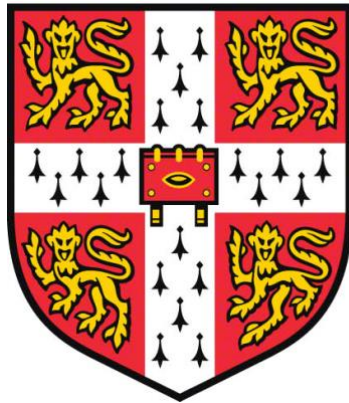


The Role of the Multiple-Demand System in Fluid Intelligence and Selective Attention: Insights from Lesion and Neuroimaging Studies



Nadene Caber Dermody

ST. CATHARINE'S COLLEGE

Medical Research Council Cognition and Brain Sciences Unit

University of Cambridge

This thesis is submitted for the degree of Doctor of Philosophy

June 2024

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text.

It is not substantially the same as any work that has already been submitted, or is being concurrently submitted, for any degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Abstract

Our ability to flexibly and rapidly coordinate our thoughts and actions in line with our goals is a defining feature of human cognition. A prominent proposal posits the multiple-demand (MD) system, a distributed network of predominantly frontoparietal regions, plays a critical role in this process. A growing body of evidence shows the MD system is consistently recruited in response to a wide range of cognitive demands and preferentially represents information currently relevant to behaviour. Yet our understanding of how these regions support complex, goal-directed behaviour is still incomplete. In this thesis, I further elucidate the role of the MD system in facilitating purposeful behaviour, focussing on its role in fluid intelligence and selective attention. Across three empirical studies, I draw on evidence from patients with focal brain lesions and healthy adults using neuropsychology, functional neuroimaging (functional magnetic resonance imaging (fMRI) and magnetoencephalography (MEG)), and multivariate pattern analysis techniques to investigate the role of the MD system in fluid intelligence and selective attention.

In Chapter 2, I analysed a large neuropsychological dataset from 174 patients with focal brain lesions to identify the brain networks critical for fluid intelligence. For this, I quantified the extent of grey matter damage and white matter network disconnection in seven canonical resting-state brain networks, and related these measures to an estimate of individual fluid intelligence loss. The data suggested that of all resting-state networks, frontoparietal network, a resting-state network closely aligned to the MD

system, had the most prominent role in fluid intelligence. Even in this relatively large patient sample, it was not possible to dissociate the contribution of frontoparietal network from a broader contribution of total lesion volume. However, I showed that damage to this network was predictive of fluid intelligence deficit even after accounting for damage to other networks.

In Chapter 3, I turned to the role of the MD system in supporting selective attention. I used fMRI, multivariate pattern analysis techniques, and MEG-fMRI fusion to investigate what task information is prioritised in the MD system when attention is directed both in space (left/right) and to particular object features (colour/shape). For this, I collected and analysed fMRI data from healthy adults ($N=30$). I found that MD codes were highly tuned to the task at hand, only reflecting behaviourally relevant information. These codes were organised to reflect both the difficulty of the current discrimination and the details of the attended visual features. Using MEG-fMRI fusion, I then compared the data to an MEG dataset from participants performing the same task, finding broad agreement between modalities. These data emphasised the role of the MD system in selective prioritisation of attended information.

In Chapter 4, I combined the focal lesion approach with MEG and multivariate pattern analysis to explore how the representation of task-related information during selective attention is affected by parietal lobe damage. In this exploratory case-series, I examined the dynamic neural prioritisation of relevant over irrelevant information at an individual level in six patients with focal parietal lesions and 10 age-, gender-, and education-matched controls. I found that patients with parietal lesions can ultimately prioritise relevant information in their neural activity, commensurate with their high behavioural accuracy. However, for some patients, slower reaction times were associated with subtle deficits in neural prioritisation. This aligns with prominent suggestions of a causal role of parietal MD regions in attentional prioritisation and demonstrates the feasibility of combining MEG and MVPA to study individual differences in the dynamic prioritisation of information after brain damage.

In conclusion, I observed that damage to frontoparietal network predicts fluid intelligence deficit, that MD regions selectively code behaviourally relevant

information, and that damage to some of this network can affect the degree to which behaviourally relevant information is prioritised in the brain. Taken together, this body of work suggests that MD regions, or frontoparietal network more broadly, play an important role in fluid intelligence and selective attention, processes fundamental to intelligent human behaviour. My work also showcases new methodological approaches and suggests interesting directions for future research into how the MD system gives rise to our extraordinary capacity for complex, flexible behaviour.

Preface

While I have used first person singular (I, me, my, myself) throughout much of this thesis, the research reported in this thesis benefitted from the contributions of multiple researchers. In addition to my supervisor, Alexandra Woolgar, and advisors, Romy Lorenz and Elizabeth Michael, I note these collaborations below.

Chapter 2 was done in collaboration with Arnaud Boré, John Duncan, and Moataz Assem. Arnaud Boré shared the pipeline for estimating extent of white matter network damage and provided support in running this pipeline. John Duncan and Moataz Assem gave advice on design and analysis.

Chapter 3 was done in collaboration with Erin Goddard and Arno Villringer. Erin Goddard shared her paradigm script and gave advice on design and analysis. Arno Villringer provided access to imaging facilities and general support at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. This chapter has been prepared for peer-review submission and therefore I use first person plural (we, us, our, ourselves) throughout this chapter.

Chapter 4 was done in collaboration with Erin Goddard, John Duncan, Polly Peers, and Olaf Hauk. Erin Goddard shared her paradigm script and gave advice on design. John Duncan, Polly Peers, and Olaf Hauk, and gave advice on design and analysis.

To provide context to the empirical chapters, each of these chapters is bookended with a Foreword and an Afterword. A list of abbreviations and acronyms is provided at the beginning of the thesis for ease of reference. Supplementary Materials for each chapter can be found at the end of each chapter. References to all chapters are located at the end of the thesis to reduce repetition. Tables and figures (including Supplementary tables and figures) are numbered sequentially throughout the thesis.

Acknowledgements

Scientific endeavours are the result of the work of many, and this thesis is no exception. I am grateful to the many that have given me guidance, support, and help over the past few years.

First of all, thank you to my wonderful supervisor, Alexandra Woolgar – you are a tremendous source of intellectual and personal wisdom, and my most trusted provider of encouragement and reassurance. I have learnt so much from you. Your support has meant so much to me over the past few years, and, particularly over these past few months, it is really what got me over the line. I cannot thank you enough.

To my brilliant advisors: Romy Lorenz and Elizabeth Michael. Romy, thank you for all your excellent suggestions – many of which have resulted in some of the more interesting analyses I have done in this work – and for the energy, enthusiasm, empathy, and kindness you brought to my supervisory team. Lizzie, thank you also for your wonderful suggestions and questions. Both I and this work have benefitted greatly from your sagacity.

I am lucky to have also benefitted from the wonderful environment that is the CBU. To John Duncan and Moataz Assem – thank you both for reading and improving my work. Thank you also to Polly Peers, Erin Goddard, and Olaf Hauk for your tremendous help and support in various aspects of the work that follows. Thank you also to our wonderful IT team, notably Anthony Parry-Jones for essentially learning everything one needs to

know about NextFlow in order to make my pipeline run. Thank you also to the many others at the CBU who have enriched my experience and my thinking.

I would also like to acknowledge a few other people who helped me with specific chapters.

For Chapter 2, Arnaud Boré, Michel Thiebaut de Schotten – thank you for answering my many, many questions. And thank you very much to my wonderful summer intern, Jana Tomašíková, who collected a substantial amount of the control data featured in this chapter.

For Chapter 3, I am indebted to many people at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. Most notably, thank you to Arno Villringer, who gave me so much freedom to run my experiment and ensured I had whatever help I needed. Jöran Lepsen, Toralf Mildner, Mona Garvert, Alice Hodapp, Paul Steinfath – thank you all for your help in shaping and helping me run this project.

Chapter 4 has particularly required the help of many. In addition to John, Polly, Olaf, and Erin, thank you so much to (in alphabetical order) Ajay Halai, Clare Cook, Estela Carmona, Federica Magnabosco, Lucy MacGregor, Martyn Blairs, Máté Aller, Michelle Naessens, Pranay Yadav, and Rebecca Williams. I have relied on your help for various parts of this project, and I would not have been able to collect any data or analyse anything without your support. Thank you also to the wonderful patients and controls who participated in this study. Participation required much time, concentration, and patience on your part. I am incredibly grateful for your participation and perseverance.

I was fortunate to be supported by the University of Cambridge Harding Distinguished Postgraduate Scholars Programme. The generosity of Claudia and David Harding not only made my PhD work possible, but the financial stability their programme provided ensured I could make my work my priority. Thank you ever so much. I also would like to express my sincere thanks to the Harding Administrative Team for their incredible support, in particular, Gabby Uncles.

A final few important thank yous.

To my lovely friends in Cambridge, Leipzig, Australia and beyond, I am very grateful for your support, your kindness, and your perspectives on the world. I am so lucky to have such wonderfully thoughtful and caring people in my life. Thank you for reminding me of what is most important. Special thank you shoutouts to Caitlin Duschene and Hannah Bostock for helping me with proofreading, not an easy task!

To my parents, Garry and Teo, my sister, Narelle, and little Ochre – thank you for being all that you are and for shaping me to be who I am.

Finally, and most importantly, I would like to thank (boy) Alex. You are my star.

Table of Contents

ABSTRACT	iii
PREFACE	vi
ACKNOWLEDGEMENTS	viii
TABLE OF CONTENTS	xii
LIST OF ABBREVIATIONS AND ACRONYMS	xiv

CHAPTER 1 INTRODUCTION	1
1.1 Overview and structure of thesis	1
1.2 Cognitive control: A distinguishing characteristic of human behaviour	4
1.3 Brain regions for control	6
1.4 The role of MD regions in supporting higher-order cognition	16
1.5 A general factor: Intelligence	27
1.6 Selective attention	32
1.7 Methods used in this thesis	38
1.8 Summary	47

CHAPTER 2 	
NEUROPSYCHOLOGICAL EVIDENCE FOR THE ROLE OF	
FRONTOPARIETAL CORTEX IN FLUID INTELLIGENCE	51
2.1 Foreword	51
2.2 Abstract	53
2.3 Introduction	54
2.4 Materials and Methods	60
2.5 Results	74
2.6 Discussion	90
2.7 Afterword	98
2.8 Supplementary Materials for Chapter 2	100

CHAPTER 3 	
SPATIAL AND FEATURE-SELECTIVE ATTENTION INTERACT TO DRIVE	
SELECTIVE CODING IN FRONTOPARIETAL CORTEX	109
3.1 Foreword	109
3.2 Abstract	111
3.3 Introduction	112
3.4 Materials and Methods	116
3.5 Results	138
3.6 Discussion	157
3.7 Afterword	163
3.8 Supplementary Materials for Chapter 3	164
 CHAPTER 4 	
HOW DO FOCAL BRAIN LESIONS AFFECT INFORMATION CODING	
DURING SELECTIVE ATTENTION?	173
4.1 Foreword	173
4.2 Abstract	176
4.3 Introduction	177
4.4 Methods	183
4.5 Results	210
4.6 Discussion	230
4.7 Afterword	244
4.8 Supplementary Materials for Chapter 4	246
 CHAPTER 5 	
GENERAL DISCUSSION	251
5.1 Summary of findings	253
5.2 Implications	256
5.3 Limitations and Future Directions	266
5.4 Conclusions	272
 REFERENCES	275

List of abbreviations and acronyms

ACC	Anterior cingulate cortex
AI/FO	Anterior insula/frontal operculum
AIC	Akaike Information Criterion
aIFS	Anterior inferior frontal sulcus
aLaF	Attended location, attended feature (relevant information)
aLuF	Attended location, unattended feature (irrelevant information)
ANOVA	Analysis of variance
AUC	Area under the curve
BA	Brodmann's Area
BF	Bayes Factor
BIC	Bayesian Information Criterion
BOLD	blood-oxygen-level-dependent (response)
CAB-NP	Cole-Anticevic Brain Network Parcellation
CCNRP	Cambridge Cognitive Neuroscience Research Panel
CI	Confidence interval
COM	Centre of mass
CON	Cingulo-opercular network
DAN	Dorsal attention network
dIPFC	Dorsolateral prefrontal cortex
DMN	Default mode network
EEG	Electroencephalography
EOG	Electrooculogram
EPI	Echo planar imaging
FB	Feedback
FDR	False discovery rate
FF	Feedforward
fMRI	Functional magnetic resonance imaging
FOV	Field of view
FPN	Frontoparietal network
FWE	Family-wise error
FWER	Family-wise error rate correction
FWHM	Full width at half-maximum

GLM	General linear model
GM	Grey matter
HCP	Human Connectome Project
HPI	Head position indicator
ICA	Independent component analysis
IFJ	Inferior frontal junction
IPS	Intraparietal sulcus
IQ	Intelligence quotient
LL	Log Likelihood
LME	Linear mixed effect models
LREC	Cambridge Local Research Committee
MD	Multiple-demand
MEG	Magnetoencephalography
MNI	Montreal Neurological Institute
MPI CBS	Max Planck Institute for Human Cognitive and Brain Sciences
MRC CBU	Medical Research Council Cognition and Brain Sciences Unit
MRI	Magnetic resonance imaging
MVPA	Multivariate pattern analysis
NART	National Adult Reading Test
P-FIT	Parietal-Frontal Integration Theory
PCA	Principal component analysis
PET	Positron emission tomography
PFC	Prefrontal cortex
pIFS	Posterior inferior frontal sulcus
PM	Premotor cortex
PRE	Psychology Research Ethics
RDMs	Representational dissimilarity matrices
ROC	Receiver Operating Characteristic
ROC AUC	Area Under the Receiver Operating Characteristic Curve
ROI	Region of interest
RSN	Resting-state network
RT	Reaction time
SALT	Sustained Attention Lighthouse Task
SD	Standard deviation
SN	Salience network

SPM	Statistical Parametric Mapping
SVM	Support Vector Machine
TE	Echo time
TFCE	Threshold-free cluster enhancement
TLV	Total lesion volume
TMS	Transcranial magnetic stimulation
TR	Repetition time
tSSS	Spatiotemporal signal-space separation
TVA	Theory of Visual Attention
uLaF	Unattended location, attended feature (irrelevant information)
uLuF	Unattended location, unattended feature (irrelevant information)
VAN	Ventral attention network
VIF	Variance inflation factors
VLPFC	Ventral lateral prefrontal cortex
WAIS	Wechsler Adult Intelligence Scale
WCST	Wisconsin Card Sorting Test
WM	White matter

Chapter 1 | Introduction

1.1 Overview and structure of thesis

This thesis explores how a network of frontoparietal regions, the multiple-demand (MD) system, supports intelligent human behaviour. Human cognition is remarkably complex and varied, and yet this network of regions repeatedly shows a common pattern of activity in response to a wide range of cognitive demands. These MD regions seem to support processes that are fundamental to complex thought and intelligence. Much work has already been done to elucidate the fascinating properties these regions share, and theories abound regarding how these regions facilitate some of the core cognitive processes that underpin complex human behaviour. However, the story of how MD regions support cognition is not yet complete. This thesis uses a number of different methodological approaches, including focal lesion models, quantification of grey matter resting-state network damage and white matter network disconnection, functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), MEG-fMRI fusion, and multivariate pattern analysis (MVPA) methods, to contribute to our understanding of the role these regions play in higher-order human cognition.

In **Chapter 1**, the present chapter, I will first briefly review the main concepts used in this thesis. I will clarify what I mean by higher-order cognition, focussing on the

Chapter 1

concepts of cognitive control, intelligence, and selective attention. I will also describe how these higher-order cognitive functions are linked within a broader theoretical framework which places the MD system at the centre. I will then describe some of the key methods used to relate human behaviour with brain activity, with a focus on those used in this thesis: focal lesion patient models, neuroimaging (fMRI, MEG, and their combination), and MVPA methods. I will discuss the advantages and limitations of these approaches in helping us understand how the brain realises cognition. I will discuss the existing gaps in our knowledge, challenges in addressing these gaps, and what this thesis aims to contribute.

The subsequent three experimental chapters report empirical work I have done to advance our understanding of how frontoparietal cortex supports intelligent behaviour.

Chapter 2 draws on a large set of neuropsychological data from 174 patients with focal brain lesions to explore the brain network(s) critical for fluid intelligence. I quantify the extent of damage to different networks, considering both grey matter damage and network disconnection through damage to white matter tracts. I then relate the extent of damage to an estimate of each patient's fluid intelligence deficit, derived with reference to a large set of age-matched controls. I show that damage to frontoparietal network, a resting-state network closely related to the MD system, predicts fluid intelligence deficit over contributions from other networks. However, in these data, that contribution could not be distinguished from the contribution of total lesion volume.

Chapter 3 uses fMRI and MEG-fMRI fusion to examine the role of the MD system in selective attention. Selective attention is a core component of intelligence and in this study, I use fMRI and MVPA methods to examine how attention affects task-related representations in the MD system. I specifically examine the interaction between two types of selective attention: spatial and feature attention, which are frequently studied separately but often used together in real life. I show that coding in the MD system is highly tuned to reflect task-relevant information, representing only information at the intersection of spatial and feature attention. In addition, I relate these fMRI findings to extant MEG work using MEG-fMRI fusion to explore whether representations in the MD system might be driving widespread focus on task-relevant information throughout the

brain. However, this analysis was unable to yield a clear timecourse for when these attention effects arose, leaving open the question of whether MD codes affect task-related representations elsewhere in the brain.

In **Chapter 4**, I take a novel approach to understanding how frontoparietal regions support higher-order cognition: combining lesion work, MEG, and MVPA methods. Building on the observation in Chapter 2 that frontoparietal regions seem to be important for higher-order cognition, and the observation in Chapter 3 that the MD system selectively prioritises the representation of task-relevant information, I ask whether and how attention processes change when the MD system is damaged. For this, I carry out an exploratory case series, in which I apply MVPA methods to MEG data collected from six patients with focal parietal lesions and ten matched controls. I examine the dynamics of neural attentional prioritisation (i.e., relative prioritisation of relevant information over irrelevant information) in individual patients and controls, and ask whether variation in this neural prioritisation relates to behavioural performance. Compared to controls, I show that one patient demonstrated subtle reductions in the relative prioritisation of relevant information against a backdrop of slower reaction times and high behavioural accuracy, suggesting that parietal damage may affect neural attentional processes. Moreover, I find preliminary evidence that inter-individual differences in attentional prioritisation may relate to inter-individual differences in behaviour, suggesting that this approach of measuring neural attentional processes may capture something meaningful about behaviour.

In **Chapter 5**, the General Discussion, I synthesise the preceding chapters and reflect on the implications of the thesis results and how they add to our understanding of how frontoparietal regions give rise to complex human behaviour.

1.2 Cognitive control: A distinguishing characteristic of human behaviour

1.2.1 Setting the stage: Cognitive control and the multiple-demand system

Consider what is involved in preparing dinner in the evening. This ostensibly simple task involves many mental processes and actions: sensory abilities (seeing or feeling for the requisite ingredients and equipment); problem-solving (deciding what to make on the basis of ingredient availability and time constraints); planning (organising how and when to get additional ingredients if required, deciding when to start cooking in order to have dinner at a reasonable hour); task decomposition (working out the steps involved in making the meal and assembling these steps into a coherent sequence of actions); working memory (keeping those steps in mind); motor skills (reaching for, preparing, and cooking ingredients); and flexible selective attention (switching the focus of attention to each of the steps in turn while ignoring potential distractors that could derail your efforts). To accomplish this task effectively, it is apparent that ‘control’ is necessary to direct these varied mental processes towards the overarching goal. Our capacity for cognitive control, that which enables us to execute this task with ease, represents a distinguishing characteristic of human cognition¹.

Cognitive control² refers to our ability to flexibly and rapidly coordinate thought, perception, and action in line with internally maintained goals. In short, cognitive control is what underpins our ability to engage in purposeful behaviour. Many processes

¹ I use a broad conceptualisation of cognition here, one that encompasses perception, thinking, and behaviour.

² Executive function and attentional control are related to cognitive control and are often used to refer to similar processes (e.g., see J. D. Cohen, 2017; Friedman & Robbins, 2022; Menon & D’Esposito, 2022). Similarly, I consider these terms as synonyms in this work.

have been linked to this concept, including, *inter alia*, planning, working memory, attention, compositionality, mental flexibility, inhibition, and conflict monitoring (e.g., see Friedman & Robbins, 2022).

These terms suggest a mechanism that is required to be selective (be able to prioritise processing of goal-related information in the environment) and flexible (be able to update what is relevant as the goal or circumstances change). Understanding the neural mechanisms underpinning these two core features of cognitive control is key to understanding how we are capable of complex, purposeful behaviour. This thesis aims to contribute to our evolving understanding of how our brains enact control.

Our starting place for understanding the neural mechanisms underpinning cognitive control comes from a reliable finding: a wide range of cognitively challenging tasks consistently engages a specific set of frontoparietal regions (notably Assem, Glasser, et al., 2020; Duncan & Owen, 2000; Fedorenko et al., 2013). These regions comprise bilateral inferior frontal sulcus, anterior insula/frontal operculum, dorsal anterior cingulate cortex/pre-supplementary motor area, and intraparietal sulcus. I refer to this network of frontoparietal regions as the ‘multiple-demand’ (MD) system to reflect its broad recruitment across demands (following Duncan, 2010, 2013). What can we make of the fact that these regions are reliably engaged when tasks become more demanding? Does its common recruitment point to a core, uniting process that underpins cognitive control, and thus complex action and thought?

Over the course of this Introduction, we will arrive at a prominent hypothesis which links the core function of the MD system to attentional integration (Duncan, 2013; Duncan et al., 2020). Within this framework, complex thoughts and behaviours are conceptualised as arising from breaking a task down into a series of simpler parts, or sub-tasks, within which attention integrates the fragments of cognition required for the current part (attentional episodes, following Duncan, 2013). Organised behaviour is thought to depend on attentional episodes being structured into an overall mental control program, with each part focussed on in turn. A wealth of research is emerging to support the idea that MD regions play an important role in this process, particularly in novel contexts. I will consider several lines of research that take us to this endpoint

and highlight where my work contributes to advancing our understanding of this process.

1.3 Brain regions for control

1.3.1 Frontoparietal regions and control: Neuropsychological evidence

The coordination of cognitive processes to give rise to purposeful behaviour has long been considered the primary function of the frontal lobes, especially prefrontal cortex (PFC)³ (see E. K. Miller & Cohen, 2001). Evolutionary, developmental, and clinical work supports this idea.

PFC is present in all mammalian species and is most developed in primates. The degree of expansion of PFC, which is particularly pronounced in humans, is strongly linked to the emergence of complex cognitive, higher-order⁴ capabilities (E. K. Miller & Cohen, 2001). Moreover, cognitive functions appear rather late in development, paralleling the late maturation of PFC relative to the rest of the brain (Cohen & Casey, 2017). Finally, PFC represents an interconnected network that links motor, sensory, and subcortical areas through extensive reciprocal projections (e.g., Selemon & Goldman-Rakic, 1988).

³ Prefrontal cortex (PFC) refers to the more anterior parts of the frontal lobes, outside of the primary motor cortex and secondary motor areas.

⁴ Higher-order or high-level cognitive functions refer to processes at a higher level of abstraction, for example, cognitive control processes, memory, language, and social behaviour. These processes stand in contrast to lower-level, more automatic processes which include, inter alia, sensory processing, reflexive motor actions, and well-learned skills (E. K. Miller & Wallis, 2009).

It is thus well situated to integrate and coordinate processing across the brain to facilitate goal-directed behaviour.

In support of this picture are the disordered, stimulus-driven behavioural profiles that are associated with PFC damage in human and non-human primates. Broadly speaking, damage to these regions is associated with a complex picture: disordered, inflexible, inappropriate, or disinhibited behaviour against a backdrop of seemingly intact (at least on superficial examination) perceptual abilities, motor skills, and performance on intelligence tests, reasonably unimpaired knowledge, and memory (e.g., Lhermitte, 1983; Lhermitte et al., 1986; Penfield & Evans, 1935; Shallice & Burgess, 1991). Such characteristics emphasise an impairment in goal-orientation, planning, focussed attention, and flexibility.

For example, following the removal of a right frontal glioma in his sister, Wilder Penfield (Penfield & Evans, 1935) describes her capable of intelligent conversation and her memory seemed unimpaired. Yet her ‘...capacity for planned action [was] clearly defective’ (p. 132):

‘...she had planned to get a simple supper for one guest [Penfield] and four members of her family. She looked forward to it with pleasure and had the whole day for preparation. This was a thing she could have done with ease ten years before. When the appointed hour arrived she was in the kitchen, the food was all there, one or two things were on the stove, but the salad was not ready, the meat had not been started and she was distressed and confused by her long continued effort alone. It seemed evident that she would never be able to get everything ready at once. With help the task of preparation was quickly completed and the occasion went off successfully with the patient talking and laughing in an altogether normal way.’ (p. 9)

Cognitive rigidity or lack of goal-directed behaviour is also characteristic following frontal damage. For instance, the Wisconsin Card Sort Test (WCST; Grant & Berg, 1948) is an executive function test that is considered to be sensitive to frontal damage (e.g., Demakis, 2003; Milner, 1963). This task requires participants to sort cards according to one of three features: colour, shape, or number. They are only told whether their sorting technique is correct, so they must learn the rule associated with sorting. This task also involves switching: at some point, participants are informed they are incorrect and are required to learn a new sorting rule. Impaired performance on this task, typically in the form of inflexibility or perseverative errors (an inability to shift from sorting according to an old rule), is commonly associated with frontal damage. Relatedly, failure to perform a task, despite understanding and remembering task instructions, what is referred to as goal neglect, is also associated with frontal damage (Duncan et al., 1996). These deficits essentially reflect an inability to engage in goal-directed behaviour and are characterised by behaviour that lacks focus. Indeed, poor performance on WCST and goal neglect are argued to reflect deficits in fluid intelligence, a process closely linked to the frontoparietal MD network (Duncan, 2013; Duncan et al., 2020; Roca et al., 2010). I will revisit this in Section 1.5.2.

Damage to regions of parietal cortex have also been associated with processes related to cognitive control, particularly attention. Most notably, unilateral damage to (commonly right) posterior parietal regions⁵ can lead to pronounced unawareness of stimuli presented on the contralesional side of space despite intact sensory functions, a

⁵ I note here that posterior parietal lesions do not exclusively, nor definitively, lead to unilateral spatial neglect (Driver & Mattingley, 1998; Karnath et al., 2001, 2011; Moore et al., 2023; Vallar, 1993). However, reduced awareness of the contralesional side is widely accepted to be one of the most prominent disorders to follow lesions to this region (e.g., in Lezak et al., 2012).

phenomenon typically referred to as unilateral spatial neglect⁶. In a milder form of neglect, extinction, patients can perceive a contralesional item presented alone yet fail to detect this item when concurrently presented with ipsilesional items. The phenomenon of extinction suggests that stimuli ‘compete’ for attentional representation and highlights the attentional nature of this syndrome. Indeed, the severity of extinction can be reduced if items are perceptually grouped to represent a single object, and thus are no longer in competition (Mattingley et al., 1997; Ward et al., 1994). I will return later to the idea that attention serves to resolve competition by prioritising certain inputs in Section 1.6.

In summary, neuropsychological studies suggest frontal and parietal regions may support control processes. Lesion studies, however, are coarse and can only provide us with some of the story. With the advent of neuroimaging, we have gleaned a more nuanced understanding of the neurobiological mechanisms of cognitive control, and a more complex picture.

1.3.2 Frontoparietal regions and control: Neuroimaging evidence

1.3.2.1 Introduction to frontoparietal networks

The role of frontoparietal regions in supporting cognitive control has been further informed by neuroimaging. Initial work sought to ascribe cognitive control processes to

⁶ I have used the term ‘neglect’ in this work to refer to contralesional spatial inattention or bias, in keeping with how this phenomenon is widely referred to in the literature. However, I note that some take issue with the term as it seems to imply, incorrectly so, intention or deliberateness on the part of the patient (e.g., Halligan & Marshall, 1998). Accordingly, for instance, Lezak et al. (2012) in their foundational text on neuropsychological assessment instead use ‘left-sided inattention’ or ‘right-sided inattention’ to refer broadly to ‘unilaterally depressed awareness’ in an effort to provide a more neutral description of the disorder.

specific frontoparietal regions (see Friedman & Robbins, 2022), consistent with a modular view of the brain where regions act as independent processors for distinct cognitive functions.

More recent work has emphasised a network-oriented perspective of the brain, arguing that the brain is better understood as organised into networks rather than modular functional regions. Individual brain regions may still have differentiated roles, but this network view suggests that cognitive constructs are supported by interactions between network nodes. In PFC, several networks have been identified based on their functional connectivity during tasks or resting-state connectivity (see Menon & D'Esposito, 2022 for a detailed summary of these main PFC networks). These PFC networks include the frontoparietal network (FPN); salience network (SN); cingulo-opercular network (CON); ventral attention network (VAN); dorsal attention network (DAN); and default mode network (DMN) (see Figure 1). These networks can be considered 'domain-general'⁷ as they are not specifically associated with any one cognitive domain.

⁷ c.f. domain-specific networks which are linked to a specific cognitive domain (e.g., visual information is processed in a specialised visual system).

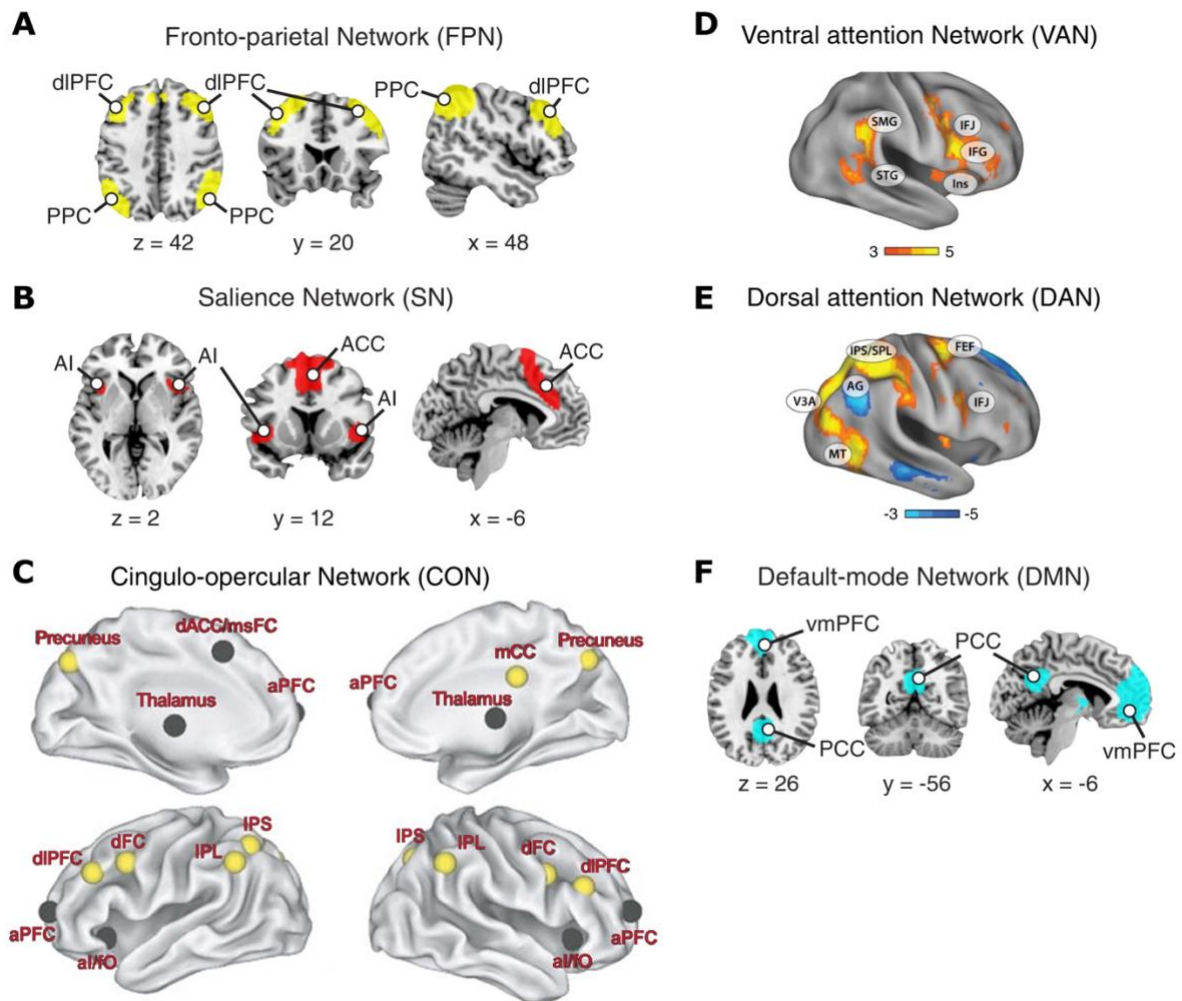


Figure 1. Six PFC networks as described in Menon & D'Esposito (2022). Panel A: Frontoparietal network (FPN), with key nodes in dorsolateral PFC (dlPFC) and posterior parietal cortex (PPC). Panel B: Salience network (SN), with key nodes in anterior insula (AI) and dorsal anterior cingulate cortex (ACC). Panel C: Cingulo-opercular network (CON, black) with key nodes in anterior insula/frontal operculum (aI/fO), dorsal ACC and medial superior frontal cortex (dACC/msFC), anterior PFC (aPFC) and thalamus, as distinguished from the FPN (yellow). Panel D: Ventral attention network (VAN), with key nodes in insula (Ins), inferior frontal junction (IFJ), supramarginal gyrus (SMG), and superior temporal gyrus (STG). Panel E: Dorsal attention network (DAN), with key nodes in frontal eye fields (FEF), inferior frontal junction (IFJ), intra-parietal sulcus and superior parietal lobule (IPS/SPL), angular gyrus (AG), visual area 3A (V3A), and middle temporal visual area (MT). Panel F: Default mode network (DMN), with key nodes in ventromedial PFC (vmPFC) and posterior cingulate cortex (PCC). Figure from Menon & D'Esposito (2022), reproduced with permission.

Among these, FPN and, to a lesser extent, SN and CON are considered important for supporting cognitive control processes. Research by Dosenbach and colleagues (Dosenbach et al., 2006, 2007) found FPN and CON are consistently active during cognitive control tasks, leading them to propose a dual-network model for top-down control (Dosenbach et al., 2008). Within this framework, FPN is thought to provide rapid signals to initiate and adjust control, whereas the CON is responsible for maintenance of the task set over time. However, more recent work suggests a more complex picture. For instance, Lorenz et al. (2018) directly compared neural activation between both networks across a large space of cognitive tasks, finding FPN and CON were not highly dissociated for tasks as predicted by the dual-network model. Indeed, Dosenbach and colleagues have more recently revised this theory, proposing a more action-oriented focus for CON (Dosenbach et al., 2024). Other work has emphasised interactions between FPN and other networks as being critical to our capacity for control. For instance, FPN and CON show greater integration (i.e., functional connectivity between these regions increased) with increasing cognitive control demands (J. R. Cohen et al., 2014). Moreover, in this work, better task performance was associated with stronger connectivity between these cognitive control networks and the task-related somatosensory-motor network, underlining the importance of network integration for cognitive efficiency. Greater separation between networks at rest has also been associated with better performance on cognitive tasks (Bertolero et al., 2018; Sripada et al., 2019) with, for example, separation in the degree of activation between FPN and DMN shown to be predictive of general intelligence (Sripada et al., 2019).

In addition, FPN is among the most functionally connected resting-state network in the brain (Cole et al., 2010; Schultz et al., 2022), an important property for a network involved in cognitive control. Moreover, its connectivity with other brain networks is highly flexible. More so than other resting-state networks, FPN dynamically updates its brain-wide functional connectivity patterns in accordance with task demands, consistent with a ‘flexible hubs’ mechanism for cognitive control (Cocuzza et al., 2020; Cole et al., 2013; Soreq et al., 2019). Recent work shows that the extent of a region’s inter-network connectivity correlates with its domain-generalty (i.e., the degree to which diverse task information can be decoded), and that the pattern of task

information held in FPN can be predicted from its functional connectivity profile along with information content present in connected brain regions (Schultz et al., 2022). Moreover, computational models suggest that rapid plasticity within network nodes, like those in FPN, allows for dynamic shifts in neural representation and attention, providing a theoretical basis for how these networks can adapt to changing cognitive demands (Manohar et al., 2019). Taken together, these findings suggest that FPN's broadly connected and highly adaptable functional connectivity is key to its capacity to flexibly represent and integrate diverse task-related information from across the brain.

1.3.2.2 The multiple-demand (MD) system

Another way to examine the neural basis for cognitive control is by considering which regions are active when participants perform challenging tasks. In human functional magnetic resonance imaging (fMRI), a strikingly reliable finding is that diverse demanding tasks recruit the same set of frontoparietal regions, likely reflecting a common need for cognitive control. This pattern of activation across diverse tasks defines the regions known as the MD system (Duncan, 2010).

Consistent coactivation across frontoparietal cortex in response to several different cognitive demands was initially reported in a meta-analysis of 20 neuroimaging studies in the 1990s (Duncan & Owen, 2000). With advancements in neuroimaging methods, the delineation of this network has been further refined (e.g., Fedorenko et al., 2013; but notably Assem, Glasser, et al., 2020). Generally, we consider the MD system to incorporate areas around inferior frontal sulcus, anterior insula, frontal operculum, dorsal anterior cingulate cortex, presupplementary motor area, and intraparietal sulcus (see Figure 2). These regions have also been referred to as the task positive network (Fox et al., 2005), frontoparietal control network (Vincent et al., 2008), task-activation ensemble (Seeley et al., 2007), or cognitive control network/networks (Cole & Schneider, 2007; Gratton et al., 2018).

The MD system was recently mapped out with a high degree of spatial specificity. Using task-based fMRI data from 449 healthy volunteers from the Human Connectome Project (HCP) database and the Glasser parcellations, Assem, Glasser and colleagues (2020) ascertained which regions showed consistently stronger activation across several hard versus easy contrasts. This work revealed a highly interconnected core MD system, reliably recruited across all diverse tasks, surrounded by a penumbra of additional regions. A comparison of Assem, Glasser, et al.'s (2020) recent definition of MD system is shown with the classic definition (Fedorenko et al., 2013) in Figure 2 (Panel A and B). Notably, Assem, Glasser, et al.'s (2020) work points to a more distributed MD system than had been observed in earlier conceptualisations⁸.

⁸ The recent, more precise anatomical definition of the MD system by Assem, Glasser et al. (2020) also notes the inclusion of subcortical regions (such as basal ganglia and thalamus), temporal regions and cerebellum as part of the MD system. Understanding the role of these regions in supporting cognitive control is still at its early stages and beyond the scope of the current work. Accordingly, reference to the MD system in this work generally refers to frontoparietal regions of the MD system.

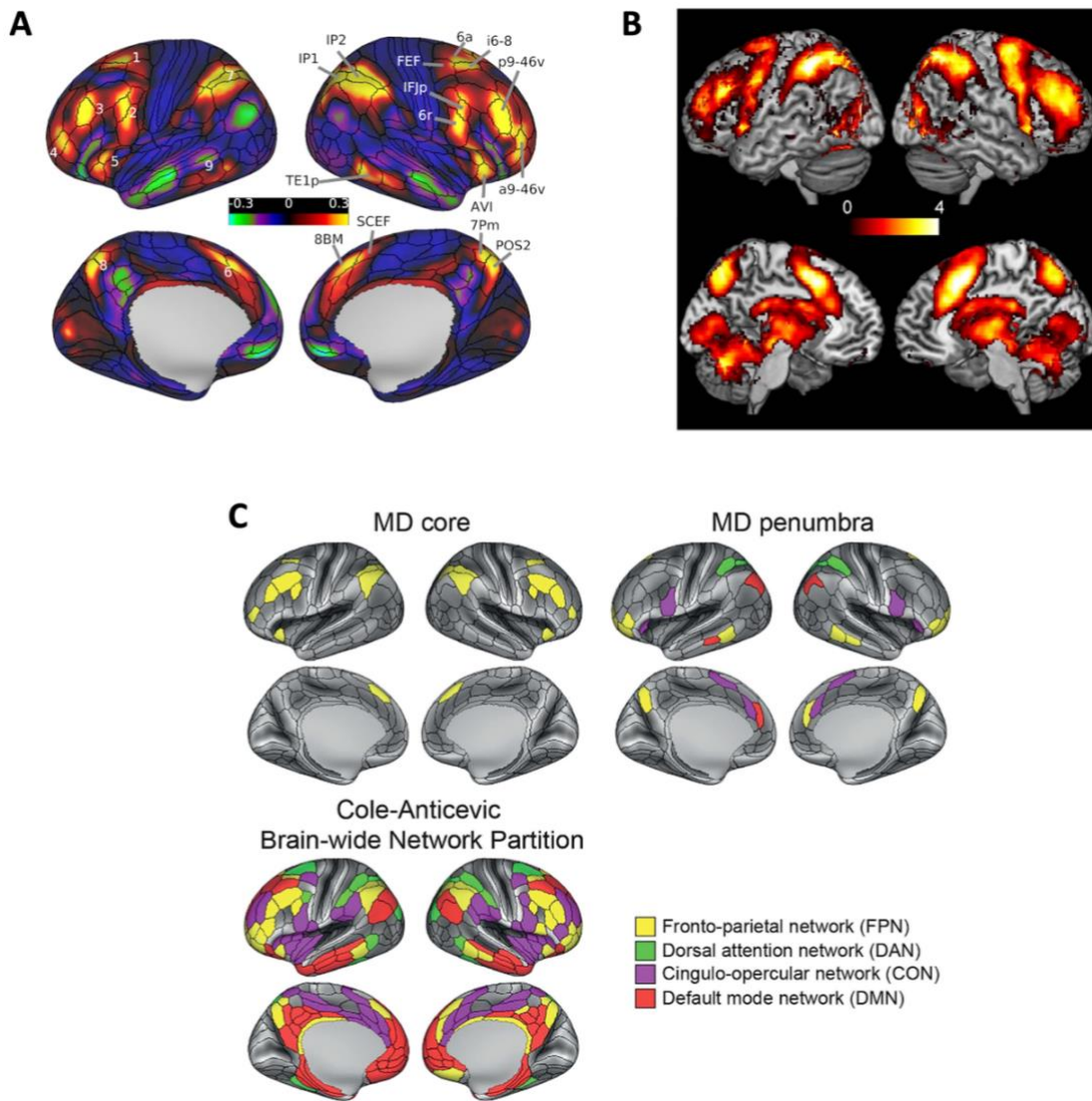


Figure 2. Panel A: MD regions, indicated in hot colours, identified in Assem, Glasser, et al. (2020). Image represents average of the three HCP group average task contrasts used in this study (i.e., Working Memory 2-back > 0-back, Relational Reasoning Hard > Easy, Math > Story). Values are beta estimates. Black contours correspond to the HCP multimodal parcellation MMP_1.0 (210V) areal borders. Numbers on the left hemisphere correspond to visually separable patches of activity distributed across the cortex. Panel B: Volumetric MD map from Fedorenko et al. (2013) computed by averaging seven hard > easy task contrasts (2-mm smoothed) displayed on a volume rendering lateral surface (above) and medial slice (below) of the Montreal Neurological Institute (MNI) template. Values are t -statistics. Fedorenko's MD definition is used in Chapter 3. Panel C: The MD system identified in Assem, Glasser, et al. (2020), shown as core (top left) and penumbra (top right), and relative to resting-state networks from Cole-Anticevic Brain Network Partitioning: CAB-NP (Ji et al., 2019) (bottom left). As shown, core MD regions are shown to overlap with CAB-NP FPN (shown in yellow). Figures from Assem, Glasser, et al. (2020), reproduced under CC BY licence.

To relate the MD system to the FPN mentioned in Section 1.3.2.1, Assem, Glasser, et al. (2020) compared the MD system with an established fMRI resting-state network parcellation (specifically, Cole-Anticevic Brain Network Parcellation: CAB-NP (Ji et al., 2019)). Here, all 10 core MD regions were encompassed within CAB-NP's FPN (see Figure 2, Panel C). Accordingly, for the purposes of this work, I will consider the resting-state FPN as being a reasonable proxy for MD system, where applicable (for instance, in Chapter 2).

In summary, we have seen that a core frontoparietal MD system is commonly engaged across a range of diverse demands. If we take activation to signify involvement, what role do these regions play in supporting cognitive control, or flexible goal-directed behaviour more broadly? Emerging consensus is that the structure of cognition is hierarchical, and that MD regions play a critical role in attentional integration, controlling cognition through the principle of compositionality.

1.4 The role of MD regions in supporting higher-order cognition

1.4.1 Attentional episodes framework

The common recruitment of MD regions across diverse demands begs the question of whether a common process underlies complexity, something that lies at the heart of control processes and the primary function of the MD system. A prominent proposal links cognitive control, intelligence, and the function of MD regions to the assembly of focussed attentional episodes, which are hierarchically structured in service of an overarching goal (Duncan, 2013; Duncan et al., 2020). I will first describe what is meant by attentional episodes and how the MD system might support cognitive control, before considering how this framework relates to intelligence and attention. For ease of reference, I will refer to this proposal as the 'attentional episode framework'.

At a foundational level, the attentional episode framework presupposes two things. First, that there is an almost infinite space of possible actions, states, and inputs available to us to enact or process. We could consider here, for example, the wealth of sensory information available to us at any one time and the multitude of ways we could think about or respond to these inputs. Second, purposeful behaviour depends on constraining this infinite possibility space in accordance with the principle of compositionality.

This proposal builds on the seminal work of Miller and colleagues (1960). They posit that behaviour is organised hierarchically, with goals central to the structure of cognition and behaviour. Within the Miller et al. (1960) framework, behaviour is articulated through a series of Test-Operate-Test-Exit (TOTE) units. Each unit focusses on meeting a specific goal by evaluating the current state, taking actions to close the gap towards the desired state, and reassessing until the goal is attained. Within Miller et al.'s (1960) framework, these elementary units can then be arranged hierarchically to meet short- and long-term goals.

