adapted in this thesis, these fall in two categories: 1) top-down projections that signal context, relevance, and expectations; 2) bottom-up adaptations that directly act on lower level regions.

5.1.3 Top-Down Prioritisation of Relevant Features

Top-down modulations of working memory are governed to a large extend by attention. In turn, attentional prioritisation facilitates the strategic weighting of task-relevant features to serve behavioural goals (Heuer & Schubö, 2017; Olivers & Roelfsema, 2020). In my studies, I observed a pronounced effect of task relevance on working-memory content maintained in working memory. This brings about

interesting questions as to how visual features are represented without sensory input and how attention can act on these representations in working memory to serve future behaviour. In visual perception, attention typically spreads across all features of an object (Baylis & Driver, 1993; Desimone & Duncan, 1995; Duncan, Humphreys, & Ward, 1997; Egly, Driver, & Rafal, 1994; Kahneman & Henik, 1981). This facilitates the binding of features into an integrated item, which is encoded into memory (Prinzmetal, Hoffman, & Vest, 1991; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). Because working memory builds on perception, it is tempting to view working-memory content as being limited to a set of integrated items that are each individually maintained (Cowan, 2000; Luck & Vogel, 1997; Pashler, 1988; Pratte, Park, Rademaker, & Tong, 2017; Vogel & Machizawa, 2004; Zhang & Luck, 2008). Indeed, item-based representations in working memory are efficient in that they group features together in an integrated format, (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001), which is adaptive because they are likely to be accessed together in daily scenarios. In **Chapter 2**, the reliable benefit of item-based prioritisation in working memory hints at prominent role of feature integration. Yet, as emphasised earlier, attentional prioritisation of working memory extends beyond item-based representations and also allows for selection of different formats, such as feature-dimensions.

More surprisingly is the similarity between the item- and dimension-based prioritisation in working memory (**Chapter 2**). Prioritising a dimension such as colour or orientation, without prioritising a specific feature value or item, is uncommon in perception. Yet, present results show that dimension-based

prioritisation is strikingly similar to item-based prioritisation in working memory. Although this is informative about the flexibility of attention and the adaptability of working-memory representations, more research is required to corroborate these findings. For example, it would be interesting to investigate whether the results generalise across different experimental conditions, such as other dimensions, challenging set-sizes, and different probe types.

5.1.4 Bottom-Up Modulation of Feature Representations

Previously encoded information in working memory interacts with ongoing perceptual and memory processes. It has been theorised that working memory leaves a short-term trace that is at least in part the result of short-term synaptic connectivity changes to cortical networks (Fitz et al., 2020; Mongillo, Barak, & Tsodyks, 2008). Even though the proposed changes are temporary, they can affect subsequent sensory processing and memory maintenance (like a matched filter; Myers, Walther, Wallis, Stokes, & Nobre, 2015; Stokes, 2015; Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008). The low-level adaptations can affect the content of our internal representations through suppression or facilitation. Critically, short-term changes in neural circuits extend beyond the interval where this is information is directly relevant (Bae & Luck, 2017; Cicchini, Anobile, & Burr, 2014; Czoschke, Fischer, Beitner, Kaiser, & Bledowski, 2019; Czoschke, Peters, Rahm, Kaiser, & Bledowski, 2020; Fischer & Whitney, 2014; Fritsche, Mostert, & de Lange, 2017). In **Chapter 4**, I investigated this influence of sensory history on working-memory representations.

On the one hand, **Chapter 4** investigated the modulatory influence of current working-memory content on processing of sens0ry input. I observed clear evidence of a repulsive bias from information that had been maintained in memory on working-memory encoding of new features. On the other hand, working memory was not biased by irrelevant sensory processing during the maintenance period. Data from **Chapter 2** also did not show an influence of task-irrelevant feature information presented during the working-memory delay on subsequent reports (but see e.g., Rademaker, Bloem, Weerd, Sack, & Rademaker, 2015). In **Chapter 4**, I argue that bottom-up changes in neural circuitry are in part due to low-level adaptations that bias subsequent sensory processing. Yet, these biases are not fully accounted for by adaptation as our empirical results suggest that the bias is more pronounced in task-relevant stimuli than those that are not relevant for the task at hand. This was the case even though overall decoding accuracy of attended and unattended orientations did not differ. This suggests that in this case top-down task-relevance can at least in part regulate the extent to which the proposed bottom-up adaptation acts on sensory processing.

5.2 Future Directions

This thesis investigated a select set of mechanisms through which working-memory representations can be adapted to perform the task at hand. One contribution of the current work is the emphasis on the representational format in which information is maintained in working memory. By expanding on current theory and neurophysiological analysis methods, these results may prompt future research in this field and contribute to a foundation on which further advances can be made. Looking forward, many interesting developments in the field could prompt novel discoveries in the field of attention and working memory. In the sections below, I preview two interesting routes through which experimental work could progress in the years to come.

5.2.1 Bottom-Up Modulation of Feature Representations

The field of cognitive neuroscience has benefitted from fast-paced methodological advances over the past decades. Accelerations are spearheaded by neurophysiological recording techniques that show increasingly better signal-to-noise ratios and increasingly advanced analysis methods that capture more variance and describe the data better. In the present thesis, I describe a set of methodological steps that increase the separation of colour and orientation features embedded in distributed patterns of MEG/EEG measurements and across time. More specifically, the supervised-learning methods described in **Chapter 3** and **4** capitalise on both spatial and temporal information and therefore have an advantage compared to conventional

methods that only take into account pattern-information at a single time point (Grootswagers, Wardle, & Carlson, 2017; Wolff, Jochim, Akyürek, Buschman, & Stokes, 2020; Wolff, Jochim, Akyürek, & Stokes, 2017). I note that such methods are beneficial when analysing signals that are stimulus-locked and thereby displays signal dynamics that are consistent across trials. Yet, these decoding methods are limited in the sense that when this former criterion is violated, classifier performance is impaired. These limitations were particularly apparent in the analyses of data described in **Chapter 2**. In this chapter, I set-out to investigate working-memory representations following the presentation of an item or dimension retro-cue by decoding the pattern information for prioritised and deprioritised features. However, current methods were unable to decode either of these features from working memory. I hypothesise that the prioritisation of features following a retro-cue is subject to a cognitive process that is initiated at a variable time after the presentation of the cue. This causes variation across trials in the time at which the internal feature is attended (Stokes & Spaak, 2016)). In dynamic neural signals, these offsets are detrimental when training a classifier, rendering chance-level accuracy.

To address exactly this issue of temporal variability in the onset of neural representations, Vidaurre et al. (2019) described a temporally unconstrained decoding analysis (TUDA). Using principles of Hidden Markov modelling (Rabiner, 1989; Vidaurre, Smith, & Woolrich, 2017), Vidaurre et al. (2019) describe the multivariate time-series data of each trial using a fixed number of states. These states describe the sequence of processes involved in visual or cognitive events and can capture the variability in onset times of each of these states. A decoder

can then be trained on each of these states, independent from the absolute onset time. Applied to retro-cue epochs, these methods could account for the temporal variance of cognitive processes and decode reactivated visual features. Altogether, this example highlights how more information can be extracted from neural time series, opening the window for further development of computational methods that extract increasingly complex information from neural time series. Ultimately, the methods contribute to a closer description of human cognition.

5.2.2 Naturalistic and Dynamic Environments

In the pursuit of reliable replicable science, experiments are in most cases conducted in controlled laboratory environments. A common critique is that such fundamental research lacks ecological validity and therefore fails to translate into real-life environments (Draschkow, Kallmayer, & Nobre, 2021; Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006; Snow & Culham, 2021). Though instrumental for controlled research studies in this field, the isolated presentation of elemental visual stimuli on computer displays can have limited applicability to other settings. Outside of the laboratory, factors such as situational context, eye movements, and body position are important factors guiding our behaviour (Smilek, Birmingham, Cameron, Bischof, & Kingstone, 2006). In the field of working-memory research, the focus is shifting towards its role in guiding prospective actions (van Ede, Chekroud, Stokes, & Nobre, 2019). The lack of ecologically valid context is particularly relevant in a compound process such as working memory, which relies on sensory experiences

but also on task demands and output formats. Hence, more interactive immersive experiments could tap into a broader spectrum of working memory function and thereby provide a more accurate account of working memory.

While it is highly challenging to conduct controlled studies in complex real-world scenario's, virtual-reality technology (VR) offers great potential for future research (Bohil, Alicea, & Biocca, 2011; Wilson & Soranzo, 2015) and rich experimental settings (e.g., Draschkow, Kallmayer, & Nobre, 2021). The concept of VR offers many benefits over conventional experimental environments. For example, VR headsets can control the full field of view, thereby providing better experimental control over the stimuli perceived by participants and undesired factors can be more easily blocked out. This also favours reproducibility of the experiment across laboratories. Furthermore, the immersive property of VR also facilitates more intuitive interactions with the digital environment (i.e., the experimental task). Finally, the VR setup makes it more feasible to collect metrics such as head rotations, head position, eye-tracking, and three-dimensional motor responses. Altogether, VR is becoming an attractive avenue from an experimental point of view.

In recent years, the potential of VR has resulted in an increased use in scientific settings. VR is already being implemented in a variety of scientific fields to add context and experimental control. Increasing availability of VR for experimental use and ongoing developments of hard- and software make this particularly interesting for cognitive studies, and I expect these methods to gain more popularity over upcoming years.

5.3 Final Concusions

It is well known that working-memory representations are essential in guiding behaviour. However, much remains to be learned about how content is represented – and how it can be accessed from – working memory. In the current thesis, I investigated modulations of feature representations through attentional control and sensory history. Doing so, I analyse and model a diverse set of behavioural responses and neurophysiological recordings. Together, these findings extend our view of how adaptable feature representations in working memory are. This is demonstrated by attentional prioritisation of relevant features or feature-dimensions and read-out of representations and their modulation by lingering sensory history. Apart from theoretical contributions, this thesis also contributes through further developing and describing computational methods that can lay the foundation for future discoveries in neuroscience and psychology. Taken together, this thesis demonstrates the flexibility through which working memory representations can be modulated to serve future behaviour adaptively.