Similarly, the attentional episode framework posits that all thought and behaviour can be segmented into a series of simpler parts or sub-tasks which are then structured and assembled to achieve a goal. Each part represents the necessary cognitive fragments needed to solve a specific subproblem, constraining the possibility space to just the relevant elements. Following Duncan (2013), we can think of these parts as attentional episodes, as successful completion of a behaviour requires attention be directed to only the elements contained within each episode.

Goal-directed behaviour can then be thought to rely on structured mental programs, whereby these attentional episodes are momentarily assembled according to the requirements of the current task. The process for creating and executing these programs involves several key steps (Duncan, 2010), notably segmentation, selective attention, and dynamic updating of the current focus of attention. That is, in order for behaviour to be achieved, sub-parts or attentional episodes of a task need to be defined and assembled (segmentation); attention then needs to focus on just the elements contained within an attentional episode (selective attention) and integrate them appropriately;

and then, after a sub-task is completed, focus needs to shift to the next set of elements relevant for the subsequent episode (dynamic updating).

The MD system is thought to play a central role in supporting this process. More specifically, Duncan posits that a core role of MD regions lies in defining and controlling the parts of the structured mental program (Duncan, 2010) and in integrating the elements needed within each episode (Duncan et al., 2020). Its involvement is considered particularly critical when tasks are novel or are more challenging. This framework explains why difficulty serves as a localiser for this network. Solving novel or difficult tasks arguably relies more heavily on the segmentation and definition of attentional episode assemblies than well-learned or simple tasks do, thus resulting in greater recruitment of the MD system.

1.4.2 Neural mechanisms for control

The attentional episodes framework proposes that the MD system integrates diverse information into a structured assembly of attentional episodes required for behaviour. Accordingly, two key predictions that stem from this framework are that the MD system should selectively code information relevant to a given task (selective attention), and flexibly update this representation (dynamic updating) as the task changes (Zheng et al., 2024).

A prominent proposal is that neurons adapt their coding to reflect what is currently relevant to behaviour, thereby flexibly coding a range of different information as required by a given task. This ‘adaptive coding hypothesis’ (Duncan, 2001) thus predicts that frontoparietal regions should be able to encode a broad array of information types, capturing the task-relevant dimensions of a task. Essentially, adaptive coding in the MD system is purported to provide the basis for flexible goal-directed behaviour, allowing for the necessary elements of attentional episodes to be dynamically integrated and utilised (Duncan, 2010). Converging evidence supports the idea that MD regions exhibit

flexibility in their coding properties that align with the adaptive coding hypothesis; I will consider some of this evidence in this section.

1.4.2.1 Frontoparietal neurons show adaptive coding, evidence from non-human primates

Single-unit recordings in non-human primates provided initial evidence that frontoparietal neurons show flexibility and selectivity in their responses that reflects behaviourally relevant distinctions. Activity in many PFC and posterior parietal neurons can distinguish between targets and non-targets (e.g., Bichot et al., 1996; Everling et al., 2002) and has been shown to dynamically reflect different aspects of the task, including cues, responses, abstract task rules, and behaviourally relevant information about target stimuli (e.g., Asaad et al., 1998; Freedman et al., 2001; Freedman & Assad, 2006; Rainer et al., 1998; e.g., Rao et al., 1997; Sakagami & Niki, 1994; Toth & Assad, 2002; Wallis et al., 2001).

For instance, Freedman and colleagues examined coding of learned categories in lateral PFC (Freedman et al., 2001). Stimuli in this task were computer-generated morphs of various cat and dog prototypes, such that images were blended to vary along a continuum (Figure 3). Monkeys, unfamiliar with cats and dogs, were trained to distinguish between these two classes of stimuli while ignoring within-category differences.

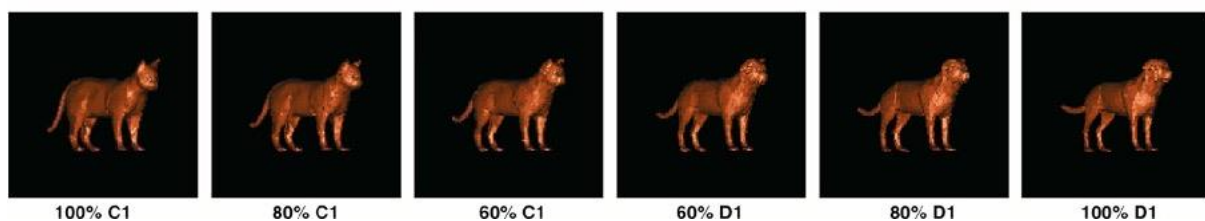


Figure 3. Examples of cat (C1) and dog (D1) stimuli morphs used in Freedman et al. (2001). Stimuli were created to vary along a continuum from 100% cat to 100% dog. Note, more prototypes of cat and dog were used in the study than are displayed here. The initial task was to discriminate between cat (the first three stimuli) and dog (the last three stimuli). Figure from Freedman et al. (2001), reproduced with permission.

Single-unit recordings from lateral PFC made during this task demonstrated category-selective activity in nearly a third of recorded neurons. That is, PFC activity tended to reflect category membership of stimuli (task-relevant distinctions), rather than distinguishing between individual stimuli within the same category (task-irrelevant distinctions) (Figure 4).

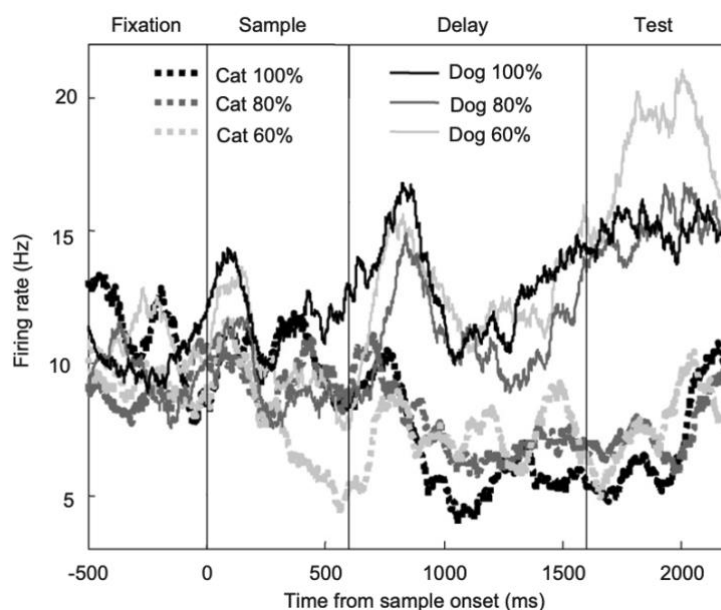


Figure 4. PFC single neuron. A spike-density histogram showing mean firing rate across different conditions for a category-sensitive PFC neuron. As shown, the neural response is more distinct between the cat-dog categories than within these individual categories. Figure from Freedman et al. (2001), reproduced with permission.

In a further demonstration of the flexibility of PFC coding properties and their link to task-relevance, one monkey was retrained to perform a different categorisation task with the same stimulus set. Single-unit recordings taken from the same site showed PFC neural activity now reflected the newly learned categorical distinctions; the old categories (which were now not relevant to the task) were no longer coded. These neurons demonstrated adaptive coding, with neural responses shaped by different inputs depending on their behavioural relevance. Moreover, this property was not constrained to PFC. Using a related paradigm, activity in posterior parietal neurons

similarly reflected the task-relevant category boundary, that also reconfigured to reflect a new task-relevant category boundary after retraining (Freedman & Assad, 2006).

Reconfiguration of task-related representations has also been shown to occur much more rapidly, for instance, within a single trial. For example, monkeys were trained on a working memory task in which they were first presented with a target object, and subsequently a second display where the target and a distractor were presented in two locations; monkeys were required to indicate the location of the target after a delay (Rao et al., 1997). Accordingly, what was considered task-relevant changed over the course of the trial (i.e., identity after the first display, location after the second display). PFC cells were found to code task identity *and* location, and moreover showed selectivity to the particular phase the monkey was (e.g., identity was only coded for the initial part of the trial, and location was more strongly coded later in the trial). Similar flexibility in coding with task switching has been seen in other studies (e.g., Asaad et al., 2000; Rainer et al., 1998). These studies show that activity in PFC and parietal cortex is strongly tuned to the particular task the monkey has learned, and moreover illustrate that representations flexibly reconfigure as the task context changes. Importantly, flexibility here also allows for selectivity: activity tends to emphasise task-relevant distinctions over irrelevant distinctions (e.g., Freedman et al., 2001; Freedman & Assad, 2006).

More recent non-human primate work demonstrates that frontoparietal neurons respond to diverse, non-linear combinations of task-related variables (e.g., Aoi et al., 2020; Mante et al., 2013; Naya et al., 2017; Rigotti et al., 2013; C. Y. Zhang et al., 2017), what has been termed ‘mixed selectivity’ (Rigotti et al., 2013). While this property is not restricted to frontoparietal neurons (e.g., McKenzie et al., 2014), mixed selective neurons are abundant in these regions (Rigotti et al., 2013). Mixed selectivity, a concept closely related to adaptive coding, is now considered central to understanding how these regions may facilitate complex behaviour. For instance, mixed selectivity allows for high-dimensional neural representations that can capture complex information within the same neural population, conferring a significant computational advantage over low-dimensional, highly specialised neural coding systems (Fusi et al., 2016). These higher-dimensional representations can then be readout in more domain-specific

regions to implement a wide range of behavioural responses. However, the flexibility that comes with increasing the dimensionality of the neural representation space incurs a cost to its generalisability, which, for example, may be important when using an existing control structure in a novel context (Badre et al., 2021; Fusi et al., 2016). Thus, control may be enacted through adaptive leveraging of low- and high- dimensional representations, depending on whether flexibility or generalisability demands dominate (Badre et al., 2021). These studies suggest a mechanism through which mixed selective neurons can build flexible cognitive structures.

To summarise, non-human primate work has demonstrated that frontoparietal neurons can represent complex combinations of task-related information and dynamically adjust their neural response to reflect current task-specific demands. The flexible and adaptive coding properties these neurons exhibit may facilitate the integration of behaviourally relevant information and thus play an important role in implementing control functions in the brain.

1.4.2.2 Adaptive coding in the human brain

In the human brain, the candidate regions for adaptive coding are the MD system. As discussed in Section 1.3.2.2, univariate fMRI studies show MD regions show consistent activation across a range of cognitive demands. This is in keeping with the adaptive coding hypothesis, which predicts neurons with highly dynamic, task-linked response properties would be active across a broad spectrum of tasks. However, common recruitment across different tasks is not necessarily indicative of adaptive coding. For instance, greater activity in MD regions in more difficult tasks could reflect a global response unrelated to adaptive coding, such as effort. Indeed, difficulty has been shown to increase the response of neurons (e.g., Spitzer et al., 1988).

In contrast to univariate effects, multivariate pattern analysis (MVPA) decoding methods⁹ allow us to examine what information is coded in the brain, and thus test whether MD regions demonstrate adaptive coding properties. I will describe MVPA decoding methods in more detail in Section 1.7.2, but in brief, these methods typically involve applying machine learning classification to patterns of neural activity elicited in response to different events, such as different stimuli (e.g., Haynes & Rees, 2006). Information coding is inferred when different events elicit reliably different patterns of neural activity. For instance, if the patterns of activation evoked in response to red and green stimuli are distinct in a given brain region, such that a classifier can reliably predict whether the stimulus was red or green based on these patterns, we would conclude this region codes colour information.

A 2016 meta-analysis of 100 MVPA analyses of fMRI data highlighted the domain-generalty of MD regions, revealing their capacity to code many different aspects of tasks, including auditory, visual, motor and rule information (Woolgar et al., 2016). In contrast, sensory and motor networks predominantly coded information relevant to their specific domains. This work underscored the broad representational capacity of MD regions.

Other work has shown that the MD system is finely tuned to preferentially represent information relevant to the current task, leveraging its flexibility for task-specific processing (e.g., Jackson et al., 2017; Jackson & Woolgar, 2018; Pischedda et al., 2017; Woolgar et al., 2011; Woolgar, Williams, et al., 2015). For instance, during a visual categorisation task where participants were asked to categorise novel objects on the

⁹ My use of ‘decoding’ and ‘MVPA methods’ in this work generally refers to supervised linear machine learning classification methods (described here in brief and in Section 1.7.2 in more detail), but other MVPA methods to examine representational content exist (e.g., multivoxel pattern correlation analysis, see Haxby et al., 2001, or representational similarity analysis, RSA, see Kriegeskorte et al., 2008). Accordingly, I generally describe representations in terms of strength (e.g., decoding accuracies) rather than in terms of coding format (e.g., representational geometry, as with RSA, see Kriegeskorte & Kievit, 2013).

basis of one of two stimulus dimensions (length or orientation), MD regions demonstrated stronger representation of the task-relevant feature than the irrelevant feature (Jackson et al., 2017). In addition, disruption of one MD node (right dorsolateral prefrontal cortex, dlPFC) with transcranial magnetic stimulation (TMS) led to a decrease in task-relevant coding in unstimulated MD regions, indicating that right dlPFC has a causal role in enhancing coding of relevant information (Jackson et al., 2021).

Moreover, MD representations appear to adjust in accordance with task requirements. For instance, when the perceptual demands of a task increase, and thus likely require more top-down control, MD regions accordingly show stronger (more readily decodable) representations of task-relevant features (Woolgar et al., 2011; Woolgar, Williams, et al., 2015), despite weaker representation of this information in earlier processing areas (Woolgar et al., 2011). Similarly, more complex rules yield stronger representation of task rules relative to simpler rules (Woolgar, Afshar, et al., 2015), and training on a task leads to diminished representations in these regions (Woolgar et al., 2011). These findings fall within a broader literature suggesting that strength of task-relevant codes in frontoparietal regions are modulated by task demands (e.g., Muhle-Karbe et al., 2017; Sobrado et al., 2022; Wisniewski et al., 2023).

In summary, in humans, the MD system is thought to facilitate goal-directed behaviour through adaptive coding, whereby neurons dynamically adjust coding to selectively represent information currently relevant for behaviour. Corroborating this, functional neuroimaging work demonstrates that these regions show capacity to code a wide array of information types and to adjust representations in response to task demands.

1.4.2.3 Top-down influence of MD coding

It is not sufficient, however, for MD regions to merely hold this task-relevant information; for the MD system to facilitate cognitive control, these regions must

interact with more specialised regions to drive the brain-wide goal-oriented response necessary for adaptive behaviour. In line with this, the biased competition hypothesis posits that selective representations in the MD system may act as a source of bias on processing across the system, driving selective processing of task-relevant information in more specialised regions, such as visual cortex (Desimone & Duncan, 1995).

As mentioned previously, one recent study (Jackson et al., 2021) perturbed activity in right dlPFC with TMS while concurrently recording neural activity using fMRI. In addition to reducing coding of relevant information in the unstimulated MD system relative to a control condition, perturbation of right dlPFC had a downstream effect on coding in higher visual cortex (namely, reduced task-related coding overall in these regions). This result provides direct support for the proposal that right dlPFC causally affects information processing in specialised brain regions. However, this distal effect was not specific to task-relevant information coding (i.e., TMS also affected coding of irrelevant visual information). This limits the conclusions that can be drawn about the role of the MD system in prioritising task-relevant information specifically.

Another way to assess the influence of the MD system on processing elsewhere in the brain is to consider the temporal dynamics of information coding, for example, by asking whether attention effects in sensory areas, such as visual cortex, follow from the preferential coding of task-relevant information in frontal lobes. Goddard et al. (2022) used magnetoencephalography (MEG) to investigate information exchange between frontal and occipital regions during a visual attention task. This task (which is described in detail in Chapters 3 and 4) required participants to indicate the colour or shape of one of two presented objects. Information exchange was investigated using information flow analysis (Goddard et al., 2016). This method uses the logic of Granger causality to track the exchange of task-related information, as captured in representational dissimilarity matrices (RDMs). In brief, the method uses partial correlations to identify instances in which a region's RDM at one time point (e.g., region X, time 0ms) can predict that of another region at a later time point (e.g., region Y, time 150ms), after partialling out variance attributable to that second region at the earlier time point (e.g., region Y, time 0ms). These instances then represent periods of information exchange.

Goddard et al. (2022) found that attention enhanced representation of the attended, or task-relevant, stimuli in occipital cortex, and that the timecourse of these effects corresponded to the onset of feedback from frontal areas (Figure 5).

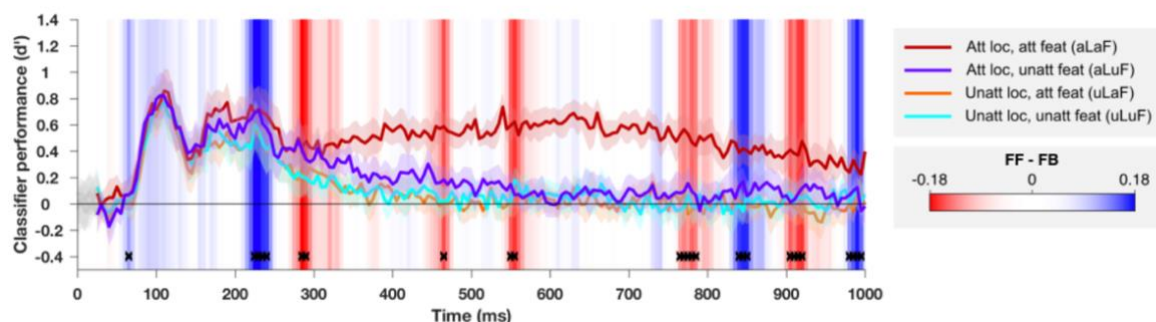


Figure 5. MEG colour decoding results in occipital cortex from Goddard et al.'s (2022) study. The colour of the background represents time points where feedforward (FF, from occipital to frontal: blue) and feedback (FB, from frontal to occipital: red) flows of information were dominant. Time points where FF minus FB were significantly different from zero are shown with black crosses towards the bottom of the plot. Each coloured line represents decoding of attended and unattended stimulus colour. Attended information, that is, the attended feature at the attended location (colour of the target when participants attended to colour), is shown in red. Unattended information, that is, the unattended feature at the attended location (colour of the target when participants attended to shape), and the attended and unattended feature at the unattended location (colour of the distractor when participants attended to colour and when they attended to shape), is shown in purple, orange, and teal, respectively. Initially, object colour was decodable irrespective of attention condition, and the flow of information was predominantly feedforward. At 285ms, the dominant direction of information flow reversed, such that information flowed predominantly in the feedback direction. Shortly after, stimulus processing became selective for the task-relevant information (red trace). Figure from Goddard et al. (2022), reproduced under CC BY licence.

These findings support the concept of top-down influence supporting attention, and specify that top-down influence may contain stimulus-specific information (as opposed to, or in addition to, general attentional rules (e.g., E. K. Miller & Cohen, 2001) or priority maps (Sprague et al., 2018; e.g., Sprague & Serences, 2013)). However, Goddard et al.'s (2022) study relied on large frontal and occipital regions of interest, making it difficult to draw spatially specific conclusions about the source of this feedback. In Chapter 3, I examine whether the MD system might be the source of feedback seen in Goddard et al. (2022). Using the same paradigm as Goddard et al. (2022) with fMRI, I

compare task-related representations from fMRI data with this extant MEG data to derive a spatially-resolved account of how task-related information is processed in the brain.

To summarise, MD regions code a wide range of task-related information and have been shown to adjust these codes to prioritise behaviourally relevant information as required by the task. MD codes are then proposed to bias processing in more specialised regions, giving rise to a focus on task-relevant information across the brain. While some work has provided tentative evidence in favour of the biased competition hypothesis, it remains an open question whether MD codes drive task-selective processing elsewhere in the brain.

1.5 A general factor: Intelligence

Let us now consider how the attentional episodes proposal intersects with broader concepts related to cognitive control. For instance, an interesting question stems from conceptualising behaviour as consisting of modular cognitive blocks that can be rearranged to meet task requirements: could this parallel the way we approach novel problem-solving or intelligence tasks?

Intelligence is an incredibly broad concept, and in relating the principle of compositionality to intelligence I do not intend to cover all its diverse interpretations. Accordingly, I will first narrow the scope of what I mean by intelligence.

The view of intelligence I will rely on here stems from one of Spearman's early, and perhaps most striking, observations: that performance across different cognitive tasks tend to be positively correlated, a phenomenon known as 'positive manifold' (Spearman, 1904). Simply put, good performance in one task is usually predictive of strong performance in other tasks. For example, in Spearman's original work, he linked the

ability to make fine sensory distinctions with overall academic achievement (Spearman, 1927).

To account for positive manifold, Spearman proposed that performance on any one test reflects the contribution of two factors: a ‘general’ or ‘*g*’ factor, indicative of a broad, underlying ability factor contributing to success across a range of different tasks, and ‘specific’ or ‘*s*’ factors, reflecting skills or abilities specific to that task (Spearman, 1904, 1927). Under this framework, positive manifold arises because all tests variably index *g*, some common element that underpins all mental tasks.

I note briefly that Spearman’s proposal that a unitary *g* explains positive manifold is not the only account of this observation. Notably, Thomson (1916) argued that positive manifold could alternatively manifest if one were to assume many independent cognitive processes are sampled by different tests. Here, positive manifold could then arise as any two tests are likely to show some degree of overlap in their comprising functions or ‘bonds’. This perspective is reflected more recently in a more modern conceptualisation of Thomson’s ‘Bonds model’ (Bartholomew et al., 2009, 2013), in process overlap theory (Kovacs & Conway, 2016), and mutualism (van der Maas et al., 2006), all of which variably explain positive manifold without postulating a single unitary concept. Similarly, using a combination of behavioural and neuroimaging methods, Hampshire et al. (2012) propose that *g* comprises at least two orthogonal components, each linked to distinct brain networks. Here, *g* arises as an emergent property of multiple distinct brain systems, each with its own capacity, rather than being driven by a single general factor.

To weigh in on the debate of whether *g* represents a unitary concept is outside the scope of this work. Rather, the focus here is on exploring the neural mechanisms underpinning our capacity to engage in flexible, complex behaviour, which can be measured with novel problem-solving tasks of the sort typically used to index *g* and the related concept of fluid intelligence. It is argued that these novel problem-solving tasks require compositionality – that is, the breaking down of a task into discrete components – which is a proposed function of the MD system. I will consider this proposal in more detail in the following sections.

1.5.1 Fluid intelligence

Fluid intelligence, closely associated with the concept of *g* (Deary et al., 2016), encompasses the ability to solve novel problems, recognise patterns, and link concepts, independent of acquired knowledge¹⁰ (Cattell, 1971; Horn & Cattell, 1966). Complex novel problem-solving tasks such as Raven's Progressive Matrices (Raven et al., 1988) and the Cattell Culture Fair (Institute for Personality and Ability Testing, 1973) are typically used to measure fluid intelligence. Performance on these tests is predictive of a wide range of outcomes, from academic and professional success to longevity and health (e.g., Gottfredson, 1997; Ree et al., 1994). The broad relevance and pervasiveness of fluid intelligence mirrors that of Spearman's *g*. Indeed, fluid intelligence tests have high *g*-saturation, suggesting these tests effectively measure the core aspects of *g* (Duncan et al., 2000).

For our purposes, I operationalise fluid intelligence as reflecting general novel problem-solving capabilities, which in turn are considered a reasonably good index of *g* (Deary et al., 2016; Duncan et al., 2000). Henceforth, reference to 'intelligence' within this work refers to fluid intelligence, and more broadly to the general factor thought to underpin positive manifold.

¹⁰ Cattell also defines a second type of intelligence, crystallised intelligence, reflecting the influence of education, cultural exposure, and general experience (Cattell, 1971; Horn & Cattell, 1966). Crystallised intelligence is conceptually similar to the 'specific' factors first proposed by Spearman (1904, 1927). According to Cattell, crystallised intelligence depends on fluid intelligence: fluid intelligence reflects one's intellectual capacity or potential to learn, while crystallised intelligence reflects the knowledge or skills gleaned through use of fluid intelligence (Cattell, 1971). Accordingly, in a context with equal opportunity, fluid intelligence will tend to correlate with crystallised intelligence. I will leave discussion of 'specific' and 'crystallised' factors here as they are not relevant to the current work.

1.5.2 Intelligence also relies on compositionality

To return to the idea of compositionality and the attentional episodes framework, it is argued that the ability to divide a complex task into simpler, separately attended parts is also what is being measured in fluid intelligence tasks (Duncan, 2017). For instance, as briefly mentioned in Section 1.3.1, goal neglect purportedly reflects difficulties in managing the complexity of a task model (Duncan et al., 2008), and it is strongly related to poor performance on fluid intelligence tasks (Duncan et al., 1996). However, alternative proposals have emphasised that different core cognitive processes underpin performance on fluid intelligence tasks. For instance, fluid intelligence has also been argued to reflect working memory capacity (Halford et al., 2007; Kyllonen & Christal, 1990) and processing speed (Kievit et al., 2016; Schubert et al., 2017; Sheppard & Vernon, 2008). Yet much of the evidence linking these processes to fluid intelligence is largely correlational. Given positive manifold, correlational data alone cannot disentangle which process might be critical to fluid intelligence.

Stronger evidence in favour of the segmentation account comes from recent work by Duncan and colleagues (Duncan et al., 2017). Here, researchers aimed to isolate and compare the impact of working memory, processing speed, and segmentation demands on fluid intelligence. They created modified versions of a fluid intelligence task (Cattell Culture Fair; Institute for Personality and Ability Testing, 1973), reducing the impact of one of the three processes of interest in each modified version. Participants with low fluid intelligence struggled on the original version of the task and on the versions where working memory and processing speed demands were reduced. Indeed, performance only improved for the version where segmentation demands were mitigated. This finding suggests that segmentation ability is the critical limiting factor of fluid intelligence.

1.5.3 Bringing it together: The MD system as the neural basis for fluid intelligence

Given that the assembly of attentional episodes is proposed as a core role for the MD system and as the core process measured in fluid intelligence tests, can we consider the MD system to then be the neural basis for fluid intelligence?

A growing body of work speaks in favour of this idea. First, individual differences in performance on fluid intelligence tasks are associated with individual differences in MD activity (e.g., Assem, Blank, et al., 2020). More convincing is lesion work that links deficits in fluid intelligence with damage to MD system. For instance, Woolgar and colleagues found fluid intelligence deficits in focal lesion patients was associated with extent of MD damage, but not with extent of damage to language-selective regions (Woolgar et al., 2018) nor non-MD damage more broadly (Woolgar et al., 2010).

However, this account is far from indisputable. Other work shows overall damage is the best predictor of fluid intelligence deficits (Cox et al., 2019), that interactions with other networks such as DMN might be important (Sripada et al., 2019), or even integrity of white matter (WM) tracts (Kievit et al., 2016) may be more significant predictors of fluid intelligence than MD system integrity. In addition, there may be functional differentiation within the MD network with respect to how it might support intelligence, as has been proposed elsewhere (e.g., Hampshire et al., 2012). For example, a recent study applied graph-based lesion-deficit mapping methods to investigate the neural basis of fluid intelligence in a large cohort of patients with unilateral brain lesions ($N=227$) (Cipolotti et al., 2023). This method models the brain as a densely interconnected graph of voxels (nodes) with weighted connections (edges), allowing for explicit testing of how lesions affect individual nodes and interactions between distributed areas. Yet their data revealed that fluid intelligence deficits are primarily associated with integrity within a set of localised, predominantly right frontal regions, casting doubt on the notion of fluid intelligence being underpinned by a broadly distributed neural network.

Overall, the findings here are mixed, complicated by uncertainty and variability in lesion location, small sample sizes, and limitations in the variety of methodological approaches used. Thus, while Woolgar et al.'s (2010, 2018) work provided preliminary evidence that the MD system has a causal role in supporting fluid intelligence, clearly more work is needed to establish a causal relationship between the MD system and fluid intelligence.

In Chapter 2, I aim to contribute to the work in this space. Specifically, I explore the impact of both grey matter (GM) damage and WM disconnection to brain networks (defined using the resting-state definitions from Yeo et al., 2011) on fluid intelligence in a large cohort of focal lesion patients ($N=174$). This approach extends the work of Woolgar et al. (2010) by examining the contribution of all brain networks, not just MD, to fluid intelligence loss. Moreover, motivated by accounts that underscore the significance of WM integrity in supporting fluid intelligence, I incorporate both GM and WM network damage in the analysis. This approach thus offers a more comprehensive measure of network damage compared to analyses based solely on GM, and permits a clearer comparison of the contribution of all brain networks to supporting fluid intelligence.

1.6 Selective attention

Thus far, we have considered a prominent proposal that identifies the structured assembly of attentional episodes as a key role of the MD system. A core component of this proposal is the concept of selective attention, by which we prioritise processing of the elements needed within each episode. Many theories of selective attention have been proposed, yet it is still unclear how broader proposals for selective attention relate to the observation that MD regions preferentially represent attended information, and dynamically adapt their representations as the task changes (see Section 1.4.2.2). In this

section, I will consider how the voluntary¹¹ deployment of covert¹² visual selective attention¹³ affects processing of information in the brain.

The field of attention research is large, so I will provide a brief overview of the theories and associated predictions that are pertinent to this work. Fundamentally, the principle underpinning these theories is that cognitive processes are inherently competitive, and attention serves to resolve this competition by selectively prioritising certain inputs for further processing.

Extending a well-known metaphor (e.g., Posner et al., 1980), we can consider voluntary visual attention as a spotlight. At the theatre, for example, the beam of light (the allocation of cognitive resources) is directed towards specific targets or elements for further processing, to the exclusion of other elements. While imperfect, this metaphor conveys the flexibility with which attention can be deployed. For instance, we can deploy our attention to a particular location within the visual field (spatial attention, e.g., the beam of light is directed to the corner of the stage where actors are engaged in a lively discussion), to a particular object (object-based attention, e.g., the spotlight is directed to an actor delivering a monologue), or to specific features (feature attention, e.g., the spotlight highlights the silvery gleam of raised swords across the stage as the actors approach a battle). In this section, I will briefly consider the differing behavioural and neural effects of these subtypes of attention.

¹¹ Voluntary attention, also known as endogenous attention, refers to our ability to intentionally direct attention to something. It is typically thought of as a top-down process. This stands in contrast to reflexive or exogenous attention, whereby attention is directed in a more stimulus-driven, bottom-up fashion.

¹² Much of attention research utilises covert attention paradigms (where attention is directed without concomitant eye movements). This approach is thought to allow for separation of the effects of attention from those of eye movements, ensuring that any observed differences in neural processing or task performance are purely attributable to attentional processes rather than physical changes in the visual field.

¹³ I use 'selective attention' and 'attention' synonymously.

1.6.1 Paying attention in different ways

The subtypes of selective attention introduced earlier are thought to have different behavioural and neural consequences. This has been studied in detail within specialised regions such as visual cortex. For example, spatial attention (e.g., covertly attend left, and await a target) increases firing rate of neurons across corresponding parts of visual cortex (e.g., Luck et al., 1997; Moran & Desimone, 1985), boosts the magnitude of visual processing signals (Di Russo et al., 2003), and is associated with increased neural activity (i.e., increase in blood-oxygen-level-dependent (BOLD) response) in the visual areas to which the attended targets are projected (e.g., Tootell et al., 1998). These neural effects are thought to reflect enhanced sensory sensitivity at the attended site, commensurate with faster reaction times (e.g., Posner, 1980) and higher accuracy (e.g., Liu et al., 2007) for stimuli presented in attended locations as compared to unattended ones.

However, many studies have shown that attention can also be directed in a non-spatial manner. Attention directed to a particular object (object-based attention) is associated with enhancement of all information within the bounds of an attended object, such that behavioural performance is improved when responding to attributes of the attended object (Brawn & Snowden, 2000; Duncan, 1984; Egly et al., 1994). Moreover, focussing on a specific object leads to increased activity in brain regions linked to the processing of that object (Baldauf & Desimone, 2014; O'Craven et al., 1999), which also spreads to enhance processing of multiple features of the attended object (O'Craven et al., 1999). Examination of patients with neglect also support the existence of an object-based attention. For instance, as mentioned in Section 1.3.1, inattention to the contralesional side can be reduced when bilateral items appear to form a single object (Mattingley et al., 1997; Ward et al., 1994).

Attention can also be directed to the selection of visual features, such as colour and motion, regardless of location (feature attention). Some feature attention studies have focussed on selection of a given feature value within a feature dimension (such as leftward versus rightward motion, what is referred to as feature-based attention, see

Carrasco, 2011; e.g., Liu et al., 2007; Treue & Martínez-Trujillo, 1999). Other studies have referred to feature attention as the selection of one feature dimension over another (such as attending to colour over motion, in what is sometimes referred to as feature-selective attention, see X. Chen et al., 2012). Like spatial and object-based attention, feature attention is associated with modulation of cortical areas specialised for processing the attended domain (e.g., Chawla et al., 1999; Liu et al., 2003; M. Schoenfeld et al., 2007). Moreover, behavioural and neurophysiological studies show feature attention has effects across the visual field (e.g., Liu et al., 2007; Sàenz et al., 2002; Treue & Martínez-Trujillo, 1999; W. Zhang & Luck, 2009). For instance, cueing the direction of motion of a subsequent target enhanced accuracy, regardless of whether stimulus appeared on the left or right hemifield (Liu et al., 2007), and attention to a particular feature increased cortical responses to the attended and unattended stimulus (Sàenz et al., 2002), indicating that feature attention leads to enhancement across the visual field, independent of spatial attention (Hillyard & Münte, 1984; W. Zhang & Luck, 2009). Moreover, spatial and feature attention are thought to have distinct effects on the tuning of neurons, with spatial attention being primarily characterised as increasing response or contrast gain and feature attention characterised as causing tuning curve sharpening (reviewed by Reynolds & Heeger, 2009), a finding also reflected in population-level codes with MEG (e.g., Goddard et al., 2022).

1.6.2 The effects of paying attention in the MD system

These subtypes of attention yield seemingly conflicting predictions of what information is upregulated in the brain. For instance, in the case of object-based attention, all features within the bounds of an object are thought to be enhanced, while with feature attention, it is enhancement of attended feature information across the visual field. Given these types of attention are thought to reflect top-down processes, as the purported source of attention, how might frontoparietal regions, or the MD system more specifically, implement these subtypes of attention?

I have already introduced the idea (Section 1.4.2.1) that frontoparietal neurons code task-relevant or attended information, including information about both an attended object's features (e.g., form or colour) or location. Moreover, as discussed in Section 1.4.2.2, MVPA work has shown that MD regions selectively represent attended information, indicating that the flexibility of the MD system could underpin selective attention. For instance, when participants were cued to one of two objects, MD regions showed stronger coding of the attended object over the unattended object (Woolgar, Williams, et al., 2015). These regions also show high attentional selectivity within a single object. Where participants were trained to categorise a single spikey object on the basis of one of two feature dimensions (orientation or length of its spikes), MD regions coded the attended feature dimension more strongly than the unattended feature dimension (Jackson et al., 2017). This work illustrates that the MD system can emphasise attended information as required by the task, including object and feature information, providing a possible mechanism for implementing selective attention more broadly. Yet an important limitation of these neuroimaging studies is that the effects of spatial and feature attention were investigated separately, leaving open the question of what information is prioritised in MD regions when attention is directed in multiple ways at once.

1.6.3 MD system when multiple types of attention are deployed

It remains unclear how the MD system codes task-related information when different subtypes of attention are deployed simultaneously. This is important to consider as we rarely attend to spatial or feature information independently. Rather, in the real world, the target of our focus is typically informed by a combination of spatial and feature cues against a backdrop of many other possible competing inputs. Elucidating how these types of attention interact to affect representations in the MD system is a critical step towards understanding how this system supports selective attention in more complex situations, as required to assemble attentional episodes in real life.

As a first step towards addressing this gap, Goddard and colleagues (2022) explored how these types of attention interact using MEG. In Section 1.4.2.3, I described their study in the context of feedback and biased competition. Here, I mention it again but this time highlighting the attention manipulation used in this study. Goddard and colleagues (2022) used MVPA to examine how spatial and feature attention interact to affect task-related representations in broadly defined frontal and occipital regions. The paradigm used in this study is described in more detail in Chapters 3 and 4, but in brief, participants were asked to indicate the colour or the shape (feature attention manipulation) of one of two objects presented left and right of fixation (spatial attention manipulation). They found spatial and feature attention interacted multiplicatively in frontal regions, leading to a highly specific representation of just the attended feature information at the attended location. While this finding suggests that broadly defined frontal regions selectively prioritise behaviourally relevant information, it remains to be established whether MD regions exhibit this specificity.

In Chapter 3, I replicate this study with fMRI to examine whether MD regions show this multiplicative representation in response to spatial and feature attention. In Chapter 4, I then use MEG to examine how focal damage to parietal regions might affect the specificity of task-related representations across the brain, and/or the timecourse over which attention prioritises relevant information.

1.7 Methods used in this thesis

In this thesis, I draw on evidence from neuropsychology, neuroimaging with multivariate analyses, and their combination. Here, I will consider the strengths and limitations of these methodological approaches.

1.7.1 Neuropsychological studies

Neuropsychological studies rely on the logic that we can understand the contribution of a brain region to behaviour by examining the behavioural changes that follow when that region is damaged. It can answer questions regarding whether a region makes a necessary contribution to a particular behaviour or cognitive process.

Accordingly, neuropsychological studies, or lesion studies, in humans and non-human primates have led to invaluable insights into brain function, greatly informing current theories on neural mechanisms. For example, I briefly mentioned in Section 1.3.1 that damage to frontal regions is characterised by disorganised, stimulus-driven, or perseverative behaviour (e.g., Lhermitte, 1983; Lhermitte et al., 1986; Penfield & Evans, 1935; Shallice & Burgess, 1991), underscoring the critical role of these regions in supporting goal-directed behaviour.

Moreover, lesion studies can support causal narratives, such as whether the integrity of a given region is necessary to support a particular behaviour. As I will also mention in Chapter 2, this is an advantage over functional neuroimaging studies, which cannot usually imply causality. For instance, while functional neuroimaging studies imply a correlative relationship between MD regions and fluid intelligence (e.g., Assem, Blank, et al., 2020), they fall short of establishing a causal link. In Chapter 2, I will take advantage of the causal insights neuropsychological studies provide to examine whether

FPN, a network closely linked to the MD system, has a critical role in supporting fluid intelligence.

However, lesion studies have some important limitations. Notably, given the circumstances under which brain lesions are acquired, many variables related to the lesion cannot be precisely controlled and thus limit the strength of the conclusions one can draw (Vaidya et al., 2019). For instance, different aetiologies could differentially affect neuropsychological outcome (e.g., Anderson et al., 1990, although see Cipolotti et al., 2015). Moreover, heterogeneity in factors relating to anatomical specificity (e.g., the extent of damage, how many structures are damaged, the volume of overall tissue damage), and in recovery (including pre-existing health conditions that may interact with recovery, length of time since insult, differences in behavioural effects following injury reflecting differences in adaptation, compensation or reorganisation) are standard in lesion studies and can make it difficult to determine whether a particular region is responsible for an observed behavioural deficit. In addition, the comparison group used in these studies can make a difference to conclusions about the role of a region (e.g., comparison to healthy participants or to patients with lesions elsewhere in the brain can affect whether patients are seen as impaired: Grafman et al 1990 vs. Anderson et al., 1991).

To illustrate the inconsistencies that can stem from lesion work, I take an example from the clinical literature. As discussed in Section 1.3.1, poor performance on executive function tests like WCST is often observed in frontal lesion patients, specifically following dlPFC lesions (e.g., Demakis, 2003; Milner, 1963), with deficits in fluid intelligence thought to underpin poor performance (Roca et al., 2010). Accordingly, WCST is widely regarded as a test of frontal function (Lezak et al., 2012). Inflexibility or perseveration on this task is commonly associated with frontal damage (e.g., Grafman et al., 1990; Milner, 1963; A. L. Robinson et al., 1980; Stuss et al., 2000), lending support to the idea that these regions are important for supporting flexible cognitive control. However, other studies have found patients with frontal lobe damage perform just as well as non-frontal patients (e.g., Anderson et al., 1991), leading some to conclude that such measures are not indicative of frontal function (Mountain & Snow, 1993). The

reasons for these discrepancies are many but include some of the factors described earlier, such as differences in the comparison group used to compare WCST performance and inconsistencies in patient variables, such as non-uniformity of lesion location and differences in chronicity and aetiology, the effects of which are exacerbated by small sample sizes (Demakis, 2003; Mountain & Snow, 1993).

Such examples highlight the importance of carefully considering the potential moderating factors in interpreting lesion studies and underscore the need for large samples. Lesion work, moreover, should be considered within the broader context of work using other methods (such as neuroimaging), where it forms one of several lines of evidence to provide convergent, complementary or divergent support for theories of brain-behavioural relationships (Vaidya et al., 2019). These caveats notwithstanding, lesion studies provide unique insights into the neural mechanisms underpinning cognition and thus play an important role in shaping our understanding of brain-behaviour relationships.

1.7.2 Multivariate pattern analysis in functional neuroimaging

In this thesis I draw on functional neuroimaging data from both fMRI and MEG. In Chapter 3, I use fMRI to address a spatial question, examining what task-related information is held in MD regions. In Chapter 4, I use MEG to look at the temporal dynamics of neural attentional processes. In both cases, I use MVPA to assess informational content.

MVPA refers to a collection of methods that rely on examining differences in the pattern of activity across multiple fMRI voxels or MEG sensors. With these methods, we can infer not only whether a region is involved in a given task, as is typically the case with univariate functional neuroimaging studies, but also what information is represented in those brain regions (e.g., Mur et al., 2009). This approach can reveal information that

is not visible using univariate approaches. For instance, if the scale over which the information is organised is small (e.g., within narrow cortical columns (Vanduffel et al., 2002)), it will be obscured by processing steps (e.g., spatial smoothing, averaging of voxels within an area of interest, etc. (Haynes & Rees, 2006)), and the requirement for close spatial alignment of these signals between subjects. In contrast, MVPA seeks to identify consistent spatial patterns at the individual subject level, with group analyses applied to the MVPA summary statistics, allowing for idiosyncratic individual subject patterns to contribute to population-level inference (Woolgar et al., 2014).

My work relies on applying machine learning linear classifiers to discriminate between idiosyncratic patterns in neuroimaging data that correspond to different experimental conditions or stimuli. If these patterns can reliably be discriminated, allowing us to predict an event from the data better than by chance, then we can ‘decode’ the event (Grootswagers et al., 2017; Hebart & Baker, 2018).

To illustrate this approach, I use a simplified example of the paradigm that forms the basis of the work in Chapters 3 and 4 (Figure 6). In this toy example, participants are shown a series of two objects, presented left and right of fixation (Figure 6, Panel A). They are asked to attend to one of the objects and identify either the colour (red or green) and shape (X-shaped or flat) of the target while their brain activity is recorded (with either fMRI, as in Chapter 3, or MEG, as in Chapter 4). Let us consider decoding of colour when participants are attending to the object on the left (shown with dotted circle in Figure 6, Panel A).

The goal of the analysis is to test whether a pattern classifier trained on a subset of the data from that individual can predict the colour (red or green) of the attended left object. Neural activity evoked during all the ‘attend left, report colour’ trials are extracted and sorted by target colour: red or green. These patterns are represented in high-dimensional space (Figure 6, Panel A, shown here in two-dimensional space with each point representing the simultaneous activation level evoked in response to the two different colours). Note that in this example, activity evoked in response to all ‘attend left, report colour’ trials are extracted, regardless of the shape of the target or features of the distractor.

Next, a linear pattern classifier is trained to discriminate between red and green objects on a subset of labelled data (Figure 6, Panel B). The classifier finds a decision boundary in multidimensional space that best separates ‘red’ from ‘green’ objects. This classifier is then tested on a held-out, unlabelled set of the data to ascertain how well the decision boundary fits the data (Figure 6, Panel C). This represents the ‘classification accuracy’. The process of testing and training on separate data sets is repeated multiple times such that all data is used equally for training and testing. This process is called n-fold cross-validation and it tests how well the classification boundary established on the training data generalises to the testing data. The final classification accuracy is the average of all the cross-classification folds and indicates how consistently patterns of activation can be used to distinguish between the events of interest. If classification accuracy is significantly above what we might expect by chance (in a binary classification, theoretical chance level is 50%, though in practice we typically establish an empirical chance level by permutation), we can then say that we can ‘decode’ colour. This entire process is repeated for each distinction of interest (colour, shape, different attentional conditions, etc.).

The neural activity of interest can be the pattern of BOLD response over voxels in a region of interest (ROI) in fMRI, or the pattern of response over MEG sensors at a particular timepoint, etc. This procedure can be repeated to examine decoding in different ROIs or at different timepoints (Figure 6, Panels D and E, respectively). This method thus allows us to examine what information the brain is representing across space and time.