Appendices

A

Supplementary Results for Chapter 2

A.1 Mixture Modelling for Orientation and Colour Dimensions

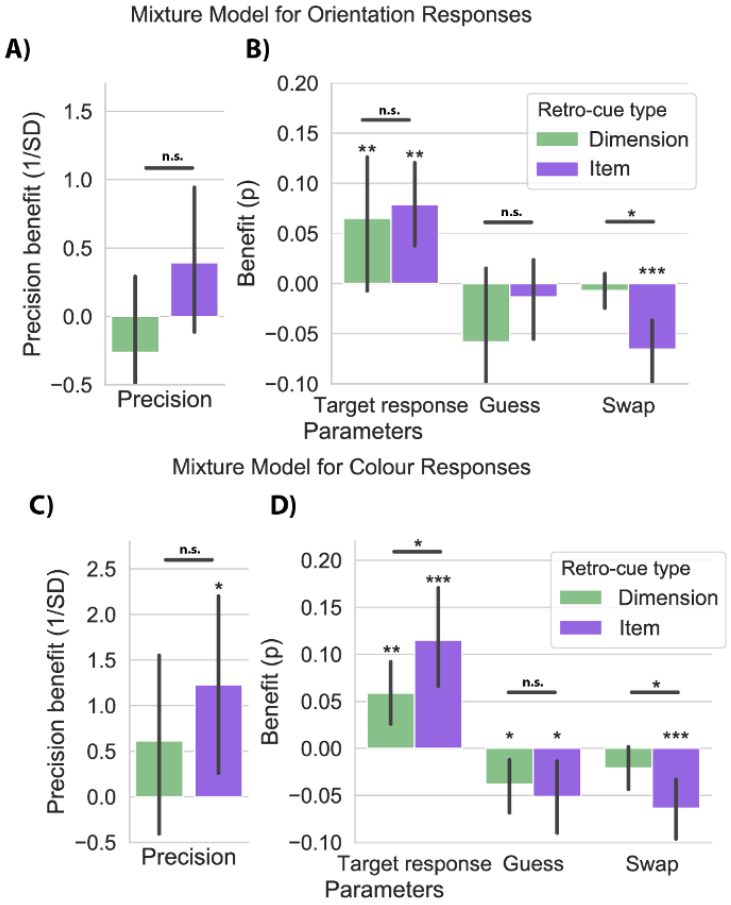


Figure A.1: Mixture modelling parameters for colour and orientation. Mixture-model estimates for the benefit of orientation recall in A-B) and colour recall in C-D). Mixture model parameters include precision (A, C), target response, guess rate, and swap rate (B, D) for trials where item retro-cues or feature retro-cues were presented compared to neutral trials. Black asterisks indicate significant differences between item – and feature retro-cueing benefits. Asterisks above bars represent significance of a two-sided t-test of the model parameter benefit against zero. Error bars indicate 95% confidence intervals. * $p < .05$, ** $p < .01$, *** $p < .001$.

A.2 Modulation of Alpha Following a Retro-cue

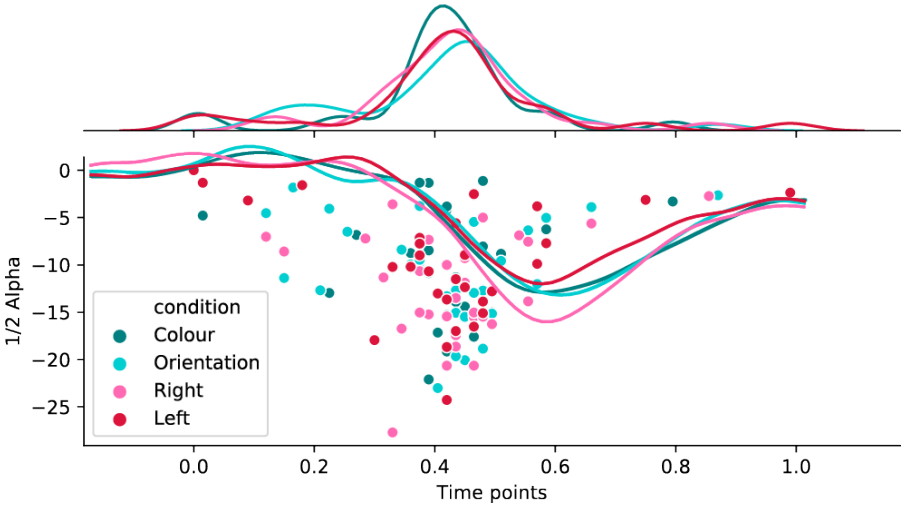


Figure A.2: Onset times of the alpha attenuation after the retro-cue. The first time point where the alpha power (neutral – informative cue) reached half of its minimum value in the interval from 0 to 1000 ms after the retro-cue onset was taken as the alpha attenuation latency. The dots represent alpha attenuation latency times for individual subjects and different cueing conditions with a density plot on the top showing the density of the dots along the x-axis for each condition. The line plot illustrates the average alpha power for each condition after subtracting their relative neutral retro-cues.

A.3 Event-related Potentials Following a Retro-cue

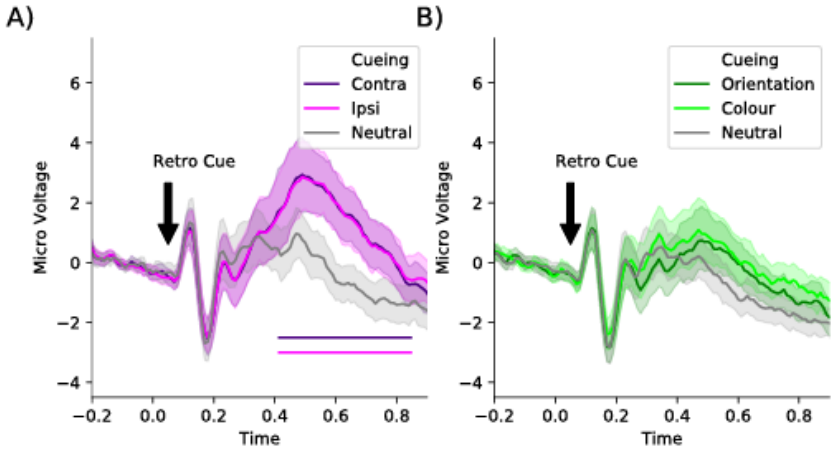


Figure A.3: Event-related potentials following informative and neutral retro-cues. Event-related potentials in left (P1, P3, P5, P7, PO3, PO7, O2) and right (P2, P4, P6, P8, PO4, PO8, O2) posterior electrodes following informative and neutral retro-cues. We plotted A) the average voltages for electrodes contralateral and ipsilateral to the cued item following item-selection cues as well as B) the average evoked potentials collapsed across posterior electrodes following colour or orientation dimension-selection cues. Grey lines show responses to neural (uninformative) retro-cues for comparison. Error bars show 95% confidence intervals. Horizontal lines show clusters of significant differences between informative cues tested against neutral cues ($t_{29} threshold of 2.045, 10000 permutations$).

A.4 Distractor Processing

Main effects and interactions for the congruence or incongruence of the distractor with the probed memory feature, tested with a 2 × 2 × 2 repeated-measures ANOVA with the factors distractor congruence, cue informativeness, and block type, separately for each of our four dependent variables: error in orientation recall **Table ??** or colour recall **Table ??**; reaction times in orientation recall **Table ??** or colour recall **Table ??**.

Table A.1: Within-Subjects Effects for *Orientation Recall: Error*.

	F	**p**	η^2
Distractor	5.515	0.026	0.160
Retro-cue × Distractor	0.256	0.617	0.009
Informativeness × Distractor	0.142	0.702	0.005
Retro-cue × Informativeness × Distractor	0.026	0.872	0.001

Table A.2: Within-Subjects Effects for *Colour Recall: Error*.

	F	**p**	η^2
Distractor	0.286	0.597	0.010
Retro-cue × Distractor	1.290	0.265	0.043
Informativeness × Distractor	0.643	0.429	0.022
Retro-cue × Informativeness × Distractor	0.295	0.591	0.010

Table A.3: Within-Subjects Effects for *Orientation Recall: RT*.

	F	**p**	η^2
Distractor	1.939	0.174	0.063
Retro-cue × Distractor	2.556	0.121	0.081
Informativeness × Distractor	0.549	0.465	0.019
Retro-cue × Informativeness × Distractor	0.652	0.426	0.022

Table A.4: Within-Subjects Effects for *Colour Recall: RT*.

	F	**p**	η^2
Distractor	0.629	0.434	0.021
Retro-cue × Distractor	0.119	0.733	0.004
Informativeness × Distractor	0.698	0.410	0.024
Retro-cue × Informativeness × Distractor	0.023	0.879	0.001

B

Supplementary Results for Chapter 3

B.1 Decoding Distractor Features

[!htb] During the delay period, participants were presented with a bilateral distractor for 100 ms which consisted of two coloured discs (of the same colour) on half of the trials or two oriented Gabor gratings (of the same orientation) on the other half. Sixteen evenly spaced colour and orientation features were defined for the distractor, which were binned into 8 bins for the following decoding analyses to ensure a similar number of trials per feature. I epoched the data from 400 ms prior to distractor onset to 800 ms post distractor onset. The same spatial-temporal decoding methods described in the manuscript were applied to the neural data. Classifier evidence for colours was an order of magnitude higher than that observed for items in the encoding frame. This may be due to the fact that the distractors only contained a single feature that was identical for the left and right side.

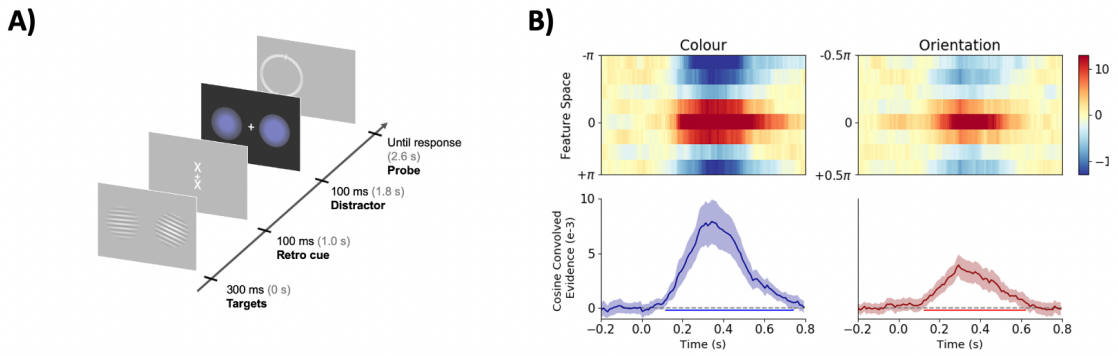


Figure B.1: Decoding the colour and orientation of the visual distractors. A) On every trial, participants were presented with a bilateral array of distractors that consisted of a single feature. On a randomly selected half of trials this was a colour and on the other half of trials this was an orientation. The decoding analyses for the distractor epoch were carried out in the same way as in the encoding epoch, colour and orientation features were binned into 8 stimulus bins and decoding analyses were applied in an identical fashion for the distractor epoch. B) Mean cosine-convolved evidence with red lines for colour, blue for orientation. Two-dimensional, decoding tuning curves with the likelihood for target colour or orientation (value "0") and all neighbouring values of theta. Error bars show 95% confidence intervals, calculated across participants (n = 30).

B.2 Classification Using Single Time-point Data

B.2.1 Decoding Time Course

When applying LDA to the data from the 17 posterior electrodes for each time point, I did not reduce the dimensionality using PCA. Other preprocessing steps are identical. Single time-point LDA decoding shows that colour decoding was first significant after 125 ms, until 595 ms (**Figure B.2**). Orientation decoding was only significant after 185 ms, until 625 ms.

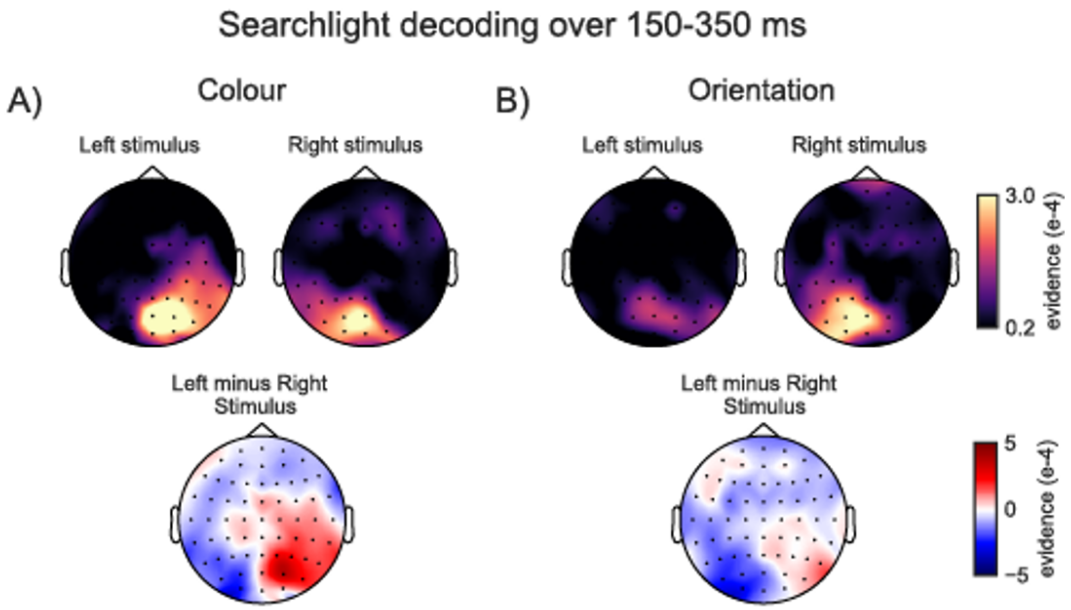


Figure B.2: Decoding performance for colours and orientations with single time-point decoding. A) Two dimensional tuning curves with the likelihood for target colour or orientation and all neighbouring values of theta. B) Mean cosine-convolved evidence with red lines for colour, blue for orientation. Cluster-permutation corrected significant time points are indicated with horizontal lines. Grey-shaded area between 150 to 350 ms is used for subsequent analyses. Error bars show 95% confidence intervals.

B.2.2 Localisation of Feature Decoding

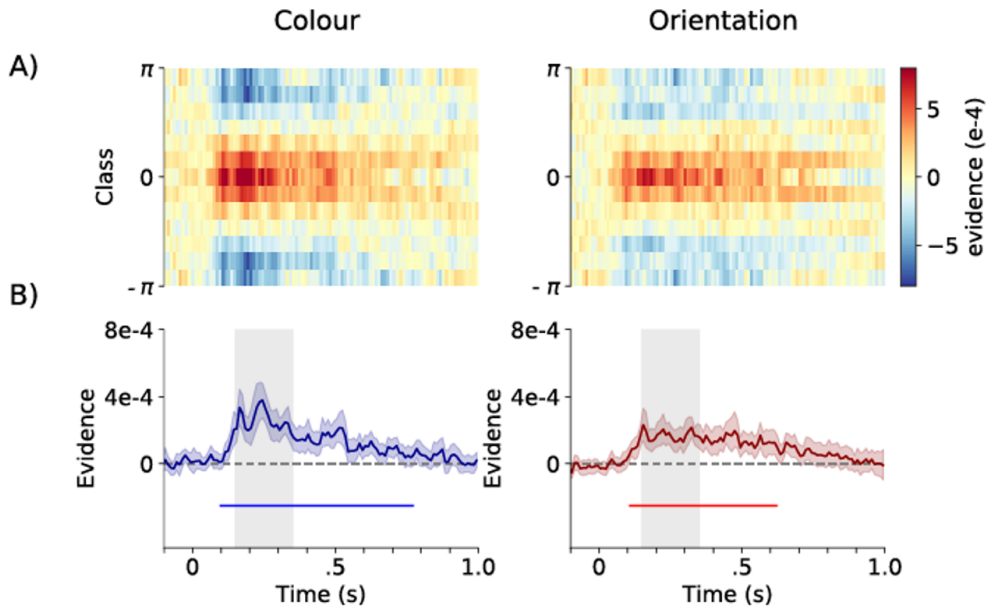


Figure B.3: Colour and orientation decoding are primarily driven by posterior electrodes contralateral to the decoded stimulus. After averaging over data within the 150- to 350-ms time window, I ran a searchlight analysis using only the spatial pattern (in contrast to **Figure 3.3** where I also incorporated spatial-temporal information). Top row topographies show which electrodes show highest mean cosine-convolved evidence for A) colours and B) orientations presented on the left and on the right. Posterior contralateral electrodes show highest evidence. Bottom panels show the difference between left and right feature decoding topographies, highlighting the lateralisation of the evidence.

B.2.3 Individual Feature Decoding

Instead of concatenating data over the window of 150 ms until 350 ms post stimulus onset, I averaged data into an average pattern and proceeded with the same analyses as depicted in (**Figure 3.4**). For colour bins, 11/12 colours could be decoded significantly against a $p < .05$ threshold (**Figure B.4**; mean $t_{29} = 3.391$; min $= 0.530$; max $= 6.040$), and for orientation all 6/12 orientations could be decoded significantly against an uncorrected single-tail threshold (mean$t_{29} = 2.436$; min $= 0.530$; max $= 4.561$). After applying Bonferroni correction, 5/12 colour bins were significant compared to 4/12 orientations.

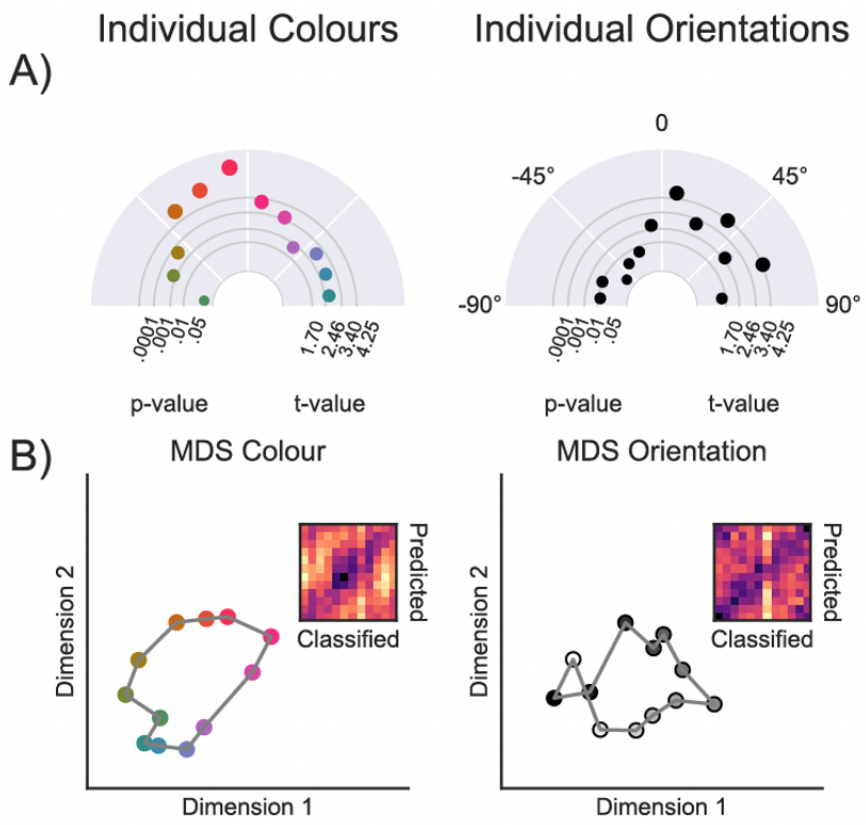


Figure B.4: Decoding of individual features between 150 and 350 ms. After averaging over data within the 150- to 350-ms time window, I ran a searchlight analysis using only the spatial pattern (in contrast to **Figure 3.3** where I also incorporated spatial-temporal information). Top row topographies show which electrodes show highest mean cosine-convolved evidence for A) colours and B) orientations presented on the left and on the right. Posterior contralateral electrodes show highest evidence. Bottom panels show the difference between left and right feature decoding topographies, highlighting the lateralisation of the evidence.

B.2.4 Trial and Subject Numbers for Decoding

B.2.5 Feature Decoding Using Alpha Power

To investigate the contribution of alpha-band power to my current results, data were bandpass filtered between 8 – 12 Hz and submitted to a Hilbert transform to compute the time-resolved magnitude of the complex analytical signal. The resulting alpha-power estimation was passed through the same classification procedure described in the methods section of the manuscript. While I did find one relatively weak

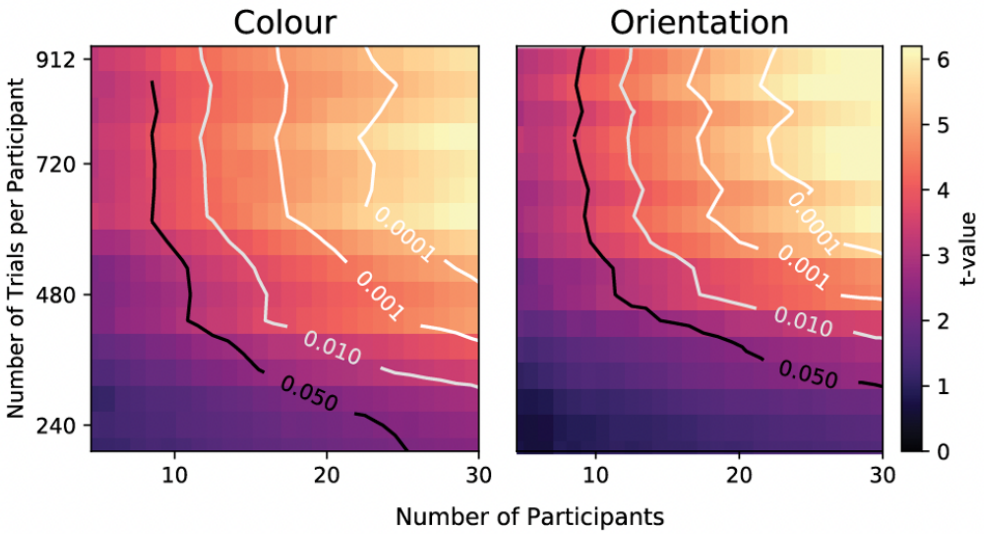


Figure B.5: Feature decoding as a function of trial and participant numbers. Average cosine-convolved evidence between 150 and 350 ms for the average of left and right feature decoding. Contours mark the .05, .01, .001, and .0001 p-value thresholds for n-1 degrees of freedom.

significant cluster in the single-timepoint decoding data relative to stimulus encoding (385 – 999 ms; **Figure B.6B**, left), this effect occurred much later than the decoding from the broadband signal. Moreover, unlike the broadband decoding results, this effect did not replicate across the other three analyses depicted in **Figure B.6**.