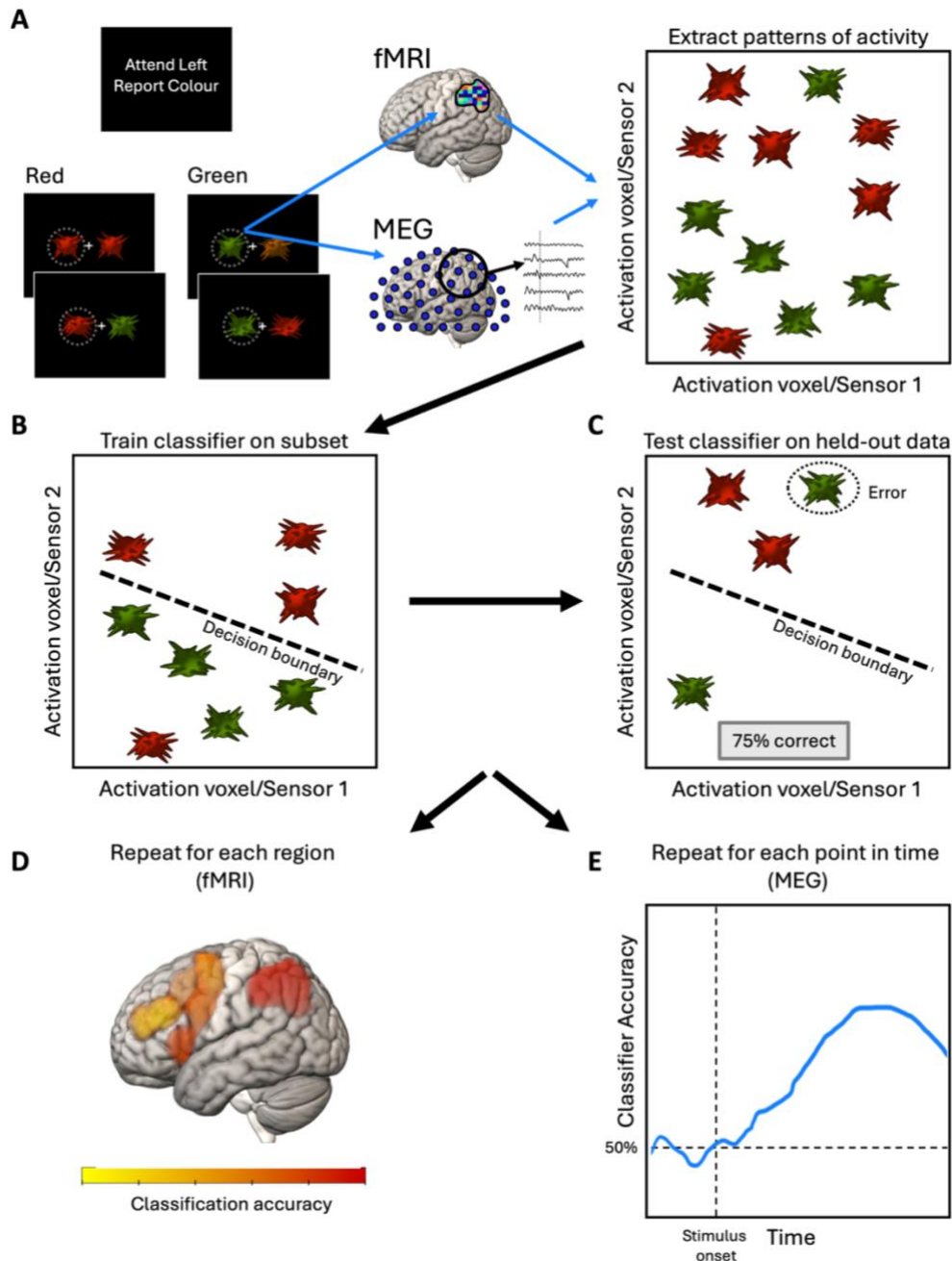


Figure 6. Schematic of decoding steps for fMRI (Chapter 3) and MEG (Chapter 4). Panel A: in this example, based on the actual task used in Chapters 3 and 4, participants categorised the attended object (left object, shown with dotted circle which was not shown to participants), as either ‘reddish’ or ‘greenish’. Brain responses are recorded with fMRI or MEG. Patterns of activation evoked by the stimulus conditions (attended left red object, attended left green object) are extracted from voxels of a region of interest (fMRI) or from sensors (MEG) and are represented in multiple dimensions (shown here in two dimensions for simplicity). Panel B: the classifier is trained to distinguish between the two colours. Panel C: the performance of the classifier is then evaluated by testing whether it can distinguish between these two colours on a held-out, unlabelled dataset. Panel B and C are repeated multiple times such that all data is trained and tested. This process is also repeated for each brain region of interest for fMRI (Panel D) or for each timepoint for MEG (Panel E). Figure adapted with permission from 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.

Powerful though these methods are at probing representational content, there are limitations. In particular, many of the general limitations of the recording method still apply. In fMRI, inference is limited by the poor temporal resolution of the BOLD response. Accordingly, neural events which transpire over milliseconds can be blurred over seconds, limiting our ability to detect transient events or explore the rapid dynamics of how information might be exchanged throughout the brain with this method alone. Rather, sensitivity here is in favour of more long-term or dominant neural events.

MEG, by contrast, has high temporal resolution, but relatively poor spatial resolution. This limitation is due in part to the ill-posed nature of the inverse problem: for any given pattern of magnetic fields measured using MEG, an infinite number of configurations of neural activity could be responsible (Sarvas, 1987). Simply put, it is impossible to unambiguously determine the sources that gave rise to the recorded measurements. Several approaches exist to estimate the sources of MEG signals, notably involving co-registration of MEG sensor data with either the individual's structural MRI or a standardised template brain (e.g., as described in Hauk et al., 2022). However, these are not perfect. For instance, a common concern with source estimation methods is source leakage, where activity estimated from one region incorrectly appears in another region of the reconstructed model. We still lack consensus on how reliable or sensitive these methods are, although efforts have been made to evaluate the resolution of different source estimation approaches (e.g., Hauk et al., 2011, 2022).

In addition to considering MVPA in fMRI and MEG separately, it is possible to integrate data from both modalities by examining similarities in how each modality represents task-related information: MEG-fMRI fusion (Cichy et al., 2014, 2016; Cichy & Oliva, 2020; Hebart et al., 2018). I use this method in Chapter 3 to estimate the timecourse of information coding in the MD system, to in turn ascertain whether frontal MD regions are reasonable candidates for the source of top-down information flow seen in Goddard et al. (2022). This approach involves representing data from each modality in a common representational space, which then allows for direct comparison between the

modalities. A graphical representation of this approach, adapted from Chapter 3, is shown in Figure 7 to illustrate.

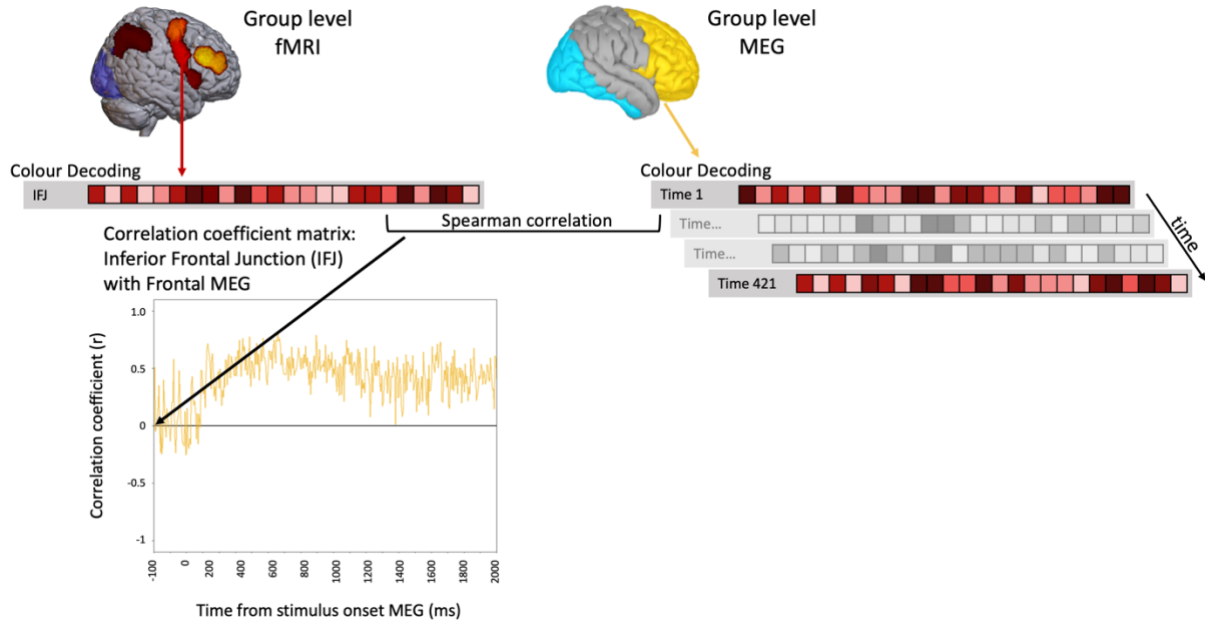


Figure 7. A graphical representation of the method for fusing the pattern of colour decoding accuracies between the inferior frontal junction (IFJ)(fMRI) and the frontal MEG region across time, adapted from Chapter 3. Each coloured square in the representational dissimilarity matrices (RDMs, here shown as vectors) represents average decoding accuracy for a given pairwise comparison, with higher decoding accuracy reflecting more distinctive neural patterns (i.e., greater dissimilarity) between conditions. The fMRI vector for a given region of interest is then correlated with each timepoint of the MEG matrix, yielding a correlation coefficient representing the correlation between the fMRI vector with the MEG results at each timepoint. The correlation coefficient vector between the IFJ and frontal regions for colour decoding is plotted in gold.

Data are typically represented in RDMs (one for each timepoint for MEG, and one for each ROI for fMRI, shown as vectors in Figure 7), where each cell in the matrix reflects the extent of dissimilarity in the patterns of neural activity elicited by a given pair of stimuli or conditions. In Chapter 3, as in other work (e.g., Moerel et al., 2021), I use decoding accuracies to reflect dissimilarity in neural patterns. These RDMs are then correlated across the two imaging modalities to ascertain the commonality in representational structure in time and space. This allows us to examine at which timepoints (in MEG) the data most closely resembled the representational pattern

recorded in each brain region (with fMRI). This is taken as an indication of when information may have emerged in this region. I will discuss this method in more detail in Chapter 3. However, as we will see in that chapter, there are limitations to this approach. In particular, this method is sensitive to signals being present in both modalities, and depends on there being variation in the distinctiveness of neural patterns across the brain.

A more general limitation is that MVPA is still inherently correlational, meaning that we cannot infer causality. For instance, the fact that we can decode colour information when it is attended in a given region does not necessarily mean that this region has a causal role in attention. Nor does it necessarily mean that this information is ‘used’ by the brain. Increasingly, researchers are starting to look for evidence that explicitly links information coding to behaviour (Zheng et al., 2024), for example exploring how patterns predict reaction times (e.g., González-García et al., 2021), change on error (A. K. Robinson et al., 2022), or are modulated by TMS (e.g., Jackson et al., 2021). In Chapter 4, I contribute a new approach to bring insight from MVPA into the causal domain. I use MVPA with MEG in focal lesion patients to examine how the timecourse over which attended information is prioritised in the brain might change following damage to parietal regions. This work helps us get closer to understanding what role representational content has in cognition.

1.8 Summary

To summarise, this thesis explores how a network of frontoparietal regions, the MD system, supports intelligent human behaviour, with a particular focus on their role in fluid intelligence and attention. It investigates key aspects of the MD system's role in human cognition across three chapters, addressing three broad questions:

Chapter 2 Which brain networks are critical to fluid intelligence?

In this chapter, I draw on a large cohort of focal lesion patients to undertake a data-driven comparison of the contribution of different networks to fluid intelligence. I examine the extent of damage across each of seven canonical resting-state brain networks (Yeo et al., 2011) by quantifying the extent of grey matter network damage and estimating the extent of white matter network damage caused by lesion damage to white matter tracts. By then relating network damage using a composite of these measures to fluid intelligence loss, I seek to ascertain which brain networks critically underpin fluid intelligence.

Chapter 3 What task information is prioritised in the MD system when spatial and feature attention are used together?

In this chapter, I use MVPA methods with fMRI data to investigate the role of the MD system during selective attention. Specifically, I examine what task information is prioritised in the MD system during a task that requires both spatial and feature attention, looking to ascertain how selectively this system responds to complex attentional demands. In addition, I use MEG-fMRI fusion to estimate when attention effects arise in different brain regions, to investigate the hypothesis that task-relevant representations in MD regions might bias processing elsewhere in the brain.

Chapter 4 How is the representation and processing of task-related information affected following damage to the MD system (specifically parietal lobe)?

In the final empirical chapter, I combine MVPA methods with MEG data collected from focal lesion patients and matched controls to examine how damage to some of the MD system (specifically, to parietal lobe) affects neural selective attention processes. Using the demanding selective attention task from Chapter 3, I examine the dynamic neural representation of task-related information in individual patients and controls. Taking a case-series approach, I explore whether and how parietal damage affects the prioritisation of relevant information and consider whether the strength of this attentional prioritisation relates to behavioural performance.

Together, these chapters aim to clarify the critical role of the MD system in supporting complex human cognition and behaviour.

Chapter 2 |

Neuropsychological evidence for the role of frontoparietal cortex in fluid intelligence

2.1 Foreword

This study explores whether frontoparietal network is critical for fluid intelligence, in comparison to other brain networks. As detailed in Chapter 1, there are several approaches one could take to explore the neural underpinnings of a cognitive process. Here, I use a focal lesion model approach to investigate the relationship between damage to large-scale resting-state networks and fluid intelligence loss in a large cohort of focal lesion patients. The primary aim is to identify specific functional networks that might be important for fluid intelligence.

Much of the previous work in this area has examined contributions of individual regions, for example through use of voxel-based methods. While such approaches can identify disparate neural correlates of intelligence, they are insensitive to the contribution of functional connectivity between regions, such as that represented in resting-state networks. The importance of examining cognitive processes in the context of networks has been emphasised in more recent studies (e.g., Santarnecchi et al., 2017), and is particularly critical to consider in the context of neuropsychology, as focal damage to part of a network can have widespread effects (Gratton et al., 2012). The work in this chapter represents an important step in this direction. It builds upon previous work by my group (namely, Woolgar et al., 2010, 2018), which has previously shown that multiple-demand (MD) damage, but not non-MD damage, was predictive of fluid intelligence loss. Here, I extend this work by considering not only the contribution of frontoparietal network (closely related to the MD system) to fluid intelligence, but also other brain networks, as defined using the resting-state network parcellations provided by Yeo and colleagues (Yeo et al., 2011).

In addition, much of the previous work has focussed solely on quantifying the influence of grey matter damage. However, focal lesions also affect white matter connections between brain regions and networks, which can lead to neural changes that extend beyond the site of injury. I thus incorporate an estimate of white matter disconnection in my metric for network damage, and additionally compare the results of predicting fluid intelligence loss using grey matter network damage and white matter network disconnection separately.

2.2 Abstract

Fluid intelligence is a fundamental component of human cognition, yet the precise neural mechanisms which subserve this process are still unclear. Here, I take a data-driven approach to examine whether damage to specific resting-state networks (RSNs) predicted fluid intelligence loss in a large cohort of focal lesion patients ($N=174$). Using a well-established seven-network RSN parcellation (Yeo et al., 2011), damage to each network was defined as a composite of two approaches: lesion tracings overlaid onto network maps to quantify volume of grey matter (GM) network damage, and onto a large sample of white matter (WM) tractograms to estimate WM disconnection within each RSN. Damage to frontoparietal network predicted fluid intelligence loss over contributions from other networks, but could not be distinguished from the contribution of total lesion volume. Comparison of the two approaches taken to defining network damage (i.e., volume of GM damage and WM disconnection) showed they yielded highly similar results, indicating they may capture similar variance. Consideration of only GM network damage predictors additionally suggested that dorsal attention network may also be important for fluid intelligence, but this effect was attributed to total lesion volume rather than specific network damage. A final exploratory analysis indicated that the relationship between frontoparietal network damage and fluid intelligence was driven mainly by frontal lesion patients, particularly those with lesions to the right hemisphere. Overall, this study suggests that of all brain networks, frontoparietal network may have the most dominant causal role in underpinning fluid intelligence. However, future work with patients exhibiting more widely distributed lesions is required to dissociate the effects of frontoparietal network damage from total lesion size and to more fully explore the contribution of different nodes of each network in supporting fluid intelligence.

2.3 Introduction

Fluid intelligence is a fundamental component of human cognition. It refers to the ability to solve novel, abstract problems independent of acquired or task-specific knowledge (Horn & Cattell, 1966). Fluid intelligence is closely related to the broader concept of ‘*g*’ or general intelligence: a common factor thought to underpin the phenomenon of positive correlations between diverse cognitive tests (Spearman, 1904). Indeed, fluid reasoning tasks, such as Raven’s Progressive Matrices (Raven et al., 1988) and the Cattell Culture Fair (Institute for Personality and Ability Testing, 1973), have high *g*-saturation, suggesting these tests effectively measure the core aspects of *g* (Duncan et al., 2000). Fluid intelligence and *g* seem to matter. Performance on fluid intelligence tasks is related to success across a broad range of outcomes, such as scholastic achievement, occupational performance, and better psychosocial or health outcomes (Colom & Flores-Mendoza, 2007; Huepe et al., 2011; Ree et al., 1994; Wrulich et al., 2014). Given its foundational role in human cognition, the neural basis for fluid intelligence has long been of central interest to cognitive neuroscience. A wide range of methodological approaches have been used to address this question, yet consensus on the precise neurobiological underpinnings of fluid intelligence is still lacking.

Neuroimaging work strongly implicates frontal and parietal regions in the neural mechanisms of fluid intelligence. For instance, the Parietal-Frontal Integration Theory (P-FIT) of intelligence (Colom et al., 2009, 2010; Jung & Haier, 2007) posits that variation in intelligence relates to variation in the structure and function of a distributed network of regions across the brain, with a particular emphasis on specific regions within prefrontal cortex (PFC) and parietal lobe. This theory, derived from a review of structural and functional correlates of intelligence across 37 neuroimaging studies (Jung & Haier, 2007), postulates that intelligence depends on functions implemented within dorsolateral PFC and parietal cortex, with additional contributions from temporal and occipital regions in early processing stages.

Relatedly, Duncan and colleagues have argued that a circumscribed set of frontoparietal regions, the multiple-demand (MD) network, is a key neural substrate of fluid intelligence (Duncan et al., 2000, 2017, 2020). These regions, comprising the posterior part of inferior frontal sulcus, anterior insular and frontal operculum, pre-supplementary motor area, dorsal anterior cingulate and intraparietal sulcus, are consistently recruited in response to a wide range of cognitively demanding tasks (Assem, Glasser, et al., 2020; Duncan & Owen, 2000; Fedorenko et al., 2013). Given the importance of fluid intelligence to a diversity of different cognitive functions and behaviour, a neural system that supports fluid intelligence would be expected to demonstrate this pattern of broad recruitment across different cognitive demands (Duncan et al., 2000). More specifically, Duncan and colleagues argue that fluid intelligence reflects the primary function of the MD system: attentional integration, that is, the segmenting of a problem into constituent parts and integrating the different components needed for each step of the problem (Duncan et al., 2017, 2020). In support of the MD account of fluid intelligence, functional magnetic resonance imaging (fMRI) studies have shown that performance on fluid intelligence tasks is linked to stronger MD activity (Assem, Blank, et al., 2020; Duncan et al., 2000; Prabhakaran et al., 1997; Woolgar et al., 2013), and measures of MD cortical volume and white matter (WM) tract integrity are associated with fluid intelligence in healthy adults across their lifespan (P.-Y. Chen et al., 2020).

Both of these theories emphasise the contribution of spatially-distributed brain networks to fluid intelligence. A parallel field has focussed on discovering the intrinsic functional connectivity of the brain, which can be captured in resting-state networks (RSNs), based on the strength of time-series correlations in spontaneous brain activity recorded at rest (e.g., Yeo et al., 2011). There is a range of different ways to delineate these networks and they have been defined with varying degrees of granularity (e.g., Ji et al., 2019; Shirer et al., 2012; Yeo et al., 2011).

While we still lack consensus on how the MD system precisely maps to various RSN classifications, the MD system shows strong anatomical similarity to frontoparietal network (FPN) (Assem, Glasser, et al., 2020; Shashidhara et al., 2019). Accordingly,

recent studies exploring the relationship between intelligence and functional connectivity within RSNs have identified FPN as a key contributor. However, additional networks have also been implicated. For instance, in an effort to merge activation findings with RSNs, Santarnecchi et al. (2017) overlaid meta-analytic fluid intelligence maps (based on an activation likelihood estimation meta-analysis of 47 fMRI and positron emission tomography (PET) studies) with group-level RSN maps, implicating both FPN and parietal regions of the dorsal attention network (DAN) to fluid intelligence.

Dynamics between RSNs have also been suggested as an important neural mechanism for intelligence. A crossover pattern of network separation, whereby FPN and default mode network (DMN) show greater separation of activation levels during challenging tasks and less during simpler ones, has been shown to be associated with higher intelligence (Sripada et al., 2019). Similarly, fluid intelligence has also been shown to be positively correlated with the degree of resting-state functional connectivity within FPN and negatively correlated with connectivity in DMN (Finn et al., 2015). This literature therefore implies contributions from both FPN and DMN in supporting fluid intelligence.

Other work suggests that functional connectivity across the brain is most predictive of general intelligence. A large study of Human Connectome Project (HCP) data (Dubois et al., 2018) ($N=1181$) examined the relationship between resting-state functional connectivity within and between brain networks and general intelligence in healthy subjects. While a model using the full functional connectivity matrix of the entire brain was most predictive of intelligence, of seven major RSNs examined (using functional networks specified in Ito et al. (2017)), some networks were more predictive than others, notably the cingulo-opercular network (CON), DMN, FPN and visual network. Overall, RSN studies suggest that there may be a broader neural mechanism of intelligence that extends beyond FPN, emphasising the critical role of widespread network integration in supporting cognitive functions, but the details of this are not fully resolved.

While structural and functional neuroimaging results are important for establishing the neural correlates of cognitive processes, they are unable to establish causality. Thus, an

important avenue to better understand the contribution of different brain networks to fluid intelligence may be to consider what happens after brain injury. Data from this field again suggests strong links between frontal and parietal regions and general intelligence. For instance, Gläscher and colleagues found general intelligence was associated with damage to left frontal and right parietal cortex in 241 patients with focal brain damage (Gläscher et al., 2009, 2010). Barbey and colleagues found fluid intelligence was associated with damage to predominantly right hemisphere frontal and parietal regions in 158 focal lesion patients (Barbey et al., 2014).

However, the story here is complicated by the fact that these lesion studies have used the Wechsler Adult Intelligence Scale (WAIS) as the basis for deriving measures of general intelligence (Gläscher et al., 2009, 2010) or fluid intelligence (Barbey et al., 2014), rather than using more direct measures of fluid intelligence such as novel problem-solving tasks as in many fMRI studies. Moreover, these lesion studies employed voxel-based methods, which relate deficits to individual voxels in the brain. Such approaches suffer from lower statistical power due to the coarse resolution of lesions and thus may be less sensitive to examination of network effects. The latter is particularly important as focal lesions necessarily disrupt the networks that they are part of, leading to suggestions that focal damage should be considered in the context of brain-wide networks (Gratton et al., 2012; Woolgar et al., 2013).

Where neuropsychological studies have taken a more network-based approach, extent of MD damage, but not non-MD damage, has been found to be predictive of fluid intelligence deficit (Woolgar et al., 2010, 2018). In Woolgar et al. (2010), the extent of damage to regions within the MD network, but not regions outside this network, was linearly predictive of fluid intelligence deficit in 80 focal cortical lesion patients, even when the effect of total lesion volume was taken into account. In the same cohort, a direct comparison of contributions of MD and language systems to fluid intelligence and verbal fluency revealed a double dissociation in which only damage to the MD system was predictive of fluid intelligence loss, providing further evidence that fluid intelligence is specifically linked to the MD system (Woolgar et al., 2018). However, due to limited sample size, these studies did not test whether damage to other brain

networks might also predict fluid intelligence deficit. Moreover, a recent study comprising 227 focal lesion patients (Cipolotti et al., 2023) found that fluid intelligence performance (measured using Raven's Progressive Matrices) did not differ between patients with and without MD lesions, and nor did extent of MD involvement predict fluid intelligence. Rather, this work implicated the importance of a set of predominantly right frontal regions in supporting fluid intelligence (see also Barbey et al., 2014; Santarnecchi et al., 2017). Thus, even the relationship between the MD network and fluid intelligence is not clearly established.

Another key limitation of the previous lesion work is that it only considered extent of grey matter (GM) damage and did not consider the extent to which networks were affected by lesions to WM. Yet this seems critical to account for. For example, cognitive and behavioural changes following stroke can reflect both structural damage at the lesion location and dysfunction in regions distal to the lesion caused by disconnection to the network (Carrera & Tononi, 2014). Moreover, there is evidence to suggest that WM connections are important to supporting general intelligence. For instance, the P-FIT model (Jung & Haier 2007) implicates not only frontoparietal regions, but also efficient information flow between key regions, particularly through arcuate fasciculus and superior longitudinal fasciculus WM tracts, as determinants of differences in general intelligence. A link between WM integrity and intelligence has also been emphasised in other neuroimaging work (e.g., P.-Y. Chen et al., 2020; Penke et al., 2012), and through voxel-based lesion studies which have examined the contribution of WM tracts (e.g., Barbey et al., 2014; Gläscher et al., 2010). In addition, some cognitive impairments have been shown to be better predicted by damage to functional connectivity interactions between regions, rather than structural damage at the lesion location (Siegel et al., 2016).

To address these limitations, here I drew on a large cohort of focal lesion patients ($N=174$) to examine the contribution of different brain networks to fluid intelligence. I used a seven-network resting-state parcellation (Yeo et al., 2011) to compare the contribution of FPN and other networks in a data-driven way. I also examined the relationship between network damage and fluid intelligence loss where network

damage is defined in two ways: through damage to GM (i.e., GM network damage) and damage to WM tracts (i.e., WM network disconnection). Following previous work (Woolgar et al., 2010), I used Cattell Culture Fair to measure current fluid intelligence, and compared this to an estimate of each patient's premorbid fluid intelligence based on their age, gender, and performance on the National Adult Reading Task (NART, with prediction equation derived from healthy controls), so as to derive an estimate of each individual's lesion-induced fluid intelligence deficit.

I addressed two main questions. First, I asked which of the seven RSNs were implicated in supporting fluid intelligence, and in particular whether damage to FPN predicts fluid intelligence loss over and above contributions from other networks. Second, I examined whether different networks were implicated if network damage was defined in terms of GM structural damage or WM disconnection. Finally, I conducted an exploratory analysis to see if additional spatial resolution could be gained by analysing subsets of patients with damage to different lobes and hemispheres.

2.4 Materials and Methods

2.4.1 Participants

I retrospectively reviewed patients that had been recruited as part of the Cambridge Cognitive Neuroscience Research Panel (CCNRP: Cambridge, UK, $N=164$) and the Institute of Cognitive Neurology Research Database (Buenos Aires, Argentina, $N=10$). My patient sample comprised 174 patients (82 females, 92 males, see Supplementary Table S 1 for patient demographic and lesion details). Mean patient age was 52.3 ± 13.4 years (Table 1). I took a mixed aetiology approach, in line with previous studies by our group (Woolgar et al., 2010, 2018) and others (e.g., Cipolotti et al., 2023), supported by findings that aetiology is not a strong predictor of fluid intelligence performance (Cipolotti et al., 2015). Accordingly, inclusion criteria were the presence of chronic, focal lesions of mixed aetiology, and availability of neuropsychological evaluations and neuroradiological data.

Table 1. Demographic information of patient and control groups

Group	Mean age, years (SD)	Mean NART IQ (SD)	Number Males (and Females)
Patients	52.3 (13.4)	108.65 (12.53)	92 (82)
Controls	49.9 (16.5)	110.62 (11.36)	47 (60)
Test	W = 9895, p = 0.376	W = 8368, p = 0.155	$\chi^2(1) = 2.12$, p = 0.145

Tests of differences between age and NART between patient and control groups were done using non-parametric Mann-Whitney test. Test of difference in proportion of males and females between patient and control samples were done using Chi-squared test. SD = standard deviation.

Patients with visual field defects, overt aphasia, pre-insult history of epilepsy, or no MRI data (structural T1 and lesion tracing) were excluded. Of the patient group included in this study, 80 had previously comprised the patient group in existing work by our group, which found that damage within the MD network selectively predicted fluid intelligence loss (Woolgar et al., 2010, 2018).

The distribution of lesions in the patient group is shown in Figure 8. The lesion was located on right in 82 patients, on the left in 76 patients and bilaterally in 16 patients. While I included all patients with complete datasets irrespective of lesion location, the majority of lesions were located in frontal and parietal lobes.

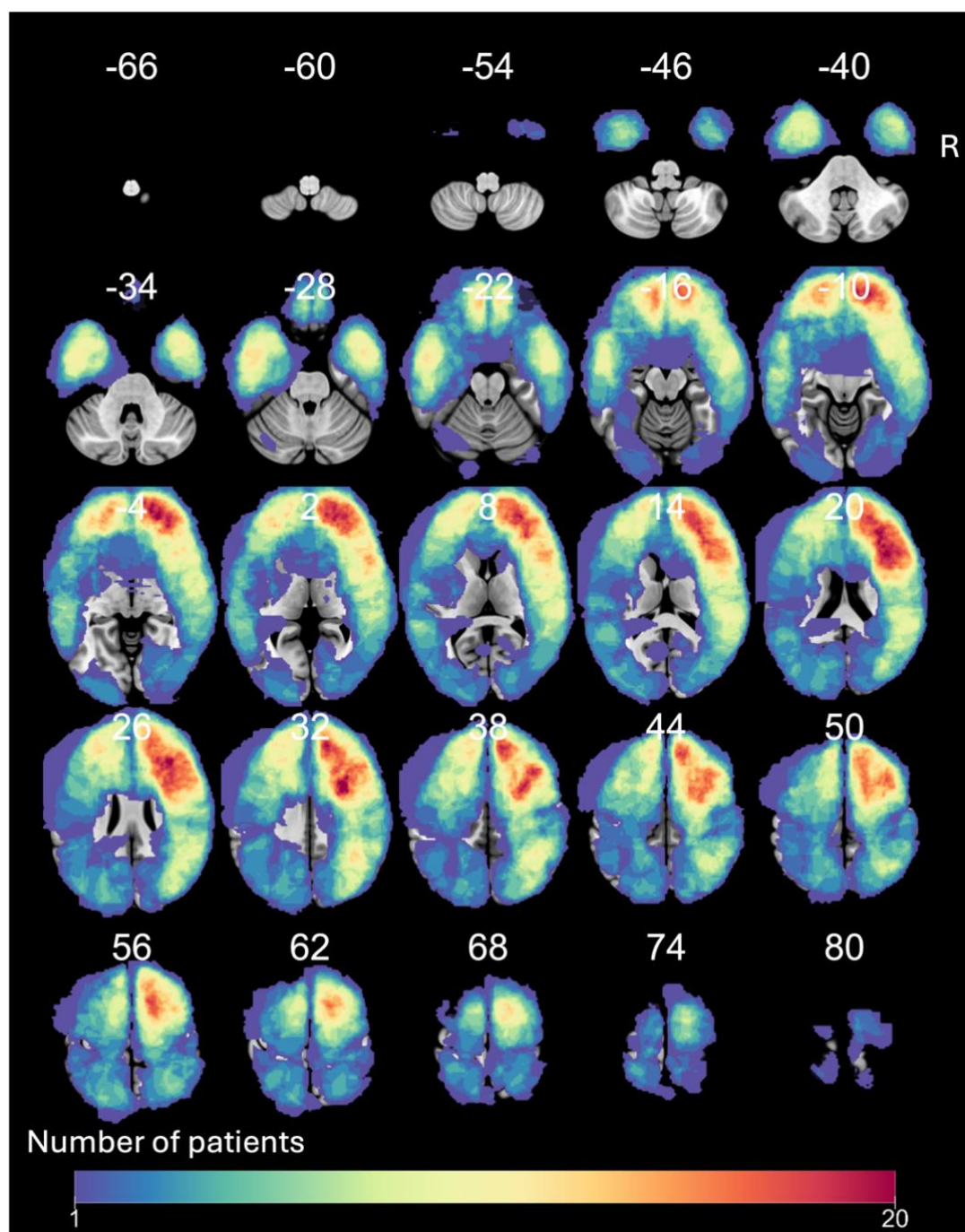


Figure 8. Map of lesion tracings for patient sample shown on standard slices from the Montreal Neurological Institute (MNI) template brain, shown in neurological convention (i.e., right hemisphere is shown on image right). Slices are numbered by z level in MNI space. Colour scale shows degree lesion overlap, with colour of each voxel representing the total number of patients with a lesion tracing that includes that voxel. Note lesion registration imperfections (particularly evident on slices 26 to 50). While lesion coverage is broad, right frontal lesions dominate.

Healthy control participants ($N=107$) were recruited from the Medical Research Council Cognition and Brain Sciences Unit (MRC CBU) Volunteer Panel. Controls were chosen to match the patient sample for distribution of age, premorbid IQ, and gender (Table 1).

All participants gave written informed consent to participate. The study was approved by the Cambridge Local Research Ethics Committee, Cambridge UK.

2.4.2 Estimation of behavioural deficit

The relationship between network damage and absolute fluid intelligence scores can be masked or mitigated by the wide fluid intelligence variation that exists in the normal population. Accordingly, network damage is related here to estimated fluid intelligence deficit, rather than absolute fluid intelligence scores, as in previous work (Woolgar et al., 2010, 2018). The estimation of fluid intelligence deficit in patients proceeded as follows.

All patients and controls completed a reading test comprising phonetically irregular words: National Adult Reading Test (NART; Nelson, 1982; Nelson & Willison, 1991). This test is often used to estimate premorbid mental ability as it correlates well with WAIS Full Scale IQ (0.72-0.81; Crawford et al., 1989; Nelson, 1982), and word reading performance is relatively spared following most non-aphasic brain disorders (Bright et al., 2002). NART IQ was calculated on the basis of number of errors made on this task, using the formula provided in the NART test manual (namely, $-1.24 * \text{NART Error} + 130.6$; Nelson & Willison, 1991).

All participants also completed a novel problem-solving task that loads highly on fluid intelligence: the Cattell Culture Fair (Scale 2 Form A) (Institute for Personality and Ability Testing, 1973). This task consists of four timed sets of problems (series completions, odd-one-out, matrices, topological relations) involving geometric figures.

Cattell Culture Fair performance was translated to a Cattell IQ score, or fluid intelligence score, using Table 5.2 from the Cattell manual.

Control data were used to estimate the relationship between age, gender, NART IQ and Cattell IQ in the healthy population. I derived a multiple regression equation from control data, predicting Cattell IQ from NART IQ, age and gender ($R^2=0.34$, $p<0.005$). Coefficients from this equation were then applied to patient data, allowing estimation of premorbid fluid intelligence for each patient. Fluid intelligence deficit in each patient was then calculated by subtracting their estimated premorbid fluid intelligence score from their current fluid intelligence score (i.e., their Cattell IQ score). Accordingly, negative deficit scores reflect worse current performance than would be expected given performance on the NART, age and gender.

2.4.3 Yeo network parcellations

To define our networks of interest, I used the cortical parcellations provided by Yeo et al. (2011) (available at https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011). Yeo et al. (2011) provide seven- and 17-network parcellations. I used the seven-network liberal mask parcellations in nonlinear Montreal Neurological Institute (MNI) 152 volume space, shown in Figure 9. The seven-network parcellation was chosen to strike a balance between reducing the number of predictors used to predict fluid intelligence deficit, while still having enough resolution to compare contributions of different RSNs. As described in Yeo et al. (2011), these parcellations were identified using functional and structural 3T data from 1000 participants, and have been shown to be highly consistent, with 97.4% of vertices assigned to the same networks when data were split into separate discovery ($N=500$) and replication samples ($N=500$).

7-Network Parcellation (N=1000)

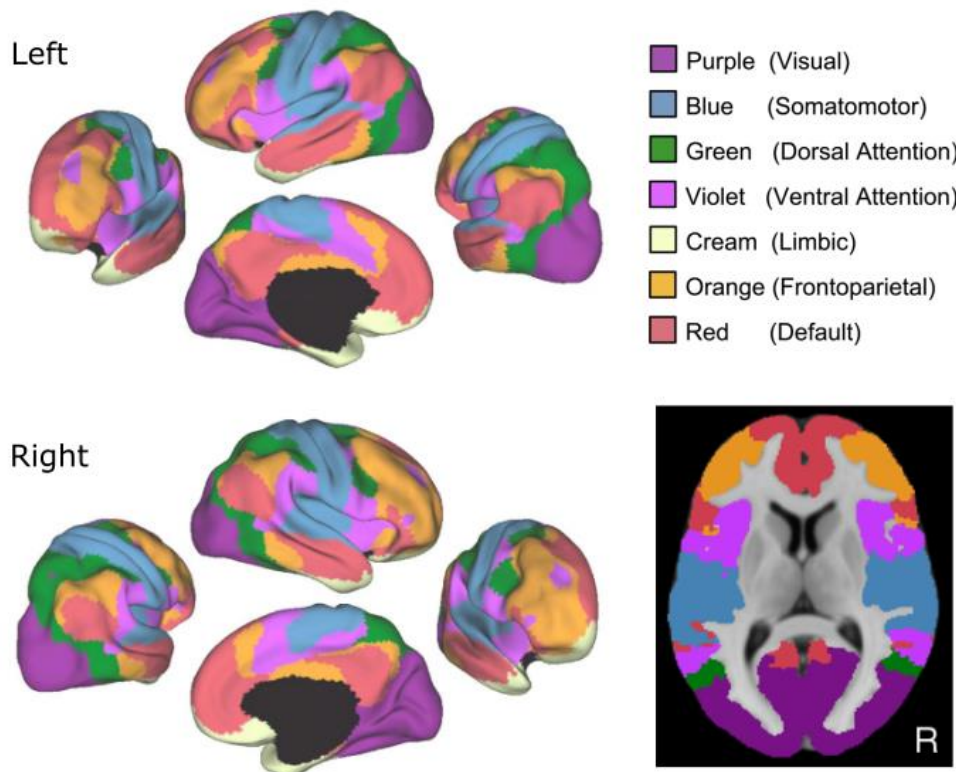


Figure 9. Seven-network parcellation from Yeo, et al. (2011). Inset in bottom right shows the mask in volumetric space. Networks are indicated by colour, as described in top right and as follows: purple = visual network, blue = somatomotor network, green = dorsal attention network (DAN), violet = ventral attention network, cream = limbic network, orange = frontoparietal network (FPN), red = default network. Figure adapted from Yeo, et al. (2011) and from https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011, reproduced with permission.

2.4.4 Structural neuroimaging data

As reported in Woolgar et al. (2010), T1-weighted gradient-recalled MRI scans (1x1x1mm resolution) were acquired for all patients. Lesions were manually traced onto an anatomical MRI by a neurologist who was blind to the experimental results. Lesion tracing was performed using MRICro or MRICroN (Rorden & Brett, 2000). Scans were normalised with SPM (www.fil.ion.ucl.ac.uk/spm) using cost-function lesion masking (Brett et al., 2001). The normalisation parameters obtained through normalising subject

TIs, were then used to normalise lesion tracings to MNI space. These tracings were then binarised by rounding values to the nearest integer.

2.4.5 Calculation of lesion damage

2.4.5.1 Overlap of lesion with cortical parcels of Yeo's networks (grey matter network damage)

To determine the degree of GM damage to each of the Yeo et al. (2011) network parcels, each patient's normalised and binarised lesion mask was multiplied against the Yeo et al. (2011) network atlas parcellation. The degree of overlap between the lesion mask and the Yeo et al. (2011) parcellations was computed by summing the number of values corresponding to each atlas label. Accordingly, if part of a patient's lesion tracing was outside of the brain (i.e., due to imperfections in lesion registration), it was excluded. This process yielded the number of voxels affected by each patient's lesion for each network. These values were then converted into volume of damage (cm³) for each of the seven networks. Volume calculations were performed using niBabel and Nilearn in Python (v3.10.6).

2.4.5.2 Overlap of lesion with network-level connections (white matter network disconnection)

Following the approach taken by Mandonnet and colleagues (Mandonnet et al., 2020; Mrah et al., 2022), I sought to estimate the degree to which each patient's lesion disconnected each of the seven Yeo networks (Yeo et al., 2011). This approach involved two steps: first, calculating the average number of WM tracts within each network across a large sample of healthy brains, and second, quantifying how many of these within-network tracts were affected by each patient's lesion.

2.4.5.2.1 *Healthy tractograms*

To estimate the number of connections within each of these networks in healthy brains, I used whole-brain tractography reconstructions in the standard MNI152 space provided by Michel Thiebaut de Schotten (private correspondence). These structural connectome data were derived from the diffusion-weighted imaging dataset of 1042 healthy participants acquired at 3 Tesla by the HCP Team. Complete processing methods are detailed in Thiebaut de Schotten et al. (2020). In brief, tractograms were produced using whole-brain deterministic tractography in native diffusion-weighted imaging space with StarTrack software (<https://www.mr-startrack.com>), and then co-registered to standard MNI space.

2.4.5.2.2 *Estimate of structural disconnection*

Using the disconets pipeline (https://github.com/scilus/disconets_flow/), based on SCIPLY tools (<https://github.com/scilus/scilpy>) (see Mrah et al., 2022), I used the Yeo seven-network atlas as a mask to filter each of the 1042 healthy HCP tractograms, and estimate the number of tracts in each network. As in previous work using this pipeline (Mandonnet et al., 2020; Mrah et al., 2022), I counted only streamlines which started and ended within a given network as comprising that network. This process yielded seven values (one per network) for each of the 1042 healthy HCP participants, where each of these values represented the total number of streamlines for that network. An example of the streamlines comprising each network for one healthy tractogram is shown in Figure 10.

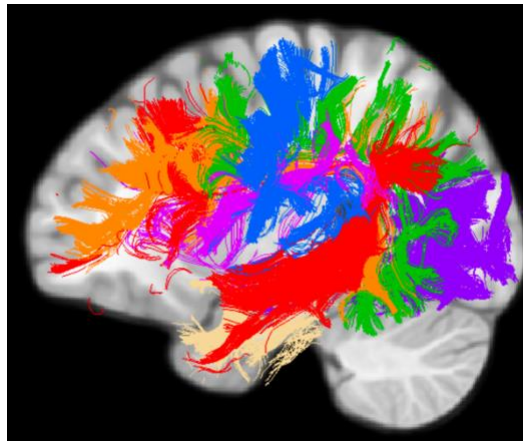


Figure 10. Streamlines parcellated into Yeo et al. (2011) networks for one healthy tractogram, shown as an example. Colours, as in Figure 9, are as follows: purple = visual network, blue = somatomotor network, green = dorsal attention network, violet = ventral attention network, cream = limbic network, orange = frontoparietal network, red = default network.

To determine the degree of structural disconnection across each of the networks in each patient, lesion tracings were first registered into the same MNI space as the tractograms using ANTs registration (<https://github.com/ANTsX/ANTs>). Each lesion tracing was then used to filter each of the 1042 Yeo-parcellated tractogram network maps to determine the number of streamlines in each network affected by the lesion. Streamlines for a given network were ‘affected’ by the lesion if any part of a streamline fell within the bounds of the lesion tracing. For each patient, this process thus generated 1042 values for each network, representing the number of streamlines affected by the lesion for each of the healthy HCP tractograms. An example of a lesion tracing overlaid over a parcellated tractogram is shown in Figure 11 for one patient against one healthy tractogram. For this patient and this tractogram, all streamlines that fell within the bounds of the lesion tracing (in red) were considered damaged by the lesion. I then computed the degree of WM disconnection for each of the seven Yeo networks by averaging the number of affected streamlines in each network across the 1042 tractograms. Thus, for each patient I obtained seven values, each of which reflected an estimate of the number of streamlines affected by their lesion for each of the seven Yeo networks. That is, this approach estimates the extent of structural disconnection caused by the lesion in each network.

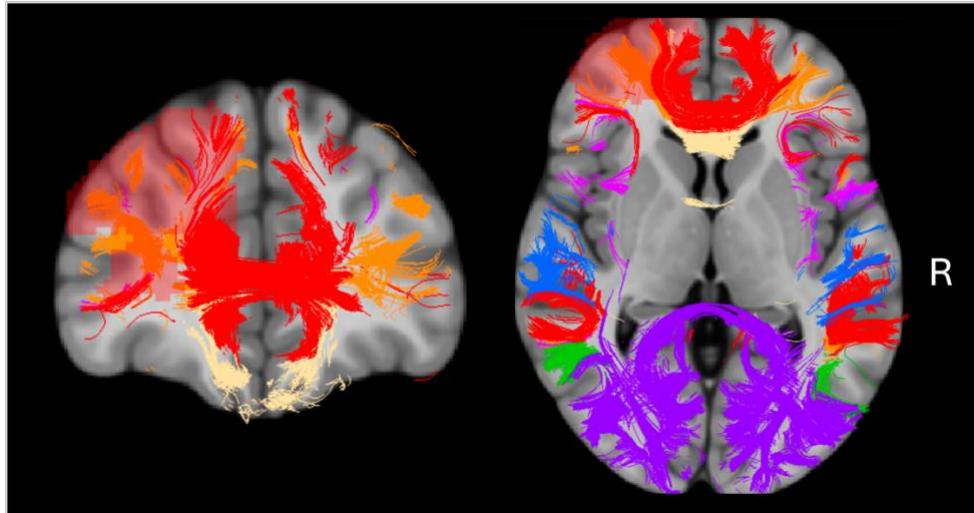


Figure 11. Streamlines parcellated into Yeo networks for one healthy tractogram, with a lesion tracing for one patient shown in red as an example. Affected or damaged streamlines for this patient in this tractogram are those which fall within the bounds of this lesion tracing. Colours, as in Figure 9, are as follows: purple = visual network, blue = somatomotor network, green = dorsal attention network, violet = ventral attention network, cream = limbic network, orange = frontoparietal network, red = default network. In this case, the majority of the streamlines affected were streamlines connecting the frontoparietal (orange) and default (red) networks.

2.4.5.3 Broad measure of damage (total lesion volume)

Total damage, reflected in total lesion volume (TLV), is often strongly predictive of cognitive change. Thus, for any network predicting fluid intelligence deficit in the main analysis, I additionally asked whether the relationship between network damage and cognitive deficit remained even after accounting for any effect of TLV. I tested this by asking whether damage to each network continued to predict fluid intelligence loss once TLV was partialled out. TLV was computed by counting the total number of voxels within the normalised and binarised lesion mask (irrespective of whether they fell within the Yeo network parcellations), and converting these values into volumetric units (cm^3).

2.4.6 Regression methods

My primary aim was to characterise the relationship between damage to the seven networks and fluid intelligence loss, where damage was defined through GM damage to, and WM disconnection of, the cortical networks.