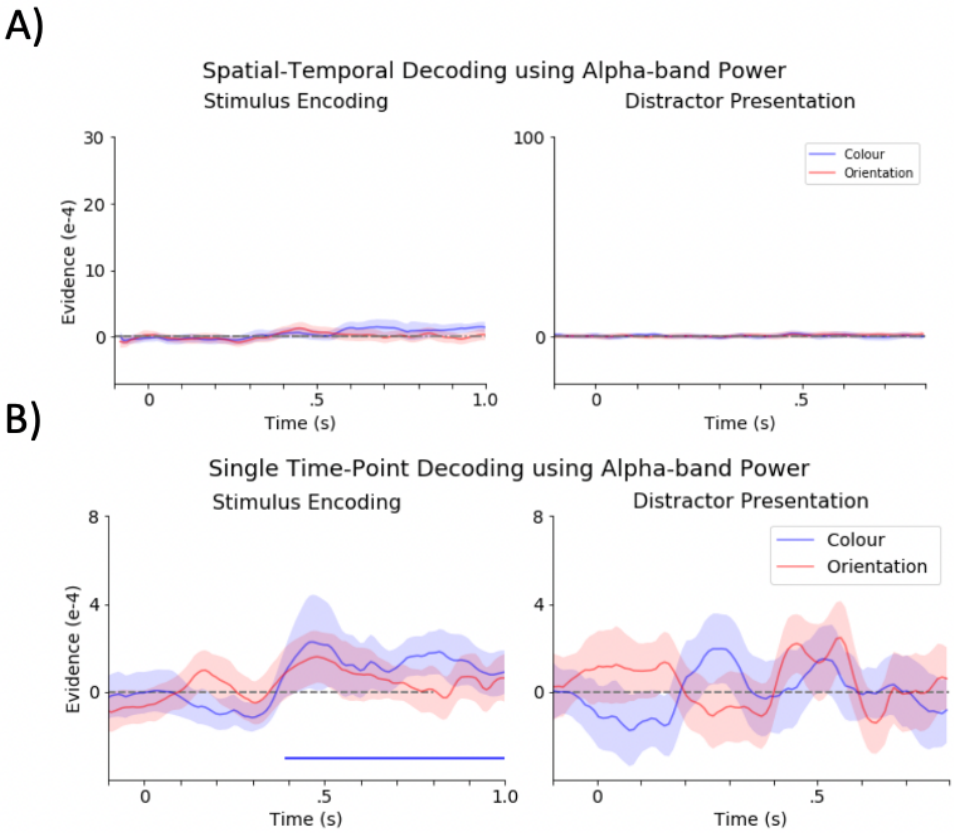


Figure B.6: LDA classification of colour and orientation using Alpha-band power in the stimulus encoding and distractor presentation epoch. A) Cosine convolved evidence in spatial-temporal LDA following stimulus presentation for colour and orientation, depicted in blue and red, respectively. The left panel shows evidence during the stimulus encoding interval and the right panel for during the distractor presentation interval. B) Analyses using Alpha-band power were repeated with single time-point classification technique (see figures channel decoding earlier in this section; appendeix B), for both the stimulus and distractor intervals. Y-axis limits reflect the same scale as the respective stimulus-decoding plots (**Figures 3.2, B.1, B.2**). Error bars indicate 95% confidence intervals.

C

Supplementary Results for Chapter 4

C.1 Decoding Distractor Features

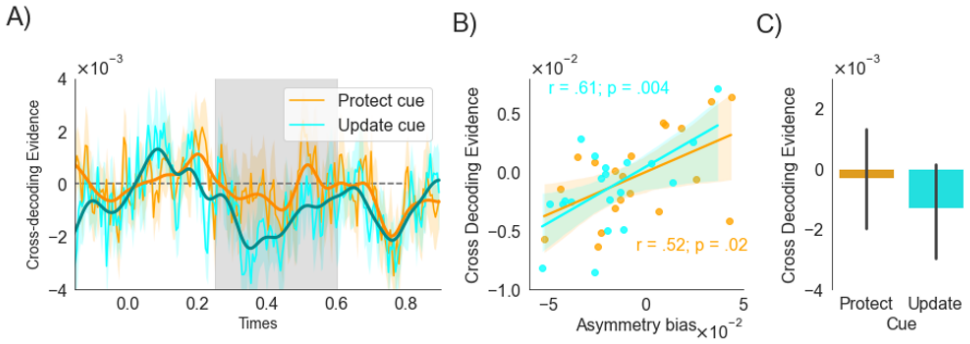


Figure C.1: Cross-decoding Evidence for Orientation 1 Upon Presentation of Orientation 2. A) Cross-decoding after presentation of the second grating orientation on trials where a protect cue (orange) or an update cue (cyan) was presented. Smoothing was applied for visualisation purposes with the same conventions as **Figure 4.6**. The grey-shaded region marks the 250 – 600 ms window used for subplot B-C. B) Scatterplot between neural bias relative to the first orientation presented on the same trial (see results section and **Figure 4.6** for details) and the cross-decoding evidence for the first orientation presented on the same trial. C) Average cross-decoding from 250 – 600 ms. Error bars indicate 95% confidence intervals.

D

Disruption Statement Due to the COVID-19 Pandemic

Like for most doctoral students during this period, the exceptional circumstances brought about by the COVID-19 pandemic, had adverse consequences on my training and limited the contents of this thesis. Some of the planned empirical research, which would have complemented my methodological training as well as the theoretical contributions of the thesis, had to be halted. Below, I highlight the major project impacted by the pandemic below.

D.1 Visuo-Motor Selection: Simultaneous EEG-fMRI

Previously described studies described processes through which working memory can be adapted for the task at hand. The planned study would extend this notion of implementing sensory information for future actions. Although this is an intuitive principle, it remains less clear how visual information is in fact deployed for future behaviour. Recent work has described that sensory information in working memory can be transferred to high-level representations that guide response planning (Myers, Stokes, & Nobre, 2017; Nobre & Stokes, 2019; Schneider, Barth, & Wascher, 2017; van Ede, Chekroud, Stokes, & Nobre, 2019) investigated the temporal relationship of the visual representation in working memory and the anticipatory motor plans. The authors presented participants with two lateralised oriented bars in different colours, of which one was probed for a precision recall at the end of the trial after a brief delay period (similar to **Figure D.1A**). The fixation cross changed colour at the end of the delay to signal which of the two items ought to be reported. Responses were made by pressing and holding a left or right button using the right or left hand, respectively. Importantly, the probe was always initialised at the vertical position on a centralised probe circle, this linked the tilt of the probed item directly to the hand required for the response. Each trial contained a leftward and rightward tilted bar that were allocated at random to the left or right side of the display. This orthogonalised the item selection and the motor selection, allowing the investigation of both processes independently. Using EEG, researchers measured posterior alpha oscillations, associated with the selection of visual information from

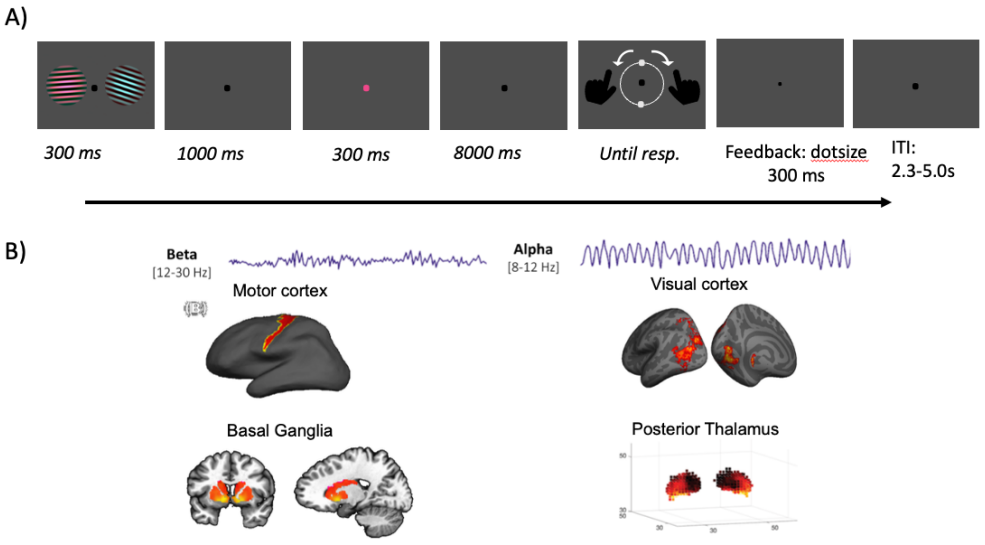


Figure D.1: Task design and data of interest. A) Task schematic for simultaneous EEG-fMRI. Participants were presented with two oriented Gabor stimuli in magenta and cyan colours, tilted in different directions. A subsequent cue could prioritise one of the orientations (33.3% magenta, 33.3% cyan) or provide no additional information (33.3% neutral). After an 8 second delay, participants adjusted a probe using their left or right hand. B) The current study will use beta oscillatory power together with motor cortex and basal ganglia ROIs to study response planning. The selection of sensory information in working memory is characterised using alpha oscillatory power together with visual, parietal, and thalamic ROIs.

memory, and beta oscillations above motor regions, as signatures of motor selection. In conclusion, van Ede et al. (**2019**) observed that, visual and motor representation were activated concurrently at the time of the memory probe.

Functionally, it has been suggested that the basal ganglia interacts with cortical motor regions to facilitate and inhibit action planning and execution, and it is commonly thought that beta oscillations are involved in this process (**Figure D.1B**; Brittain, Sharott, & Brown, 2014). Similarly, attentional prioritisation of visual information in working memory also relies on interconnections between cortical and subcortical regions (c.f., Saalmann, Pinsk, Wang, Li, & Kastner, 2012). My plan was to investigate the neural circuitry involved in the concurrent selection of visual

information and motor plans based on working-memory representations.

In order to reveal the neural circuitry related to sensory vs. motor selection in working memory, the study would combine simultaneous EEG and fMRI. The analysis would have focused on identifying changes in brain activity reflected in the BOLD signal that correlated, on a trial-by-trial basis, with changes in beta and alpha lateralisation associated with motor and sensory selection in working memory. Simultaneous EEG-MRI would provide the unique opportunity to use oscillatory power and relate this to widely distributed networks in cortex, basal ganglia, and thalamus during selection of information in working memory.

The task design is illustrated in **Figure D.1A**. The design builds on the study by van Ede et al. (**2019**). After encoding of two orientations, a retro-cue would signal the item to be reported, prompting spatial selection of the relevant stimulus and accompanying alpha lateralisation as well as motor-plan selection and accompanying beta lateralisation. A long interval of 8 s after the retro-cue would provide a sufficient window for measuring changes in brain activation accompanying sensory and motor selection.

I planned to apply empirical mode decomposition (Quinn, Lopes-dos-Santos, Dupret, Nobre, & Woolrich, 2021) on the EEG data to isolate single-trial alpha power (8-12 Hz) in posterior electrodes and beta power (12-30 Hz) in central electrodes. The single-trial oscillatory power would be related to single-trial evoked BOLD responses in previously described cortical and subcortical regions using a GLM (**Figure D.2A**). The main prediction is that fluctuations in beta power can

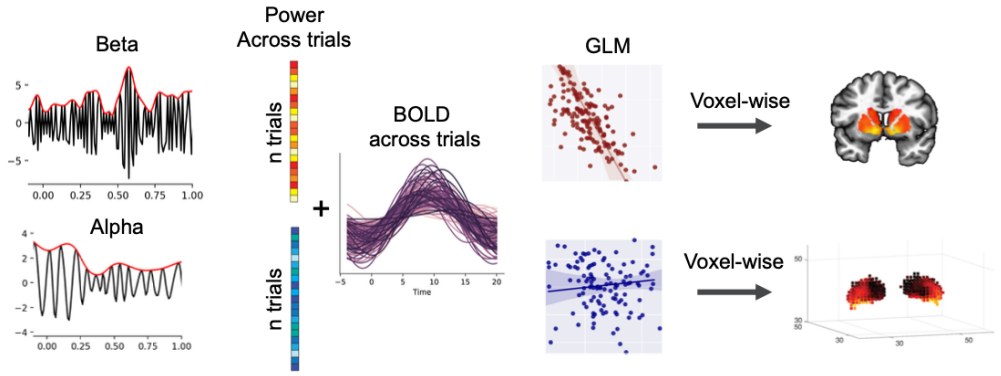


Figure D.2: Analysis relating variance in neural oscillations to BOLD response magnitude. Going from left to right, I first obtain the alpha- and beta-power envelope by applying empirical mode decomposition. A single-trial power estimate can then be obtained by averaging this envelope over 0.5 to 1.0 seconds post-cue (in line with **Chapter 2**; van Ede et al., 2019). Using a general linear model, alpha- and beta-power estimates can be related to fluctuations of BOLD responses across trials. In particular, I expect strongest modulation of BOLD amplitude by beta in the basal ganglia and of alpha in the pulvinar.

predict BOLD response amplitude in basal ganglia and motor cortex while at the same time, alpha power predicts BOLD response amplitude.

Together, this project could investigate mechanisms related to visual working-memory representations and their translation to response formats. The usage of simultaneous EEG-fMRI would allow me to study distributed systems that have, until now, been beyond reach. The planned work would complement the previously discussed work by describing the memory-guided action planning in addition to concurrent prioritisation of task-relevant sensory information.

D.1.1 State of Progress

After approval of this project late 2019, I tested the task design behaviourally and confirmed that participants were able to do the task and showed a retro-cue benefit. Early 2020, I successfully piloted the task in EEG, fMRI, and simultaneous EEG-fMRI. After setting up the preprocessing and analysis pipeline for both data

types, I verified that the data were suitable to address my hypothesis. Unfortunately, the subsequent research scans that were scheduled early April 2020 were cancelled due to the COVID-19 outbreak. Thus far, safety restrictions at the testing facilities have halted data collection for the current research. The project has been unable to advance further, and empirical results were therefore not be included in this thesis.

Works Cited

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory Is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Attwell, D., & Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, *21*(10), 1133–1145. https://doi.org/10.1097/00004647-200110000-00001

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in Spatial Working Memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780–790. https://doi.org/10.1037/0096-1523.24.3.780

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, *85*(3), 139–152. https://doi.org/10.1037/0033-295X.85.3.139

Baddeley, A. D. (1992). Working Memory. *Science*, *255*(Jan), 556–559. https://doi.org/DOI:10.1126/science.1736359

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature reviews. Neuroscience*, *4*(10), 829–39. https://doi.org/10.1038/nrn1201

Bae, G. Y., & Luck, S. J. (2017). Interactions between visual working memory representations. *Attention, Perception, and Psychophysics*, *79*(8), 2376–2395. https://doi.org/10.3758/s13414-017-1404-8

Bae, G. Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *Journal of Neuroscience*, *38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bae, G. Y., & Luck, S. J. (2019). Reactivation of Previous Experiences in a Working Memory Task. *Psychological Science*, *30*(4), 587–595. https://doi.org/10.1177/0956797619830398

Bae, G. Y., & Luck, S. J. (2020). Serial dependence in vision: Merely encoding the previous-trial target is not enough. *Psychonomic Bulletin and Review*, *27*(2), 293–300. https://doi.org/10.3758/s13423-019-01678-7

Barbosa, J., & Compte, A. (2020). Build-up of serial dependence in color working memory. *Scientific Reports*, *10*(1), 1–7. https://doi.org/10.1038/s41598-020-67861-2

Baylis, G. C., & Driver, J. (1993). Visual Attention and Objects: Evidence for Hierarchical Coding of Location. *Journal of Experimental Psychology*, *19*(3), 451–470. https://doi.org/10.1037//0096-1523.19.3.451

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision*, *9*(10), 7.1–11. https://doi.org/10.1167/9.10.7

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–4. https://doi.org/10.1126/science.1158023

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*(6), 1622–1631. https://doi.org/10.1016/j.neuropsychologia.2010.12.023

Becker, S. I., Harris, A. M., Venini, D., & Retell, J. D. (2014). Visual search for color and shape: When is the gaze guided by feature relationships, when by feature values? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 264–291. https://doi.org/10.1037/a0033489

Berlucchi, G. (2009). *Chapter 13 The contributions of neurophysiology to clinical neurology. an exercise in contemporary history* (3rd ed., Vol. 95). Elsevier B.V. https://doi.org/10.1016/S0072-9752(08)02113-1

Bettencourt, K. C., & Xu, Y. (2015). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, (November), 1–11. https://doi.org/10.1038/nn.4174

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-15199-7

Bocincova, A., & Johnson, J. S. (2018). The time course of encoding and maintenance of task-relevant versus irrelevant object features in working memory. *Cortex*. https://doi.org/10.1016/j.cortex.2018.10.013

Boettcher, S. E., Gresch, D., Nobre, A. C., & van Ede, F. (2021). Output planning at the input stage in visual working memory. *Science Advances*, *7*(13), 1–15. https://doi.org/10.1126/sciadv.abe8212

Boettcher, S. E., Nobre, A. C., & van Ede, F. (2020). Functional biases in attentional templates from associative memory. *Journal of Vision*, *20*(13), 1–10. https://doi.org/10.1167/jov.20.13.7

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, *12*(12). https://doi.org/10.1038/nrn3122

Born, R. T., & Tootell, R. B. H. (1992). Middle temporal visual area. *Nature*, *357*(June), 497–499.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 1–34. https://doi.org/10.1167/11.5.4.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brittain, J. S., Sharott, A., & Brown, P. (2014). The highs and lows of beta activity in cortico-basal ganglia loops. *European Journal of Neuroscience*, *39*(11), 1951–1959. https://doi.org/10.1111/ejn.12574

Brouwer, G. J., & Heeger, D. J. (2009). Decoding and reconstructing color from responses in human visual cortex. *Journal of Neuroscience*, *29*(44), 13992–14003. https://doi.org/10.1523/JNEUROSCI.3577-09.2009

Bruckmaier, M., Tachtsidis, I., Phan, P., & Lavie, N. (2020). Attention and Capacity Limits in Perception: A Cellular Metabolism Account. *Journal of Neuroscience*, *40*(35), 6801–6811. https://doi.org/10.1523/JNEUROSCI.2368-19.2020

Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. https://doi.org/10.1016/j.neuron.2015.09.017

Chelazzi., L., Miller., E., Duncan., J., & Desimone., R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(May), 345–347.

Christophe, T. B., Hebart, M. N., & Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience*, *32*(38), 12983–12989. https://doi.org/10.1523/JNEUROSCI.0184-12.2012

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2010). A Taxonomy of External and Internal Attention. *Annual Review of Psychology*, *62*(73), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Chunharas, C., Rademaker, R., Brady, T., & Serences, J. (2019). Adaptive memory distortion in visual working memory. (743941). https://doi.org/10.31234/osf.io/e3m5a

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(21), 7867–7872. https://doi.org/10.1073/pnas.1402785111

Cicchini, G. M., Mikellidou, K., & Burr, D. (2017). Serial dependencies act directly on perception. *Journal of Vision*, *17*(14), 1–9. https://doi.org/10.1167/17.14.6

Cicchini, G. M., Mikellidou, K., & Burr, D. C. (2018). The functional role of serial dependence. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1890). https://doi.org/10.1098/rspb.2018.1722

Cichy, R. M., Ramirez, F. M., & Pantazis, D. (2015). Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *NeuroImage*, *121*, 193–204. https://doi.org/10.1016/j.neuroimage.2015.07.011

Clifford, C. W., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1454), 1705–1710. https://doi.org/10.1098/rspb.2000.1198

Clifford, C. W. (2014). The tilt illusion: Phenomenology and functional implications. *Vision Research*, *104*, 3–11. https://doi.org/10.1016/j.visres.2014.06.009

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*(18), 7426–7435. https://doi.org/10.1523/jneurosci.18-18-07426.1998

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191. https://doi.org/10.1037//0033-2909.104.2.163

Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*, 87–185.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin and Review*, *24*(4), 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Czoschke, S., Fischer, C., Beitner, J., Kaiser, J., & Bledowski, C. (2019). Two types of serial dependence in visual working memory. *British Journal of Psychology*, *110*(2), 256–267. https://doi.org/10.1111/bjop.12349

Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Visual objects interact differently during encoding and memory maintenance. *Attention, Perception, and Psychophysics*, *82*(3), 1241–1257. https://doi.org/10.3758/s13414-019-01861-x

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *362*(1481), 761–72. https://doi.org/10.1098/rstb.2007.2086

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

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

Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When Natural Behavior Engages Working Memory. *Current Biology*, *31*(4), 869–874.e5. https://doi.org/10.1016/j.cub.2020.11.013

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

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology*, *7*(2), 255–261. https://doi.org/10.1016/S0959-4388(97)80014-1

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(1), 32–39. https://doi.org/10.1037/0096-1523.10.1.32

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

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin and Review*, *12*(6), 1127–1133. https://doi.org/10.3758/BF03206454

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research*, *43*(9), 1035–1045. https://doi.org/10.1016/S0042-6989(03)00084-1

Engel, S., Zhang, X., & Wandell, B. (1997). Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature*, *388*(6637), 68–71. https://doi.org/10.1038/40398

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R., Tuholski, S., Laughlin, J., & Conway, A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology. General*, *128*(3), 309–331. https://doi.org/10.1037//0096-3445.128.3.309.Working

Fischer, C., Czoschke, S., Peters, B., Rahm, B., Kaiser, J., & Bledowski, C. (2020). Context information supports serial dependence of multiple visual objects across memory episodes. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15874-w

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Current Biology*, *16*(5), 738–743. https://doi.org/doi:10.1038/nn.3689

Fitz, H., Uhlmann, M., Van Den Broek, D., Duarte, R., Hagoort, P., & Petersson, K. M. (2020). Neuronal spike-rate adaptation supports working memory in language processing. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(34), 20881–20889. https://doi.org/10.1073/pnas.2000222117

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science*, *29*(3), 437–446. https://doi.org/10.1177/0956797617737385

Fornaciai, M., & Park, J. (2019). Neural dynamics of serial dependence in numerosity perception. *Journal of Cognitive Neuroscience*, *32*(1), 141–154. https://doi.org/10.1162/jocn_a_01474

Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017). Alpha-Band Activity Reveals Spontaneous Representations of Spatial Position in Visual Working Memory. *Current Biology*, *27*(20), 3216–3223.e6. https://doi.org/10.1016/j.cub.2017.09.031

Fougnie, D., & Alvarez, G. A. (2011). Object features fail independently in visual working memory: Evidence for a probabilistic feature-store model. *Journal of Vision*, *11*(12), 3–3. https://doi.org/10.1167/11.12.3.Introduction

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting" account. *Perception and Psychophysics*, *58*(1), 88–101. https://doi.org/10.3758/BF03205479

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141. https://doi.org/10.1016/j.tics.2013.01.010

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502), 312–316. https://doi.org/10.1126/science.291.5502.312

Fritsche, M., & de Lange, F. P. (2019). The role of feature-based attention in visual serial dependence. *bioRxiv*, *19*, 1–13. https://doi.org/10.1101/584789

Fritsche, M., Mostert, P., & de Lange, F. P. (2017). Opposite Effects of Recent History on Perception and Decision. *Current Biology*, *27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Fritsche, M., Solomon, S. G., & De Lange, F. P. (2021). Brief stimuli cast a long-term trace in visual cortex. *bioRxiv*, 2021.02.10.430579. https://doi.org/10.1101/2021.02.10.430579

Fritsche, M., Spaak, E., & de Lange, F. P. (2020). A bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *eLife*, *9*, 1–32. https://doi.org/10.7554/eLife.55389