I initially planned to include both GM damage and WM disconnection in each network as separate predictors in a single regression predicting fluid intelligence loss. However, since GM damage and WM disconnection are related measures, I first checked whether multicollinearity might be a concern for a regression comprising all of these predictors. For this I checked whether any predictors were highly correlated (e.g., $R > 0.8$ or 0.9 , Field et al., 2012) and examined variance inflation factors (VIF) for each predictor comprising a model. VIF reflect the degree of multicollinearity between predictors, and VIF greater than 10 indicate that multicollinearity may be a concern (Bowerman & O'Connell, 1990). A correlation matrix with all predictors showed that there was a high degree of collinearity between GM and WM disconnection measures of damage for a given network (see Figure 16 in Results). Accordingly, I took several different approaches to address the primary research question while avoiding multicollinearity concerns, detailed as follows. All statistical analyses were performed using R (v4.3.2) in RStudio (v2023.12.0).

2.4.6.1 Analysis 1. Damage to which networks predicts fluid intelligence loss? Combining GM and WM measures of network damage

Given the high degree of correlation between the GM damage and WM disconnection to each network, I first took a composite score of damage, collapsing both measures into one damage score per network. I calculated these scores by first converting the GM damage and WM disconnection scores for a given network into z-scores, and then averaging across these two scores. This yielded seven composite network damage scores for each patient. I then defined a linear model with these seven predictors, with fluid

intelligence loss the dependent variable. This model identified networks that significantly predicted fluid intelligence loss over and above the contributions of other networks.

Multicollinearity concerns precluded defining a model with all composite network damage predictors and TLV (largest VIF>20). Therefore, to test whether damage to any of the networks explained variance in fluid intelligence loss over and above that explained by TLV, I tested whether any network identified as a significant predictor of fluid intelligence loss in the first model remained predictive once TLV was partialled out. To do this, for each network identified as a significant predictor of fluid intelligence loss in the main analysis, I defined a model predicting fluid intelligence which comprised TLV and that network. This allowed me to test whether the addition of the network significantly explained more variance in fluid intelligence loss than TLV alone and vice-versa.

2.4.6.2 Analysis 2. Comparison of GM network damage and WM network disconnection.

Next, I asked whether the two methods of calculating network damage converged on implicating the same or different networks as underpinning fluid intelligence. As the two predictors were highly correlated in each network, I could not meaningfully ask which was the most predictive in each network. Therefore, I sought to explore their similarity and differences by conducting separate multiple regressions: one for GM network damage and one for WM disconnection scores. Each model comprised damage to each of the seven networks as predictors, defined according to the applicable metric, converted into z-scores. Fluid intelligence loss was the dependent variable in both models.

As for Analysis 1, I also sought to ascertain whether damage to these networks remained predictive of fluid intelligence loss once a broader measure of lesion damage was accounted for using multiple regression. This was done separately for GM and WM

definitions of network damage. Again, multicollinearity concerns precluded me from defining a model comprising all networks and TLV (largest VIF >20 for GM damage and WM disconnection models). Therefore, to test whether networks identified as significant predictors remained predictive once TLV was accounted for, I carried out a similar procedure as for the collapsed networks. That is, I defined separate models comprising only TLV and each of the networks identified as significant predictors of fluid intelligence deficit in the GM damage or WM disconnection models.

2.4.6.3 Analysis 3. Relationship between network damage and lesion location – exploratory analyses

As a final step, I sought to gain additional spatial resolution by examining whether lesions in a particular lobe or hemisphere might be driving our results for any identified network(s) of interest. To do so, I examined whether the relationship between network damage and fluid intelligence loss differed according to primary lesion site (frontal, parietal, occipital, and temporal lobe) and lesion side (left and right hemisphere). I restricted this analysis to networks identified as significant predictors of fluid intelligence deficit in Analysis 1. I examined whether the relationship between damage to these networks and fluid intelligence deficit differed according to the primary lesion site and side. For these analyses, network damage was defined using the composite GM damage and WM disconnection score for that network, as defined in Analysis 1.

I first classified patients into four groups, according to whether their lesion affected their frontal, parietal, temporal, or occipital lobe. Patients with subcortical lesions or with lesions affecting more than one lobe were excluded from this analysis. For each network identified as a significant predictor of fluid intelligence deficit in Analysis 1, I correlated extent of damage to that network and fluid intelligence deficit in each subgroup of patients. This analysis indicated whether the strength of this relationship differed between patient groups. As in Analyses 1 and 2, I also tested whether significant effects observed here remained significant once the effects of TLV were partialled out.

Conclusions from this analysis were limited by small sample sizes in some of these sub-groups.

To explore differences between lesion side, I focussed my analysis on frontal patients, for whom I had the most participants. I created two sub-groups of frontal patients: patients with lesions in their left and right hemisphere. Patients with lesions affecting both hemispheres were excluded from this analysis. I again examined the relationship between network damage and fluid intelligence deficit in each frontal patient sub-group, focussing only on networks identified as significant predictors of fluid intelligence deficit in Analysis 1.

2.5 Results

2.5.1 Network coverage

The distribution of lesions in the patient group for each of the seven networks is depicted in Figure 12. Few patients in our sample had damage within the visual network (i.e., Network 1). There appeared to be reasonable coverage of the other networks.

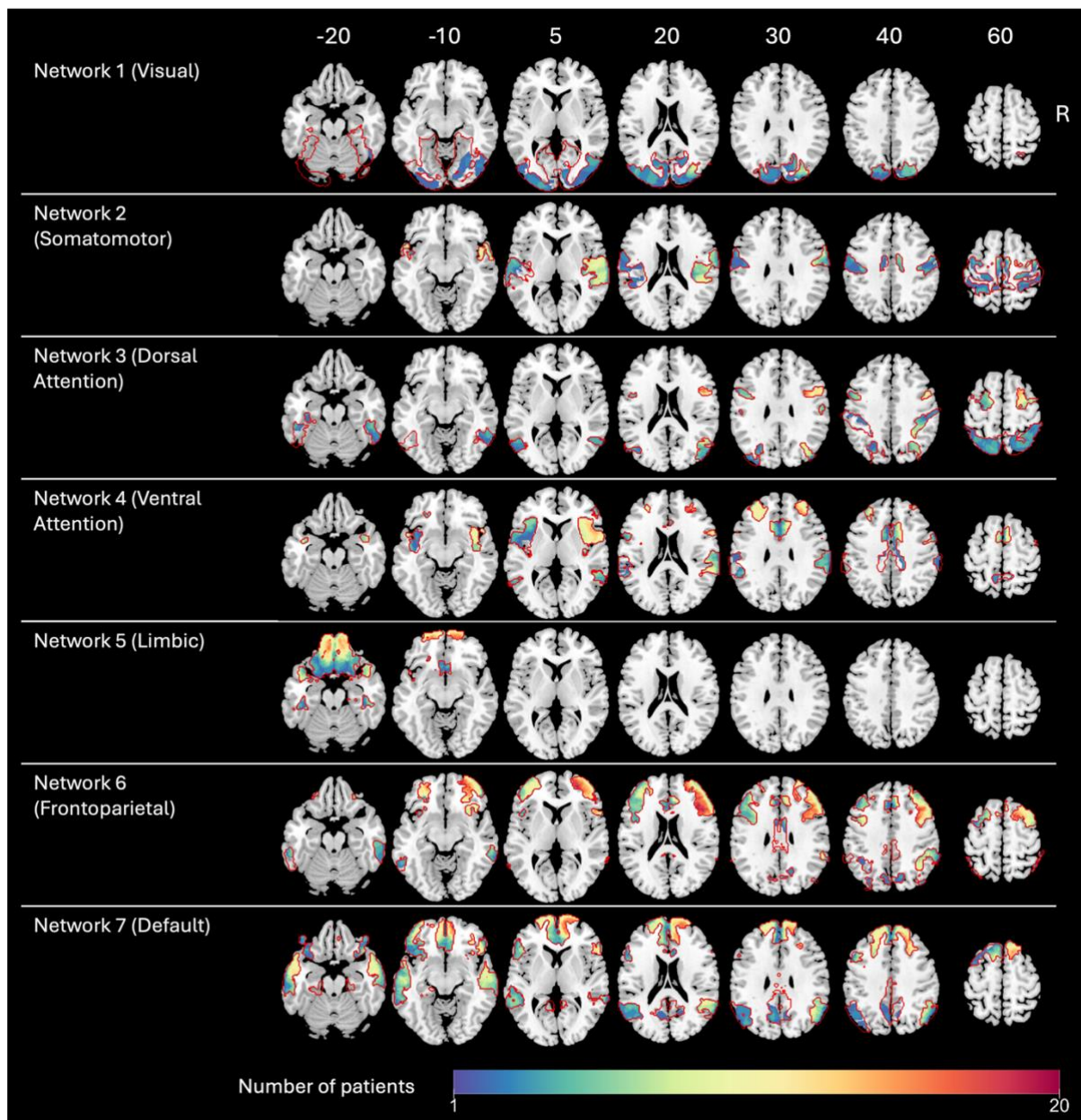


Figure 12. Map of overlap of lesion tracings for patient sample shown on standard slices from the MNI template brain, shown for each network parcel separately. Axial slices are numbered by z level in MNI space, shown in neurological convention (i.e., right hemisphere is shown on image right). Outline of network parcel shown in red. Colour scale shows degree of lesion overlap, with colour of each voxel representing the total number of patients with a lesion tracing that includes that voxel.

2.5.2 Calculation of behavioural deficit

I used control data to establish the relationship between fluid intelligence (Cattell IQ) and age, gender and NART IQ using multiple regression. The regression model was statistically significant, explaining 34.33% of the variance in control Cattell IQ ($F(3,103)=17.95$, $p<0.0001$). NART IQ ($\beta=0.78$, $p<0.0001$) was positively associated with fluid intelligence, while age showed a negative relationship ($\beta=-0.46$, $p<0.0001$), indicating fluid intelligence declines with age. While gender did not have a significant association with Cattell IQ, it was included in the model for completeness ($\beta=-1.38$, $p=0.616$). As an indication of model fit, predicted Cattell IQ scores (calculated using leave-one-out cross-validation) showed a moderate, significant correlation with current scores in controls ($R=0.54$, $p<0.0001$; Figure 13).

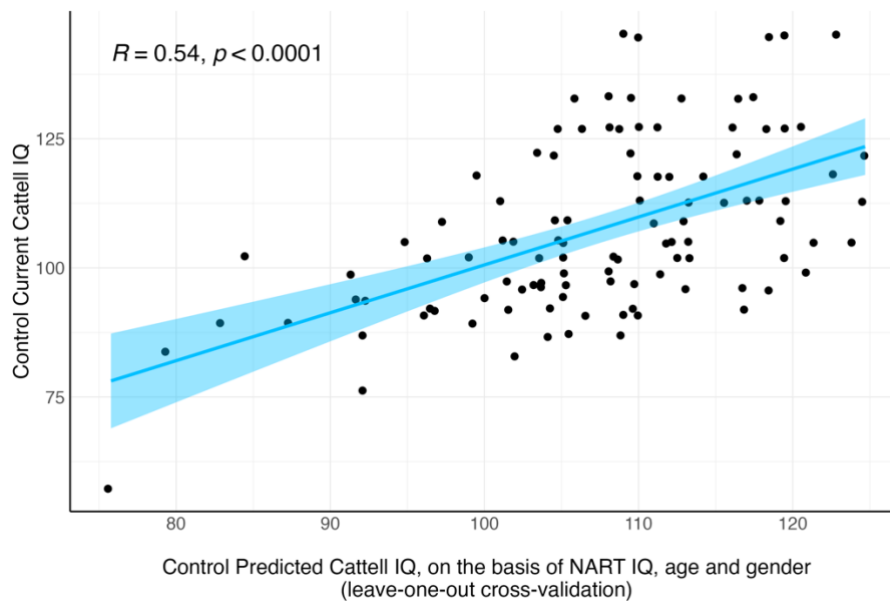


Figure 13. Scatterplot of current Cattell IQ scores against predicted Cattell IQ scores for the control cohort. To more accurately reflect the fit of the model, predicted Cattell IQ scores shown above were predicted on the basis of NART IQ, age and gender using leave-one-out cross-validation. Predicted scores show a moderate, significant correlation with actual scores.

Regression coefficients from this model were then applied to patient data to estimate each patient's premorbid fluid intelligence. Fluid intelligence deficit for each patient was then determined by subtracting estimated premorbid fluid intelligence scores from current fluid intelligence scores (i.e., Cattell IQ). Mean fluid intelligence deficit in the patient cohort was -11.99 ± 15.80 .

2.5.3 Analysis 1. Damage to FPN selectively predicts fluid intelligence loss, but effects cannot be separated from TLV

I first asked whether composite damage (reflecting both GM network damage and WM network disconnection) to any of the networks predicted fluid intelligence deficit. Figure 14 shows the raw correlation coefficients across all predictors (Panel A) and raw correlation plots (Panels B to I) between each predictor and fluid intelligence deficit. Negative fluid intelligence deficit scores indicate poorer performance on fluid intelligence tasks than expected given performance on the NART, thus negative correlations here reflect that more damage for a given network is associated with greater fluid intelligence loss. As shown in Figure 14 (Panels B to I), of all network predictors, FPN and TLV independently explain the most variance in fluid intelligence deficit ($R^2=0.11$ and 0.10 , respectively).

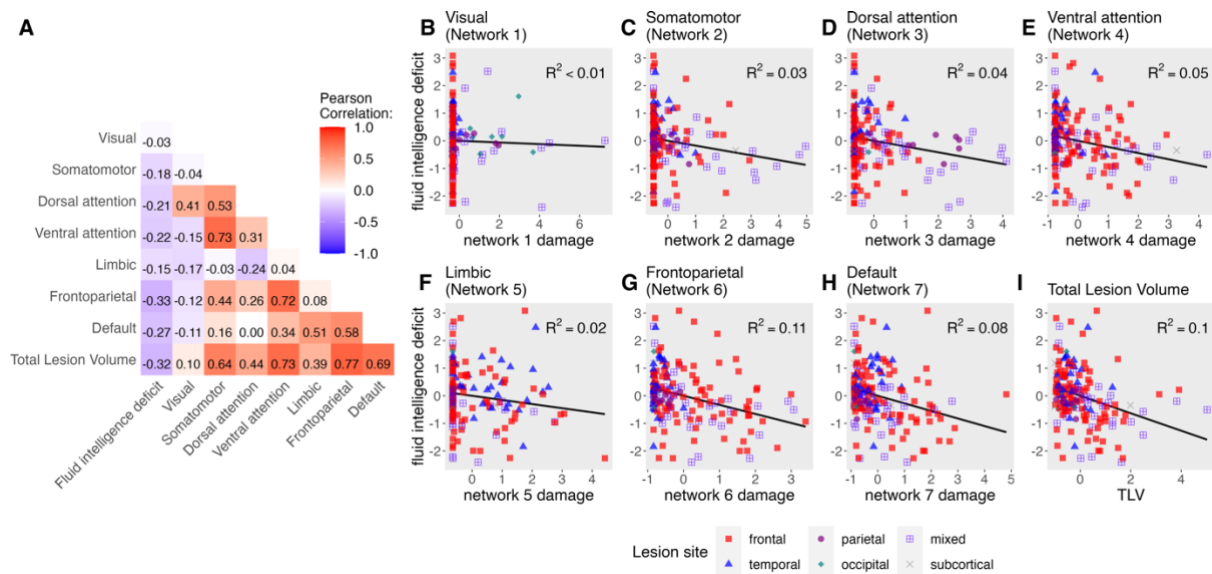


Figure 14. Panel A. Pearson correlations between variables of interest: extent of damage to each network (composite damage scores), total lesion volume, and fluid intelligence deficit. Values shown are Pearson correlation coefficients (r), and colours show strength of correlation (red = positive relationship, blue = negative relationship). Panels B to I. Individual regression lines show the raw correlation between fluid intelligence deficit and damage to each individual network (Panels B to H) and total lesion volume (TLV, Panel I). Network damage is defined according to a composite of GM damage and WM disconnection to this network and values are z-scored. The amount of variance explained (R^2) in each correlation is shown in the upper right of each plot. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

For inference, I entered the composite score from all seven networks into a multiple regression, with behavioural deficit as the dependent variable. This model was not subject to issues of multicollinearity (all VIF < 4 , all correlations were below 0.8). Model coefficients and results of this model are shown in Table 2 and are shown graphically as partial regression plots in Figure 15. FPN was the only significant predictor of fluid intelligence deficit, after accounting for all other networks ($\beta = -0.27$, $p = 0.036$).

Table 2. Beta estimates for collapsed GM and WM version of model with all networks included.

fluid intelligence deficit			
Predictors	β	SE β	p
Visual	-0.01	0.08	0.914
Somatomotor	0.02	0.12	0.861
Dorsal Attention	-0.19	0.10	0.060
Ventral Attention	0.04	0.14	0.778
Limbic	-0.14	0.09	0.110
Frontoparietal	-0.27	0.13	0.036
Default	-0.06	0.11	0.584
Observations	174		
R^2 / R^2 adjusted	0.155 / 0.119		

Note, predictors were scaled before being entered into model. Significant p-values are indicated in bold and indicate the corresponding network was identified as a significant predictor of fluid intelligence loss (after accounting for other predictors). Only frontoparietal network was identified as a significant predictor of fluid intelligence deficit, after accounting for all other networks.

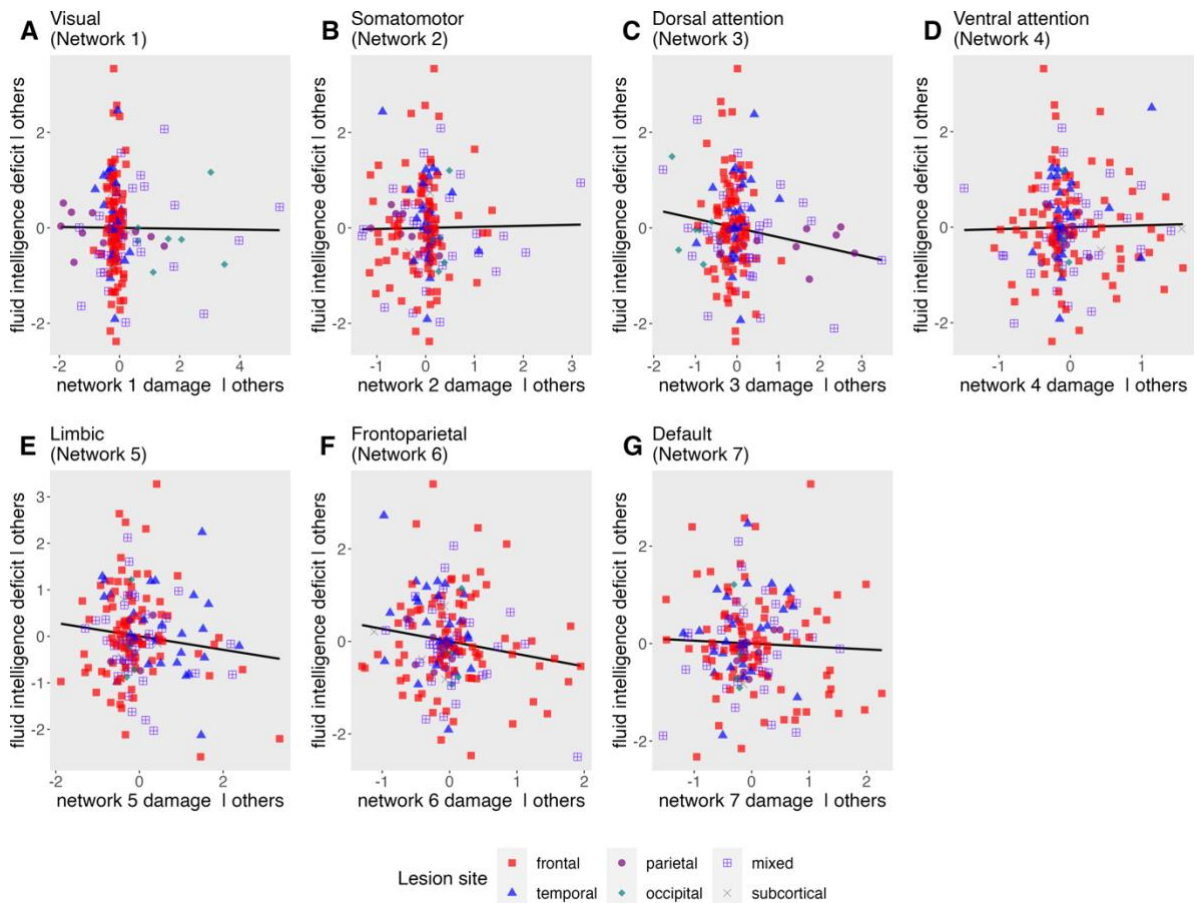


Figure 15. Partial regression plots (or added-variable plots) for each network for the model comprising all networks, collapsed over GM damage and WM disconnection. For example, Panel A shows relationship between damage to visual network (Network 1) and fluid intelligence deficit, after accounting for contributions from all other networks. Individual regression lines show the relationship between fluid intelligence deficit and damage to each individual network, controlling for the presence of damage in all other networks. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

Next, I asked whether the predictive effect of damage to FPN was over and above any predictive effect of TLV. TLV was a significant predictor of fluid intelligence deficit on its own ($\beta = -0.32$, $p < 0.0001$, $R^2 = 0.10$, $\text{adj-}R^2 = 0.09$, see Figure 14, Panel I). In a model comprising both FPN damage and TLV (VIF for both predictors < 3), neither predictor significantly predicted fluid intelligence deficit alone (TLV: $\beta = -0.15$, $p = 0.171$, FPN: $\beta = -0.21$, $p = 0.065$), suggesting these predictors explained similar variance. Indeed, extent of FPN explained 59.31% of the variance in TLV ($F(1,172) = 250.70$, $p < 0.0001$, see Figure 14, Panel A).

In summary, of all the networks, damage to FPN was most predictive of fluid intelligence deficit. However, extent of damage to this network also predicted (and was predicted by) TLV, making it difficult to separate the unique contributions of these predictors.

2.5.4 Analysis 2. GM damage to FPN and DAN, but not WM disconnection, predict fluid intelligence loss

The results thus far point to the possible importance of FPN in supporting fluid intelligence, with damage to this network significantly predictive of fluid intelligence loss, albeit not above TLV.

Here, I asked whether GM and WM methods for calculating network damage suggest the same or different networks of importance. Figure 16 shows raw Pearson correlations between each of the networks, separated into GM and WM measures of damage, with TLV and fluid intelligence deficit. GM and WM definitions of damage in each network were highly correlated (range of Pearson r for GM and WM of each network: 0.70 (somatomotor network) to 0.95 (visual network)).

Chapter 2

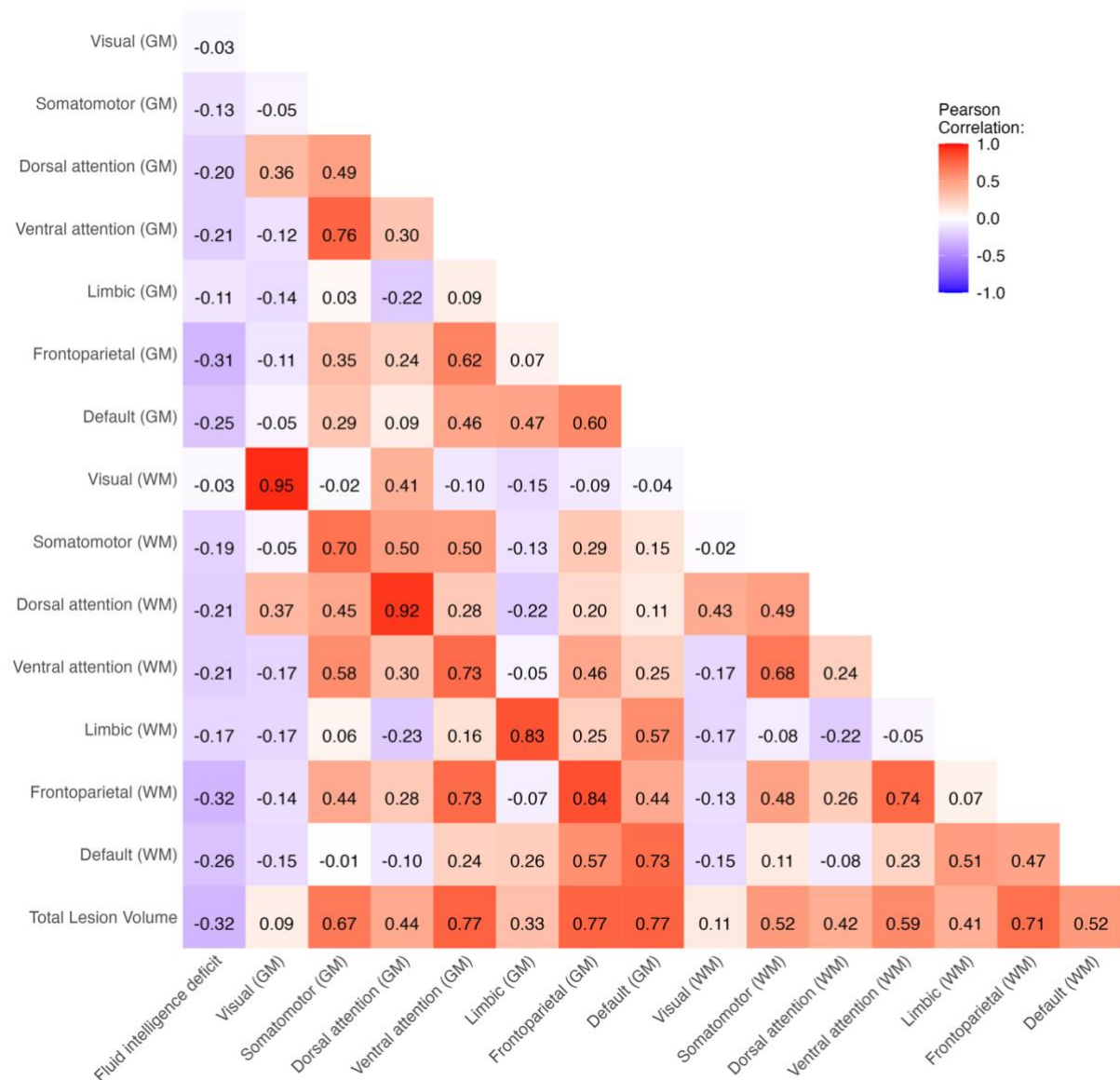


Figure 16. Pearson correlations between all variables of interest: extent of GM damage and WM disconnection to each network, total lesion volume (TLV) and fluid intelligence deficit. Values shown are Pearson correlation coefficients (r), and colours show strength of correlation (red = positive relationship, blue = negative relationship). Raw correlation plots showing the relationship between damage to each network and fluid intelligence deficit are shown in Supplementary for each network according to degree of GM damage (Figure S 1) or degree of WM disconnection (Figure S 2).

In the multiple regression using GM network damage (VIF for all GM variables <4), damage to FPN ($\beta=-0.25$, $p=0.033$) and DAN ($\beta=-0.12$, $p=0.047$) were retained as significant predictors of fluid intelligence deficit. Interestingly, despite the high correlation with GM values, in the multiple regression using WM disconnection (VIF for all WM variables <4), no significant predictors of fluid intelligence deficit were identified (all p -values > 0.05), although they trended in a similar direction. Table 3 shows the model coefficients and results of GM damage and WM disconnection models, and the relationships are shown graphically as partial regression plots for GM damage and WM disconnection separately in Figure 17 and Figure 18, respectively.

Table 3. Beta estimates for GM damage model (top left) and WM disconnection model (bottom right) with all networks included.

Predictors	fluid intelligence deficit					
	β	SE β	p	β	SE β	p
Visual (GM)	-0.00	0.08	0.974			
Somatomotor (GM)	0.09	0.13	0.488			
Dorsal attention (GM)	-0.20	0.10	0.047			
Ventral attention (GM)	-0.04	0.14	0.792			
Limbic (GM)	-0.12	0.09	0.166			
Frontoparietal (GM)	-0.25	0.11	0.033			
Default (GM)	-0.03	0.11	0.762			
Visual (WM)				-0.02	0.08	0.826
Somatomotor (WM)				-0.01	0.11	0.942
Dorsal attention (WM)				-0.19	0.10	0.058
Ventral attention (WM)				0.02	0.13	0.852
Limbic (WM)				-0.15	0.09	0.086
Frontoparietal (WM)				-0.23	0.12	0.066
Default (WM)				-0.10	0.10	0.302
Observations		174			174	
R ² / R ² adjusted		0.134 / 0.097			0.161 / 0.126	

Note, predictors were scaled before being entered into the model. Significant p-values are indicated in bold and indicate the corresponding network was identified as a significant predictor of fluid intelligence loss (after accounting for other predictors). When examining GM damage to networks only, dorsal attention and frontoparietal networks were identified as significant predictors of fluid intelligence loss, once damage to other networks were accounted for. While not significant, the pattern was similar for WM network damage with the largest beta estimates for dorsal attention and frontoparietal networks.

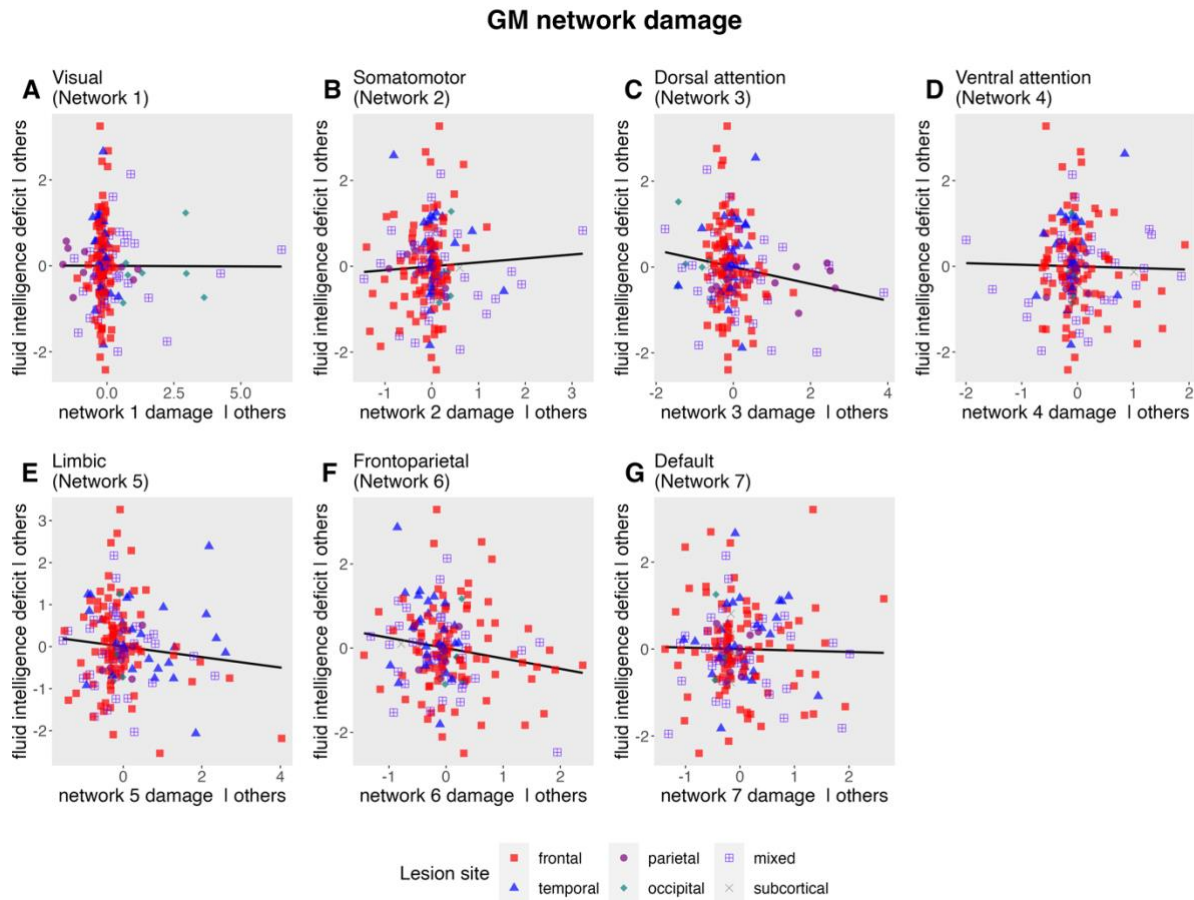


Figure 17. Partial regression plots (or added-variable plots) for each network for the model comprising all networks, where network damage is defined by degree of lesion overlap with GM cortical networks. For example, Panel A shows relationship between GM damage to visual network (Network 1) and fluid intelligence deficit, after accounting for contributions from all other networks. As in Figure 15, individual regression lines show the relationship between fluid intelligence deficit and damage to each individual network, controlling for the presence of damage in all other networks. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

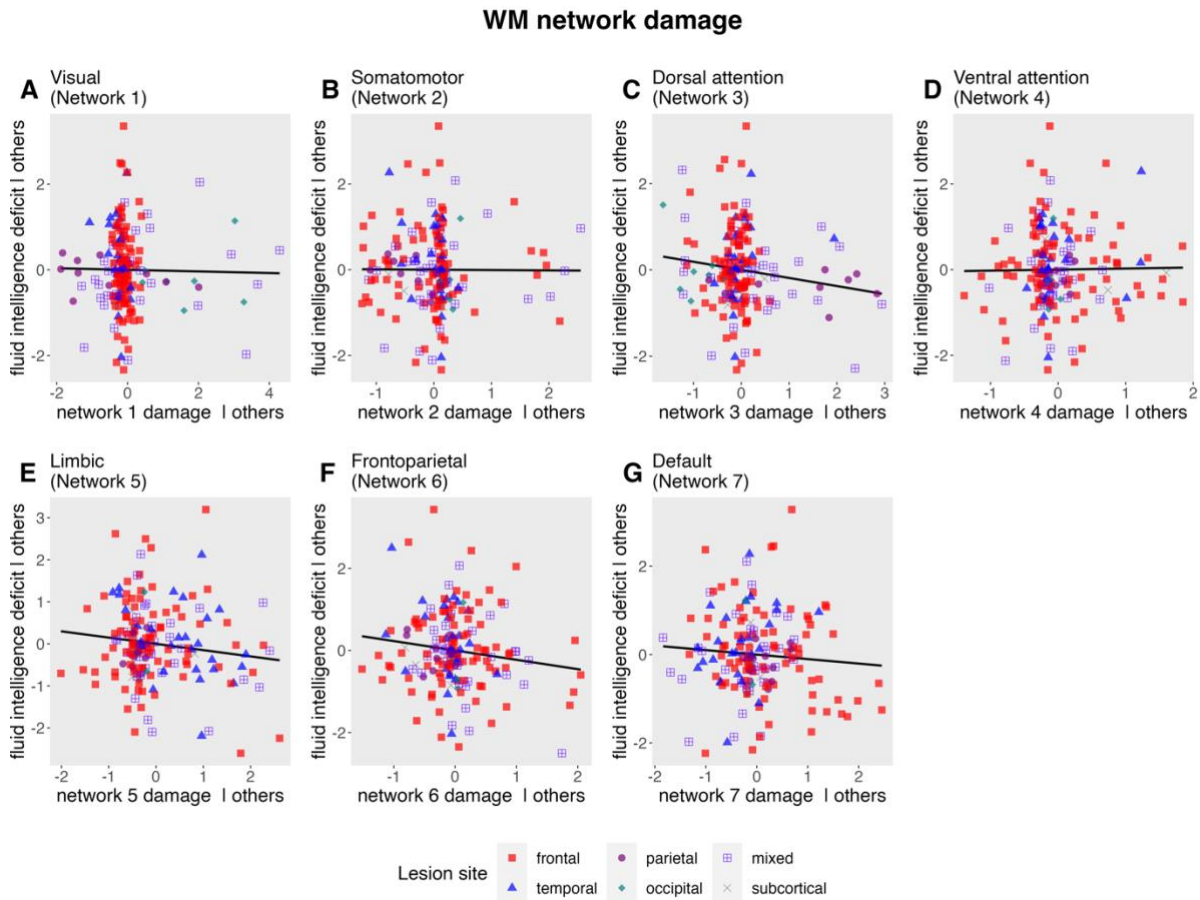


Figure 18. Partial regression plots (or added-variable plots) for each network for the model comprising all networks, where network damage is defined by degree of WM disconnection caused by the lesion. For example, Panel A shows relationship between degree of WM disconnection within visual network (Network 1) and fluid intelligence deficit, after accounting for contributions from all other networks. As in Figure 15, individual regression lines show the relationship between fluid intelligence deficit and damage to each individual network, controlling for the presence of damage in all other networks. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

Again, I sought to ascertain whether damage to GM networks identified as being predictive of fluid intelligence loss (i.e., FPN and DAN) were predictive over and above broader measures of damage (i.e., TLV). In the model comprising FPN GM damage and TLV (VIF for both predictors <3), neither predictor significantly predicted fluid intelligence deficit (TLV: $\beta=-0.19$, $p=0.100$, GM FPN: $\beta=-0.17$, $p=0.130$), suggesting that I could not separate the contributions of these predictors.

In the model comprising DAN GM damage and TLV (VIF for both predictors <2), TLV remained a significant predictor of fluid intelligence loss after accounting for GM damage to DAN (TLV: $\beta=-0.28$, $p=0.001$, DAN: $\beta=-0.07$, $p=0.371$). For completeness, I computed a final model where fluid intelligence loss was predicted on the basis of all three of these predictors (i.e., GM damage to FPN, DAN and TLV; VIF for all predictors <3). Here, no single predictor was found to significantly explain fluid intelligence loss (all p -values > 0.089).

In summary, these results again point to the contribution of FPN to fluid intelligence, and suggest that the contribution of this network is captured by considering its GM effects as well as, or better than, considering the extent of WM disconnection. This analysis also suggested a possible contribution of DAN to fluid intelligence. TLV was predictive over and above the contribution of DAN, but TLV and the contribution of FPN could not be separated.

2.5.5 Analysis 3. Extent of FPN damage predicts fluid intelligence deficit in frontal patients

The FPN is distributed across several brain regions. Having identified it as the most predictive network for causing fluid intelligence deficit after damage, I next carried out a series of exploratory analyses to examine whether damage to any particular part of this network was most critical. For this, I examined the relationship between extent of FPN damage and fluid intelligence loss across sub-groups of patients, split according to the

primary lesion site: frontal ($N=88$), parietal ($N=12$), temporal ($N=28$), and occipital ($N=7$) lobe (Figure 19). There was only statistical evidence for a relationship between FPN damage and fluid intelligence deficit for frontal patients ($r=-0.28$, false discovery rate (FDR) correction across four comparisons, $\text{adj-}p=0.030$; all other $\text{adj-}p$ -values > 0.984). As the sample sizes for remaining patient sub-groups were small, interpretation of lack of results for these groups is limited.

To test whether the result for frontal patients remained significant after TLV was accounted for, I defined a model comprising FPN damage and TLV (VIF for both predictors < 3). As in analysis 1, neither predictor significantly predicted fluid intelligence deficit once the other was accounted for (TLV: $\beta=-0.19$, $p=0.330$, FPN: $\beta=-0.16$, $p=0.370$).

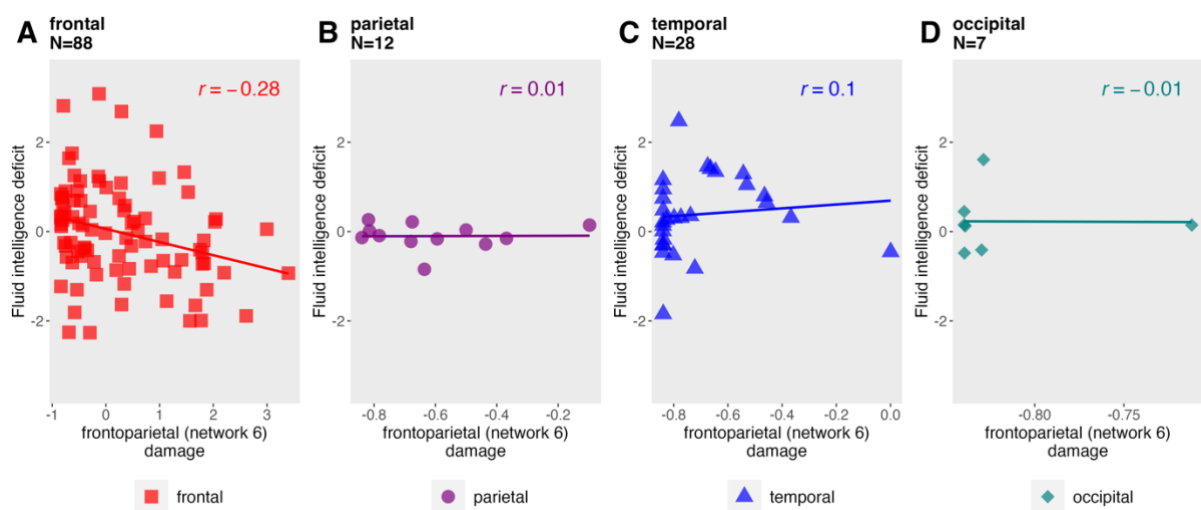


Figure 19. Individual regression lines show correlation between fluid intelligence deficit and extent of frontoparietal network damage in patient sub-groups: patients with frontal damage ($N=88$), parietal damage ($N=12$), temporal damage ($N=28$), and occipital damage ($N=7$). Network damage is defined according to a composite of GM damage and WM disconnection to this network and values are z-scored. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond). Pearson correlation coefficients are shown in upper right of each plot. The relationship between frontoparietal network damage and fluid intelligence deficit was only significant for frontal lesion patients ($\text{adj-}p=0.030$, all other $\text{adj-}p$ -values > 0.984).

I next sought to ascertain whether the relationship between FPN damage and fluid intelligence loss held for both left ($N=33$) and right ($N=40$) frontal lesion patients. A visual representation of this analysis hinted that a correlation between these variables might only pertain to right frontal patients (see Figure 20). However, after multiple comparison correction (FDR correction across two tests), the relationship was not significant for either group (left frontal $r = -0.03$, $\text{adj-}p=0.887$; right frontal $r = -0.32$, $\text{adj-}p=0.094$).

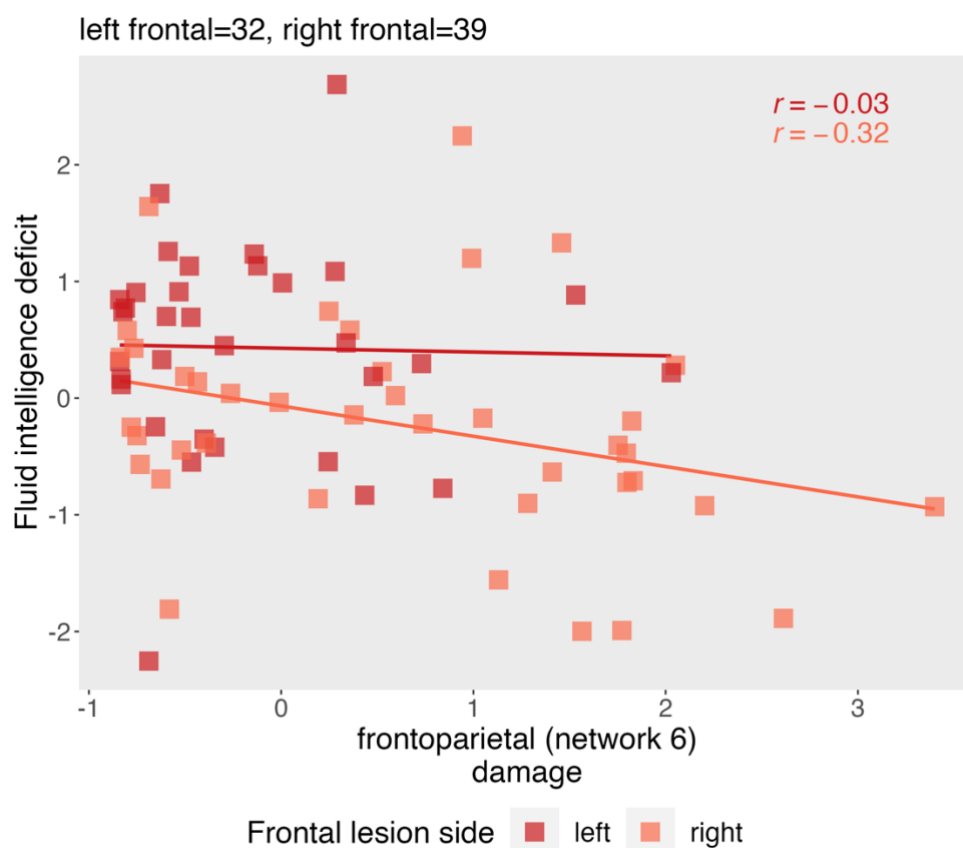


Figure 20. Individual regression lines show correlation between degree of frontoparietal network damage and fluid intelligence loss in left (darker red) and right (coral) frontal lesion patients. Network damage is defined according to a composite of GM damage and WM disconnection to this network and values are z-scored. Individual points represent individual frontal lesion patients, denoted in different colours according to whether their lesion was on the left hemisphere (dark red, $N=33$) or right hemisphere (coral, $N=40$). Pearson correlation coefficients are shown in upper right. The correlations shown here were not significant after correction for multiple comparisons (FDR, across two tests; for left frontal lesion patients $\text{adj-}p\text{-value} = 0.887$, for right frontal lesion patients $\text{adj-}p\text{-value} = 0.094$).

2.6 Discussion

In a large cohort of focal lesion patients, I used a data-driven approach to examine whether damage to specific RSNs predicted fluid intelligence loss. Damage to RSNs was defined as a composite of two approaches to estimating network damage: overlaying lesion tracings onto GM cortical RSN maps (Yeo et al., 2011), and onto a large sample of WM tractograms to estimate disconnection within each RSN. I observed that damage to only FPN predicted fluid intelligence loss. Although this relationship was observed over and above contributions from other networks, its effects could not be separated from TLV. While the two approaches I took to defining network damage (i.e., volume of GM damage and WM disconnection) were highly correlated, thus motivating the decision to use the composite measure of network damage, I also examined whether these methods differed in their sensitivity in predicting fluid intelligence loss. Results using these approaches trended in the same direction, although were significant only in the case of GM damage. When considering only GM network damage, damage to both FPN and DAN were predictive of fluid intelligence loss after accounting for contributions from other networks, but these effects did not survive once TLV was taken into account. In a final exploratory analysis to gain additional spatial resolution on the primary result, I showed tentative evidence that integrity of frontal, possibly right hemisphere, regions of FPN are critical for fluid intelligence. I will consider these results in more detail below.