Gazzaley, A. (2011). Influence of early attentional modulation on working memory. *Neuropsychologia*, *49*(6), 1410–1424. https://doi.org/10.1016/j.neuropsychologia.2010.12.022

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in cognitive sciences*, *16*(2), 129–135. https://doi.org/doi:10.1016/j.tics.2011.11.014

Gekas, N., McDermott, K. C., & Mamassian, P. (2019). Disambiguating serial effects of multiple timescales. *Journal of Vision*, *19*(6), 1–14. https://doi.org/10.1167/19.6.24

Geng, J. J., Di Quattro, N. E., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1993–2007. https://doi.org/10.1037/xhp0000430

Ghose, G. M., & Maunsell, J. H. (2002). Attentional modulation in visual cortex depends on task timing. *Nature*, *419*(6907), 616–620. https://doi.org/10.1038/nature01057

González-García, C., Formica, S., Liefooghe, B., & Brass, M. (2020). Attentional prioritization reconfigures novel instructions into action-oriented task sets. *Cognition*, *194*(April 2019), 104059. https://doi.org/10.1016/j.cognition.2019.104059

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(267), 1–13. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

Griffin, I. C., & Nobre, A. C. (2003). Orienting attention to locations in internal representations. *Journal of cognitive neuroscience*, *15*(8), 1176–1194. https://doi.org/10.1162/089892903322598139

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

Hafed, Z. M., & Clark, J. J. (2002). Microsaccades as an overt measure of covert attention shifts. *Vision Research*, *42*(22), 2533–2545. https://doi.org/10.1016/S0042-6989(02)00263-8

Hajonides, J. E., Nobre, A. C., van Ede, F., & Stokes, M. G. (2021). Decoding visual colour from scalp electroencephalography measurements. *NeuroImage*, *237*(April), 118030. https://doi.org/10.1016/j.neuroimage.2021.118030

Hajonides, J. E., van Ede, F., Stokes, M. G., & Nobre, A. C. (2020). Comparing the prioritization of items and feature-dimensions in visual working memory. *Journal of Vision*, *20*(8), 1–12. https://doi.org/10.1167/JOV.20.8.25

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Hermann, K., Rosenthal, I., Singh, S., Pantazis, D., & Conway, B. R. (2020). Temporal dynamics of the neural mechanisms for encoding hue and luminance contrast uncovered by magnetoencephalography. *BioRxiv*, 1–37. https://doi.org/https://doi.org/10.1101/2020.06.17.155713

Heuer, A., Ohl, S., & Rolfs, M. (2020). Memory for action: a functional view of selection in visual working memory. *Visual Cognition*, *28*(5-8), 388–400. https://doi.org/10.1080/13506285.2020.1764156

Heuer, A., & Schubö, A. (2016a). Feature-based and spatial attentional selection in visual working memory. *Memory and Cognition*, *44*(4), 621–632. https://doi.org/10.3758/s13421-015-0584-5

Heuer, A., & Schubö, A. (2016b). The focus of attention in visual working memory: Protection of focused representations and its individual variation. *PLoS ONE*, *11*(4), 1–19. https://doi.org/10.1371/journal.pone.0154228

Heuer, A., & Schubö, A. (2017). Selective weighting of action-related feature dimensions in visual working memory. *Psychonomic Bulletin and Review*, *24*(4), 1129–1134. https://doi.org/10.3758/s13423-016-1209-0

Heuer, A., Schubö, A., & Crawford, J. D. (2016). Different Cortical Mechanisms for Spatial vs. Feature-Based Attentional Selection in Visual Working Memory. *Frontiers in Human Neuroscience*, *10*(August), 1–9. https://doi.org/10.3389/fnhum.2016.00415

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual Working Memory Modulates Rapid Eye Movements to Simple Onset Targets. *Psychological Science*, *24*(5), 790–796. https://doi.org/10.1177/0956797612459767

Huang, L. (2020). Distinguishing target biases and strategic guesses in visual working memory. *Attention, Perception, and Psychophysics*, *82*(3), 1258–1270. https://doi.org/10.3758/s13414-019-01913-2

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://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*(1), 106–154.2. https://doi.org/10.1523/JNEUROSCI.1991-09.2009

Hyvärinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. https://doi.org/10.1109/72.761722

James, W. (1890). *The Principles of Psychology.* Dover Publications. https://doi.org/10.2307/2107586

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, *9*(5), 690–696. https://doi.org/10.1038/nn1691

Jazayeri, M., & Movshon, J. A. (2007). A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, *446*(7138), 912–915. https://doi.org/10.1038/nature05739

Kahneman, D., & Henik, A. (1981). Perceptual Organization and Attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (First Edit, pp. 181–213). Routledge.

Kim, S., Burr, D., Cicchini, G. M., & Alais, D. (2020). Serial dependence in perception requires conscious awareness. *Current Biology*, *30*(6), R257–R258. https://doi.org/10.1016/j.cub.2020.02.008

King, J. R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kiyonaga, A., Scimeca, J. M., Bliss, D. P., & Whitney, D. (2017). Serial Dependence across Perception, Attention, and Memory. https://doi.org/10.1016/j.tics.2017.04.011

Kleinschmidt, A., Lee, B. B., Requardt, M., & Frahm, J. (1996). Functional mapping of color processing by magnetic resonance imaging of responses to selective P- and M-pathway stimulation. *Experimental Brain Research*, *110*(2), 279–288. https://doi.org/10.1007/BF00228558

Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*(5), 3155–3164. https://doi.org/10.1152/jn.00086.2007

Koles, Z. J. (1998). Trends in EEG source localization. *Electroencephalography and Clinical Neurophysiology*, *106*(2), 127–137. https://doi.org/10.1016/S0013-4694(97)00115-6

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception and Psychophysics*, *50*(3), 267–284. https://doi.org/10.3758/BF03206750

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*(November), 4. https://doi.org/10.3389/neuro.06.004.2008

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Kumada, T. (2001). Feature-based control of attention: Evidence for two forms of dimension weighting. *Perception and Psychophysics*, *63*(4), 698–708. https://doi.org/10.3758/BF03194430

Kuo, B.-C., Stokes, M. G., & Nobre, A. C. (2011). Attention Modulates Maintenance of Representations in Visual Short-term Memory. *Journal of Cognitive Neuroscience*, *24*(1), 51–60. https://doi.org/10.1162/jocn_a_00087

Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large Capacity Storage of Integrated Objects Before Change Blindness. *Vision Research*, *43*, 149–164.

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex*, *27*(10), 4881–4890. https://doi.org/10.1093/cercor/bhw283

Lennie, P. (2003). The Cost of Cortical Computation. *Current Biology*, *13*, 493–497. https://doi.org/10.1016/S0960-9822(03)00135-0

Lepsien, J., Thornton, I., & Nobre, A. C. (2011). Modulation of working-memory maintenance by directed attention. *Neuropsychologia*, *49*(6), 1569–1577. https://doi.org/10.1016/j.neuropsychologia.2011.03.011

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of cognitive neuroscience*, *24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Lewis-Peacock, J. A., & Norman, K. A. (2014). Competition between items in working memory leads to forgetting. *Nature Communications*, *5*(5768), 1–10. https://doi.org/10.1038/ncomms6768

Libby, A., & Buschman, T. J. (2021). Rotational dynamics reduce interference between sensory and memory representations. *Nature Neuroscience*, *24*(5), 715–726. https://doi.org/10.1038/s41593-021-00821-9

Liu, T., & Hou, Y. (2011). Global feature-based attention to orientation. *Journal of Vision*, *11*(10), 1–8. https://doi.org/10.1167/11.10.1

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-Based Attention Modulates Orientation-Selective Responses in Human Visual Cortex. *Neuron*, *55*(2), 313–323. https://doi.org/10.1016/j.neuron.2007.06.030

Lopes da Silva, F. (2013). EEG and MEG: Relevance to neuroscience. *Neuron*, *80*(5), 1112–1128. https://doi.org/10.1016/j.neuron.2013.10.017

Luck, S. J., & Vogel, E. K. (1997). 36846. *Nature*, *390*(1996), 279–281.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Machado, R., Neto, D. A., Bartels, A., Israelita, H., & Einstein, A. (2021). Disrupting short-term memory in premotor but not visual cortex affects serial dependence in visuomotor integration. *bioRxiv*, *307*, 1–52.

Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychonomic Bulletin and Review*, *14*(6), 1072–1078.

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory and Cognition*, *36*(1), 43–52. https://doi.org/10.3758/MC.36.1.43

Malebranche, N., & Olscamp, P. J. (1997). *Malebranche: The Search After Truth and Elucidations of the Search After Truth* (T. Lennon, Ed.). Cambridge University Press. https://doi.org/10.5840/schoolman198360243

Malmivuo, J. (2012). Comparison of the properties of EEG and MEG in detecting the electric activity of the brain. *Brain Topography*, *25*(1), 1–19. https://doi.org/10.1007/s10548-011-0202-1

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martínez-Trujillo, J. C., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*(2), 365–370. https://doi.org/10.1016/S0896-6273(02)00778-X

Matsukura, M., & Vecera, S. P. (2015). Selection of multiple cued items is possible during visual short-term memory maintenance. *Attention, Perception, and Psychophysics*, *77*(5), 1625–1646. https://doi.org/10.3758/s13414-015-0836-2

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*(1), 431–441. https://doi.org/10.1523/jneurosci.19-01-00431.1999

Miller, E. K., Erickson, C. a., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, *16*(16), 5154–5167. https://doi.org/10.1.1.41.2959

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). Plans and the structure of behaviour. *Systems Research for Behavioral Science: A Sourcebook*, 369–382. https://doi.org/10.2307/411065

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Mostert, P., Albers, A. M., Brinkman, L., Todorova, L., Kok, P., & de Lange, F. P. (2018). Eye movement-related confounds in neural decoding of visual working memory representations. *eNeuro*, *5*(4), 1–14. https://doi.org/10.1523/ENEURO.0401-17.2018

Müller, H. J., Heller, D., & Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception and Psychophysics*, *57*(1), 1–17. https://doi.org/10.3758/BF03211845

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention Restores Discrete Items to Visual Short-Term Memory. *Psychological Science*, *24*(4), 550–556. https://doi.org/10.1177/0956797612457782

Myers, N. E., Stokes, M. G., & Nobre, A. C. (2017). Prioritizing Information during Working Memory: Beyond Sustained Internal Attention. *Trends in Cognitive Sciences*, *21*(6), 449–461. https://doi.org/10.1016/j.tics.2017.03.010

Myers, N. E., Stokes, M. G., Walther, L., & Nobre, A. C. (2014). Oscillatory brain state predicts variability in working memory. *Journal of Neuroscience*, *34*(23), 7735–7743. https://doi.org/10.1523/JNEUROSCI.4741-13.2014

Myers, N. E., Walther, L., Wallis, G., Stokes, M. G., & Nobre, A. C. (2015). Temporal Dynamics of Attention during Encoding versus Maintenance of Working Memory: Complementary Views from Event-related Potentials and Alpha-band Oscillations. *Journal of Cognitive Neuroscience*, *27*(3), 492–508. https://doi.org/10.1162/jocn

Navalpakkam, V., & Itti, L. (2007). Search Goal Tunes Visual Features Optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Niklaus, M., Nobre, A. C., & van Ede, F. (2017). Feature-based attentional weighting and spreading in visual working memory. *Scientific Reports*, *7*, 1–10. https://doi.org/10.1038/srep42384

Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *Journal of cognitive neuroscience*, *16*(3), 363–73. https://doi.org/10.1162/089892904322926700

Nobre, A. C., & Kastner, S. (2014). Attention: time capsule 2013. *The Oxford Handbook of Attention*, (July), 1201–1222. https://doi.org/10.1093/oxfordhb/9780199675111.013.040

Nobre, A. C., & Stokes, M. G. (2019). Premembering Experience: A Hierarchy of Time-Scales for Proactive Attention. *Neuron*, *104*(1), 132–146. https://doi.org/10.1016/j.neuron.2019.08.030

Nobre, A. C., & Stokes, M. G. (2020). Memory and attention: the back and forth. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (6th Editio, pp. 291–300). Cambridge: MIT Press.

Nobre, A. C., & Van Ede, F. (2018). Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience*, *19*(1), 34–48. https://doi.org/10.1038/nrn.2017.141

Nobre, A. C., & van Ede, F. (2020). Under the mind's hood: What we have learned by watching the brain at work. *Journal of Neuroscience*, *40*(1), 89–100. https://doi.org/10.1523/JNEUROSCI.0742-19.2019

Oberauer, K. (2019). Working memory and attention - A conceptual analysis and review. *Journal of Cognition*, *2*(1), 1–23. https://doi.org/10.5334/joc.58

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243–1265. https://doi.org/10.1037/0096-1523.32.5.1243

Olivers, C. N., & Roelfsema, P. R. (2020). Attention for action in visual working memory. *Cortex*, *131*, 179–194. https://doi.org/10.1016/j.cortex.2020.07.011

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception and Psychophysics*, *64*(7), 1055–1067. https://doi.org/10.3758/BF03194756

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601–605. https://doi.org/10.1038/s41586-021-03390-w

Panichello, M. F., DePasquale, B., Pillow, J. W., & Buschman, T. J. (2019). Error-correcting dynamics in visual working memory. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-11298-3

Papadimitriou, C., White, R. L., & Snyder, L. H. (2017). Ghosts in the Machine II: Neural correlates of memory interference from the previous trial. *Cerebral Cortex*, *27*(4), 2513–2527. https://doi.org/10.1093/cercor/bhw106

Park, Y. E., Sy, J. L., Hong, S. W., & Tong, F. (2017). Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychological Science*, *28*(12), 1773–1785. https://doi.org/10.1177/0956797617719949

Pascucci, D., Mancuso, G., Santandrea, E., Libera, C. D., Plomp, G., & Chelazzi, L. (2019). *Laws of concatenated perception: Vision goes for novelty, decisions for perseverance* (Vol. 17). https://doi.org/10.1371/journal.pbio.3000144

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, *44*(4), 369–378. https://doi.org/10.3758/BF03210419

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature reviews. Neuroscience*, *6*(2), 97–107. https://doi.org/10.1038/nrn1603

Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *Journal of Neuroscience*, *33*(2), 532–543. https://doi.org/10.1523/JNEUROSCI.3345-12.2013

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2

Persichetti, A. S., Thompson-Schill, S. L., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2015). Functional magnetic resonance imaging adaptation reveals a noncategorical representation of hue in early visual cortex. *Journal of Vision*, *15*(6), 1–19. https://doi.org/10.1167/15.6.18

Pertzov, Y., Manohar, S., & Husain, M. (2017). Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(4), 528–536. https://doi.org/10.1037/xlm0000328

Pilling, M., & Barrett, D. J. (2016). Dimension-based attention in visual short-term memory. *Memory and Cognition*, *44*(5), 740–749. https://doi.org/10.3758/s13421-016-0599-6

Poch, C., Capilla, A., Hinojosa, J. A., & Campo, P. (2017). Selection within working memory based on a color retro-cue modulates alpha oscillations. *Neuropsychologia*, *106*(July), 133–137. https://doi.org/10.1016/j.neuropsychologia.2017.09.027

Posner, M. I. (1980). Orienting of Attention. *Attention in a Social World*, *31*(1), 3–25. https://doi.org/10.1093/acprof

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38. https://doi.org/10.1016/j.neuroscience.2005.06.005

Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 6–17. https://doi.org/10.1037/xhp0000302

Priebe, N. J., & Lisberger, S. G. (2002). Constraints on the source of short-term motion adaptation in macaque area MT. II. Tuning of neural circuit mechanisms. *Journal of Neurophysiology*, *88*(1), 370–382. https://doi.org/10.1152/jn.2002.88.1.370

Prinzmetal, W., Hoffman, H., & Vest, K. (1991). Automatic Processes in Word Perception: An Analysis From Illusory Conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(4), 902–923. https://doi.org/10.1037/0096-1523.17.4.902

Quax, S. C., Dijkstra, N., van Staveren, M. J., Bosch, S. E., & van Gerven, M. A. (2019). Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage*, *195*(March), 444–453. https://doi.org/10.1016/j.neuroimage.2019.03.069

Quinn, A., Lopes-dos-Santos, V., Dupret, D., Nobre, A. C., & Woolrich, M. (2021). EMD: Empirical Mode Decomposition and Hilbert-Huang Spectral Analyses in Python. *Journal of Open Source Software*, *6*(59), 2977. https://doi.org/10.21105/joss.02977

Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Rademaker, R. L., Bloem, I. M., Weerd, P. D., Sack, A. T., & Rademaker, R. L. (2015). Journal of Experimental Psychology : Human Perception and Performance The Impact of Interference on Short-Term Memory for Visual Orientation The Impact of Interference on Short-Term Memory for Visual Orientation.

Rademaker, R. L., Park, Y. E., Sack, A. T., & Tong, F. (2018). Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 925–940. https://doi.org/10.1037/xhp0000491

Rademaker, R. L., Van De Ven, V. G., Tong, F., & Sack, A. T. (2017). The impact of early visual cortex transcranial magnetic stimulation on visual working memory precision and guess rate. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175230

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, *393*(6685), 577–579. https://doi.org/10.1038/31235

Ramkumar, P., Jas, M., Pannasch, S., Hari, R., & Parkkonen, L. (2013). Feature-specific information processing precedes concerted activation in human visual cortex. *Journal of Neuroscience*, *33*(18), 7691–7699. https://doi.org/10.1523/JNEUROSCI.3905-12.2013

Rerko, L., & Oberauer, K. (2013). Focused, unfocused, and defocused information in working memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(4), 1075–1096. https://doi.org/10.1037/a0031172

Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory and Cognition*, *42*(5), 712–728. https://doi.org/10.3758/s13421-013-0392-8

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*(5), 1736–1753. https://doi.org/10.1523/jneurosci.19-05-01736.1999

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Reynolds, J. H., & Pasternak, T. (2000). Attention Increases Sensitivity of V4 Neurons increases the magnitude of the neuronal response elic. *Neuron*, *26*, 703–714.

Rohenkohl, G., Gould, I. C., Pessoa, J., & Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision*, *14*(4), 1–13. https://doi.org/10.1167/14.4.8

Rosenthal, I. A., Singh, S. R., Hermann, K. L., Pantazis, D., & Conway, B. R. (2021). Color Space Geometry Uncovered with Magnetoencephalography. *Current Biology*, *31*(3), 515–526. https://doi.org/10.1016/j.cub.2020.10.062

Rosner, B. (1983). Percentage Outlier Points for Generalized ESD Many- Procedure. *Technometrics, 25*(2), 165–172.

Rossi, A. F., & Paradiso, M. A. (1995). Feature-specific effects of selective visual attention. *Vision Research*, *35*(5), 621–634. https://doi.org/10.1016/0042-6989(94)00156-G

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., Fox, P. T., & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: A coordinate-based meta-analysis. *NeuroImage*, *60*(1), 830–846. https://doi.org/10.1016/j.neuroimage.2011.11.050

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., & Kastner, S. (2012). The pulvinar regulates information transmission between cortical areas based on attention demands. *Science (New York, N.Y.)*, *337*(6095), 753–6. https://doi.org/10.1126/science.1223082

Flanker task, attention modulates alpha and in turn gamma frequencies during the period between cue and stimulus

Sadil, P., Cowell, R. A., & Huber, D. E. (2021). The yin-yang of serial dependence effects: every response is both an attraction to the prior response and a repulsion from the prior stimulus. *PsyArXiv*.

Sàenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research*, *43*(6), 629–637. https://doi.org/10.1016/S0042-6989(02)00595-3

Sahan, M. I., Sheldon, A. D., & Postle, B. R. (2019). The Neural Consequences of Attentional Prioritization of Internal Representations in Visual Working Memory. *Journal of Cognitive Neuroscience*, 1–28. https://doi.org/10.1162/jocn_a_01517

Sandhaeger, F., von Nicolai, C., Miller, E. K., & Siegel, M. (2019). Monkey EEG links neuronal color and motion information across species and scales. *eLife*, *8*, 1–21. https://doi.org/10.7554/eLife.45645

Savoy, R. L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica*, *107*(1-3), 9–42. https://doi.org/10.1016/S0001-6918(01)00018-X

Schneegans, S., & Bays, P. M. (2017). Neural Architecture for Feature Binding in Visual Working Memory. *The Journal of Neuroscience*, *37*(14), 3913–3925. https://doi.org/10.1523/JNEUROSCI.3493-16.2017

Schneider, D., Barth, A., & Wascher, E. (2017). On the contribution of motor planning to the retroactive cuing benefit in working memory: Evidence by mu and beta oscillatory activity in the EEG. *NeuroImage*, *162*(April), 73–85. https://doi.org/10.1016/j.neuroimage.2017.08.057

Scolari, M., Edward, F., & Serences, J. T. (2014). Feature- and Object-Based Attentional Modulation in the Human Visual System. *Oxford handbook of attention* (pp. 573–600). https://doi.org/10.1093/oxfordhb/9780199675111.013.009

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Shin, H., Zou, Q., & Ma, W. J. (2017). The effects of delay duration on visual working memory for orientation. *Journal of Vision*, *17*(14), 1–24. https://doi.org/10.1167/17.14.10

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586. https://doi.org/10.1146/annurev.ne.18.030195.003011

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores? *PLoS ONE*, *3*(2), 2–10. https://doi.org/10.1371/journal.pone.0001699

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

Snow, J. C., & Culham, J. C. (2021). The Treachery of Images: How Realism Influences Brain and Behavior. *Trends in Cognitive Sciences*, *25*(6), 506–519. https://doi.org/10.1016/j.tics.2021.02.008

Song, J. H., & Jiang, Y. (2006). Visual working memory for simple and complex features: An fMRI study. *NeuroImage*, *30*(3), 963–972. https://doi.org/10.1016/j.neuroimage.2005.10.006

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348. https://doi.org/10.1016/j.tics.2008.05.007

Souza, A. S., & Oberauer, K. (2016). In search of the focus of attention in working memory: 13 years of the retro-cue effect. *Attention, Perception, and Psychophysics*, *78*(7), 1839–1860. https://doi.org/10.3758/s13414-016-1108-5

Spitzer, H., Desimone, R., & Morant, J. (1988). Increased Attention Enhances Both Behavioural and Neuronal Performance. *Nature*, *240*(4850), 338–340. https://doi.org/10.1126/science.3353728

Sprague, T. C., Ester, E. F., & Serences, J. T. (2016). Restoring Latent Visual Working Memory Representations in Human Cortex. *Neuron*, *91*(3), 694–707. https://doi.org/10.1016/j.neuron.2016.07.006

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89. https://doi.org/10.1016/j.tics.2013.12.001

Sreenivasan, K. K., & D'Esposito, M. (2019). The what, where and how of delay activity. *Nature Reviews Neuroscience*, *20*(8), 466–481. https://doi.org/10.1038/s41583-019-0176-7

Srinivasan, M. V., Laughlin, S. B., & Dubs, A. (1982). Predictive coding: a fresh view of inhibition in the retina. https://doi.org/10.1098/rspb.1982.0085

St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *36*(23), 6186–92. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Stocker, A. A., & Simoncelli, E. P. (2008). A Bayesian model of conditioned perception. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, *20*, 1409–1416.