2.6.1 Damage to FPN predicts fluid intelligence loss, with some suggestion of contributions from DAN, but cannot be separated from TLV

The primary analysis examined whether damage to specific RSNs, where network damage reflected both GM damage and WM disconnection, predicted fluid intelligence deficit. Only the extent of damage to FPN significantly predicted a deficit in fluid

intelligence over and above contributions from other networks. This result aligns well with findings from neuroimaging and lesion studies which suggest a critical role for frontoparietal regions in supporting fluid intelligence (Barbey et al., 2014; Duncan et al., 1995; Finn et al., 2015; Gläscher et al., 2009, 2010; Santarnecchi et al., 2017; Woolgar et al., 2010, 2018). However, when damage to FPN was considered in the context of TLV, neither predictor significantly predicted fluid intelligence deficit, suggesting that they explained similar variance. Indeed, FPN damage and TLV were highly correlated in this study (Pearson $r = 0.77$, see Figure 14). Accordingly, these results cannot claim a critical link between FPN damage and fluid intelligence deficit over and above TLV as these factors were tightly linked in this cohort.

The results also hinted at the possible importance of DAN to fluid intelligence. When only GM network damage was considered, damage to both FPN and DAN were predictive of fluid intelligence loss over and above contributions from other networks. The relationship between GM damage to DAN and fluid intelligence loss was eliminated once TLV was taken into account, however, and unlike for FPN, TLV was retained as a predictor over and above DAN. This indicated that the observed effects of DAN on fluid intelligence may be attributable to overall lesion volume rather than specific network damage. While previous work has linked DAN to fluid intelligence (e.g., Santarnecchi et al., 2017), this study provides only tentative evidence in support of these findings.

TLV was a significant predictor of fluid intelligence deficit in this study. This aligns with other work that has indicated lesion volume predicts cognitive outcomes (e.g., Puy et al., 2018; Reber et al., 2021; Thye & Mirman, 2018). A common approach when testing whether a region or network makes a specific contribution to a cognitive process is to treat TLV as a nuisance regressor (e.g., Cipolotti et al., 2023; Woolgar et al., 2010), an approach that was also taken in this study. The assumption underpinning this decision is that larger lesions likely impact more brain structures and accordingly cause more, or more substantial, cognitive deficits. Therefore, to establish the particular significance of a specific region or network, the relationship between damage to that area and cognitive effects should remain even after controlling for TLV. However, recent work has highlighted that, in the context of stroke at least, the relationship between lesion size,

location and cognitive deficit can vary across the brain (e.g., post-stroke frontal or anterior lesions are more likely to be larger than post-stroke lesions in posterior or subcortical areas), thus controlling for lesion size may not always be appropriate (Sperber, 2022). Indeed, as already discussed, TLV and extent of FPN damage were highly correlated in this work, making it difficult to draw conclusions about their separate contributions, if any. Future studies with a large sample of patients with more widely distributed lesions, and careful consideration of whether or how to account for lesion size, will help to address the question of whether FPN and/or TLV regardless of location are specifically critical for fluid intelligence.

Previous RSN work has suggested that interactions between FPN and DMN are associated with fluid intelligence (e.g., Finn et al., 2015; Sripada et al., 2019). While this study lacked sufficient coverage of some networks (notably visual network, also discussed below), coverage of DMN was adequate (see Figure 12) and thus the data permit a robust test of this hypothesis. In contrast to this previous RSN work, I did not find evidence to suggest DMN is causally related to fluid intelligence. Damage to DMN was not found to be predictive of fluid intelligence deficit in this study once contributions from other networks were taken into account.

2.6.2 Different measures of network damage explain similar variance

Focal lesions can extend beyond the cortical layer and affect underlying WM tracts, potentially leading to more widespread network disruption than indicated by GM damage alone. Yet few studies consider the contribution of WM to fluid intelligence. To address this, I defined network damage for each patient's lesions according to the volume of GM damaged within each network and the number of affected tracts within each network (i.e., WM network disconnection). However, network damage calculated by these two approaches was highly correlated (see Figure 16), such that the unique contributions of GM and WM network damage could not be considered within a single

model due to multicollinearity concerns. While separate models to predict fluid intelligence deficit based on GM network damage and WM network disconnection yielded different results (namely, only GM damage to FPN and DAN were predictive of fluid intelligence loss), the results for both models trended in the same direction (Table 3). At most, these results suggest that defining network damage according to WM disconnection using the approach taken here is less sensitive than a more traditional approach of examining extent of GM network damage.

That both approaches yielded such similar results was surprising. I had anticipated that estimating the extent of WM network disconnection would capture more widespread network damage and thus provide a more sensitive measure of network damage than considering extent of GM network damage alone. Moreover, previous work had highlighted the importance of WM integrity to fluid intelligence (e.g., Barbey et al., 2014; P.-Y. Chen et al., 2020; Gläscher et al., 2010). However, these findings do accord with a recent study that predicted cognitive changes two weeks post-stroke using lesion information (e.g., damage at the site of the lesion) and an estimate of structural disconnection caused by the lesion, finding that these two approaches explained comparable amounts of variance in cognitive performance (Salvalaggio et al., 2020). While I consider more chronic changes in this work, my results also suggest that the information provided by my approach to estimating structural disconnection may be similar to that provided by examining the impact of GM damage directly.

However, it is important consider some limitations to the approach I took to estimate WM network disconnection. First, as in previous work using the same pipeline (Mandonnet et al., 2020; Mrah et al., 2022), I only considered streamlines where both termini were located within the same network. While this approach captures widespread network dysfunction across various regions comprising a single network, it does not account for dysfunction caused by between network disconnection. Yet focal lesions can also disrupt interactions between networks through disconnection of the WM tracts that link these networks, potentially impairing the function of both networks involved. Examination of how inter-network disconnection relates to fluid intelligence

deficit is beyond the scope of the current work; however, this would be an interesting avenue for future work.

In addition, the current approach of overlaying lesions onto RSNs is unable to capture changes to functional connectivity and how this might affect fluid intelligence. Yet functional network reorganisation has been reported following focal brain damage (e.g., Gratton et al., 2012), and changes to functional connectivity have been shown to predict some concomitant cognitive changes better than lesion information in the sub-acute stage (Siegel et al., 2016). While some network recovery occurs from the sub-acute to chronic stage, individuals vary in the extent to which this recovers, with network recovery linked to cognitive recovery (Siegel et al., 2018). The approach I have taken does not consider how such changes to functional network dynamics might affect fluid intelligence. Accordingly, future work which considers different ways to measure functional connectivity changes, for example using resting-state or task-based fMRI, might provide a more nuanced account of how focal lesions affect functional brain networks and, in turn, how these changes impact fluid intelligence.

2.6.3 Damage to frontal, particularly right, FPN seems critical for fluid intelligence, but data are limited

FPN was identified as a key network supporting fluid intelligence. Since this network is widely distributed throughout the brain, I sought to garner further spatial specificity by asking whether damage to specific parts of FPN predicted fluid intelligence deficit. Comparison of patients with lesions in frontal, temporal, parietal and occipital lobe revealed a significant relationship between FPN damage and fluid intelligence deficit for frontal lesion patients only, although again this result could not to be separated from TLV. While this result points to the possible importance of frontal regions of FPN to supporting fluid intelligence, the small sample sizes for the remaining patient subgroups (N=7-28 for other lobes) preclude definitive conclusions regarding the absence of similar effects in these regions.

A recent study involving a large cohort ($N=227$) of focal lesion patients reported that fluid intelligence is underpinned by a set of predominantly right frontal regions (Cipolotti et al., 2023). To explore if a similar effect could be detected in this cohort, I examined whether the relationship between FPN damage and fluid intelligence deficit was observed for both left and right frontal lesion patients. The relationship was not statistically significant for either group following adjustment for multiple comparisons, and conclusions are further limited by small sample sizes ($N=32$ and 39 , for left and right frontal lesion groups, respectively). However, I observed a trend toward significance for right, but not left, frontal lesion patients. These findings hint that damage to right frontal regions of FPN might be important for fluid intelligence, in line with Cipolotti et al.'s (2023) proposal, but more data are needed to draw conclusions regarding the contributions of specific regions within FPN to fluid intelligence.

2.6.4 Strengths, caveats and future directions

To my knowledge, this study is the first large-scale investigation of the impact of network-level damage on fluid intelligence in focal lesion patients. This work extends previous focal lesion work in this area in two key aspects. First, while WM has previously been implicated in fluid intelligence, few focal lesion studies consider contributions from WM damage. This study examined WM damage in the context of network disconnection. Moreover, WM network disconnection was estimated by examining tract damage across a large sample of healthy tractograms ($N=1042$), far more than has been used in other studies that have similarly estimated structural disconnection (e.g., $N=20-176$, Mandonnet et al., 2020; Mrah et al., 2022; Salvalaggio et al., 2020). Thus, estimates of WM disconnection here are more robust than have been previously reported in other work. Interestingly, however, the addition of WM damage estimation to the analysis did not reveal additional or distinct networks compared to the more traditional GM approach.

Second, neuropsychological studies often involve modest sample sizes, yet as lesion and recovery characteristics can vary across patients, large samples are needed to ascertain whether a specific region is responsible for an observed deficit. This study recruited a large number of focal lesion patients and accordingly permitted a more data-driven approach to examining the relationship between network damage and fluid intelligence.

Nonetheless, several of my conclusions were limited by sample size and the distribution of lesions, particularly when patients were broken down into sub-groups, suggesting that even bigger samples will be needed in the future. In particular, the distribution of lesions across the various networks was not equal. In particular, I had poor coverage of the visual network and limited coverage of the somatomotor network (networks 1 and 2, see Figure 12 and Figure 14). Indeed, most lesions were located within frontal lobe. The tight coupling between lesion location and network damage is also reflected in the challenge of distinguishing contributions of FPN damage from that of TLV. While coverage of other networks was reasonable, it is important to note that where data are limited, minimal conclusions can be drawn from the observed lack of effects. Future work would benefit from recruiting patients with a broader sample of lesions, particularly outside of FPN.

As a final caveat, I note that I used the Yeo et al. (2011) seven-network parcellations in this work, yet many other RSN classifications exist. While there is some overlap in these delineations and in the labels used across different parcellations, there is also divergence. For instance, parts of the FPN of Yeo et al.'s (2011) parcellation fall within what is labelled the dorsal and ventral attention network in another widely used parcellation (Shirer et al., 2012). Moreover, while FPN is considered to show broad anatomical overlap with the MD system (e.g., as discussed in Assem, Glasser, et al., 2020), it has not formally been compared to the Yeo et al. (2011) FPN parcellation. Accordingly, the results in this study can only speak to the role of FPN more broadly, rather than the MD system specifically, in supporting fluid intelligence. If there is divergence between these networks, then the RSN approach may not be maximally sensitive to the contribution of MD damage. Future research could build on these findings by exploring various RSN parcellations, incorporating functional activation-

based network definitions, and/or including precise anatomical comparison between RSN and activation-based definitions.

2.6.5 Conclusions

The study reviewed 174 patients with chronic focal brain lesions to identify specific networks that might be important for fluid intelligence. While the results provide some suggestion of a tight link between fluid intelligence and FPN, particularly frontal FPN, damage to FPN closely reflected TLV. Network damage was quantified by examining GM damage and by estimating WM network disconnection. However, both approaches yielded similar results, indicating that incorporating WM disconnection, as estimated here, did not add additional information over and above the traditional method of quantifying only GM network damage. Overall, this study suggests that of all RSNs, FPN, particularly its frontal aspects, critically underpins fluid intelligence. However, future work with patients exhibiting more widely distributed lesions is required to dissociate the contributions of FPN and TLV and provide more conclusive evidence on the neural basis of fluid intelligence.

2.7 Afterword

This study investigated the relationship between damage to specific brain networks and fluid intelligence loss in a large cohort of focal lesion patients. The strengths of this study have been highlighted, but to restate: the study involved a large cohort of focal lesion patients ($N=174$), examined contributions from several RSNs using a reliable parcellation (Yeo et al., 2011), and examined and contrasted contributions from GM network damage and WM network disconnection to predicting fluid intelligence deficit. However, the conclusions from this work were tentative. Damage to (possibly right) frontal regions of FPN might be critical to fluid intelligence, but effects could not be separated from TLV as neither predictor was significant after partialling out the other. Similarly, DAN may contribute to fluid intelligence, but in this case its effects were likely attributable to overall lesion volume rather than specific network damage since TLV remained a significant predictor even after partialling out damage to DAN. Finally, high similarity in the estimates of GM network damage and WM network disconnection, at least as calculated in this study, suggested these measures captured similar variance.

The tentative conclusions drawn from this work, despite the large sample size, highlight some of the difficulties of conducting neuropsychological studies. Retrospective examinations of patient data are not uncommon for large lesion studies. Large patient studies often acquire patient data over years as patients are recruited after they have sustained their injury, and many aspects of the process (e.g., manual lesion tracings, acquiring scans, neuropsychological testing) are labour intensive and can involve a large team of clinicians and researchers. In the context of this study, retrospective examination meant that I had little control over the variability of lesion location in my cohort. As stated, the majority of patients in this study had frontal lesions. In an ideal situation, as I have mentioned as an avenue for future work, patients would be recruited with more widely distributed lesions thus permitting a more definitive test of whether FPN subtends fluid intelligence.

Moreover, I had access to a limited amount of neuropsychological data. Accordingly, while I could test whether damage to FPN relates to fluid intelligence deficit, I could not rule out that damage to FPN (or damage per se, as reflected in TLV) affects a more general process that in turn affects performance on fluid intelligence tasks. That is, even with a stronger claim that FPN damage is linked to fluid intelligence deficit, I would struggle to make a strong conclusion about the specificity of the cognitive deficit. On the other hand, it is important to note that even with a large neuropsychological battery, a clear claim that FPN is linked to fluid intelligence deficit specifically might be difficult, as selective cognitive deficits following focal damage are rare. Thus, this is likely a limitation of many neuropsychological investigations and underscores the importance of convergent evidence from various methodologies to substantiate strong claims about neural mechanisms.

2.8 Supplementary Materials for Chapter 2

Table S1. Details of patients included in study

Case Number	Gender	Age	In Woolgar PNAS 2010	Lesion side	Broad lesion area
1	F	25	0	left	frontal
2	F	28	0	left	frontal
3	M	29	1	left	frontal
4	M	34	1	left	frontal
5	F	35	1	left	frontal
6	F	42	1	left	frontal
7	F	42	0	left	frontal
8	F	43	1	left	frontal
9	M	43	1	left	frontal
10	M	46	1	left	frontal
11	M	50	0	left	frontal
12	F	51	0	left	frontal
13	F	51	0	left	frontal
14	F	53	0	left	frontal
15	F	54	1	left	frontal
16	M	56	1	left	frontal
17	M	57	0	left	frontal
18	F	58	1	left	frontal
19	M	58	0	left	frontal
20	M	59	0	left	frontal
21	F	61	1	left	frontal
22	M	61	0	left	frontal
23	F	64	1	left	frontal
24	F	64	0	left	frontal
25	M	65	0	left	frontal
26	M	65	0	left	frontal

Frontoparietal cortex in fluid intelligence

27	M	65	1	left	frontal
28	F	67	1	left	frontal
29	F	67	0	left	frontal
30	F	68	0	left	frontal
31	F	70	0	left	frontal
32	M	70	0	left	frontal
33	M	70	0	left	frontal
34	M	23	0	right	frontal
35	F	27	1	right	frontal
36	F	29	1	right	frontal
37	F	35	0	right	frontal
38	F	39	1	right	frontal
39	F	40	0	right	frontal
40	M	41	1	right	frontal
41	M	42	1	right	frontal
42	M	44	0	right	frontal
43	M	44	0	right	frontal
44	M	45	0	right	frontal
45	F	46	1	right	frontal
46	F	46	0	right	frontal
47	M	46	1	right	frontal
48	F	47	1	right	frontal
49	M	48	0	right	frontal
50	F	50	1	right	frontal
51	M	50	1	right	frontal
52	F	52	0	right	frontal
53	M	52	0	right	frontal
54	F	53	1	right	frontal
55	M	53	1	right	frontal
56	M	55	1	right	frontal

Chapter 2

57	F	56	0	right	frontal
58	M	58	0	right	frontal
59	M	60	1	right	frontal
60	F	61	1	right	frontal
61	M	61	0	right	frontal
62	M	62	1	right	frontal
63	M	64	1	right	frontal
64	M	65	1	right	frontal
65	M	65	0	right	frontal
66	M	67	1	right	frontal
67	F	68	1	right	frontal
68	F	68	0	right	frontal
69	M	68	1	right	frontal
70	M	68	0	right	frontal
71	M	68	1	right	frontal
72	F	69	0	right	frontal
73	F	70	0	right	frontal
74	M	27	1	bilateral	frontal
75	F	38	1	bilateral	frontal
76	F	41	0	bilateral	frontal
77	F	41	1	bilateral	frontal
78	M	41	1	bilateral	frontal
79	M	42	0	bilateral	frontal
80	M	46	1	bilateral	frontal
81	F	54	0	bilateral	frontal
82	F	58	0	bilateral	frontal
83	F	59	1	bilateral	frontal
84	F	59	1	bilateral	frontal
85	M	61	0	bilateral	frontal
86	F	65	1	bilateral	frontal

Frontoparietal cortex in fluid intelligence

87	M	66	0	bilateral	frontal
88	M	76	0	bilateral	frontal
89	M	64	1	left	occipital
90	M	65	0	left	occipital
91	M	67	1	left	occipital
92	M	68	1	left	occipital
93	M	54	1	right	occipital
94	M	63	1	right	occipital
95	M	65	1	right	occipital
96	F	36	0	left	parietal
97	F	44	1	left	parietal
98	F	47	0	left	parietal
99	F	48	1	left	parietal
100	M	49	1	left	parietal
101	F	52	1	left	parietal
102	F	72	0	left	parietal
103	F	25	0	right	parietal
104	F	44	1	right	parietal
105	F	48	1	right	parietal
106	F	53	0	right	parietal
107	M	71	0	right	parietal
108	M	21	1	left	temporal
109	F	25	0	left	temporal
110	M	28	0	left	temporal
111	M	29	1	left	temporal
112	F	36	1	left	temporal
113	F	37	1	left	temporal
114	M	41	1	left	temporal
115	M	42	0	left	temporal
116	F	49	1	left	temporal

I17	F	50	0	left	temporal
I18	F	52	0	left	temporal
I19	M	53	1	left	temporal
I20	M	55	1	left	temporal
I21	F	57	1	left	temporal
I22	M	58	1	left	temporal
I23	M	60	0	left	temporal
I24	F	61	0	left	temporal
I25	F	65	0	left	temporal
I26	F	65	1	left	temporal
I27	M	67	0	left	temporal
I28	F	20	1	right	temporal
I29	F	22	0	right	temporal
I30	M	28	0	right	temporal
I31	F	29	0	right	temporal
I32	M	33	0	right	temporal
I33	M	35	1	right	temporal
I34	M	36	1	right	temporal
I35	M	44	1	right	temporal
I36	M	24	0	left	subcortical
I37	M	42	0	left	subcortical
I38	M	44	0	left	subcortical
I39	F	63	0	left	subcortical
I40	M	69	0	left	subcortical
I41	F	45	0	right	subcortical
I42	M	39	0	left	mixed*
I43	M	49	1	left	mixed
I44	F	51	0	left	mixed
I45	F	56	0	left	mixed
I46	M	68	0	left	mixed

Frontoparietal cortex in fluid intelligence

147	M	68	1	left	mixed
148	M	70	1	left	mixed
149	F	35	0	right	mixed
150	F	43	0	right	mixed
151	F	43	0	right	mixed
152	M	44	0	right	mixed
153	M	45	0	right	mixed
154	F	46	0	right	mixed
155	M	48	1	right	mixed
156	F	51	0	right	mixed
157	F	52	0	right	mixed
158	M	53	0	right	mixed
159	M	54	0	right	mixed
160	M	55	0	right	mixed
161	F	58	0	right	mixed
162	M	58	0	right	mixed
163	F	60	0	right	mixed
164	M	60	1	right	mixed
165	F	63	0	right	mixed
166	F	65	1	right	mixed
167	F	65	0	right	mixed
168	M	65	0	right	mixed
169	F	67	0	right	mixed
170	M	68	1	right	mixed
171	M	69	0	right	mixed
172	M	69	1	right	mixed
173	M	75	0	right	mixed
174	M	49	1	bilateral	mixed

*Mixed group refers to patients with lesions affecting more than one region.

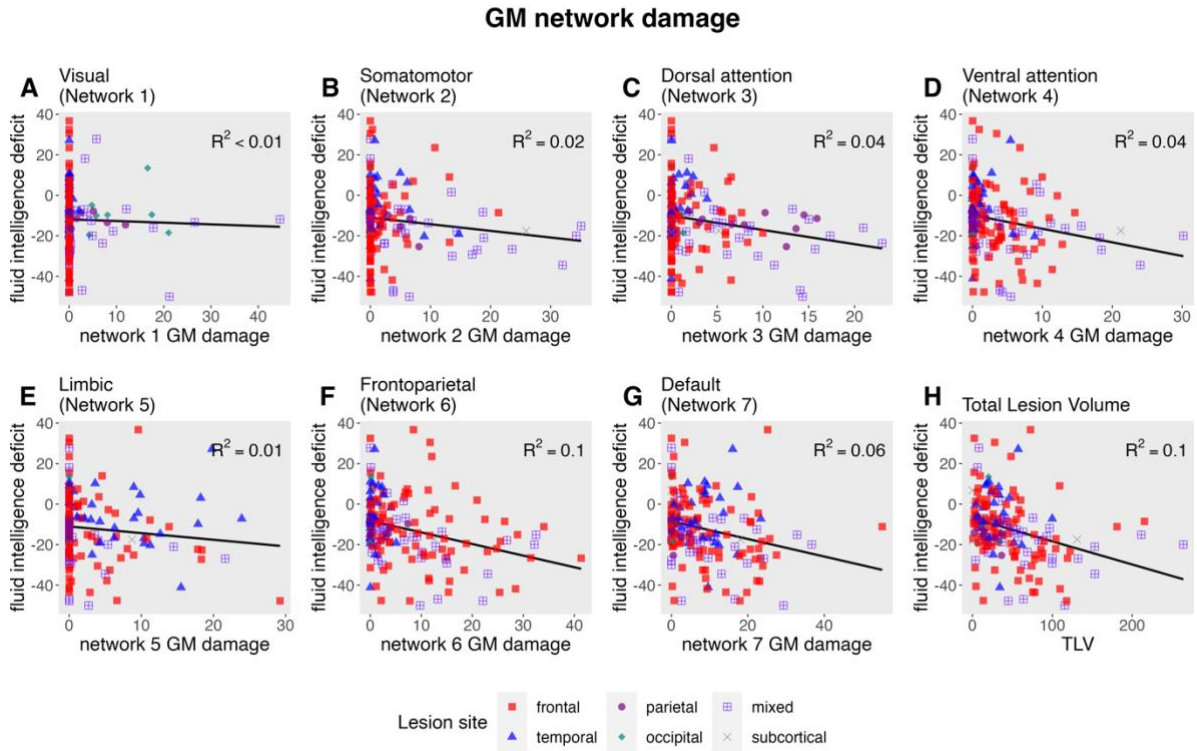


Figure S1. Individual regression lines show the correlation between fluid intelligence deficit and GM damage to each individual network and TLV (shown as volume of damage, cm^3). Pearson correlation coefficients for the relationship are shown in the upper right of each plot. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

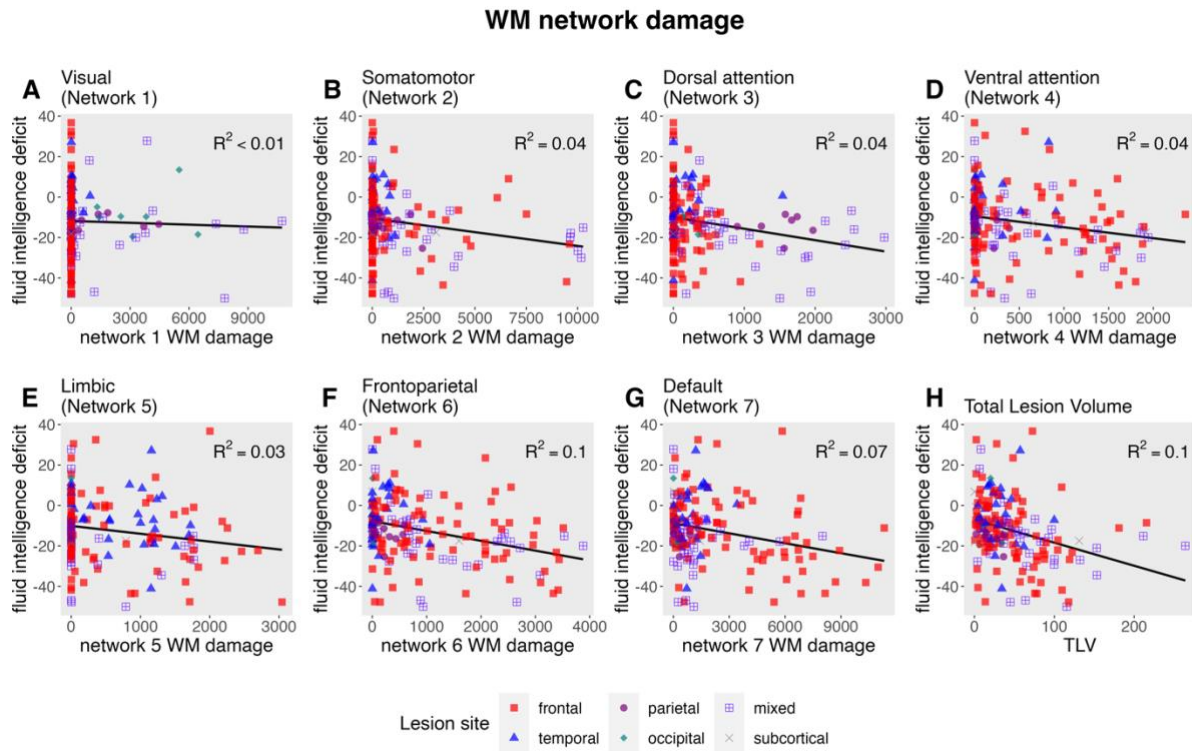


Figure S 2. Individual regression lines show the correlation between fluid intelligence deficit and WM disconnection to each individual network (shown as average number of tracts damaged, cm^3) and TLV (shown as volume of damage, cm^3). Pearson correlation coefficients for the relationship are shown in the upper right of each plot. Individual points represent individual patients, with colours and shapes reflecting the primary site of the patient's lesion: frontal (red square), temporal (blue triangle), parietal (purple circle), occipital (teal diamond), mixed (purple open square), subcortical (grey cross). Note, 'mixed' here refers to patients with lesions affecting more than one region.

Chapter 3 |

Spatial and feature-selective attention interact to drive selective coding in frontoparietal cortex

3.1 Foreword

The study in this chapter looks at the role of multiple-demand (MD) regions in selective attention. Specifically, I consider how the interaction between two types of selective attention – spatial and feature attention – affect task-related representations in the MD system in healthy adults using functional magnetic resonance imaging (fMRI) with multivariate pattern analysis methods. As this paradigm had previously been used in an magnetoencephalography (MEG) context by my group (Goddard et al., 2022), I also relate my fMRI data to this existing MEG data to explore whether MD regions might be

Chapter 3

driving focus on behaviourally relevant information across the brain. I collected the data for this study at the Max Planck Institute for Human Cognitive and Brain Sciences (MPI CBS) in Leipzig, Germany with my advisor, Romy Lorenz.

This work has been prepared for peer-review submission and will soon be available in as a pre-print with the following title and author information:

Dermody, N., Lorenz, R., Goddard, E., Villringer, A., Woolgar, A. Spatial and feature-selective attention interact to drive selective coding in frontoparietal cortex

3.2 Abstract

Attention enables the selective processing of relevant information. Two types of selective attention, spatial and feature attention, have separable neural effects but in real life are often used together. Here, we asked how these types of attention interact to affect information coding in a frontoparietal ‘multiple-demand’ (MD) network, essential for attentional control. Using functional magnetic resonance imaging (fMRI) with multivariate pattern analysis, we examined how covert attention to object features (colour or shape) and spatial locations (left or right) influences coding of task-related stimulus information. We found significant above-chance decoding of only the attended feature of the attended object, reflecting spatial and feature attention interacted multiplicatively on information coding in MD and visual regions. We also uncovered a multidimensional neural representation of task-relevant information, with stimulus information (e.g., colour) and discrimination difficulty (distance from the decision boundary) coded in separate dimensions. Comparing fMRI with extant magnetoencephalography (MEG) data revealed attention affected colour coding in MD regions before visual areas; this temporal distinction was less clear for shape coding. Rather than boosting processing of whole objects or relevant features across space, our results suggest neural activity reflects precise tuning to relevant information, indicating a highly selective control process that codes behaviourally relevant information across multiple dimensions.

3.3 Introduction

Everyday experience tells us that there are limits to the amount of information we can process at any one time. To find a friend in a crowded bar, we can ignore aspects of the environment irrelevant to our goal and zero in on what is relevant, focussing on likely locations and paying attention to particular features (e.g., hair colour). Selective attention describes the process by which a subset of task-relevant inputs is ‘selected’ for prioritised processing from the wealth of information available to us. Behaviourally, attention enhances perception of the attentional target relative to unattended stimuli (e.g., see Carrasco, 2011). At the neural level, attention enhances the signal of the attended information (e.g., Duncan, 2001; Noudoost et al., 2010; Reynolds & Heeger, 2009; Thiele & Bellgrove, 2018). In non-human primates, attending to a particular stimulus can increase the response (e.g., firing rate) of neurons responding to that stimulus (e.g., Burrows & Moore, 2009; Luck et al., 1997; Luo & Maunsell, 2018; McAdams & Maunsell, 1999; Thiele et al., 2016), reduce response variability of individual neurons and shared variability across neurons (M. R. Cohen & Maunsell, 2009; Kanashiro et al., 2017; Mitchell et al., 2007), and is associated with a reduction in the response of cells tuned to an unattended stimulus (Moran & Desimone, 1985; see Thiele & Bellgrove, 2018 for a review). In humans, attention to a particular spatial location, object or feature is reflected in increased activation in related cortical (e.g., Heinze et al., 1994; O’Craven et al., 1999; Sàenz et al., 2002) and subcortical areas (e.g., O’Connor et al., 2002), and reduced neural variability (Arazi et al., 2019). In addition, multivariate decoding of human neuroimaging data repeatedly shows stronger representation of attended information relative to unattended (e.g., Jackson et al., 2017; Jackson & Woolgar, 2018; Jehee et al., 2011; Keller et al., 2022; Y. Li et al., 2018; Woolgar, Williams, et al., 2015) that is sustained over time (Barnes et al., 2022; Battistoni et al., 2020; Goddard et al., 2022; Grootswagers et al., 2021; Moerel et al., 2022).

The process of prioritising relevant information is thought to depend on frontoparietal cortex (e.g., E. K. Miller & Cohen, 2001). In particular, in humans, a network of frontoparietal ‘multiple-demand’ (MD) regions (Duncan, 2010; Duncan & Owen, 2000)

is heavily implicated in driving preferential focus on task-relevant information. This collection of strongly interconnected frontoparietal regions is active during a wide range of cognitively demanding tasks (Assem, Blank, et al., 2020; Fedorenko et al., 2013) and has also been referred to as the task positive network (Fox et al., 2005), frontoparietal control network (Vincent et al., 2008), task-activation ensemble (Seeley et al., 2007), or cognitive control network/networks (Cole & Schneider, 2007; Gratton et al., 2018) comprising ‘flexible hubs’ that rapidly update their brain-wide functional connectivity patterns in accordance with task demands (Cole et al., 2013). MD regions are thought to prioritise task-relevant information through a process of adaptive coding (Duncan, 2001). Under this framework, frontoparietal neurons flexibly adapt to code behaviourally relevant information, which in turn act as a source of bias, driving widespread congruence in representations across the brain (Desimone & Duncan, 1995; E. K. Miller & Cohen, 2001).

In line with this, MD regions demonstrate flexible coding properties, including the capacity to represent a wide range of different types of task information (Woolgar et al., 2016) and an adaptive response to changing task demands (e.g., Woolgar, Afshar, et al., 2015; Woolgar et al., 2011; Woolgar, Williams, et al., 2015). Moreover, MD regions have been shown to preferentially represent attended-to or task-relevant information, with previous work in the visual domain, for instance, demonstrating preferential representation of the attended object or object feature (e.g., Jackson et al., 2017; Jackson & Woolgar, 2018; Woolgar, Williams, et al., 2015). Moreover, using brain stimulation to intervene on neural activity, one MD node (right dorsolateral prefrontal cortex) was shown to have a causal role in upregulating coding of task-relevant object features elsewhere in the network (Jackson et al., 2021).

In these studies, attention is typically directed to a particular spatial location, target object or target feature, reflecting some of the ways in which visual selective attention can be deployed. These subtypes of selective attention are thought to differentially affect the scope of information prioritised for further processing. Briefly, spatial attention, whereby attention is directed to a particular location, is associated with selection of all information at the attended location for further processing (e.g., Kastner

et al., 1999). Object-based attention theories posit that attention directed to a particular object leads to enhancement of all information within the bounds of an attended object (Duncan, 1984; O’Craven et al., 1999), including multiple features of that object (Duncan, 1984; Egly et al., 1994). In contrast, feature-based attention, whereby attention is directed to a particular feature (e.g., the colour red), is associated with enhanced processing of that feature across the visual field (see Liu, 2019; Maunsell & Treue, 2006 for reviews of related literature), with associated improvements in behavioural performance (e.g., greater detection, discrimination of the attended feature) and more interference from distractors possessing a similar feature, regardless of location (e.g., Rossi & Paradiso, 1995; Sàenz et al., 2003; W. Zhang & Luck, 2009). Attention can alternatively be directed to an entire feature dimension (e.g., to make fine discriminations of colour while ignoring shape), sometimes known as feature-selective attention, resulting in enhanced separation of task-relevant categories along the attended dimension (X. Chen et al., 2012; Jackson et al., 2017; see also Wisniewski et al., 2023).

Yet in real life, we rarely deploy attention so singularly. Instead, the target of our attention is usually informed by a combination of spatial and feature cues against a backdrop of many other possible competing inputs. Searching for a particular brand of rolled oats in our local supermarket, for example, typically requires attentional direction to both the general area in which they are typically located (e.g., the bottom shelf of the cereal aisle), to the most prominent feature of the item (e.g., shape: rectangular box, or colour: black), while ignoring information irrelevant to this task. Despite a wealth of behavioural data on the interaction between different attentional subtypes, there are few neuroimaging studies examining the specificity of task-relevant representations when multiple subtypes of attention are combined.

Guided by seminal theories, an additive effect might be predicted when both spatial and feature are deployed, resulting in enhancement of both irrelevant features of attended objects (due to object-based attentional spread) and relevant features of unattended objects (due to field-wide feature attention effects), relative to irrelevant features of irrelevant objects. Instead, we recently reported a striking multiplicative interaction of

spatial and feature-selective attention (Goddard et al., 2022). Using multivariate decoding of magnetoencephalography (MEG) signals to track the dynamics of information coding in the human brain, we found that information across the visual field was initially decodable in occipital cortex irrespective of attentional condition. However, frontal cortex maintained a selective representation of only task-relevant information that was selected by both types of attention (i.e., the task-relevant feature at the selected location). This highly tuned response in turn appeared to Granger-cause a sustained selective response in occipital cortex. In occipital cortex, consequently, only coding of the task-relevant feature at the selected location was sustained through time, while representation of all other aspects of the visual display returned to baseline.

In this study, we sought to clarify the spatial basis for this effect using functional magnetic resonance imaging (fMRI). Given the proposed role of the MD system in attentional control, we asked whether the same selective focus on information at the intersection of spatial and feature attention would be seen in the MD system, or whether this system would demonstrate a pattern of information coding more in line with seminal theories of object-based and feature-based attention. Participants completed the same covert attention task as in our previous MEG study (Goddard et al., 2022). This task manipulated both spatial and feature attention, and we used multivariate pattern analysis (MVPA) to decode visual stimulus information, in MD regions and elsewhere in the brain, under each attentional condition. The primary question we asked was: do spatial and feature attention interact in their modulation of stimulus information coding in MD regions? In addition, as our paradigm comprised stimuli which varied incrementally across feature space, we sought to characterise how the attended information was organised with respect to feature and task dimensions. We then compared this result to the visual cortex and whole brain using a roaming searchlight. Finally, we directly compared the fMRI data to extant MEG using MEG-fMRI fusion (e.g., Cichy et al., 2014, 2019; Cichy & Oliva, 2020) to characterise the timecourse of these spatially-resolved effects.

3.4 Materials and Methods

3.4.1 Participants

Thirty participants (16 female, 14 male, mean age = 27.20, SD = 4.16) took part in the fMRI experiment. All participants were right-handed native German speakers, had normal or corrected to normal vision, including self-reported normal colour vision. The study was approved by the local ethics committee at the Medical Faculty of the University of Leipzig. All participants gave written informed consent prior to the study and were reimbursed for their time.

3.4.2 Task Design

3.4.2.1 Stimuli

We used the visual attention paradigm developed by Goddard et al. (2022) with task instructions translated into German. While our design sought to closely replicate that of Goddard et al. (2022), we made some slight modifications and thus describe the stimuli and paradigm here in full. The main differences from Goddard et al. (2022) were that we provided feedback at the end of each block instead of trial-wise (except as noted), we used longer stimulus presentations (150ms instead of 92ms), and that we encountered some technical difficulties with colour calibration meaning that our colours were not identical to the originals, as described in the following.

Stimuli comprised a set of 16 novel ‘spiky’ objects which varied along two dimensions (colour and shape, see Figure 21, Panel A). The bounding box for individual objects was 621 pixels (10 degrees of visual angle (dva)) wide by 621 pixels (10 dva) high. Their total size varied with their spikes, but the spikes never reached the border of the object image.

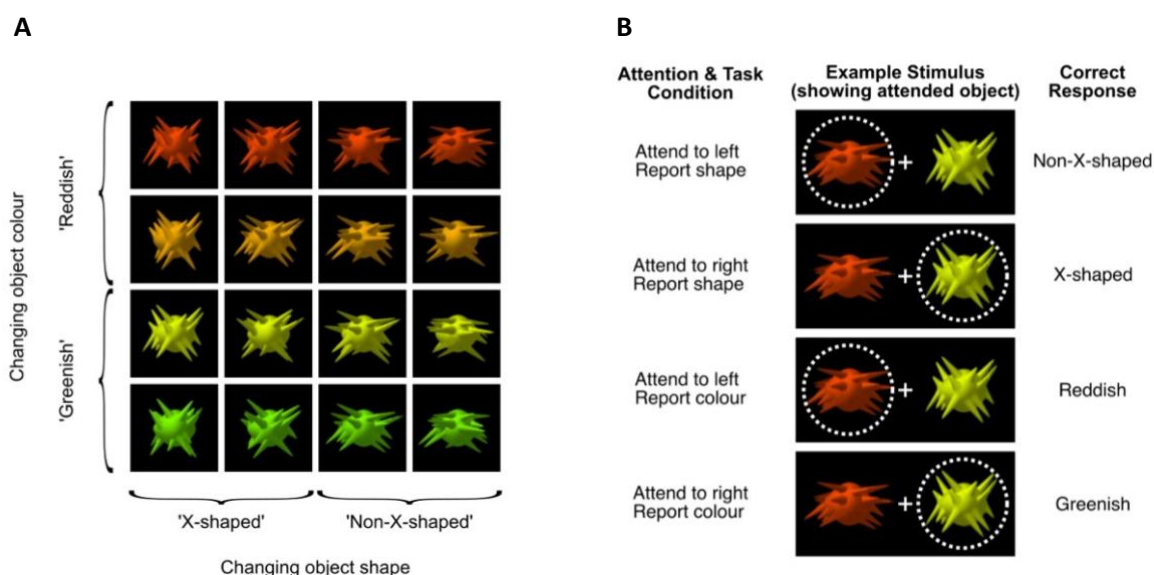


Figure 21, Panel A: Visual stimuli used in the experiment. Each object varied systematically across two dimensions: colour and shape. In the colour task, participants categorised the attended object as either 'reddish' or 'greenish'. In the shape task, participants categorised the attended object as either 'X-shaped' or 'non-X-shaped', based on the orientations of the object's spikes. Panel B: Visual stimuli as presented in the scanner. Participants were asked to covertly attend to one side of space, and report (via button press) either the shape or colour of the object in the attended space. The dotted circle shows the locus of spatial attention and was not shown to participants.

The spike orientation statistics were varied to create four shape classes: 'strongly X-shaped', 'weakly X-shaped', 'weakly non-X-shaped', 'strongly non-X-shaped'. To discourage participants using a single spike to judge shape, each class comprised 100 slight variations. For the shape-based task, participants classified the target object as 'X-shaped' or 'non-X-shaped' (Figure 21, Panel A).

We also varied colour to create four classes: 'strongly red', 'weakly red', 'weakly green', 'strongly green'. For the colour-based task, participants classified the target object as either 'reddish' or 'greenish' (Figure 21, Panel A). The four feature values along each dimension meant that for both tasks, stimuli were either far from the decision boundary (e.g., strongly red: 'easy' trials), or close to the decision boundary (e.g., weakly red: 'hard' trials). Our intention was that the four colours would be equally spaced along the equiluminant plane joining the strongly red and strongly green coordinates in $u'v'$ space. However, due to technical difficulties with the screen calibration on the projector

system, the colours were not fully calibrated for some participants. For the first set of participants ($N=17$), the chromacity coordinates for the colours were as follows: strongly red: 110.34, 41.38, weakly red: 62.07, 48.28, weakly green: 33.10, 52.41, strongly green: -15.17, 58.31. For the remaining participants ($N=13$), luminance was set to 48cd m^2 for all four colours, with chromaticity coordinates for this group as follows: strongly red: 33.03, 28.48, weakly red: 18.21, 30.90, weakly green: 9.31, 32.34, strongly green: -5.52, 34.76. The slight changes in colour between these two sub-groups of participants was included as a predictor in all analyses but did not have a meaningful impact.

3.4.2.2 Task

On each trial, two stimuli were presented simultaneously against a black screen, one left and one right of a grey fixation cross, with height and width of $1dva$ (Figure 21, Panel B). Participants were instructed at the beginning of every run to covertly attend to one side of space and report either the colour ('reddish' or 'greenish') or the shape ('X-shaped' or 'non-X-shaped') of the attended object. Participants responded by pressing one of two response keys on a button box using the index and middle fingers of their right hand.

3.4.2.3 Training

To ensure familiarity with the stimulus categories (i.e., 'reddish' and 'greenish' for colour, and 'X-shaped' and 'non-X-shaped' for shape) and task instructions, participants completed a training session on a computer located outside the fMRI scanner. Each participant completed a minimum of two training blocks, one for the colour task and one for the shape task, each comprising 30 trials. If participants scored below 70% for either the colour or shape task block, they completed additional blocks of that task type

until their performance reached 70% or above. Training took approximately 20 minutes.

3.4.3 fMRI Acquisition

Functional MRI data were acquired using a Siemens 3.0-Tesla Skyra scanner (Siemens, Erlangen) with 32-channel head coil at the Max Planck Institute for Human Cognitive and Brain Sciences (MPI CBS) in Leipzig, Germany. We used a high resolution interleaved T2*-weighted echo planar imaging (EPI) acquisition sequence covering the whole brain. The following parameters were used: repetition time (TR) 1000ms; echo time (TE) 30ms; multi-band acceleration factor of 3; 45 interleaved slices of 3.0mm slice thickness with a 10% interslice gap; flip angle of 61 degrees; field of view (FOV) 204mm. AutoAlign was used to correct for movement between acquisition runs. We also acquired T1-weighted MPRAGE structural images for all participants (resolution 1.0 x 1.0 x 1.0mm).

3.4.4 Procedure

Stimuli were presented using MATLAB with Psychophysics Toolbox-3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) and were back-projected onto a screen viewed through a head-coil mounted mirror. Participants, lying supine, viewed the screen from a distance of 95cm. Covert attention was ensured using an EyeLink 1000 Plus remote tracking system; participants were required to fixate on the fixation cross for a variable duration before the task would proceed (details as follows). The eye tracker tracked the participant's right eye.

Participants performed eight runs in total, where each run comprised a unique combination of the four possible spatial and feature attention manipulations (i.e.,

attend to one of the two locations: left or right, and one of the two features: colour or shape) and the two response mappings (e.g., button one for 'reddish', button two for 'greenish'). Run order was counterbalanced across participants and arranged such that the attended location and attended feature alternated between runs. Starting location and attended feature were also counterbalanced across participants. Each participant's imaging data were collected in a single session of approximately 120 minutes.

At the start of each run, participants first completed an eye tracker calibration task. Task instructions and response mappings for the run were then displayed along with a graphical reminder of the attended feature categories.

Each run comprised 256 trials, representing all possible pairs of the 16 objects. Within each run, objects were presented in a pseudo-random order so that objects of each shape and colour were equally likely to precede objects of all shapes and colours. Each trial began with a grey fixation cross (present throughout the entire run), and the trial proceeded once fixation was verified with the eye tracker. Participants were required to fixate on the cross for at least 300ms before the stimulus would appear. Stimuli were then presented for 150ms. Finally, participants were shown the grey fixation cross on a blank screen for a variable duration. This period comprised the participant response time (maximum of 2500ms, after which time the trial proceeded; percentage of missed trials across entire study = 0.69%) and a variable inter-trial interval (500-1000ms) during which fixation was verified for the subsequent trial. A variable inter-trial interval was used to mitigate expectancy effects.

Participants only received feedback (average accuracy and reaction time (RT)) at the end of each run, with the exception of five participants who additionally received feedback after every trial. For these participants, after making a response, the fixation cross changed colour for 200ms to indicate whether their response was correct (white) or incorrect (blue). Variation in feedback scheme was included as a factor in all our analyses but failed to explain any variance in the behavioural or neural data.

3.4.6 Behavioural data analysis

All analyses were conducted using R (v4.2.2) in RStudio (v2023.3.0). Statistical analyses to determine whether participant behaviour (accuracy and RT) varied across task (attend to colour or shape) and feature steps (e.g., difference in accuracy between classifying ‘weakly red’ versus ‘strongly red’ object as ‘red’), were performed by fitting linear mixed-effects models (Baayen, 2008) using the *lme4* package (v1.1-32, Bates et al., 2015) and then reducing these models to identify key effects using the *lmerTest* package (v3.1-3, Kuznetsova et al., 2017) in R.

Our full models were fitted by maximum likelihood for accuracy (*Proportion Correct*) and mean reaction time (*Average RT*). Predictors for both models comprised task (*Task*: attend to colour or shape), feature steps (*Feature Step*: e.g., for colour task, colours one to four), and their interaction (Table 4). *Subject* was defined as a random effect in both models. We only considered task-relevant information (i.e., the colour or shape of the target, in attend colour and attend shape trials, respectively) and the analysis was collapsed across attended location (left or right of fixation). For RT, only correct and feasible (i.e., within 0.15-2.5secs) trials were analysed. In addition, both models included other factors that were not of scientific interest but might capture variance (i.e., stimulus colour set and feedback scheme).

We used model selection to determine which fixed and random effect factors best estimated our accuracy or RT data, using the *step* function of the *lmerTest* package (Kuznetsova et al., 2017). This function performs a backward elimination of all effects included in our full model, starting with random effects followed by fixed effects, commencing with higher order interactions first. Our final model was that identified through this process and comprised only non-eliminated effects, namely those considered to be significant predictors of our outcome variable. Full and final models for Accuracy and RT along with model fit parameters are shown in Table 4.

These final models were re-fitted using a linear model for accuracy (as random effects were eliminated through the step procedure described previously) and with restricted

maximum likelihood for RT, to allow for significance testing of fixed effects (Meteyard & Davies, 2020). Further post-hoc tests, where necessary, were conducted using the emmeans package (v1.8.5), using the Satterthwaite method to calculate degrees of freedom where random effects were specified. P-values were adjusted for multiple comparisons using the Bonferroni method.

Extraneous variables (i.e., stimulus colour set and feedback scheme) in both accuracy and RT models were eliminated through our step model selection procedure, indicating that the inclusion of these factors did not explain significant additional variance or improve model fit (Table 4).

Table 4. Full and final linear mixed effects models for behavioural analysis

Accuracy								
Sampling units:	Total number of observations = 240 N participants = 30							
Model name	Fixed effects	Random effects	AIC	Model fit			Model test	
				BIC	LL	df	X ²	df
Accuracy full model	propCorrect ~ Task * FeatureStep + Calibration_Type + Feedback_Given	(1 Sub)	-350.98	-312.7	186.49	11	1.67	2
Accuracy final model ¹	propCorrect ~ Task * FeatureStep	-	-353.32	-321.99	185.66	9		

Reaction Time								
Sampling units:	Total number of observations = 240 N participants = 30							
Model name	Fixed effects	Random effects	AIC	Model fit			Model test	
				BIC	LL	df	X ²	df
Reaction time full model	meanRT ~ Task * FeatureStep + Calibration_Type + Feedback_Given	(1 Sub)	-676.6	-638.32	349.3	11	0.005	1
Reaction time final model	meanRT ~ Task * FeatureStep	(1 Sub)	-678.6	-643.79	349.3	10		

For above, propCorrect= accuracy, reflected as proportion correct; Task = task condition (e.g., attend colour or attend shape); FeatureStep = the feature step (i.e., for colour and shape, respectively: one = strongly red or strongly non-X-shaped, two = weakly red or weakly non-X-shaped, three = weakly green or weakly X-shaped, four = strongly green or strongly X-shaped); meanRT = mean reaction time; Sub = subject identifier. The remaining two variables – Calibration_Type (type of colour calibration used or stimulus colour set), Feedback_Given (reflecting feedback scheme, i.e., whether trial-by-trial feedback was given or not) – were included to ascertain the effect of acquisition variability on our results. In both final models Calibration Type and Feedback given were eliminated, indicating they did not significantly improve model fit.

AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, df = degrees of freedom, X² = Chi-square. Note: ‘*’ indicates an interaction with all lower order effects included in model. Terms in brackets represent random effect terms. Final models were chosen using the ‘step’ function from lmerTest. This performs a backward elimination of random-effect terms followed by backward elimination of fixed-effect terms in linear mixed models. ¹No random effects specified in this final model, so Accuracy final model re-fitted as linear model.

3.4.7 fMRI data analysis

3.4.7.1 fMRI preprocessing

MRI data were preprocessed using SPM 12 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm) in MATLAB 2019b. Functional MRI data were converted from DICOM to NIFTI format, spatially realigned to the first image, and slice time corrected. EPI images were then slightly smoothed (4mm FWHM Gaussian kernel) to improve signal-to-noise ratio, as in our previous MVPA work (e.g., Jackson et al., 2017; Woolgar et al., 2015).

Structural images were co-registered to the mean EPI, segmented into grey and white matter, and then normalised to tissue specific templates. This was done to generate individual normalisation parameters, which we used to deform template-defined regions of interest (ROIs) to native space.

3.4.7.2 Regions of interest (ROIs)

We used the 13 frontoparietal MD ROIs from the parcellated map provided by Fedorenko et al. (2013), available online at imaging.mrc-cbu.cam.ac.uk/imaging/Mdsystem. These ROIs are activation-based and represent regions which demonstrate consistent recruitment across a diverse range of demanding cognitive tasks. The regions include bilateral anterior cingulate cortex (ACC; centre of mass (COM): 0 15 46.2, volume: 18.6 cm³), bilateral anterior insula/frontal operculum (AI/FO; COM: ±34.3 19 2.2, 7.9cm³), bilateral inferior frontal junction (IFJ; COM: ±43.8 3.8 32, 10.2cm³), bilateral anterior inferior frontal sulcus (aIFS; COM: ±34.6 47.1 18.5, 5.0cm³), bilateral posterior inferior frontal sulcus (pIFS; COM ±40.1 31.6 26.5, 5.7cm³), bilateral intraparietal sulcus (IPS; COM: ±29.4 -55.7 46.2, 34.0cm³), and bilateral premotor cortex (PM; COM: ±27.5 -2.3 56, 9.1cm³).

Left and right visual cortex ROIs, Brodmann's Area (BA) 17/18 (COM: -13.3, -80.6, 2.7 and 15.5 -79.2 3, respectively; volume: 56 cm³ and 54 cm³, respectively), were defined from the Brodmann's template of MRIcro (Rorden & Brett, 2000). Coordinates are in MNI152 space (McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, QC, Canada).

MD and visual cortex ROIs were deformed to the native space by applying the inverse of each participant's normalisation parameters. This allowed us to perform pattern classification analysis on an ROI basis using the native space data for each participant.

3.4.7.3 fMRI MVPA analysis

3.4.7.3.1 *General linear model (GLM)*

To estimate activation patterns for MVPA, we implemented a general linear model for each participant using the realigned, slice-time-corrected, smoothed, native space EPIs with SPM12. To account for trial-by-trial differences in RT (Todd et al., 2013), trials were modelled as epochs lasting from stimulus onset until response and thus each epoch varied in length (Grinband et al., 2008; Henson, 2007; Woolgar et al., 2014), and convolved with the hemodynamic response of SPM12. We also included six movement parameters (rotation around and translation in x, y, z) and corrected for serial autocorrelation using the FAST approach (Olszowy et al., 2019).

We modelled the data with 16 regressors per run. These regressors represented eight regressors (one for each of the four shapes, and one for each of the four colours) per object (left and right of fixation) in the display. Each trial contributed to the estimation of four regressors: the shape and colour of the object on the left, and the shape and colour of the object on the right. In combination with the attentional condition in each run (attend left/right and report shape/colour), this is equivalent to modelling the target colour, target shape, distractor colour, and distractor shape on each trial. Error trials were not modelled and were excluded from the analysis.

3.4.7.3.2 *Stimulus information decoding*

The primary aim of this study was to examine the effect of spatial and feature attention on attended and unattended stimulus information in our ROIs. To address this, we used MVPA to discriminate patterns of activation pertaining to the different feature values of the attended and unattended feature of the object in the attended and unattended location.

We performed MVPA using the Decoding Toolbox version 3.991 (Hebart et al., 2015), which wraps the LIBSVM library (Chang & Lin, 2011). We trained separate classifiers to discriminate between each of the six possible pairings of the four feature values (1v2, 1v3, 1v4, 2v3, 2v4 and 3v4) for shape and colour separately, in each of the four attention conditions. The four attention conditions were as follows: attended location attended feature (aLaF, e.g., the colour of the attended object in attend colour blocks), attended location unattended feature (aLuF, e.g., the shape of the attended object in attend colour blocks), unattended location attended feature (uLaF, e.g., the colour of the unattended object in attend colour blocks), and unattended location unattended feature (uLuF, e.g., the shape of the unattended object in attend colour blocks).

To illustrate how the decoding analysis proceeded, the decoding analysis of colour one from colour two in the aLaF condition is described in detail. For a given ROI, the pattern of beta values was extracted from each of the four relevant beta images (one for each colour) in each of the four relevant runs (four attend colour runs, comprising two attend left and two attend right runs, and both possible response mappings), yielding 16 multivoxel vectors. All the voxels in each ROI contributed to each vector, without feature selection. We used a linear support vector machine (LibSVMC, cost parameter $C = 1$) to classify between the four vectors pertaining to colour one ('strongly red') and the four vectors pertaining to colour two ('weakly red'). As the four vectors comprising each classification included both stimulus locations, the classifier generalised across stimulus location. We split the data into training and testing chunks using leave-one-run-out cross-validation and iterated around all combinations so that all runs contributed equally to training and testing. The four classification accuracies from the

four cross-validation folds were then averaged to give a single accuracy score for that participant, ROI, attention condition, and pair of colours. We repeated this procedure for each pairwise comparison and each of the four attention conditions separately. For all decoding analyses, we use percentage correct to express classifier accuracy, where chance performance has a theoretical value of 50%.

3.4.7.3.3 *Model selection and statistical analyses*

As with our behavioural analyses, statistical analyses to examine the effect of spatial and feature attention on stimulus information coding, across ROIs and across stimulus space, were performed by fitting linear mixed-effects models using the *lme4* package (v1.1-32, Bates et al., 2015) and then reducing these models to identify key effects using the *lmerTest* package (v3.1-3, Kuznetsova et al., 2017) in R.

Our full models for both research questions comprised all fixed and random effects we considered to be of theoretical importance (see Table 5 for details of our full and final models for MD and visual regions). In addition, where appropriate, variables that might capture additional variance were also included (i.e., stimulus colour set and feedback scheme). In all cases, models were fit to the accuracy of the classification (i.e., strength of information coding) of the given feature being decoded (*Feature Decoded*: colour/shape, averaged across all pairwise classifications), across each of the different attention conditions (*Location Attended * Feature Attended*: aLaF, aLuF, uLaF, uLuF). We then tailored our full models to our research question by specifying additional fixed effects as appropriate. To address the question of whether the spatial and feature attention effects varied across MD ROIs, we additionally included *ROIs* as a predictor in our full MD model. To examine whether the effect of attention varied across stimulus space, we additionally included *Feature Step Size* as a predictor to both MD and visual region models. This predictor reflected the physical discriminability of the stimuli with three levels: small (pairwise comparisons of one feature step size: lv2, 2v3, 3v4), medium (pairwise comparisons of two feature step sizes: lv3, 2v4), large (pairwise comparisons

of three feature step sizes: lv4). All predictors were included as fixed effects with all possible interaction terms included. As we reasoned that classification accuracy may vary across subjects as a function of attention condition, attention condition and subject were defined as random effects. As we sought to compare models primarily on the basis of fixed effects, full models were generated using maximum likelihood as is recommended in the literature (e.g., see Meteyard & Davies, 2020). We used model selection to determine which fixed and random effect factors best estimated decoding accuracy, using the *step* function of the lmerTest package (Kuznetsova et al., 2017). The final models comprised effects that were identified as significant predictors of classification accuracy. As in the behavioural analysis, final models were re-estimated with restricted maximum likelihood, to allow for significance testing of fixed effects (Meteyard & Davies, 2020). Post-hoc analyses were performed using the emmeans package (v 1.8.5), using the Satterthwaite method to calculate degrees of freedom. We used Bonferroni adjustment to correct for multiple comparisons.

Table 5 shows the full (i.e., initially specified) and final models, along with model fit parameters, in MD and visual regions. As with the behavioural results, our step model selection procedure eliminated extraneous variables (i.e., stimulus colour set and feedback scheme) from our final models, indicating that the inclusion of these factors did not significantly explain any additional variance or improve model fit (see Table 5).

Table 5. Full and final models specified for decoding analysis for MD and visual regions and final model selected

How do spatial and feature attention modulate the coding of stimulus information?								
MD regions	Sampling units: Total number of observations = 10080 N participants = 30							
	Model name	Fixed effects	Random effects	AIC	BIC	LL	df	Model test X ² df
	Full model	Decoding Accuracy ~ Location Attended * Feature Attended * Feature Decoded * ROIs + CalibrationType + feedbackGiven	(Location Attended * Feature Attended Sub)	88096	88594	-43979	69	12.93 26
	Final model	Decoding Accuracy ~ Location Attended + Feature Attended + Feature Decoded + ROIs + Location Attended: Attended:Feature Attended + Location Attended: Feature Decoded + Feature Attended: Feature Decoded + Location Attended: ROIs + Feature Attended: ROIs + Location Attended: Attended:Feature Attended: Feature Decoded + Location Attended: Attended:Feature Attended: ROIs	(Location Attended * Feature Attended Sub)	88057	88367	-43986	43	

How do spatial and feature attention modulate the coding of stimulus information?

Visual regions		Sampling units: Total number of observations = 1440 N participants = 30						
Model name	Fixed effects	Random effects	Model fit				Model test	
			AIC	BIC	LL	df	X ²	df
Full model	Decoding Accuracy ~ Location Attended * Feature Attended * Feature Decoded + CalibrationType + feedbackGiven	(Location Attended * Feature Attended Sub)	12747	12857	-6352.3	21	4.03	4
Final model	Decoding Accuracy ~ Location Attended + Feature Attended + Feature Decoded + Location Attended:Feature Attended + Feature Attended: Feature Decoded	(Location Attended * Feature Attended Sub)	12743	12832	-6354.4	17		

Does the effect of selective attention vary across stimulus space?

MD regions		Sampling units: Total number of observations = 10080 N participants = 30						
Model name	Fixed effects	Random effects	Model fit				Model test	
			AIC	BIC	LL	df	X ²	df
Full model	Decoding Accuracy ~ Location Attended * Feature Attended * Feature Step Size * Feature Decoded	(Location Attended * Feature Attended Sub)	87899	88152	-43915	35	N/A	
Final model	Final model = full model							

Does the effect of selective attention vary across stimulus space?

Visual regions	Sampling units: Total number of observations = 1440 N participants = 30							
	Model name	Fixed effects	Random effects	AIC	BIC	LL	df	Model test X² df
	Full model	Decoding Accuracy ~ Location Attended * Feature Attended * Feature Step Size * Feature Decoded	(Location Attended * Feature Attended Sub)	12749	12933	-6339.4	35	29.93 18
	Final model	Decoding Accuracy ~ Location Attended + Feature Attended + Feature Decoded + Location Attended:Feature Attended + Feature Attended: Feature Decoded	(Location Attended * Feature Attended Sub)	12743	12832	-6354.4	17	

For above, Decoding Accuracy = average decoding accuracy for given comparison, Location Attended = binary variable indicating whether decoding comparison was for the object in the attended location, Feature Attended = binary variable indicating whether decoding comparison was for the attended feature, Feature Decoded = binary variable indicating whether decoding of colour or shape information, ROIs = each of the MD ROIs included in this analysis, Feature Step Size = reflecting the physical discriminability of the stimuli with three levels: small (pairwise comparisons of one feature step size: 1v2, 2v3, 3v4), medium (pairwise comparisons of two feature step sizes: 1v3, 2v4), large (pairwise comparisons of three feature step sizes: 1v4), Sub = subject identifier. For the first research question, Calibration_Type (type of colour calibration used or stimulus colour set) and Feedback_Given (reflecting feedback scheme, i.e., whether trial-by-trial feedback was given or not) were included to ascertain the effect of acquisition variability on our results. In both these final models, Calibration Type and Feedback given were eliminated, indicating their inclusion did not significantly improve model fit. Accordingly, they were not included in the second research question.

AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion, LL = LogLikelihood, df = degrees of freedom, X² = Chi-square. Note: ‘:’ indicates interaction between only the terms specified (i.e., without lower-order effects), while ‘*’ indicates an interaction with all lower order effects included in model. Terms in brackets represent random effect terms. Final models were chosen using the ‘step’ function from lmerTest. This performs a backward elimination of random-effect terms followed by backward elimination of fixed-effect terms in linear mixed models as explained in the text.

3.4.7.4 Generalisation analyses

As a secondary analysis, we sought to probe the organisational structure of task-related representations in MD regions (see Supplementary Results for equivalent analysis in visual regions). To investigate this, we ran a series of cross-classification decoding analyses using only task-relevant (aLaF) information.

First, we examined whether we could decode ‘difficulty’ (i.e., ‘easy’ from ‘hard’ stimuli) from our data, independent of feature category, for colour and shape. To illustrate the procedure, we describe the process for the colour task. We first trained our classifier to distinguish between two features on the same side of the decision boundary that reflected two levels of difficulty (e.g., reddish: colour one (strongly red, easy) and colour two (weakly red, hard)). We then tested this classifier on whether it would correctly classify the difficulty of the two features on the other side of the decision boundary (e.g., greenish: colour four (strongly green, easy) and colour three (weakly green, hard)). This procedure was repeated in the reverse (i.e., training to distinguish between easy green and hard green, and testing on easy red and hard red), and separately for shape information. This analysis allowed us to examine whether a representation of ‘difficulty’, as defined by distinguishing between ‘hard’ and ‘easy’ trials on one side of the decision boundary, generalised across feature space.

Second, we tested whether we could decode the feature categories (i.e., for colour: red or green, for shape: X-shaped or non-X-shaped), independent of difficulty level. To illustrate the procedure, we again describe the process for one of the colour classifications in full. We first trained our classifier to distinguish between two features on opposite sides of the decision boundary but equidistant from it, that is, matched in terms of difficulty (e.g., colour one (strongly red, easy) and colour four (strongly green, easy)). We then tested this classifier to distinguish between the remaining two features (e.g., colour two (weakly red, hard) and colour three (weakly green, hard)), and vice versa. We repeated the same procedure for shape information. This analysis allowed us to examine whether the distinction between feature categories (e.g., ‘red’ and ‘green’, as

defined by using only ‘hard’ stimuli), generalised across difficulty (e.g., to capture the distinction of ‘red’ versus ‘green’ as captured using only ‘easy’ stimuli).

3.4.8 Statistical testing, comparison to chance

We used the Bayes Factor package in R (Morey & Rouder, 2018) to quantify evidence for above-chance (H1) or at-chance (null) classification accuracy for each of our pairwise comparisons. We defined priors using a half-Cauchy prior for the alternative, with a default width of 0.707 and an interval ranging from a standardised effect size of 0.5 to infinity (Jeffreys, 1998; Rouder et al., 2009; Wetzels et al., 2011), as in our previous work (Moerel et al., 2021; Teichmann et al., 2021). We used a point null centred on chance (Morey & Rouder, 2011). We report the Bayes Factor (BF10), which indicates the likelihood that the evidence favours the alternative hypothesis relative to the null hypothesis. For instance, a BF10 of 3 indicates that the data were three times more likely to occur under the alternative than the null hypothesis. BF10s > 3, 10, 30, and 100 are considered ‘moderate’, ‘strong’, ‘very strong’, and ‘extreme’ evidence in support of the alternative hypothesis (i.e., decoding accuracy above chance), respectively (Wagenmakers et al., 2018). Conversely, BF10s < 1/3, 1/10, 1/30 and 1/100 are considered ‘moderate’, ‘strong’, ‘very strong’, and ‘extreme’ evidence in support of the null hypothesis (i.e., at-chance decoding accuracy).

3.4.9 MEG-fMRI fusion analysis

Recent MEG data from the same paradigm indicated a significant influence of frontal information coding on the response of occipital cortices at later timepoints (Goddard et al., 2022). Moreover, the timecourse with which this feedback dominated information flow in the brain corresponded to the onset of multiplicative attention effects. As a final analysis, we sought to formally relate the spatial data from the current fMRI study to the timecourse data from this previous MEG study, using MEG-fMRI fusion (Cichy et al., 2014, 2019; Cichy & Oliva, 2020). By combining the two datasets and looking for the commonalities between them, we could estimate the timecourse of information coding in the MD system and in turn establish whether frontal MD regions are reasonable candidates for the source of top-down information flow seen in MEG.

Following the methods developed by Cichy and colleagues (see Cichy et al., 2014, 2019; Cichy & Oliva, 2020), we created representational matrices that were equivalent across the two methods. More specifically, we related the accuracy of the classifier in decoding colour or shape (between each feature step in each of the attention conditions) across the two imaging modalities to ascertain the commonality in representational structure in time and space. This allowed us to examine at which timepoints (in MEG) the data most closely resembled the representational pattern recorded in each brain region (with fMRI).

The extant MEG data were analysed in source space and partitioned into broadly defined frontal and occipital regions (see Goddard et al. (2022) for more details on MEG acquisition and analysis). The approach used to analyse the MEG data was similar to the approach described above to analyse our fMRI data. For the MEG data, the classifier was trained to discriminate between these patterns at each timepoint, for left and right objects in the two (frontal and occipital) ROIs separately.

To facilitate comparison with our fMRI data, we reconfigured the group-level results of the MEG decoding analyses into four matrices: one for colour decoding and one for shape decoding, for each of the two MEG ROIs. Each matrix comprised 421 x 24 entries.

This represented average decoding accuracy for the 24 possible pairwise classifications (i.e., six pairwise comparisons of feature values for each of the four attention conditions), for each time point (421 time points: 5ms bins, covering 100ms prior to and 2000ms after stimulus onset).

For the fMRI data, we similarly generated 24-item vectors representing group-level decoding accuracy for the six pairwise classifications x four attention conditions. We generated one such vector for each of the eight ROIs (seven MD regions, collapsed across hemisphere, and one visual region: BA17/18), for colour decoding and shape decoding separately.

Next, we used Spearman rank-order correlation to compare the group-level fMRI data for each region with the group-level MEG data at each timepoint. For example, to look at the relationship between the pattern of colour decoding accuracies between fMRI IFJ and MEG Frontal, we computed a Spearman's rank-order correlation between the fMRI vector for IFJ coding of colour and each row of the frontal MEG colour coding matrix (each row comprising a timepoint in the MEG analysis) (see Figure 22). This yielded a 421-item vector reflecting the degree of correlation between the pattern of colour decoding for IFJ (fMRI) and the frontal region (MEG) across time. A graphical depiction of this process is shown in Figure 22, with the final 421-item correlation coefficient vector plotted across time. We repeated this process such that each fMRI MD region was correlated with the frontal MEG timecourse, and fMRI BA17/18 was correlated with the occipital MEG timecourse. Colour and shape decoding results were correlated separately (i.e., colour decoding was correlated with colour decoding only, not with shape).

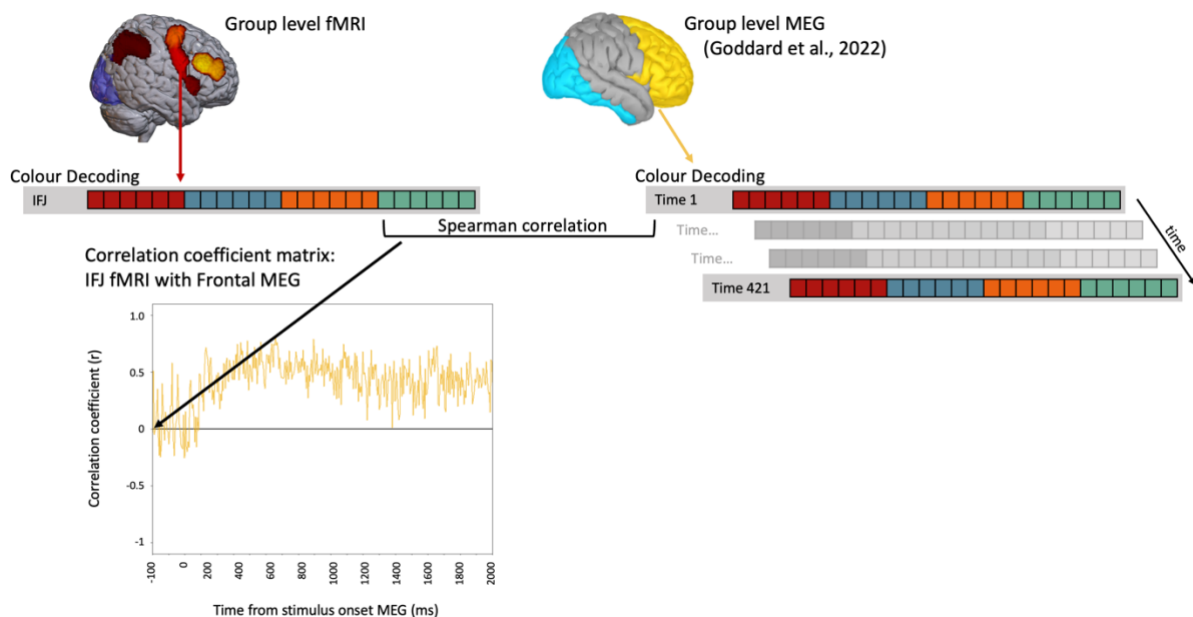


Figure 22. Graphical representation of the method for fusing the pattern of colour decoding accuracies between inferior frontal junction (IFJ) (fMRI) and the frontal MEG region across time. Each square represents the average decoding accuracy for the given pairwise comparison in the applicable attention condition (described earlier). Colours reflect from which attention condition the given feature was decoded: red = attended feature in the attended location (aLaF), blue = attended feature in the unattended location (aLuF), orange = unattended feature in the attended location (uLaF), teal = unattended feature in the unattended location (uLuF). The fMRI vector was correlated with each timepoint of the MEG matrix, as shown above, yielding a correlation coefficient vector representing the correlation between the fMRI vector with the MEG results across time. The correlation coefficient vector between the IFJ and frontal regions for colour decoding is plotted in gold above.

To test which of our MEG-fMRI correlations were significant, we first generated a null distribution from the data. More specifically, we permuted each fMRI vector, allowing for the group-level decoding accuracies to be randomly assigned to any of the pairwise classifications under any attention condition. We then recalculated the correlation between the MEG matrices and these permuted fMRI vectors as described previously. We repeated this process 1000 times to produce a null distribution of reshuffled correlation values against which we compared our original observed correlations. Thus, for each MEG-fMRI correlation (e.g., colour decoding: fMRI ACC with MEG frontal), we

generated 1001 p-value timecourses (one from the observed data and 1000 from the permuted data).

We then used a threshold-free cluster enhancement (TFCE) correction (Smith & Nichols, 2009) on our p-values to correct for multiple comparisons at a cluster level. For a given MEG-fMRI comparison, we determined the length of the longest significant cluster (defined as the longest continuous string of uncorrected p-values less than 0.05) in each of the permuted timecourses pertaining to this comparison, thus forming a null distribution of cluster lengths. We then compared the length of each significant p-value cluster (continuous string of p-values less than 0.05) in observed data for this comparison to this null distribution. A cluster in the observed data was considered significant according to TFCE if it was longer than the top 5% of lengths in the null distribution ($p < 0.05$). We repeated this process for each MEG-fMRI comparison.

3.5 Results

3.5.1 Behavioural accuracy and reaction time (RT)

The final models selected through our model selection process for accuracy and RT shown in Table 4. Both of our final models for accuracy and RT included task (attending to colour or shape), feature step size (colours/shapes one to four) and their interaction, indicating the importance of these predictors in our models.

For accuracy, the main effect of task was not significant ($F(1,232)=2.31$, $p=0.130$), indicating participants were on average as accurate at identifying colours (mean accuracy = 0.84 ± 0.21 , Figure 23, Panel A) as shapes (mean accuracy = 0.86 ± 0.13 , Figure 23, Panel B). For RT, however, this main effect of task was significant ($F(1,210)=7.20$, $p=0.008$), indicating participants were on average faster to correctly identify colours (mean RT= 0.71 ± 0.13 s, Figure 23, Panel C) than shapes (0.72 ± 0.10 s, Figure 23, Panel D).

Our final models also permitted examination of the effect of feature step size on accuracy and RT, and the interaction of feature step with task. For accuracy, the main effect of feature step size was significant ($F(3,232)=87.87$, $p<0.0001$). However, this was superseded by a significant interaction between task and feature step ($F(3,232)=21.25$, $p<0.0001$). Post-hoc tests indicated that for the shape task, participants were more accurate when classifying ‘easy’ stimuli (i.e., feature values further from the decision boundary, values one (strongly non-X-shaped) and four (strongly X-shaped)) compared with ‘hard’ stimuli (i.e., feature values closer to the decision boundary, values two (weakly non-X-shaped) and three (weakly X-shaped)) (all p -values <0.0001 , Figure 23, Panel B). For the colour task, while participants were on the whole more accurate in classifying colours further from, relative to closer to, the decision boundary (all p -values <0.05), they were also significantly less accurate in classifying colour two (weakly red) than any other colour (all p -values <0.0001) (Figure 23, Panel A).

Similarly, while the main effect of feature step size on RT was also significant ($F(3,203)=118.95$, $p<0.0001$), this was superseded by a significant interaction between task and

feature step size ($F(3,203)=19.68$, $p<0.0001$), indicating RT differed as a function of both feature step size and task. Nevertheless, post-hoc analyses showed that for both colour and shape tasks, participants were faster to classify stimuli further from the decision boundary than stimuli closer to the decision boundary (all p -values for colour <0.0001 , all p -values for shape <0.05 , see Figure 23, Panels C and D).

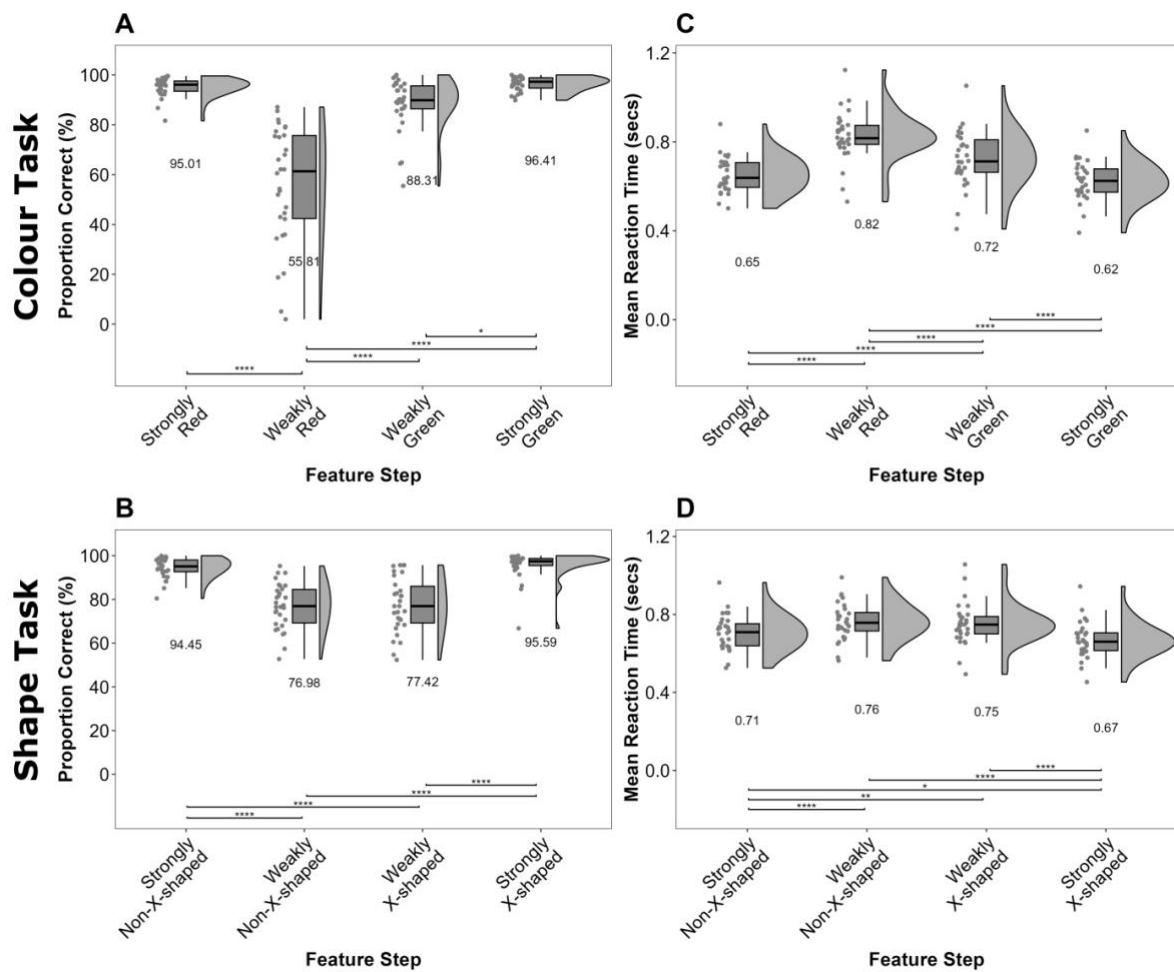


Figure 23. Panels A and B: Proportion correct across all participants, across relevant feature steps, for colour (A) and shape (B) trials. Panel C and D: Average reaction time (in seconds) for correct trials across all participants, across relevant feature steps, for colour (C) and shape (D) trials. Scatter at left depicts mean accuracy or mean reaction time for the given task for each participant. Boxplots indicate mean accuracy or mean reaction time (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value within $1.5 \times$ applicable quartile range. Data distribution is indicated at right for each condition (half violin). Significant pairwise comparisons marked with Bonferroni adjusted p -value codes as follows: * < 0.05 , ** < 0.01 , *** < 0.001 , **** < 0.0001 .

3.5.2 Stimulus information coding

3.5.2.1 MD regions selectively code of task-relevant information

Our key question was whether spatial and feature attention modulated the coding of stimulus information and, if so, whether the two subtypes of attention interacted in modulating the neural code. As shown in Table 5, we found that both spatial attention ($F(1,29)=90.13$, $p<0.0001$) and feature attention ($F(1,29) = 59.09$, $p<0.0001$) boosted stimulus information processing. However, these effects were modulated by a significant two-way interaction between the two subtypes of attention $F(1,29)=67.59$, $p<0.0001$). This was driven by a multiplicative interaction in which information coding was boosted only if a stimulus feature was both relevant to the task and at the focus of spatial attention ($BF_{10}>100$ for the attended feature information at the attended location, i.e., aLaF, with $BF_{10}s<0.1$ for other attention conditions, see Figure 24).

Effect of spatial and feature attention in frontoparietal cortex

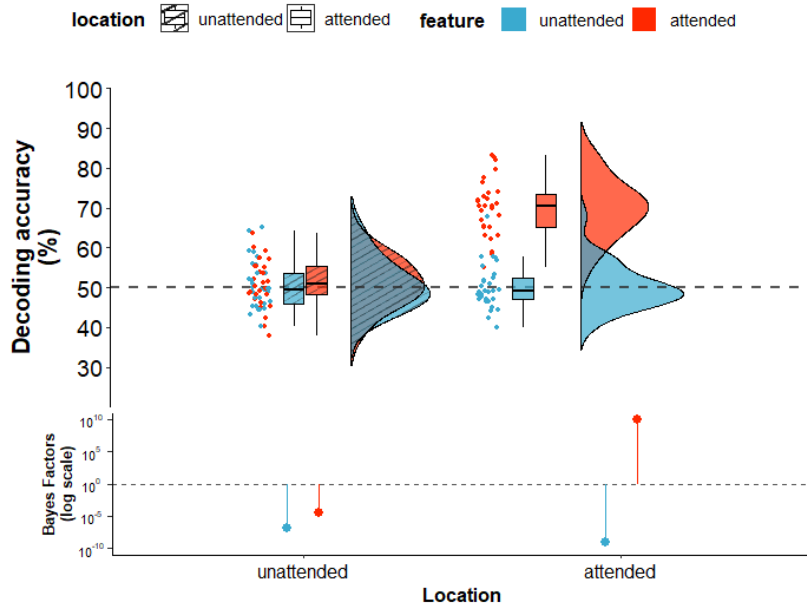


Figure 24. Average decoding accuracy for feature information, averaged across MD ROIs and task (colour and shape), split by spatial and feature attention conditions. Decoding accuracy reflects average accuracy in decoding colour or shape information in the object that was in the attended (solid colour) or unattended (lined colour) location, when the feature in question was attended (red) or unattended (blue). Scatter at left depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value within 1.5 x applicable quartile range. Data distribution is indicated at right for each condition (half violin). The strength of evidence for above-chance ($H1$) or at-chance ($H0$) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

The key two-way interaction was modulated by two three-way interactions that did not change the interpretation. First, there was a three-way interaction between spatial attention, feature attention and MD region ($F(6,9932)=3.51$, $p=0.002$). This reflected variation in the strength of the two-way interaction over MD regions with the most prominent effects in IFJ and ACC and smallest effects in PM (see Figure 25). Nevertheless, the multiplicative interaction was significant in all regions separately (all p -values ≤ 0.0001), with selective boosting of only attended feature information at the attended location across all regions (all $BF10s > 100$ for aLaF information; anecdotal evidence for above-chance decoding of uLaF in the IPS; else $BF10s < 0.3$ for other attention conditions across all MD regions, see Figure 25).

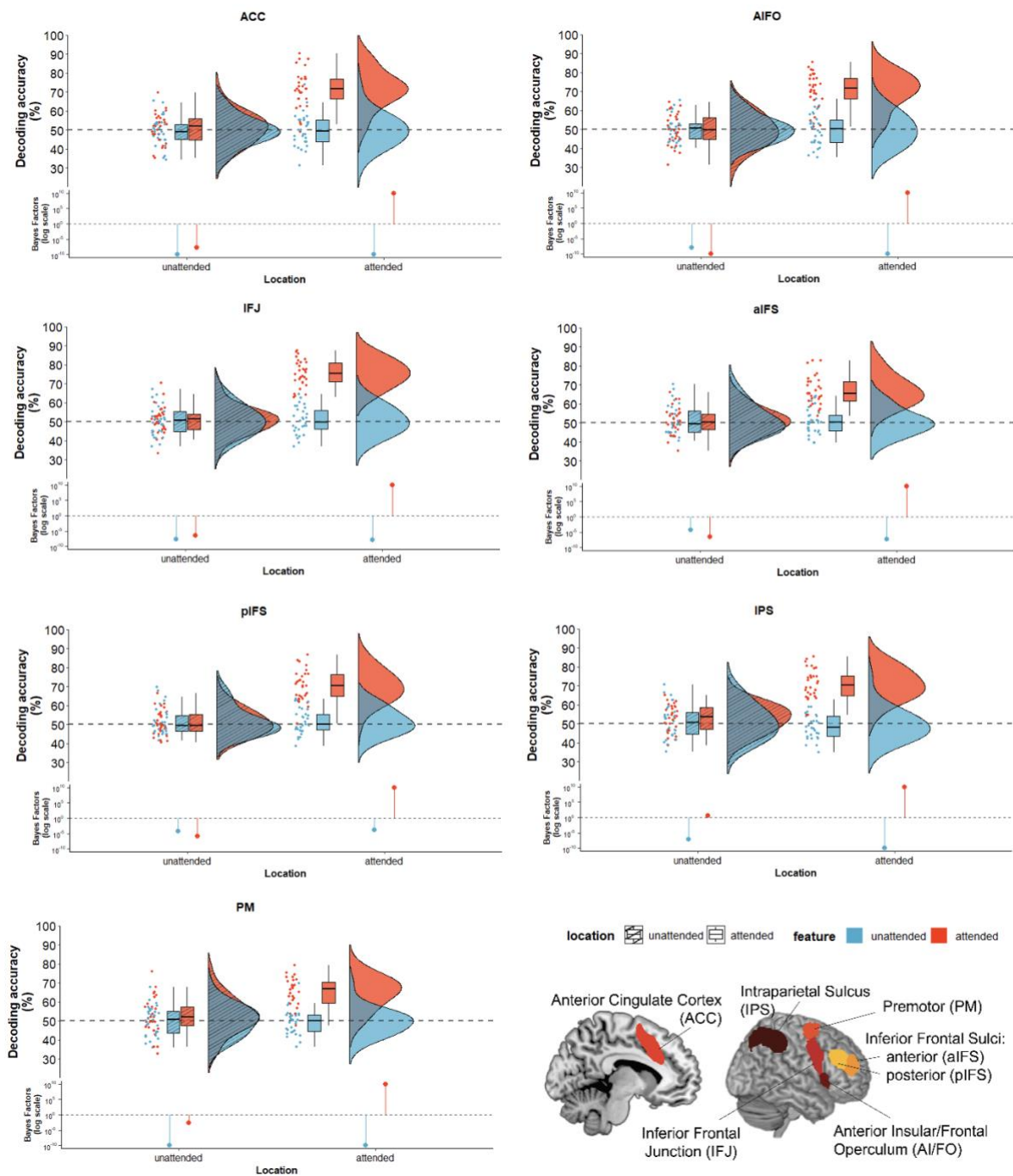


Figure 25. Average decoding accuracy for feature information averaged over task (colour or shape), across the different spatial and feature attention conditions, shown here for each of the MD ROIs. Decoding accuracy here reflects average accuracy in decoding colour or shape information in the object that was in the attended (solid colour) or unattended (lined colour) location, when the feature in question was attended (red) or unattended (blue). Scatter at left depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). Data distribution is indicated at right for each condition (half violin).

Second, there was a three-way interaction between spatial attention, feature attention and feature decoded ($F(1,9932)=4.81$, $p=0.028$) (see Figure 26). This reflected a larger multiplicative effect of the two types of attention on coding of object colour, relative to shape. Nevertheless, the two-way interaction was significant for each feature separately (colour: $F(1,35.18)=72.24$, $p<0.001$; shape: $F(1,35.18)=51.40$, $p<0.001$), and extreme evidence for above-chance decoding was found for only the attended feature of the attended object ($BF10s>100$; $BF10s<0.38$ for other attention conditions).

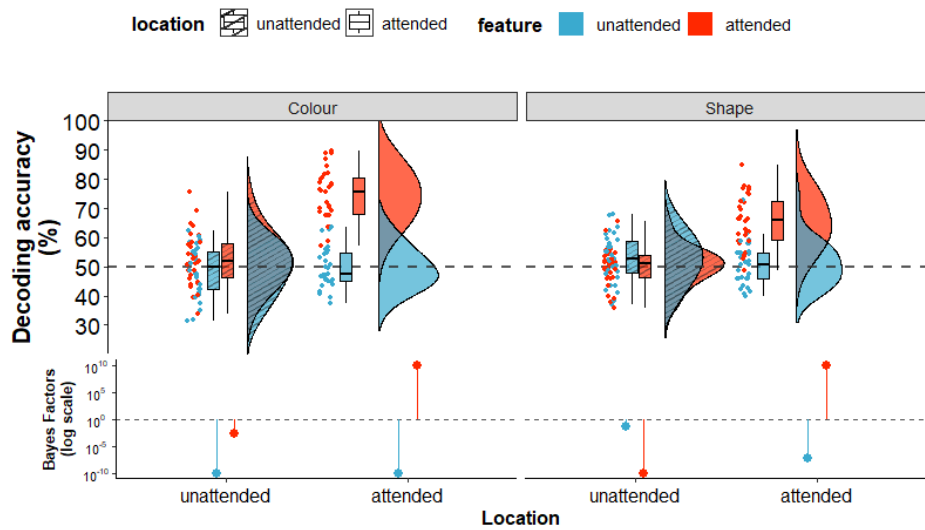


Figure 26. Average decoding accuracy for colour (left panel) or shape (right panel) information, for the object in the attended (solid colour) or unattended (lined colour) location, when the feature in question was attended (red) or unattended (blue), collapsed across all MD regions. The strength of evidence for above-chance ($H1$) or at-chance ($H0$) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

3.5.2.2 Task-relevant information is organised across two dimensions in MD regions

The results from our first analysis indicate a multiplicative effect of spatial and feature attention on representations in MD regions such that only information about the attended feature of the attended object was enhanced. In previous work with similar

stimuli, we found that spatial and feature selective attention had qualitatively different effects on coding of stimulus information, with spatial attention tending to emphasise large feature differences, and feature-selective attention tending to enhance small feature differences (Goddard et al., 2022), in line with a normalisation model of Reynolds and Heeger (2009). Here, as a secondary analysis, we sought to ascertain whether or not the selective boost in the representation of task-relevant information was uniform across stimulus space. We additionally sought to characterise the resulting organisation of task-related representations in MD regions.

Our results suggested that the interacting effect of spatial and feature attention on information coding varied with the physical discriminability of the stimuli (three-way interaction spatial attention * feature attention * step size, $F(2,9940)=31.07$, $p<0.0001$, Figure 27). This effect was more pronounced for colour than shape information (four-way interaction, $F(2,9940)=6.03$, $p=0.002$), but significant for both colour and shape information separately (colour: $F(2,9940)$ $p<0.0001$)= 31.36 ; shape: $F(2,9940)=5.74$, $p=0.003$).