Stokes, M., & Spaak, E. (2016). The Importance of Single-Trial Analyses in Cognitive Neuroscience. *Trends in Cognitive Sciences*, *20*(7), 483–486. https://doi.org/10.1016/j.tics.2016.05.008

Stokes, M. G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology*, *24*(17), 1985–1988. https://doi.org/10.1016/j.cub.2014.07.030

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Suárez-Pinilla, M., Seth, A. K., & Roseboom, W. (2018). Serial dependence in the perception of visual variance. *Journal of Vision*, *18*(7), 1–24. https://doi.org/10.1167/18.7.4

Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M., & Richmond, B. J. (2008). Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Computational Biology*, *4*(5). https://doi.org/10.1371/journal.pcbi.1000073

Sutterer, D. W., Coia, A. J., Sun, V., Shevell, S. K., & Awh, E. (2021). Decoding chromaticity and luminance from patterns of EEG activity. *Psychophysiology*, *e13779*(January), 1–21. https://doi.org/10.1111/psyp.13779

Sutterer, D. W., Foster, J. J., Adam, K. C., Vogel, E. K., & Awh, E. (2018). Item-specific delay activity demonstrates concurrent storage of multiple items in working memory. *bioRxiv*, 1–25. https://doi.org/10.1101/382879

Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the signal space separation method. *Brain Topography*, *16*(4), 269–275. https://doi.org/10.1023/B:BRAT.0000032864.93890.f9

Team, J. (2020). JASP (Version 0.14.1)[Computer software].

Teichmann, L., Grootswagers, T., Carlson, T. A., & Rich, A. N. (2019). Seeing versus knowing: The temporal dynamics of real and implied colour processing in the human brain. *NeuroImage*, *200*(June), 373–381. https://doi.org/10.1016/j.neuroimage.2019.06.062

Teichmann, L., Quek, G. L., Robinson, A. K., Grootswagers, T., Carlson, T. A., & Rich, A. N. (2020). The Influence of Object-Color Knowledge on Emerging

Object Representations in the Brain. *Journal of Neuroscience*, *40*(35), 6779–6789. https://doi.org/10.1523/JNEUROSCI.0158-20.2020

Teng, C., & Kravitz, D. J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour*, *3*(8), 827–836. https://doi.org/10.1038/s41562-019-0640-4

Thielen, J., Bosch, S. E., van Leeuwen, T. M., van Gerven, M. A., & van Lier, R. (2019). Evidence for confounding eye movements under attempted fixation and active viewing in cognitive neuroscience. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-54018-z

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, *26*(37), 9494–9502. https://doi.org/10.1523/JNEUROSCI.0875-06.2006

Töllner, T., Conci, M., Müller, H. J., & Mazza, V. (2016). Attending to multiple objects relies on both feature- and dimension-based control mechanisms: Evidence from human electrophysiology. *Attention, Perception, and Psychophysics*, *78*(7), 2079–2089. https://doi.org/10.3758/s13414-016-1152-1

Treisman, A. M. (1988). Features and Objects: The Fourteenth Bartlett Memorial Lecture. *The Quarterly Journal of Experimental Psychology Section A*, *40*(2), 201–237. https://doi.org/10.1080/02724988843000104

Treisman, A. M., & Gelade, G. (1980). A Feature-integration Theory of Attention. *Cognitive Psychology*, *12*, 97–136. https://doi.org/10.1097/00075198-199608000-00002

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141. https://doi.org/10.1016/0010-0285(82)90006-8

Treue, S. (2004). Perceptual enhancement of contrast by attention. *Trends in Cognitive Sciences*, *8*(10), 435–437. https://doi.org/10.1016/j.tics.2004.08.001

Treue, S., & Martínez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Van Den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785. https://doi.org/10.1073/pnas.1117465109

Van Moorselaar, D., Olivers, C. N. L., Theeuwes, J., Lamme, Victor, A., & Sligte, I. G. (2015). Forgotten But Not Gone : Retro-Cue Costs and Benefits in a

Double-Cueing Paradigm Suggest Multiple States in Visual Short-Term Memory. *Journal of Experimental Psychology*, *41*(6), 1755–1763. https://doi.org/10.1037/xlm0000124

van Bergen, R. S., & Jehee, J. F. (2019). Probabilistic Representation in Human Visual Cortex Reflects Uncertainty in Serial Decisions. *Journal of Neuroscience*, *39*(41), 8164–8176. https://doi.org/10.1523/JNEUROSCI.3212-18.2019

van Ede, F. (2018). Mnemonic and attentional roles for states of attenuated alpha oscillations in perceptual working memory: a review. *European Journal of Neuroscience*, *48*(7), 2509–2515. https://doi.org/10.1111/ejn.13759

van Ede, F. (2020). Visual working memory and action: Functional links and bi-directional influences. *Visual Cognition*, *28*(5-8), 401–413. https://doi.org/10.1080/13506285.2020.1759744

van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2018). Decoding the influence of anticipatory states on visual perception in the presence of temporal distractors. *Nature Communications*, *9*(1), 1449. https://doi.org/10.1038/s41467-018-03960-z

van Ede, F., Chekroud, S. R., Stokes, M. G., & Nobre, A. C. (2019). Concurrent visual and motor selection during visual working memory guided action. *Nature Neuroscience*, *22*(3), 477–483. https://doi.org/10.1038/s41593-018-0335-6

van Ede, F., Niklaus, M., & Nobre, A. C. (2017a). Temporal Expectations Guide Dynamic Prioritization in Visual Working Memory through Attenuated α Oscillations. *The Journal of Neuroscience*, *37*(2), 437–445. https://doi.org/10.1523/JNEUROSCI.2272-16.2017

van Ede, F., Niklaus, M., & Nobre, A. C. (2017b). Temporal expectations guide dynamic prioritization in visual working memory through attenuated α oscillations. *Journal of Neuroscience*, *37*(2), 437–445. https://doi.org/10.1523/JNEUROSCI.2272-16.2016

van Ede, F., & Nobre, A. C. (2021). Toward a neurobiology of internal selective attention. *Trends in Neurosciences*, *44*(7), 513–515. https://doi.org/10.1016/j.tins.2021.04.010

Vidaurre, D., Myers, N. E., Stokes, M., Nobre, A. C., & Woolrich, M. W. (2019). Temporally Unconstrained Decoding Reveals Consistent but Time-Varying Stages of Stimulus Processing. *Cerebral Cortex*, *29*(2), 863–874. https://doi.org/10.1093/cercor/bhy290

Vidaurre, D., Smith, S. M., & Woolrich, M. W. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), 12827–12832. https://doi.org/10.1073/pnas.1705120114

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R., Jones, E., Kern, R., Larson, E., ... Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748–751. https://doi.org/10.1038/nature02447

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. https://doi.org/10.1038/nature04171

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of Features, Conjunctions and Objects in Visual Working Memory. https://doi.org/10.1037//0096-1523.27.1.92

von Helmholtz, H. (1867). *Treatise on Physiological Optics* (Volume III).

Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Research*, *39*(23), 3960–3974. https://doi.org/10.1016/S0042-6989(99)00101-7

Wallis, G., Stokes, M. G., Cousijn, H., Woolrich, M., & Nobre, A. C. (2015). Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. *Journal of Cognitive Neuroscience*, *27*(10), 2019–2034. https://doi.org/https://doi.org/10.1162/jocn_a_00838

Wan, Q., Cai, Y., Samaha, J., & Postle, B. R. (2020). Tracking stimulus representation across a 2-back visual working memory task: Tracking 2-back representation. *Royal Society Open Science*, *7*(8). https://doi.org/10.1098/rsos.190228rsos190228

Wan, Q., Menendez, J. A., & Postle, B. R. (2021). Rotational remapping between differently prioritized representations in visual working memory.

Webster, M. A. (2015). Visual Adaptation. *Annual Review of Vision Science*, *1*, 547–567. https://doi.org/10.1146/annurev-vision-082114-035509

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision II Unique hues. *Journal of the Optical Society of America A*, *17*(9), 1545. https://doi.org/10.1364/josaa.17.001545

Wei, Z., Wang, X. J., & Wang, D. H. (2012). From distributed resources to limited slots in multiple-item working memory: A spiking network model with normalization. *Journal of Neuroscience*, *32*(33), 11228–11240. https://doi.org/10.1523/JNEUROSCI.0735-12.2012

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48–64. https://doi.org/10.1037/0096-3445.131.1.48

Whissell, C., Abramson, C. I., & Barber, K. R. (2013). The search for cognitive terminology: An analysis of comparative psychology journal titles. *Behavioral Sciences*, *3*(1), 133–142. https://doi.org/10.3390/bs3010133

White, A. L., & Carrasco, M. (2011). Feature-based attention involuntarily and simultaneously improves visual performance across locations. *Journal of Vision*, *11*(6), 1–10. https://doi.org/10.1167/11.6.15

Wilson, C. J., & Soranzo, A. (2015). The Use of Virtual Reality in Psychologyion: A Case Study in Visual Percept. *Computational and Mathematical Methods in Medicine*, *2015*, 2–4. http://dx.doi.org/10.1155/2015/

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, *9*(september), 1–12. https://doi.org/10.3389/fnsys.2015.00123

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., & Stokes, M. G. (2019). Drifting codes within a stable coding scheme for working memory. *bioRxiv*, 714311. https://doi.org/10.1101/714311

Wolff, M. J., Jochim, J., Akyürek, E. G., Buschman, T. J., & Stokes, M. G. (2020). Drifting codes within a stable coding scheme for working memory. *PLoS biology*, *18*(3), e3000625. https://doi.org/10.1371/journal.pbio.3000625

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *20*(6), 1–6. https://doi.org/10.1523/jneurosci.20-06-j0002.2000

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95. https://doi.org/10.1038/nature04262

Ye, C., Hu, Z., Ristaniemi, T., Gendron, M., & Liu, Q. (2016). Retro-dimension-cue benefit in visual working memory. *Scientific Reports*, *6*, 1–13. https://doi.org/10.1038/srep35573

Yu, Q., & Shim, W. M. (2017). Occipital, parietal, and frontal cortices selectively maintain task-relevant features of multi-feature objects in visual working memory. *NeuroImage*, *157*(May), 97–107. https://doi.org/10.1016/j.neuroimage.2017.05.055

Yu, X., & Geng, J. J. (2019). The attentional template is shifted and asymmetrically sharpened by distractor context. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 336–353. https://doi.org/10.1037/xhp0000609

Zamboni, E., Ledgeway, T., McGraw, P. V., & Schluppeck, D. (2016). Do perceptual biases emerge early or late in visual processing? Decision-biases in motion perception. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1833). https://doi.org/10.1098/rspb.2016.0263

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–5. https://doi.org/10.1038/nature06860