Effect of spatial and feature attention in frontoparietal cortex

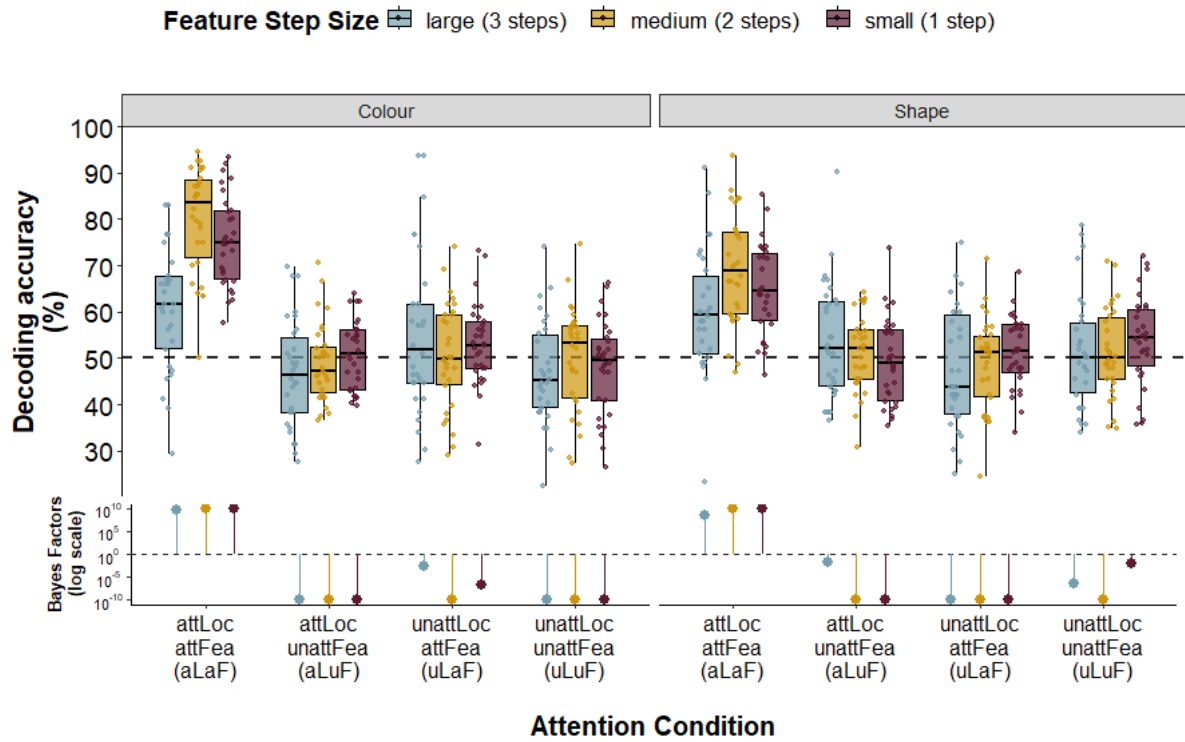


Figure 27. Average decoding accuracy of colour (left panel) and shape (right panel) information in each of the spatial and feature attention condition for each feature step size, collapsed across all MD regions. Feature step sizes are small (pairwise comparisons of one feature step size: 1v2, 2v3, 3v4), medium (pairwise comparisons of two feature step sizes: 1v3, 2v4), large (pairwise comparisons of three feature step sizes: 1v4). Scatter depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). The strength of evidence for above-chance (H1) or at-chance (H0) decoding was tested using Bayesian *t*-tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

The differential attention effect over step size appeared to be driven by differences in the coding of task-relevant (aLaF) information as all other conditions were at chance (mean decoding accuracy 60.2-80.0% and BFs>100 for aLaF conditions; for all other attention conditions, mean decoding accuracy 46.7-55.0%; all BF10s<0.34; Figure 27). Within the aLaF condition, post-hoc tests revealed that medium (two step) physical discriminations were more readily decoded than small (one step) (colour: $t(9940)=3.90$, $p<0.001$; shape: $t(9940)=3.63$, $p<0.001$), or large (three step) (colour: $t(9940)=12.11$, $p<0.0001$; shape: $t(9940)=5.77$, $p<0.0001$) discriminations. In addition, small physical

discriminations were decoded more readily than large ones (colour: $t(9940)=9.77$, $p<0.0001$); shape: $t(9940)=3.25$, $p=0.004$). This suggested that attention boosted the representation of task-relevant information in MD regions differentially according to the physical discriminability of the stimuli, with advantage for medium and then small physical discriminations over large ones.

In this analysis, medium and large discriminations always crossed the stimulus decision boundary used by the participant, while a subset of the ‘small’ discriminations (1v2 and 3v4) did not. To check whether this difference contributed to our results, we re-ran the analysis including only the pairs that straddled the decision boundary (i.e., only the 2v3 pairwise comparison was included for the small step size, see Supplementary Figure S 4). This did not change the overall pattern of results; while small step sizes were now as decodable as large step sizes, medium step size was still more decodable than small and large step sizes.

Since the medium step size classification always discriminated between one perceptually easier and one perceptually harder item, we next asked to what extent difficulty was an organising dimension for the MD representations in this task in our first cross-classification analysis. Indeed, we found item difficulty (distance from classification boundary) generalised over each half of feature space (e.g., red and green). We found strong to extreme evidence for the coding of stimulus difficulty for both colour (Figure 28, Panel A) and shape (Figure 28, Panel B) information across all MD regions (colour mean accuracy range = 64.2-73.6%, all $BF10s>100$; shape mean accuracy range = 58.2-67.7%, all $BF10s>46.80$), with the exception of shape information in IPS, where the evidence of stimulus coding was anecdotal (mean accuracy = 57.0%, $BF10=1.82$).

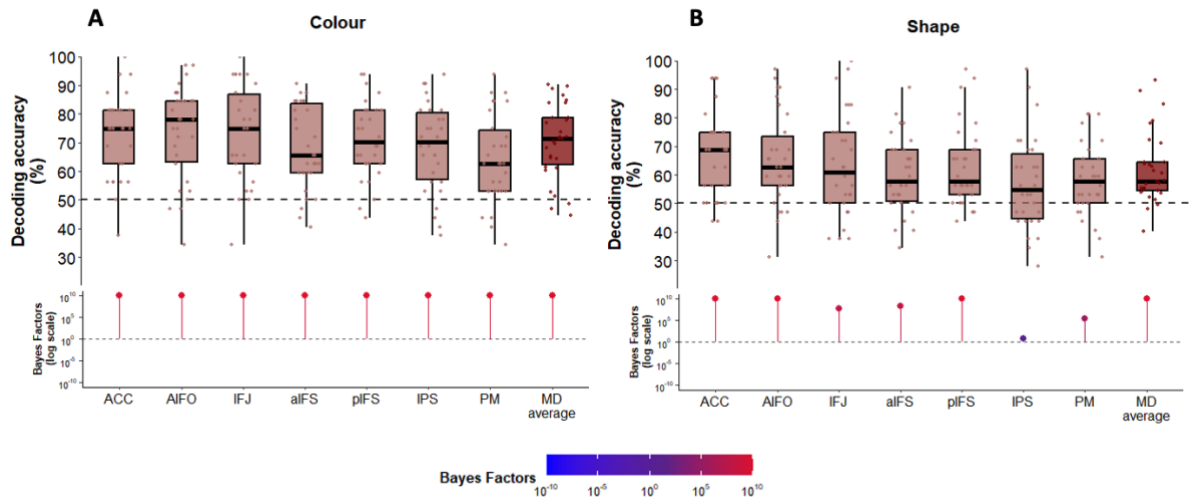


Figure 28. Decoding of difficulty for each MD region (collapsed across hemisphere, light red), and averaged across MD ROIs (dark red), generalised over the two colour categories (Panel A) and shape categories (Panel B). For task-relevant information (i.e., aLaF) only. ‘Difficulty’ here represents decoding of ‘easy’ from ‘hard’ stimuli, and vice versa, independent of feature category, for colour and shape separately. The strength of evidence for above-chance (H1) or at-chance (H0) decoding was tested using Bayesian *t*-tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

Our second cross-classification analysis examined whether a representation of stimulus category (for colour: red or green, for shape: X-shaped or non-X-shaped), generalised over difficulty. We found very strong evidence for the coding of colour over difficulty level for MD regions overall (darker bar in Figure 29, Panel A: mean accuracy = 55.2%, $BF_{10}=32.31$), and strong evidence in the IPS, pIFS, IFJ and PM regions separately (accuracy range = 54.4–56.7%, $BF_{10}s>9.71$). In contrast, we found moderate evidence for the null in the equivalent analysis for shape information in MD regions overall (darker bar in Figure 29, Panel B: mean accuracy = 52.9%, $BF_{10}=0.23$), with only aIFS showing moderate evidence for coding of shape over difficulty (mean accuracy = 55.1%, $BF_{10}=8.48$). These results suggest that responses in MD regions reflect both task-relevant stimulus categorisations (especially for colour, e.g., red vs green) and the difficulty of those discriminations (easy vs hard).

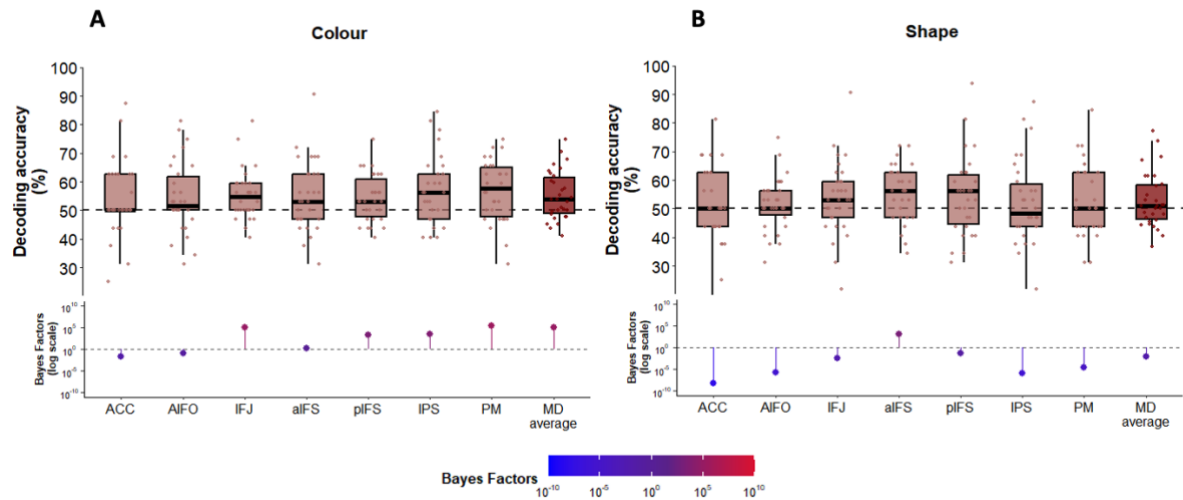


Figure 29. Decoding of colour category information (Panel A) and shape category information (Panel B) for each MD region (collapsed across hemisphere, light red), and averaged across MD ROIs (dark red), generalised over difficulty. For task-relevant information (i.e., aLaF) only. This analysis allowed us to examine whether the distinction between feature categories (e.g., ‘red’ and ‘green’, as defined by using only ‘hard’ stimuli), generalised across difficulty (e.g., to capture the distinction of ‘red’ versus ‘green’ as captured using only ‘easy’ stimuli). The strength of evidence for above-chance (H_1) or at-chance (H_0) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

3.5.2.3 Selective prioritisation of task-relevant information extends to visual cortex

Next, we sought to contextualise the attentional effects seen in MD regions by examining the effects of attention on stimulus processing across the brain. To look at this, we ran the same set of analyses in visual cortex as an ROI (BA17/18), and on a whole brain basis using a roaming searchlight.

In visual cortex, we first asked whether spatial and feature attention modulated the coding of stimulus information, and if so, whether these two subtypes of attention interacted in a multiplicative manner, as we had seen in MD regions. As shown in Table 5, there were significant main effects of spatial attention ($F(1,29)=16.21$, $p<0.001$), and feature attention ($F(1,29)=14.00$, $p<0.0001$) that were modulated by a significant two-way interaction between spatial and feature attention ($F(1,29)=16.73$, $p<0.001$), that did not differ between colour and shape information (see final model in Table 5). As we had observed in MD regions, information coding was strongest for the attended feature of the attended object (Figure 30), with only this task-relevant information decodable above chance ($BF_{10}>100$ for aLaF; BF_{10} s for all other attention conditions <0.1).

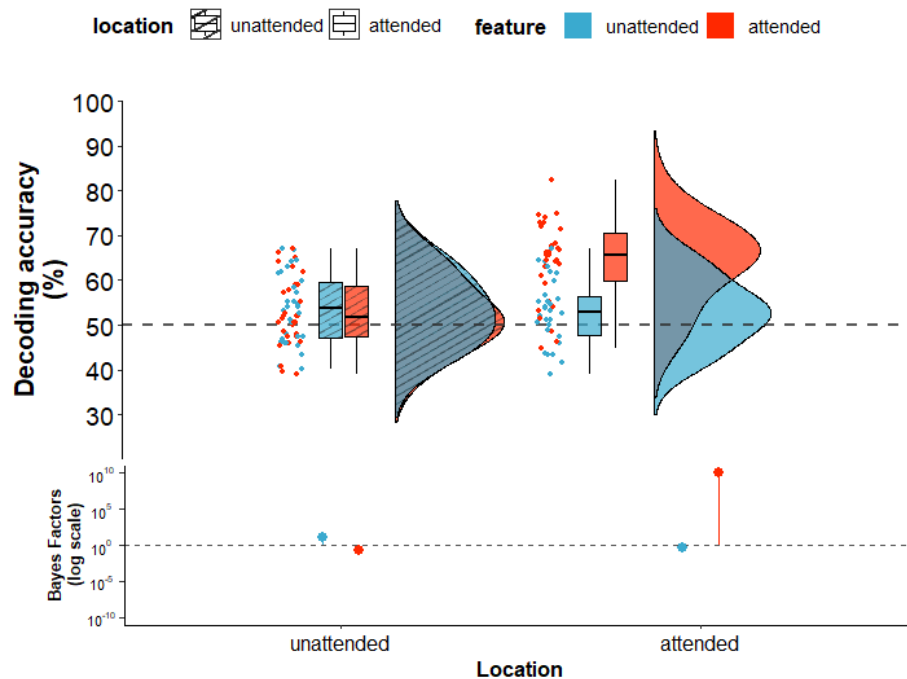


Figure 30. Average decoding accuracy for feature information for visual cortex (BA17/18) split into spatial and feature attention conditions. Decoding accuracy here reflects average accuracy in decoding colour or shape information in the object that was in the attended (solid colour) or unattended (lined colour) location, when the feature in question was attended (red) or unattended (blue). Scatter at left depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). Data distribution is indicated at right for each condition (half violin). The strength of evidence for above-chance (H_1) or at-chance (H_0) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

For completeness, we also ran whole-brain searchlight analyses (Kriegeskorte et al., 2006) to examine the extent to which the multiplicative effect was specific to the MD and visual cortex, and whether evidence of irrelevant information coding could be found in other brain regions (see Supplementary Materials for this chapter). We found widespread representation of task-relevant information throughout the brain, with the strongest decoding clusters overlapping with MD regions. In contrast, we did not find representation of irrelevant information (i.e., decoding accuracy averaged across all pairwise comparisons comprising a given attention condition, aLuF, uLaF or uLuF, with each attention condition tested separately) in any brain region.

3.5.2.4 Organisation of information in visual cortex

Next asked whether these selective attention effects varied across stimulus space in visual cortex. We anticipated that bottom-up factors (physical discriminability) might lead to a stronger representation of larger feature differences (i.e., greater decoding of features further apart in feature space) than smaller differences in visual cortex. On the other hand, top-down effects (e.g., feedback), might tend to drive representations towards the strong discrimination of medium step sizes, as seen in MD regions.

In fact, although the pattern of effects was visually similar to those of MD regions (Figure 31), there was no statistical evidence that attention effects varied across stimulus space for visual cortex. Table 5 shows the full model initially specified in our linear mixed-effects analysis and the final model selected through our analysis. Inclusion of stimulus discriminability (or levels of step size difference) did not explain any additional variance above and beyond that explained by spatial attention, feature attention, colour or shape information, or their interactions (compare full and final models specified in Table 5; Feature Step Size was not included in final model). As for the MD analysis, we also tested the same effect after restricting the analysis to only include comparisons that straddle the decision boundary; the results of this modified analysis were broadly the same, with no effect of stimulus discriminability on the spatial and feature interaction (see Supplementary Figure S 5). Although our results did not suggest that attention effects varied across stimulus space for visual cortex, for completeness, a cross-classification analysis to examine the generalisability of difficulty and feature categories in the visual cortex, as we had done in MD regions, is included in Supplementary Materials (see Supplementary Figure S 6 and Figure S 7). In brief, our results here suggested that difficulty did act as an organising principle for the representation of stimulus information in visual cortex, although these effects were less strong and less consistent than seen for MD regions. Overall, the qualitative pattern of the effect across feature space was similar, but weaker, than in MD regions.

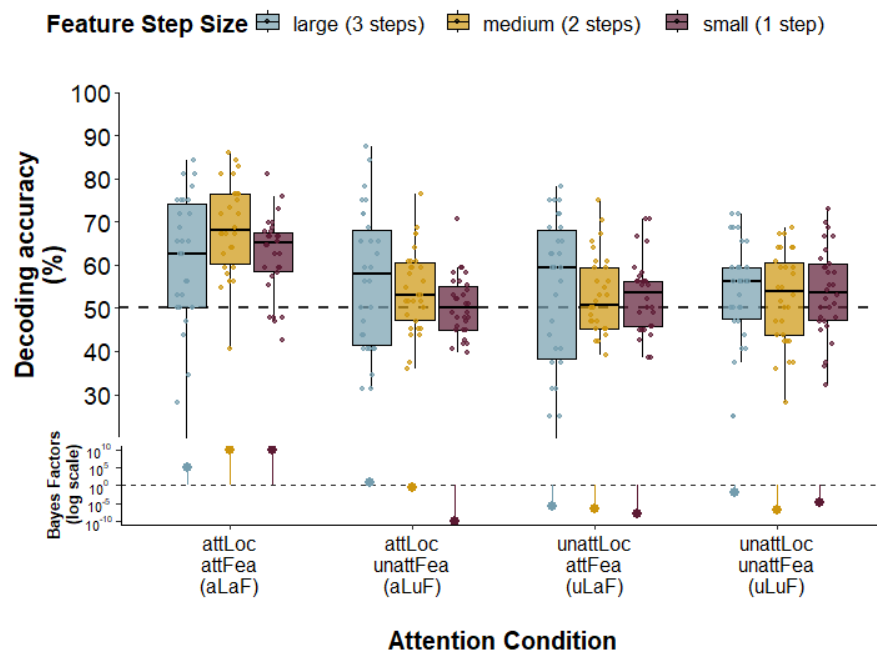


Figure 31. Average decoding accuracy of stimulus information (averaged over colour and shape information, as not significantly different between these types of information) for visual cortex (BA17/18). Shown for each attention condition for each feature step size. Feature step sizes are small (pairwise comparisons of one feature step size: 1v2, 2v3, 3v4), medium (pairwise comparisons of two feature step sizes: 1v3, 2v4), large (pairwise comparisons of three feature step sizes: 1v4). While attended information (aLaF) was more decodable than other attention conditions, inclusion of feature step size in this interaction did not explain additional variance (that is, interaction between feature step size and attention conditions was not significant). Scatter depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). The strength of evidence for above-chance (H1) or at-chance (H0) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

3.5.4 Temporal dynamics of task-relevant information prioritisation (MEG-fMRI fusion)

Our results thus far suggested broadly distributed, highly specific task-related representations in response to spatial and feature attention. We were interested in the timecourse of these effects, and particularly whether the similarity in responses between MD and visual regions could be attributed to top-down effects. However, the limited temporal resolution of fMRI precluded detailed analysis of this question in the fMRI data alone. Accordingly, we sought to relate our group-level fMRI results to group-level MEG timecourse data (obtained separately using the same paradigm, Goddard et al., 2022) to estimate when our fMRI results occurred. For this, we compared the MD fMRI data to frontal MEG data, and visual cortex (BA17/18) fMRI data to occipital MEG data (see Methods). We ran and reported this analysis for colour and shape decoding separately, since we had observed different dynamics of colour and shape coding in the MEG data (Goddard et al., 2022).

For colour information (Figure 32), with the exception of IFJ where there was significant correspondence from around 300ms, the pattern of decoding accuracies in most frontal MD regions (i.e., ACC, AI/FO, aIFS, pIFS, PM) matched the frontal MEG data for a sustained period starting from approximately 200ms after stimulus onset. In contrast, the pattern of decoding accuracies in BA17/18 matched that of the MEG occipital ROI around 100ms later, from approximately 300ms. This pattern of results is in line with selective processing of relevant information emerging first in frontal MD regions and later in visual cortex. The timecourse also matched the timecourse with which feedback from the frontal ROI was observed for colour coding in the MEG data (i.e., from around 285ms, Goddard et al, 2022).

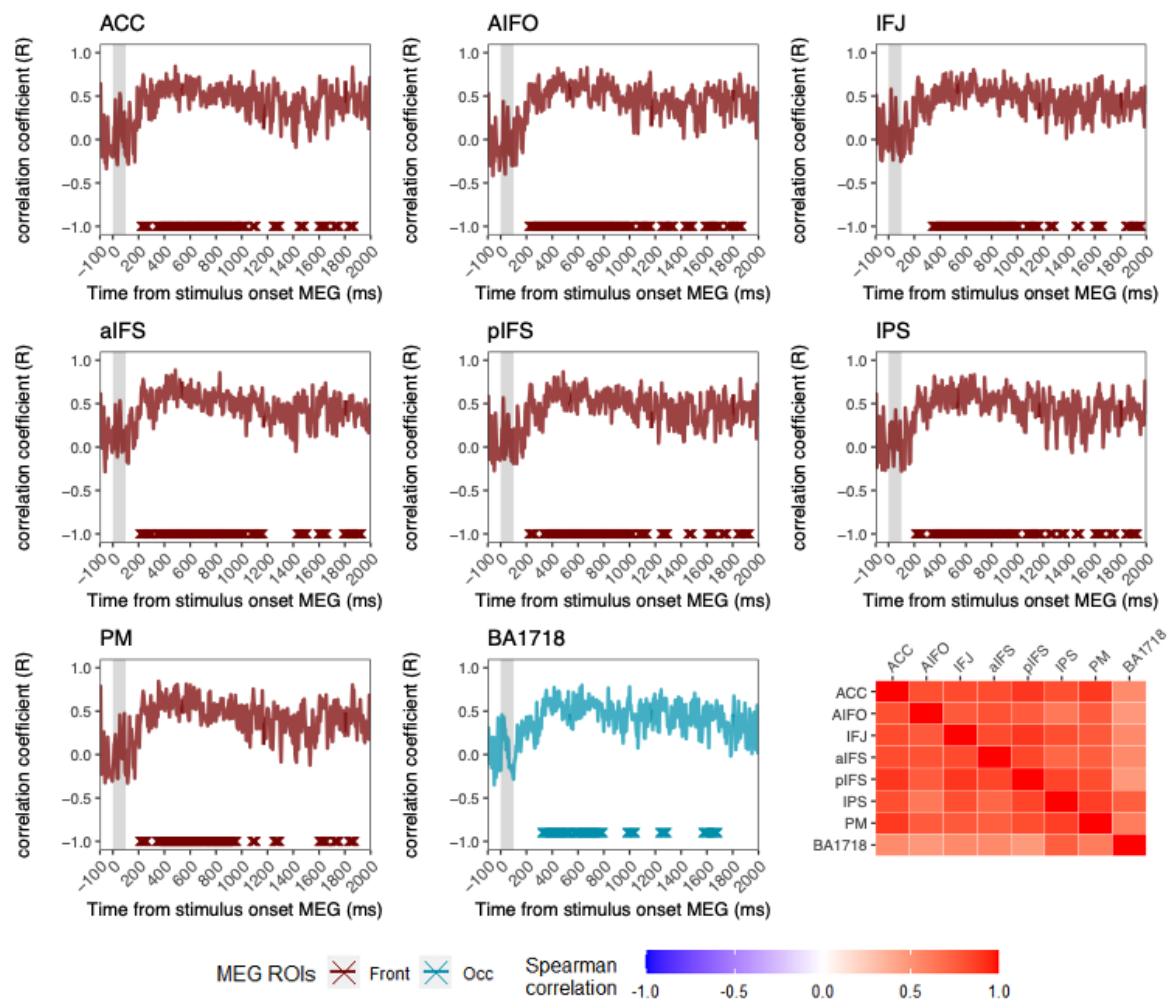


Figure 32. The timecourse of colour decoding in each ROI, based on MEG-fMRI fusion. Spearman correlation coefficients, describing the fit between the fMRI and MEG data at each timepoint, are plotted across time. The shaded panel represents the period during which stimuli were displayed. Crosses at the base of the line plots represents significant clusters of correlation coefficients after TFCE for multiple comparisons. The final plot depicts the degree of Spearman correlation across the MD decoding accuracy vectors. The pattern decoding accuracies from each MD region showed a sustained correspondence with frontal MEG data from approximately 200ms after stimulus onset for all MD regions, except for IFJ where this commenced from 300ms. For visual regions, the pattern of decoding accuracies for BA17/18 matched occipital MEG data from approximately 300ms.

For shape decoding (Figure 33), the fusion profiles were similar but the MEG-fMRI correspondence was weaker and fewer timepoints reached significance. For most frontal MD regions (i.e., ACC, AI/FO, aIFS, PM), the onset of significant commonality between the imaging modalities was around 600ms, with the exception of pIFS and IFJ where there was significant correspondence from around 350 and 400ms respectively. For visual regions, the pattern of decoding accuracies for BA17/18 matched that of the MEG occipital ROI earlier, from approximately 250ms. For shape decoding, therefore, there was no evidence that selective processing of attended information arose in the MD system before the visual system. In the MEG data, significant feedback from frontal to occipital cortex was observed from 185ms (Goddard et al, 2022). This suggests that the fMRI results for both visual and MD regions appear to primarily reflect a pattern of information coding that occurred later in time, perhaps after attention effects had already started to modulate task-related representations throughout the brain.

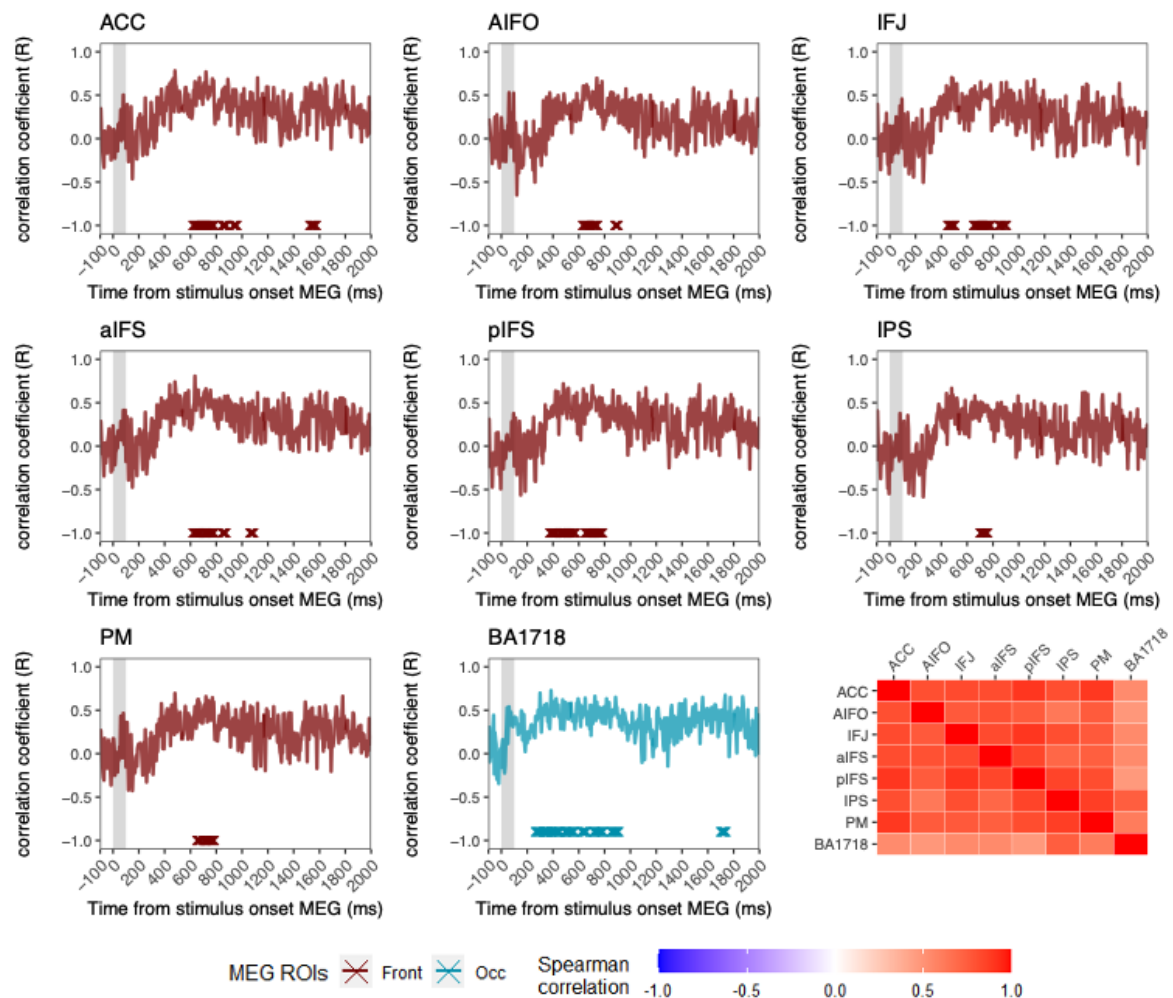


Figure 33. The timecourse of shape decoding in each ROI, based on MEG-fMRI fusion. Spearman correlation coefficients, describing the fit between the fMRI and MEG data at each timepoint, are plotted across time. The shaded panel represents the period during which stimuli were displayed. Crosses at the base of the line plots represents significant clusters of correlation coefficients after TFCE for multiple comparisons. The final plot depicts the degree of Spearman correlation across the MD decoding accuracy vectors. The pattern decoding accuracies for MD regions ACC, AIFO, aIFS, IPS, PM showed sustained correspondence with frontal MEG data from approximately 600ms. This was earlier for MD regions pIFS (350ms) and IFJ (400ms). For visual regions, the pattern of decoding accuracies for BA17/18 matched occipital MEG data far earlier than seen for MD and frontal MEG data, from approximately 250ms.

3.6 Discussion

In this study we used fMRI to understand how spatial and feature attention interact to affect task-related representations in MD regions and the visual cortex. We found that spatial and feature attention interacted multiplicatively. The MD system was highly specific in its representation of task-related information, representing only information about the attended feature of the attended object. We found similar specificity in task-related coding in visual regions. We found no evidence of irrelevant information decoding in MD or visual ROIs, or in our whole-brain searchlight, even when that irrelevant information pertained to attended objects or task-relevant features.

Our data align with recent inverted encoding model work in fMRI showing stronger and more reliable reconstruction of only the task-relevant feature of the attended object in visual working memory (Yu & Shim, 2017) and attention paradigms (Chen et al., 2021). In that work, only the task-relevant feature of the attended object could be reconstructed in frontal, parietal and occipital cortices during a task where participants were asked to recall either the colour or orientation of one of two presented gratings after a delay (Yu & Shim, 2017). Similarly, only the task-relevant feature of the attended object could be reliably reconstructed from visual areas and IPS during a task where participants reported the colour or orientation of one of three gratings (Chen et al., 2021). Interestingly, Chen et al. (2021) found some representation of the irrelevant feature of the attended object in intraparietal regions, perhaps pointing to a spatially global enhancement by feature attention, but reconstruction of these irrelevant features was weak and inconsistent. This study only examined task-related representations in occipital and posterior parietal regions. Our study corroborates and extends these studies by demonstrating the dominance of task-relevant features in frontoparietal regions in an attention task that involves novel objects.

Our findings also extend recent MEG work which reported the multiplicative effect of spatial and feature attention on decoding of object information in this and a similar task (Barnes et al., 2022; Goddard et al., 2022). As well as replicating this effect in a new modality, our results provide spatial information. Our data suggest strong dominance

of task-relevant information in MD regions, as predicted for regions strongly engaged in difficult tasks (Assem, Glasser, et al., 2020; Fedorenko et al., 2013) and which are proposed to adjust their responses to carry the most relevant information (Erez & Duncan, 2015; Jackson et al., 2017; Woolgar, Williams, et al., 2015). Moreover, in our searchlight analysis, which is free from a priori spatial hypotheses, task-relevant information was most strongly coded in the MD system. However, our data also suggest that the multiplicative effect of these two types of attention is not limited to the MD cortex, but can be seen in widespread areas of the brain including in visual regions, perhaps reflecting the dominance of attended information across the system.

Time-resolved data also show initial decoding of all stimulus information (relevant and irrelevant) that is not reflected in our results. In time-resolved data, stimulus information that is ignored (Goddard et al., 2022; Moerel et al., 2021, 2022) or not used to drive behaviour (A. K. Robinson et al., 2022) is typically encoded briefly. For instance, Goddard et al. (2022) showed brief initial coding of all stimulus information in occipital regions before resolving to a sustained preferential coding of only task-relevant information from around 400ms. Similarly, other work exploring the temporal dynamics of attention effects suggest an earlier dominance of task-related representations in frontoparietal regions, that later comes to dominate processing in visual regions (Hebart et al., 2018). The coarse temporal resolution of fMRI may have precluded us from seeing these early neural effects. Thus, we suspect that the widespread and exclusive representation of task-relevant information seen in this study reflects more sustained attentional processes whereby MD regions, and perhaps the neural system more broadly, ultimately prioritise a highly specific task-relevant representation.

These results are seemingly at odds with object-based and feature-based attention theories (e.g., Duncan, 1984; Sàenz et al., 2002). For the former, an object-grouping effect would be expected, such that attention paid to a particular feature of an object facilitates processing of other, unattended, features of the same object. Commensurate with this, behavioural studies indicate better performance when reporting multiple features of a single object than of different objects, even when accounting for spatial

selection (Duncan, 1984), and greater interference from distractors in the same object grouping than across different object groupings (e.g., Egly et al., 1994). Neuroimaging work also points to enhanced representation of all attributes of the attended object (Barnes et al., 2022; Jiang et al., 2016; O’Craven et al., 1999; M. A. Schoenfeld et al., 2014). In our case, we would expect an object-based effect to amount to greater representation of unattended features of the target object, relative to attended and unattended features of the unattended object. Our findings are inconsistent with this prediction: we found no evidence of coding of the unattended feature of the attended object. Moreover, our data do not support feature-based accounts, whereby a spatially global representation of all attended feature information would be predicted (Jehee et al., 2011). We found no evidence of neural coding of the attended feature of the unattended object.

Recent work by Barnes et al. (2022) provides an alternative lens through which to reconcile the apparent conflict between our results and predictions stemming from these theories. Barnes et al. (2022) utilised a selective attention task that varied in its demands, yielding a more complex picture of the specificity of task-related representations in the brain. The less demanding version of their task resulted in data that fit with predictions of object-based attention, with strong coding of both relevant and irrelevant feature information of the attended object. When discriminations were made more challenging with multiple competing objects, however, they observed the multiplicative effect. Relatedly, other studies point to weaker representation of distractor information (W. Zhang & Luck, 2009) and selective prioritisation of attended visual stimuli (Woolgar, Williams, et al., 2015) when stimuli are more difficult to perceive. In line with such observations, load theory (Lavie, 2005; Murphy et al., 2016) posits that load (e.g., increasing perceptual difficulty) affects the degree to which distractors are processed. Together, these findings suggest a potential explanation for why our findings might diverge from what is predicted by object-based accounts: the brain may have different ways of operating depending on the demands of the presenting challenge. Our attention task closely resembled the more demanding attentional task used by Barnes et al. (2022) and, similar to Barnes et al. (2022), we did not see an advantage for the irrelevant feature of the attended object. Our findings may therefore

reflect the context of competition between competing stimulus information in our task, culminating in only task-relevant information being selected for processing, and a different result may have obtained if the task were easier. We also cannot rule out that the brain may store irrelevant information differently to relevant information, potentially involving mechanisms that are less visible to neuroimaging (Barbosa et al., 2021; Stokes, 2015).

As secondary research questions, we also asked how the representation of task-relevant information was structured in MD regions. Our task involves stimuli which varied across multiple feature steps. We have previously found that coding of attended information in MD regions is stronger when stimuli are physically more similar, and therefore more difficult to perceive (Woolgar et al., 2011; Woolgar, Williams, et al., 2015) (although note (Wen et al., 2018)). This might predict that stimuli closer to the decision boundary would be the most decodable, in a reversal of the typical distance-to-bound effect observed in visual cortex (Ritchie & Carlson, 2016). Indeed, in our data, pairs of stimuli close to the decision boundary (2v3) could be discriminated more readily than pairs of stimuli far from the decision boundary (1v4), even though the physical difference between stimuli 1 and 4 was greater. However, rather than being intermediate, medium physical discriminations (i.e., comprising stimuli two feature steps apart, 1v3 and 2v4) were the most discriminable of all combinations. Therefore, a single axis reflecting the physical discriminability of the stimuli (or its inverse) could not account for the results.

Instead, we found that the representation of task-relevant information was captured by two organising axes. One reflected the stimulus categories (e.g., red vs green) and the other, the difficulty of the perceptual decision (easy vs hard). We obtained this result despite modelling out differences in RT at the single trial level (Grinband et al., 2008; Henson, 2007; Todd et al., 2013; Woolgar et al., 2014). Our findings here fit within a body of literature showing that prefrontal cortex (PFC) responses reflect the behavioural relevance of stimuli as much or more than their physical properties (e.g., Cromer et al., 2010; Erez & Duncan, 2015; Kadohisa et al., 2013; Sakagami & Niki, 1994; Wisniewski et al., 2023), and work emphasising high-dimensionality in PFC patterns that are related to behaviour (Badre et al., 2021; Bartolo et al., 2020; Rigotti et al., 2013). Our data add

to this picture by confirming that MD responses reflect both an abstracted representation of difficulty that generalised over stimulus information, and representation of categorical colour information (which in this case equates with decision, although not motor response), generalised over differences in difficulty.

As a final step, we related our fMRI results to extant MEG data to estimate the timecourse over which the recorded attentional effects arose in MD and visual cortices. We found statistical commonality between the signals recorded from fMRI and MEG, with fMRI signals in both MD and visual regions tending to reflect sustained activity in MEG (most consistently between 400-700ms after stimulus onset). For colour, we found that the multiplicative effect arose earlier in MD regions (around 200ms after stimulus onset) than in visual regions (around 300ms), consistent with a top-down feedback account. Moreover, these timings matched the timecourse with which patterns of activity in frontal lobe Granger-caused patterns in visual cortex in the MEG data (Goddard et al., 2022). However, for shape, the match between fMRI and MEG data was weaker, particularly for MD regions, with the onset of significant commonality later for MD cortex (ranging from 350 to 600ms) than for visual regions (around 250ms). Accordingly, our results do not confirm consistently or clearly whether selective processing of attended information arose in the MD system before or after the visual system.

A limitation of the MEG-fMRI fusion approach is that it can only find different commonality timecourses for different brain regions if the patterns of activation recorded in those brain regions are distinct. Since MD and visual fMRI results were similar in our data, correlating these patterns with whole-brain MEG data would have yielded commonality traces that were necessarily similar as well. To alleviate this, we compared our fMRI data to source-space estimated MEG data from two lobes (frontal and occipital). Nonetheless, this analysis was still limited by the extent to which frontal and occipital MEG patterns were correlated with one another. In future work, MEG-fMRI fusion could be made more fruitful by comparing timecourses of distinct patterns of information. Detailed analysis of the relative timecourses of similar information arising in different brain regions may require acquisition of data that are truly (simultaneously) resolved in both space and time (e.g., multi-unit recordings). In

addition, using methods like transcranial magnetic stimulation (TMS) to perturb the system and then observe its impact on neural coding and behaviour will be crucial to elucidate the causal role of MD regions in facilitating goal-directed attention and feedback processes.

3.6.1 Conclusions

In this study we used fMRI and MEG-fMRI fusion to understand how spatial and feature attention interact to affect task-related representations in MD regions and elsewhere in the brain. Our data strongly suggest that in the context of interacting spatial and feature attention in a difficult perceptual categorisation task, the brain is highly specific in its representation of task-related information, representing only information about the attended feature of the attended object. Second, we found that MD representations of the selected information reflected both categorical task-relevant stimulus decisions and task difficulty. Third, we found that while the effect was strongest in MD regions, the multiplicative attention effect was widespread throughout the brain, and applied even to visual regions. While we sought to estimate timecourses for these regions to elucidate when attentional effects arose in these regions, our MEG-fMRI fusion did not provide a clear picture of when and where multiplicative effects first arose.

In summary, our data shed light on the interacting effects of spatial and feature attention, a context more closely resembling the way different subtypes of attention are used in real life. Rather than boosting processing of whole objects or relevant features across space, as might be predicted when examining the effects of spatial and feature attention alone, neural activity in our data appeared to reflect all-or-nothing tuning to behaviourally relevant information. These results emphasise the importance of considering interacting attentional demands to understand the mechanisms underpinning selective attention.

3.7 Afterword

The multiplicative attention effect in this study was highly selective to task-relevance and very pronounced across the brain, which was somewhat surprising. However, we reasoned that the task was difficult, and behaviourally relevant information is more likely to be represented in MD regions when a task is more difficult. Moreover, this result accords with the selective and sustained representation of relevant information seen across the brain in Goddard et al. (2022) at a later time point; fMRI, with its sluggish response, is more likely to capture more sustained neural activity.

However, this widespread selective representation of relevant information in my fMRI results made fusion with the extant MEG data more difficult. I did not find clear evidence that MD regions consistently represented task-relevant information before domain-specific regions, which would be consistent with a biased competition account. Rather, the timecourse for MD and visual regions were often similar.

In retrospect, however, it is unsurprising that this analysis could not delineate clearly between the timecourses of MD and visual regions. To relate the spatial information of fMRI to the timecourse of MEG, I correlated the representation of information in MD fMRI data with that of frontal MEG data, and similarly for visual fMRI data and occipital MEG data. However, our MD and visual fMRI results were similar, with the multiplicative effect strongly evident in both. Thus, the fusion results could only really point to when in time we see this multiplicative effect in the brain, rather than providing a more nuanced spatial or temporal account of how and when this effect manifested. Accordingly, I find the utility of fMRI-MEG fusion in similar studies to this one to be limited.

Overall, these results should be considered within the wider literature, where there is converging evidence that MD regions selectively prioritise behaviourally relevant information. Moreover, this work adds to this literature by demonstrating that MD neural codes do not just reflect task difficulty, but also capture information related to task itself.

3.8 Supplementary Materials for Chapter 3

3.8.1 Searchlight methods

Our MVPA ROI results clearly showed evidence of a multiplicative interaction between spatial and feature attention on stimulus coding, such that attended feature information of the attended object was selectively enhanced in both MD and visual regions. To identify any additional brain regions which also showed this effect, we re-ran our decoding analysis using a roaming searchlight (Kriegeskorte et al., 2006). Since the data used in a searchlight analysis overlaps with that of the ROIs, we would expect some convergence between the two results. However, the benefit of a searchlight allows us to determine whether any additional regions show similar effects, when looking at information coding on a finer spatial scale.

For each participant, data were iteratively extracted from a sphere centred on each voxel in the brain of radius 3 voxels (~10mm). As in our previous work (e.g., Jackson & Woolgar, 2018), a linear support vector machine was trained and tested as for our ROI analysis previously, but using data from each sphere. The classification accuracy value for that sphere was then assigned to the central voxel. This yielded whole-brain accuracy maps for each individual, for each attention condition. To allow for group-level analysis, accuracy maps were normalised using normalisation parameters extracted at the preprocessing stage, and smoothed using an 8 mm full-width at half-maximum (FWHM) Gaussian kernel. A one-sample t-test was conducted at each voxel to identify where classification was significantly above chance. The results were family-wise error (FWE) corrected at $p=0.0000001$, with an extent threshold of 20 voxels. All coordinates are given in MNI152 space (McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, QC, Canada).

3.8.2 Searchlight results

Our searchlight results indicated that decoding of the attended stimulus information was widespread (Table S2 and Figure S 3). While our significant decoding clusters overlapped with MD regions (see Figure S 3), it extended across most of frontal and parietal cortex. In contrast, the same stimulus information was not decodable when it was not relevant for the task (aLuF, uLaF and uLuf conditions), even at a low threshold of $p < 0.001$ uncorrected.

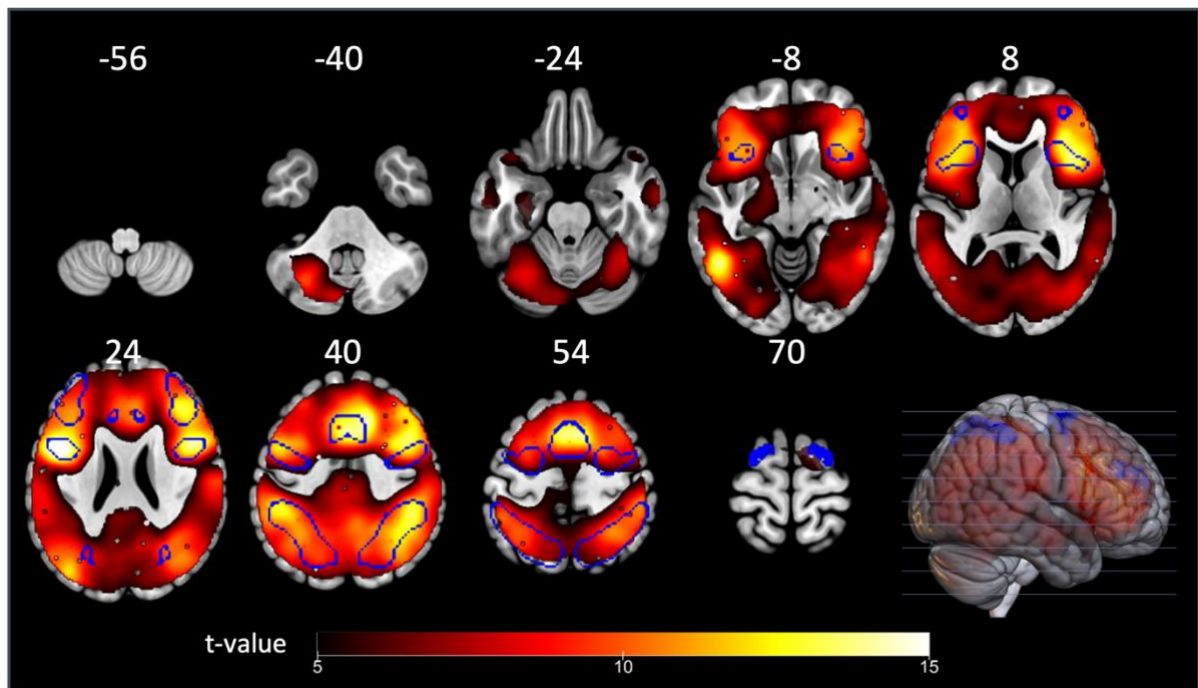


Figure S 3. Searchlight decoding results for task-relevant (aLaF) stimulus information only. Outline of MD regions are overlaid in blue. Results shown are thresholded at FWE, $p = 0.0000001$, t value minimum = 5.00. No stimulus information was decodable for other attention conditions.

Table S2. Searchlight results

Attention condition	Cluster	Hemisphere	Coordinates			BA	t-score	Cluster Extent
			x	y	z			
aLaF	Paracingulate, cingulate (anterior), frontal pole, frontal gyri,	Peak is L, but bilateral cluster	-48	6	26	4, 6, 8, 9, 10, 11, 24, 32, 38, 43, 44, 45, 46, 47, 48	15.96	29032
	Involves supramarginal gyrus, angular gyrus, some cingulate (more posterior), precuneus, tiny bit of left temporal gyri (mostly middle) and lateral occipital cortex, some cerebellum (left)	Peak is L, but bilateral cluster	-50	-56	-8	1, 2, 3, 4, 5, 7, 18, 19, 20, 21, 22, 23, 26, 37, 39, 40, 41, 42, 43, 48	14.05	20565
	Right temporal regions: inferior, middle temporal gyri, temporal occipital fusiform cortex,	Right	50	-50	-10	19, 20, 21, 22, 37	10.08	2063
	Right cerebellum	Right	-6	-80	-38	N/A	8.18	29

The table shows peak decoding for each attention condition separately. There were no voxels showing significant decoding of any attention condition comprising unattended information. Results are FWE corrected at $p=0.0000001$, with an extent threshold of 20 voxels.

3.8.3 MD decoding across feature step size, across decision boundary only

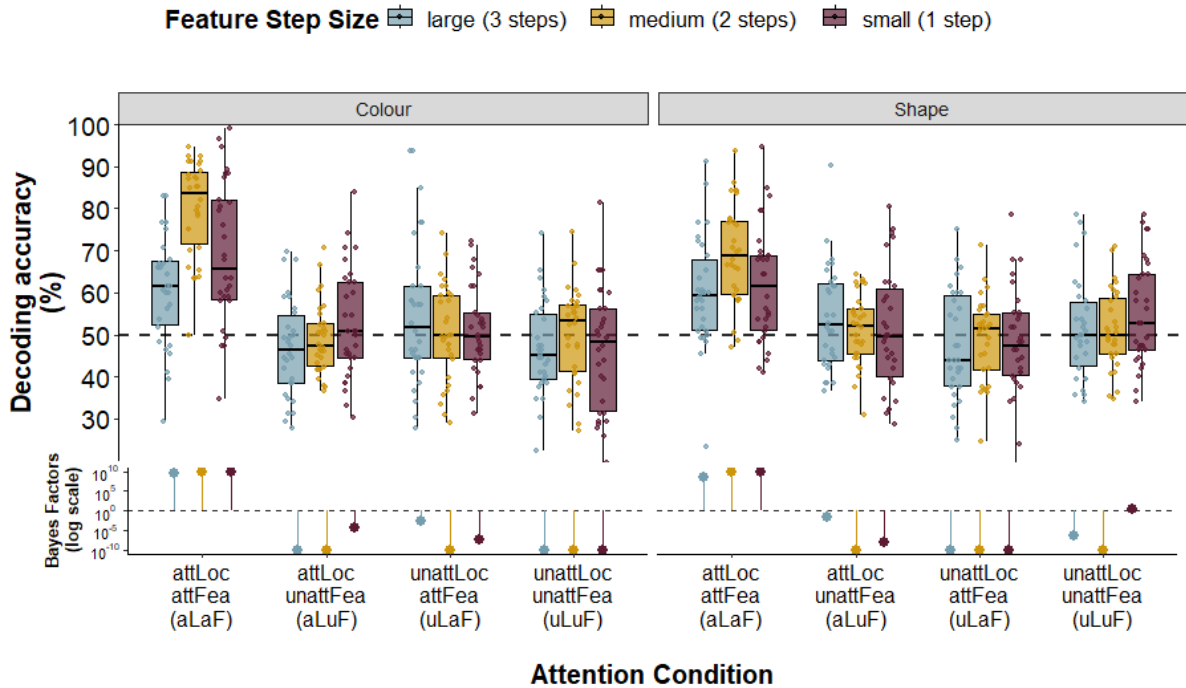


Figure S 4. Average decoding accuracy of colour (left panel) and shape (right panel) information in each of the spatial and feature attention condition for each feature step size, collapsed across all MD regions. In contrast to Figure 27, feature step size comparisons in this plot comprise only comparisons that cross the decision boundary. This represents a difference for the small feature step size (maroon bars) only. That is, in this plot small comprises pairwise comparisons 2v3 (one feature step size). Medium (pairwise comparisons of two feature step sizes: 1v3, 2v4), and large (pairwise comparisons of three feature step sizes: 1v4) feature step sizes are as defined for Figure 27. Scatter depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). The strength of evidence for above-chance (H_1) or at-chance (H_0) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

3.8.4 Visual cortex decoding across feature step size, across decision boundary only

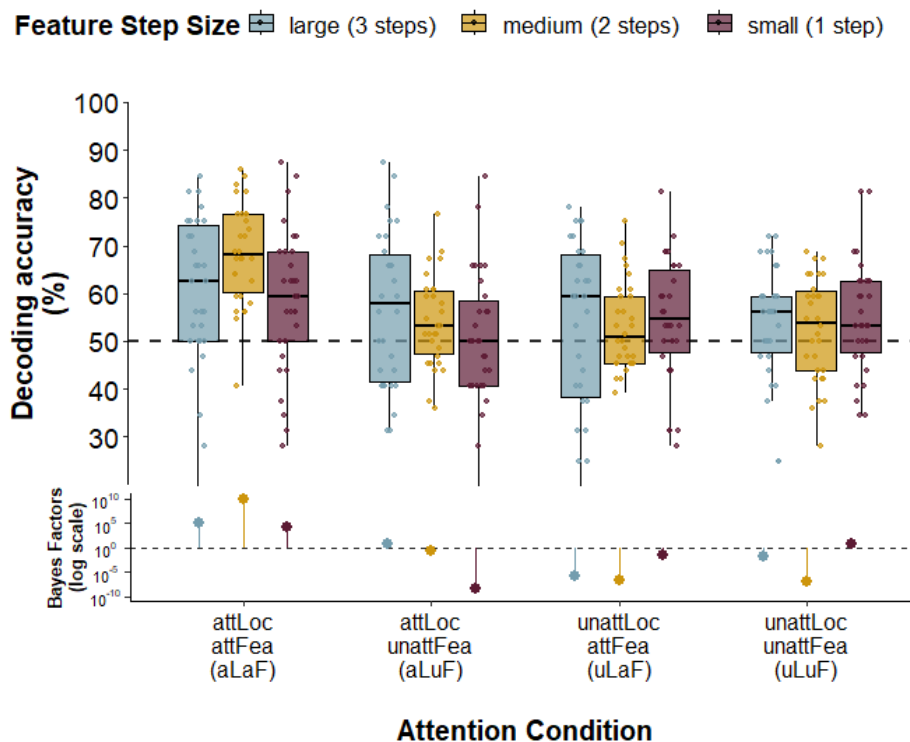


Figure S 5. Average decoding accuracy of stimulus information (averaged over colour and shape information, as not significantly different between these types of information) for visual cortex (BA17/18). In contrast to Figure 3I, feature step size comparisons in this plot comprise only comparisons that cross the decision boundary. That is, in this plot small feature step size comprises one feature step (pairwise comparison 2v3), medium feature step size comprises two feature steps (pairwise comparisons 1v3, 2v4), and large feature step size comprises three feature steps (pairwise comparison 1v4). Again, while attended information (aLaF) was more decodable than other attention conditions, inclusion of feature step size in this interaction did not explain additional variance (that is, interaction between feature step size and attention conditions was not significant). Scatter depicts decoding accuracy of each participant for the given condition. Boxplots indicate median decoding accuracy for the given condition (central line), box hinges indicate first and third quartiles, and box whiskers range from hinge to largest and smallest decoding accuracy value (that is not greater than 1.5 x applicable quartile range). The strength of evidence for above-chance (H1) or at-chance (H0) decoding was tested using Bayesian t-tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

3.8.5 Cross-classification analysis, visual regions

As stated in our main analyses, our results for visual cortex did not point to a differential effect of selective attention across stimulus space, as was evident in MD cortex. However, for completeness, we asked whether the organisation of feature information in visual cortex was similar to that seen in MD regions.

We ran the same leave-one-run-out cross-classification decoding analysis to decode difficulty that was described earlier, but for visual cortex. For example, to decode difficulty for the colour task, we trained our classifier to distinguish between two features comprising one decision class but two levels of difficulty (e.g., colour one (strongly red, easy) from colour two (weakly red, hard)). We then tested this classifier to distinguish between the two features that lay on the other side of the decision boundary (e.g., colour four (strongly green, easy) from colour three (weakly green, hard)). This procedure was repeated in the reverse (i.e., training to distinguish between easy green and hard green, and testing on easy red and hard red), and separately for shape information. This process allowed us to examine whether we could decode ‘difficulty’ (i.e., ‘easy’ from ‘hard’ stimuli) from our data, independent of feature category, for colour and shape.

While the effect was not as striking as seen for MD regions (i.e., decoding of difficulty was not consistent all features and regions, nor was the strength of evidence in favour of above-chance decoding as strong as for MD regions), our results showed that difficulty also acted as an organising principle for the representation of stimulus information in visual cortex (see Figure S 6). Bayesian t-tests revealed strong to very strong evidence of above-chance generalisation of difficulty coding across colour categories and across shape categories in lateral occipital complex (LOC) and across colour categories in BA17/18 (mean accuracy range = 57.8-62.8%; BF10s > 17.46). There was anecdotal evidence that difficulty did not generalise over shape categories in BA17/18 (mean accuracy = 53.8%, BF10=0.37).

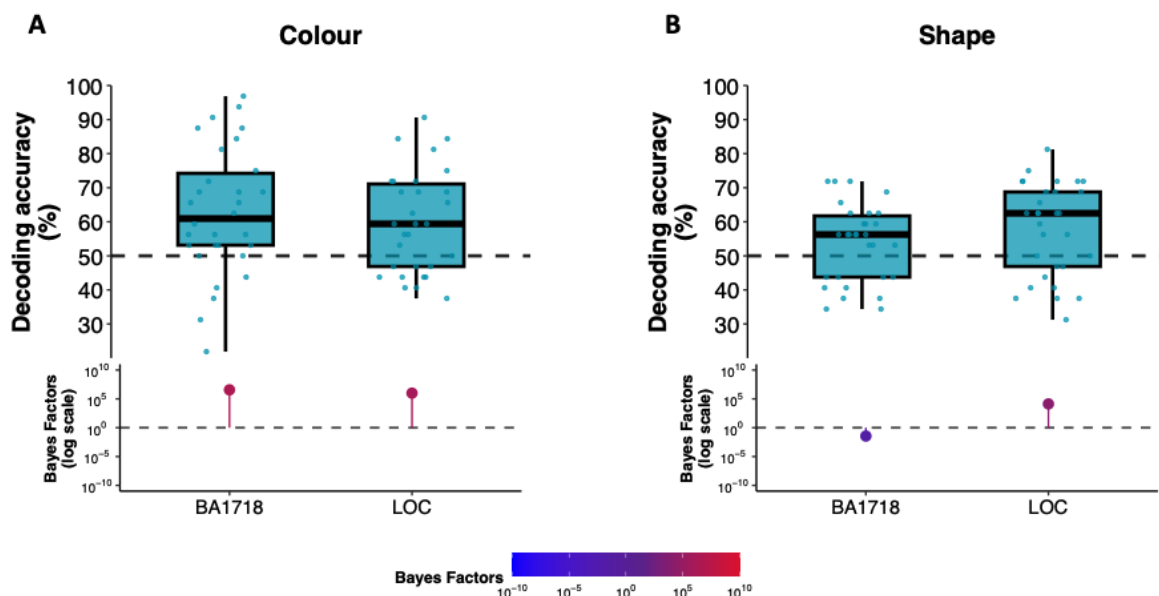


Figure S 6. Decoding of difficulty in visual regions (collapsed across hemisphere), generalised over the two colour categories (Panel A) and shape categories (Panel B). For task-relevant information (i.e., aLaF) only. 'Difficulty' here represents decoding of 'easy' from 'hard' stimuli, and vice versa, independent of feature category, for colour and shape separately. The strength of evidence for above-chance (H_1) or at-chance (H_0) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

We also tested whether we could decode the feature categories (i.e., for colour: red or green, for shape: X-shaped or non-X-shaped), independent of difficulty level. As per the analysis for MD regions, we trained our classifier to distinguish between two features matched for difficulty but on opposite sides of the decision boundary (e.g., colour one (strongly red, easy) from colour four (strongly green, easy)). We then tested whether this classifier could distinguish between the remaining two features (e.g., colour two (weakly red, hard) from colour three (weakly green, hard)), and vice versa. We repeated the same procedure for shape information. As shown in Figure S 7, we found moderate evidence for generalisable coding of shape feature categories in visual regions (mean accuracy range = 55.7-56.3%; $BF_{10s} > 5.12$), but very strong to extreme evidence for the null in the equivalent analysis for colour feature categories (mean accuracy range = 51.4-51.8%, $BF_{10s} < 0.01$). This was the opposite of the results we found in MD regions overall,

where we found very strong evidence for coding of colour categories, but not shape, over difficulty level.

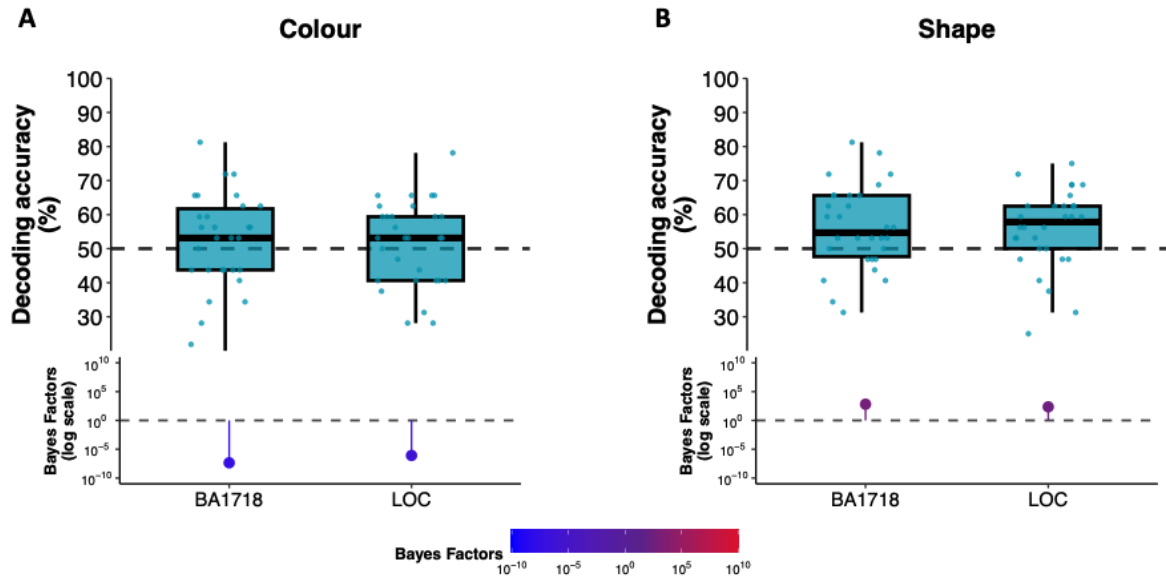


Figure S7. Decoding of colour category information (Panel A) and shape category information (Panel B) in visual regions (collapsed across hemisphere), generalised over difficulty. For task-relevant information (i.e., aLaF) only. The strength of evidence for above-chance ($H1$) or at-chance ($H0$) decoding was tested using Bayesian t -tests and is shown here as Bayes Factors for each of these conditions on a logarithmic scale.

Chapter 4 |

How do focal brain lesions affect information coding during selective attention?

4.1 Foreword

In Chapter 3, I demonstrated that stimulus representations in the multiple-demand (MD) network are highly selective to the task at hand. This adds to a large base of correlative evidence that frontoparietal regions selectively represent information at the focus of attention, which may in turn drive focussed representations elsewhere in the brain. However, we are far from being able to make causal claims about the role of the MD network in this process. For that, we need to see how processes change when the brain is perturbed or damaged. In this study, I aim to get closer to the goal of a causal

functional account by using a focal lesion approach combined with neuroimaging. I use the selective attention paradigm and methods of Chapter 3 and apply them to magnetoencephalography (MEG) data collected from focal lesion patients and healthy age-, gender- and education-matched controls. My aim is to explore whether damage to parietal lobe, thought to play an important role in spatial attention, affects the process of attentional prioritisation.

There were several factors to consider in setting up this study. First, the selective attention paradigm is demanding, and I was concerned that patients and older controls would not be able to complete this task to the level of accuracy required for my planned multivariate pattern analysis (which was conducted using correct trials only). Accordingly, to reduce task demand, I split the task across two separate MEG sessions (described in Section 4.4.4). However, this then begged the question of whether my planned MEG analysis could be conducted from data acquired across two separate sessions. To ensure this task could be completed by patients and older controls, and as an initial test of whether my decoding analysis could be run using data collected across separate sessions, I performed behavioural pilots with two focal lesion patients (who were otherwise not eligible for the main study) ($N=2$) and MEG pilots with older control participants ($N=7$). My pilot participants were able to complete the task to a reasonable level of accuracy ($>70\%$, data not shown), suggesting that this paradigm was suitable for my intended cohort. I developed my analysis pipeline using my MEG pilot data (pipeline described from Section 4.4.6.1) and sense-checked my methods for combining data from across two sessions. For instance, in volumetric source space, I checked whether visually evoked responses and motor evoked responses could be localised to expected regions (i.e., spreading from early visual areas, and left motor areas in a right-handed participant, respectively). These initial checks indicated that my approach for combining data across two sessions was reasonable.

A combination of COVID-19 disruptions and personnel changes in the management of the focal lesion patient panel at the MRC Cognition and Brain Sciences Unit meant that few suitable patients were available to participate in my study at the time of recruitment. I had initially planned to recruit large numbers of patients with lesions

affecting MD regions to different extents so that I could relate neural and behavioural profiles to lesion size and location, but this was not possible. Instead, I focussed my study on parietal lobe as an initial proof-of-concept. My inclusion criteria were broadly defined: while all patients have parietal lesions, the volume and extent of these lesions varies dramatically and are by no means restricted to parietal lobe. Of course, variation in lesion location, and indeed in the effect lesions can have on cognition and behaviour, is a common limitation in lesion studies, and also limits the possible conclusions here. In addition, very few studies have used multivariate methods to analyse neural data from focal lesion patients, and to my knowledge, none have done so with MEG. As I was unsure how factors associated with lesions (e.g., size of lesion, skull damage due to craniotomy) would affect source space reconstruction, I have restricted the analyses in these chapter to sensor space.

As such, this study represents a proof-of-concept investigation, where I explore the dynamic coding of task-related information at an individual level, and test whether strength of rule or relevant information coding differs in individual patients relative to matched controls. I hope that this work lays the groundwork for a larger study in the future which will have more power to study these effects in more detail.

4.2 Abstract

Neuroimaging data implicate frontoparietal cortex in prioritising coding of attended information, while damage to these regions (e.g., from stroke) leads to altered attentional behaviour. A key proposal is therefore that selective information coding in frontoparietal cortex is critical for driving brain-wide focus. If so, we should observe altered patterns of neural prioritisation in patients with frontoparietal brain injury. To test this, I used a selective attention paradigm with magnetoencephalography (MEG) to examine attentional prioritisation dynamics in patients with focal brain lesions. Six patients with chronic unilateral parietal lesions (three male, three female; three left, three right hemisphere; aged 55-75), and ten matched controls (six male, four female; aged 53-76), covertly attended to one of two bilaterally presented objects, and reported the attended object's colour or shape via button press. Using multivariate pattern analysis, I quantified the neural prioritisation of relevant over irrelevant visual information in each individual. Five of the six patients showed neural evidence of attentional prioritisation, with task-relevant information selectively prioritised at a level similar to controls, in accordance with their high behavioural accuracy. However, patients were slower than controls, and inter-individual differences in reaction time were related to inter-individual differences in attentional prioritisation strength in sensors over occipital regions, providing preliminary evidence linking strength of neural prioritisation to behavioural performance. The remaining patient showed more overt neural differences, including less effective prioritisation in sensors over parietal and frontal lobe of the affected hemisphere and occipital regions in both hemispheres compared to controls, but was still able to do the task. Patients with parietal lesions can successfully prioritise attended information in brain and behaviour, but subtle deficits in neural prioritisation may be reflected in slower behavioural performance.

4.3 Introduction

Our capacity for goal-directed behaviour depends on selective attention: our ability to prioritise certain (relevant) inputs for further processing to the exclusion of others. Converging evidence implicates a core set of frontoparietal regions, referred to as the multiple-demand (MD) network (following Duncan, 2010, 2013), in driving flexible and selective processing of task-relevant information (Duncan et al., 2020; Zheng et al., 2024). These regions have been shown to code many different task features, including auditory, visual, motor and rule information (Woolgar et al., 2016). Moreover, these representational codes in MD regions adjust to preferentially represent task-relevant information in accordance with task demands, including spatial attention (e.g., Woolgar, Williams, et al., 2015) and feature attention (e.g., Jackson et al., 2017). Chapter 3 showed that these representational codes are highly specific, prioritising coding of only information that lies at the intersection of spatial and feature attention.

As reviewed previously in this thesis, preferential coding of the MD system is then thought to direct focus on task-relevant information across the brain, reweighting the influence of incoming information streams to favour behaviourally relevant information (Desimone & Duncan, 1995). For example, a recent magnetoencephalography (MEG) study (Goddard et al., 2022) showed that highly selective representations of task-relevant information in frontal lobes Granger-caused representations in occipital cortex to become biased in favour of task relevance. Chapter 3 further showed that MD regions may be reasonable candidates for the source of bias in this task. However, the majority of this evidence is correlational – even Granger-causality is time-lagged correlation – and causality cannot be inferred.

In contrast, research involving focal lesion patients can elucidate causal relationships, by demonstrating the necessity of particular brain regions for particular neural and cognitive functions. For example, unilateral damage to (particularly inferior) parietal cortex is often associated with unilateral inattention to the contralesional hemifield (e.g., Chatterjee & Coslett, 2003; Molenberghs et al., 2012; Mort et al., 2003; Vallar &

Calzolari, 2018). In its most striking presentation, unilateral spatial neglect, patients can demonstrate a pronounced lack of awareness for information on the contralesional side of space despite normal visual perception. Neglect is most commonly observed following acute damage to the right hemisphere (leading to left-sided neglect) and chronic forms of neglect are typically associated with right hemisphere damage (Manly & Mattingley, 2004), although neglect and significant attentional biases can also follow left hemisphere lesions (Kleinman et al., 2007; Peers et al., 2005; Stone et al., 1993).

In most cases, the most severe form of neglect resolves (Karnath et al., 2011), but more subtle spatial attention deficits can persist, particularly in the face of competing inputs. For example, in extinction, patients may show normal recognition and report of contralesional items when they appear alone yet fail to detect contralesional items when concurrently presented with ipsilesional items (i.e., the ipsilesional item ‘extinguishes’ awareness of the contralesional item). Intriguingly, these patients do seem to extract some information from the impaired side of space. For instance, extinction increases with increasing similarity between the contralesional and ipsilesional stimuli (Baylis et al., 1993), and despite lack of awareness, visual information presented contralesionally can influence behaviour (e.g., Marshall & Halligan, 1995; McGlinchey-Berroth et al., 1993). However, it is still unclear how much information from the contralesional side is processed and indeed whether ipsilesional attention is unaffected in patients with spatial bias. While some behavioural studies indicate patient performance is equal to or better than controls for their ipsilesional side (e.g., Kinsbourne, 1993), other studies suggest that patients show interference from irrelevant information when attending to this side of space (e.g., Duncan et al., 1999; Snow & Mattingley, 2006).

Therefore, while prior research indicates that damage to parietal regions can impact attention processes, the neural changes mediating this relationship are not well understood. For example, we are missing a clear description of the effects of damage on the dynamics of task-related neural representations during attention tasks. Multivariate pattern analysis (MVPA) methods permit a glimpse into the kind of information the brain processes as it completes a task, and how this information may be exchanged and modified between regions across time. In this chapter, I use MEG and MVPA methods

to explore how lateralised damage to parietal lobe affects the dynamic coding of rule information, and the prioritisation of attended spatial and feature information. I use the demanding visual attention paradigm from Chapter 3 to examine the effect of parietal damage on the representation of rule information (i.e., which side of space to attend to, which feature to attend to) and on task-related stimulus representations during spatial and feature attention in sensor space (whole brain, and in sensors over occipital, parietal and frontal regions). As an initial proof-of-concept study, I focus on exploring how damage affects coding of this information in the affected and unaffected hemisphere, and possible relationships with behavioural performance. This lays the groundwork for future larger studies which may also examine how these effects are modulated by spatial attention biases.

Despite its potential explanatory power, there is limited precedent for the use of MVPA to examine how the nature and content of neural representations change following brain damage. Only a handful of studies have utilised MVPA with functional magnetic resonance imaging (fMRI) for this purpose (e.g., Fischer-Baum et al., 2017; Lee et al., 2017; R. Li et al., 2021; Purcell et al., 2019), and these studies focus on changes to language processes post-stroke. To my knowledge, no study has applied MVPA methods to task-based MEG data collected from focal lesion patients. In addition, many of the tasks used for MVPA are cognitively demanding (necessarily so for tasks targeting the MD system), and are often lengthy (due to counterbalancing requirements, and to ensure adequate number of condition repetitions for classification), raising feasibility questions. Fatigue is a common symptom after brain injury (Cumming et al., 2016) and previous work has indicated that tasks with complex rules or multiple response buttons might prove too taxing for stroke patients (e.g., Niessen et al., 2020). Therefore, this initial study provides useful data on the feasibility of using these methods to decode task information in individual focal lesion patients and controls.

4.3.1 Hypotheses

In this study, I focus on patients with chronic unilateral parietal lesions (three left and three right hemisphere). Accordingly, I did not anticipate florid neglect, but rather a more subtle attentional deficit that might be detectable on a task involving bilaterally presented stimuli. The task I used involves attending to a conjunction of spatial and feature information, and thus allows us to probe the representation of four pieces of task-related information: relevant information (the attended feature of the target object), and three types of irrelevant information (the unattended feature of the target object, and the un/attended feature of the distractor object).

Work in Chapter 3 and from Goddard et al. (2022) has delineated a clear spatial and temporal account of how task-related information is represented in this task. I expected that older control participants, and patients for whom no attentional deficits were detected, might demonstrate a decoding trace similar to the pattern shown in Goddard et al. (2022) in which there was an initial representation of all stimulus information, followed by a sustained selective representation of only relevant information maintained across time. Therefore, in this study, I quantify the degree to which relevant information is prioritised over irrelevant information as my measure of ‘attentional prioritisation’. I quantify attentional prioritisation in sensors over occipital, parietal and frontal regions, in left and right hemispheres. For simplicity, I refer to sensors over any part of the ipsilesional hemisphere as being over the ‘affected’ hemisphere, and sensors over any part of the contralesional hemisphere as being over the ‘unaffected’ hemisphere.

Parietal damage could affect neural attentional processes in the affected and unaffected hemisphere in different ways. First, I anticipated that the affected hemisphere in patients might show reduced attentional prioritisation, relative to average of both hemispheres in controls. Reduced attentional prioritisation in the affected hemisphere could reflect weaker processing on the damaged side of the brain.

In the unaffected hemisphere, I anticipated several possibilities. Patients could show stronger attentional prioritisation in their unaffected hemisphere, relative to controls. This, combined with reduced attentional prioritisation in their affected hemisphere, could indicate a compensatory boost in attentional processing in the unaffected hemisphere (e.g., through homologous regions). Alternatively, patients could show reduced attentional prioritisation in their unaffected hemisphere relative to controls. This could reflect a widespread reduction in attentional prioritisation, commensurate with parietal damage affecting attention processes more globally. I also anticipated that patients could show stronger attentional prioritisation in their unaffected relative to affected hemisphere, compared to the equivalent difference in attentional prioritisation between hemispheres in controls. A greater difference in prioritisation between hemispheres than controls could be indicative of either a circumscribed effect of damage over the affected side, and/or a compensatory process.

Finally, I considered that I might observe a slower timecourse for the prioritisation of relevant information in patient brains. It is difficult to determine exactly what such a result would indicate. Against a background of good behavioural accuracy, this result could reflect altered processing, for example through use of a more inefficient neural pathway due to reorganisation or compensatory mechanisms, or use of alternative cognitive strategies. Alternatively, slower neural prioritisation could reflect neural inefficiencies more broadly.

Given that individual patients could align to different predictions, I examined dynamic coding of task-related information in individual participants and took a case-series approach to compare individual patients against controls.

4.3.2 Aims of this study and research questions

The aim of this study was to provide an exploratory proof-of-concept examination of how attention processing changes after damage to parietal lobe. I asked how focal parietal damage affects the content of neural representations, the timecourse over which attended information is prioritised in neural signals, and whether and how this relates to behaviour. First, I examined whether strength of rule decoding and strength of attentional prioritisation in each patient differed from that of controls using a case-series approach. Second, to examine whether individual differences in neural measures captured something meaningful about behavioural performance, I tested whether variation in the strength of attentional prioritisation was related to behavioural performance on the task.

To anticipate the results, five of six patients showed neural evidence of attentional prioritisation, with task-relevant information selectively prioritised at a level similar to controls. One patient showed less effective neural prioritisation of relevant information compared to controls. While patients were as accurate as controls, they were slower. Individual differences in neural prioritisation of relevant information in sensors over occipital regions were related to individual differences in reaction times.

4.4 Methods

4.4.1 Participants

Six patients (three male, three female; age range: 55 to 75) with unilateral, primarily parietal, chronic lesions of mixed aetiology were recruited from the Cambridge Cognitive Neuroscience Research Panel (CCNRP) at the MRC Cognition and Brain Sciences Unit (MRC CBU) at the University of Cambridge, as detailed in the following. The CCNRP is a database of volunteers with chronic focal brain lesions. A summary of lesion aetiology, location and volume for each patient is shown in Table 6 and distribution of lesion locations is shown in Figure 34.

Table 6. Patient demographic details and description of lesion

Patient	Gender	Age	Lesion aetiology	Primary regions affected	Lesion volume (cm ³)	% lesion and parietal ROI* overlap
sub-01	REDACTION: Gender, age, and lesion aetiology of six patients have been removed. Personal data removed for confidentiality reasons.			Right lateral occipital cortex, angular gyrus, middle temporal gyrus	25.84	48.68%
sub-02				Right lateral occipital cortex, superior parietal lobule, angular gyrus, supramarginal gyrus	26.54	88.09%
sub-03				Right superior parietal lobule, supramarginal gyrus, postcentral gyrus, precentral gyrus	15.38	54.18%
sub-04				Left lateral occipital cortex, superior parietal lobule, supramarginal gyrus, postcentral gyrus	31.35	48.84%
sub-06				Left superior parietal lobule, supramarginal gyrus, postcentral gyrus, parietal opercular cortex, central opercular cortex, precentral gyrus, inferior frontal gyrus, middle frontal gyrus, frontal operculum, insula cortex	122.87	17.05%
sub-07				Left lateral occipital cortex, superior parietal lobule, precuneous, angular gyrus, supramarginal gyrus, postcentral gyrus	35.8	67.65%

*Parietal region of interest (ROI) defined as a composite of Broadman Areas 1, 2, 3, 5, 7, 39, 40, 43. See Figure 34.

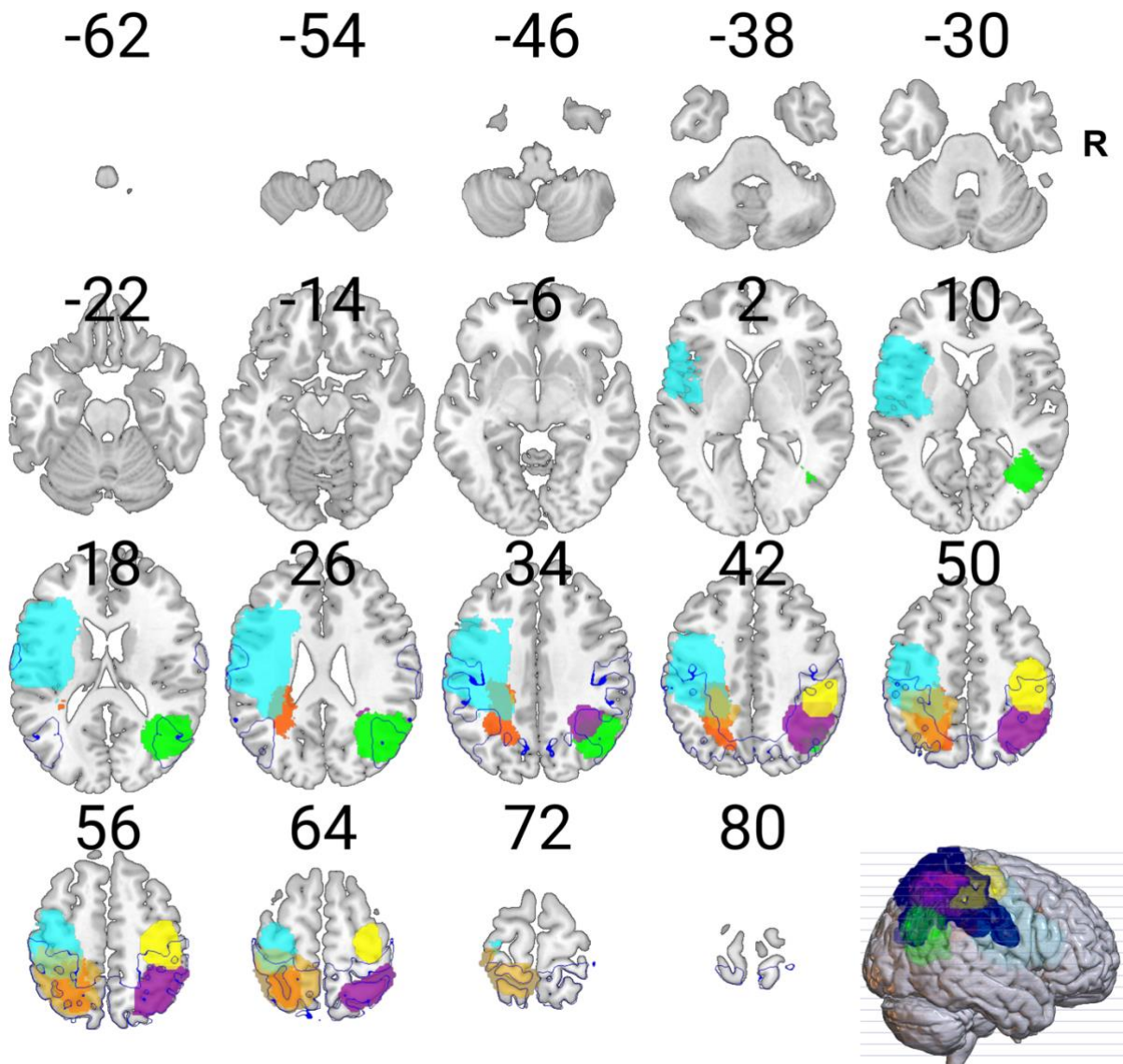


Figure 34. Lesion location for each of the six patients. Axial slices are numbered by z level in Montreal Neurological Institute (MNI) space shown in neurological convention (i.e., right hemisphere is shown on image right). Bottom panel shows overlap with parietal ROI (in blue). Patient colours are as follows: sub-01 = green; sub-02 = purple; sub-03 = yellow; sub-04 = orange; sub-06 = cyan; sub-07 = mustard. Parietal ROI used for selection of patients was defined as a composite of Broadman Areas 1, 2, 3, 5, 7, 39, 40, 43 and is shown as a blue outline on axial slices and as the darker blue patch on the rendered brain, bottom right.

4.4.1.1 Patient recruitment

I identified suitable patients on the basis of database records and contacted them directly to ascertain their interest. My criteria for determining suitability (for patients, and controls where applicable) were as follows: good colour vision, good vision (correction permissible with contacts only due to eye tracker considerations), less than 80 years of age, no irremovable metal on their person (e.g., such as titanium cranioplasty, substantial metal dental work), no significant cognitive or sensory impairments that would make participation difficult or unpleasant (e.g., hemiopia, marked fatigue), no comorbidities with other neurological conditions (e.g., learning difficulties, neurodegenerative disease). To increase my potential patient pool, I did not specify handedness as a criterion, however all recruited patients were right-handed.

I then identified possible parietal lesion patients in the CCNRP database through two methods. First, in most cases, CCNRP patient records state MRI suitability and broad regions affected by the patient's lesion (identified by the patient's medical team). I filtered all CCNRP records to identify patients who met the following criteria: a parietal lesion (as indicated on patient records), still listed as active on the panel, and no clear indication in notes that they would be ineligible for this study. From this process, 29 possible participants were identified. Of these patients, seven were contacted as they were deemed suitable given their lesion location (predominantly parietal), age (less than 80 years) and database notes which indicated they might be amenable to attend the MRC CBU for lengthy cognitive testing. Three patients were recruited through these means. Of the remaining four, one was deceased, one refused due to family obligations, one could not be contacted, and one was ineligible (colour blindness).

Second, to ensure inclusion of patients whose lesion locations might not have been fully documented in CCNRP patient records, I also identified parietal cases by comparing their normalised lesion tracings to a broadly defined parietal region of interest (ROI, covering BA areas 1, 2, 3, 5, 7, 39, 40, 43; see Figure 34) using a custom MATLAB pipeline that utilises SPM 12. Many patients on the CCNRP have an anatomical MRI that has been manually traced and normalised into Montreal Neurological Institute (MNI) space

($N=237$). From this procedure, 85 patients were identified as having any percentage of their lesion falling within the defined parietal ROI. Of these, 35 patients were no longer active on the panel (12 were deceased, 23 had left the panel or were deemed unsuitable for the panel). Of the remaining 50, 20 were deemed ineligible (due to irremovable metal, colour blindness, notes of significant cognitive or sensory difficulties, had very low parietal lobe involvement ($<5\%$) combined with other complications, had already been identified as a possible participant, or were older than 80 years of age). An additional four patients were identified as possible participants but ruled out for the time being as they were currently participating in another study at the MRC CBU that precluded them from participating in another study simultaneously. As a number of these patients had not been involved with the panel for more than five years, I first contacted the remaining 26 possible participants by letter; 10 patients responded of which eight indicated they would still like to be contacted for further research studies. Of these, for four patients, further inspection of their records suggested they may not be suitable (sensitivity to lights, complex aetiology combined with small parietal involvement, small parietal involvement, reports of difficulties with ocular tracking). I contacted the remaining four by phone, all of whom agreed to participate, bringing the total number of recruited patients to seven.

Of the seven recruited patients, one further individual (sub-05) was excluded after the home visit (Session 1, Neuropsychological Session). It was determined after further investigation that she did not have a focal brain lesion and was therefore ineligible for this study. The other six patients completed all aspects of the study.

4.4.1.2 Control recruitment

Due to heterogeneity in the patient sample, controls were chosen to closely match individual patients, rather than at a group level. Control participants were chosen carefully to match patients as closely as possible with regard to age (within three years either side of patient's age), gender, education (considering number of years of

education, highest level of educational attainment, profession). Ten age-, gender- and education-matched controls (six male, four female; age range: 53 to 76) were recruited from the MRC CBU Volunteer Panel at the MRC CBU.

Demographic details of patients and their matched controls are shown in Table 7. Each patient was matched to two control participants. Two patients (sub-03 and sub-07) were similar in terms of demographics and thus two control participants were matched to these patients. Matching of controls on the basis of education involved consideration of several education factors (e.g., number of years of education, highest level of educational attainment, profession); given the small sample size and degree of personal information involved, these details are not provided below.

Table 7. Broad demographic details of patient and control participants.

Patient	Patient gender	Patient age	Matched controls	Control gender	Control age
sub-01	REDACTION: Gender and age of six patients have been removed.		sub-51	REDACTION: Gender and age of ten control participants have been removed.	
sub-02			sub-56		
sub-03			sub-53		
sub-07			sub-58		
sub-04	Personal data removed for confidentiality reasons.		sub-50	Personal data removed for confidentiality reasons.	
sub-06			sub-55		
			sub-54		
			sub-59		
			sub-52		
			sub-57		

Before each testing session, participants gave written informed consent and were paid for their time and their travel expenses. A structural MRI scan was collected from participants who did not have an existing recent (within seven years) structural MRI scan on the panel database ($N=9$). These participants gave additional written informed

consent before completing the structural scan. Ethics approval was obtained from the Cambridge Local Research Committee (LREC) and the Psychology Research Ethics (PRE) Committee at the University of Cambridge (LREC 95/031, PRE.2002.01, PRE.2018.101).

4.4.2 Neuropsychological tasks

To characterise the general cognitive profile of participants and assess spatial bias using measures separate to the main experimental task, all participants completed several neuropsychological tests of attention and working memory as detailed in Table 8.

To measure spatial bias, participants completed the CamTrip task, a computerised extinction-like task that was recently developed at the MRC CBU. For this task, participants respond via button press to intermittently appearing targets (the letter ‘T’), while ignoring distractors (the letter ‘X’). Targets and distractors are presented left or right of fixation, or bilaterally. Compared to traditional clinical extinction tasks, CamTrip involves identification, rather than detection, of unilaterally or bilaterally presented targets, thus offering greater sensitivity in detecting and quantifying spatial bias. Moreover, CamTrip is informed by the principles of the Theory of Visual Attention (TVA), described in Peers et al. (2005), and provides outcome measures on the degree of spatial bias to left or right hemifield and top-down control.

Attention and working memory were assessed using a tablet-based application, OCS-BRIDGE (<https://ocs-bridge.com/>). OCS-BRIDGE offers a battery of validated cognitive and mood screening measures in an electronic format, including the Oxford Cognitive Screen (Demeyere et al., 2015). The selected OCS-BRIDGE tests administered in this study included measures of attention (Hearts Cancellation, Sustained Attention Lighthouse Task (SALT)), visual perception and acuity (Line Bisection Test with visual acuity screen), and working memory (forwards and backwards digit span). All tests were completed on an iPad (iPad Pro, 2020). For these tests, impairment is defined as performance falling below (or exceeding, as applicable) the fifth percentile

Chapter 4

(approximately two standard deviations) from mean performance of a healthy reference population.

Finally, to check colour vision, participants also completed the Ishihara's Tests for Colour-Deficiency (24 Plates Edition). All participants demonstrated adequate colour vision on this task.

Table 8. Neuropsychological tests and outcome measures

Test	Task description	Outcome measures used
ATTENTION		
CamTrip	Participants respond via button press to intermittently appearing targets (the letter 'T'), while ignoring distractors (the letter 'X'). Targets and distractors are presented left or right of fixation, or bilaterally	<i>Left and right maintenance</i>: proportion correct in detecting bilateral targets relative to unilateral targets (left or right, respectively). Scores < 1 here reflect poorer performance in detecting bilateral relative to unilateral targets and thus scores well below 1 may suggest extinction. <i>Bias</i>: relative extent to which performance is preserved on a particular side of space in the presence of competing target information on the other side of space. Scores close to 0 reflect strong right bias; scores close to 1 reflect strong left bias. <i>Alpha</i>: measure of top-down control, reflects extent to which non-target information can be ignored. Higher values reflect poorer top-down control.
Hearts Cancellation	Participants find visual targets distributed across the screen amongst visually similar distractors.	<i>Spatial bias</i> : difference in detection of targets on left relative to right side of space (positive values reflect left bias).
Sustained Attention Lighthouse Task (SALT)	Participants touch the screen to indicate detection of changes in the brightness of one of two clouds displayed on both sides of a central lighthouse, while simultaneously counting the number of times the lighthouse light flashes.	<i>Target spatial bias</i>: difference in detection of brightness change for target on left relative to right side of space (positive values reflect left bias). <i>Reaction time spatial bias</i>: difference in response times to the detection of brightness change for target on left relative to right side of space (positive values reflect left bias).
VISUAL PERCEPTION		
Line Bisection Test with visual acuity screen	Line bisection: Participants estimate mid-point of several horizontal lines.	<i>Bias</i> : bias in estimate of centre of line.
WORKING MEMORY		
Digit Span (Forwards and Backwards)	Participants recall (forwards) or manipulate (backwards) a verbally presented series of numbers of increasing length	<i>Number of digits recalled</i>

4.4.3 Stimuli

I used the same stimuli as described in Chapter 3, with minor adjustments to the RGB and chromaticity coordinates of the colours to ensure proper calibration to the projector used in this study. Details of the stimuli are provided again for ease of reference. Stimuli comprised a set of novel objects which varied along two dimensions (colour and shape, see Figure 35, Panel A), presented on a black background. Individual objects were 10 degrees of visual angle (dva) high. Their total size varied with their spikes, but the spikes never reached the border of the object image.

The spike orientation statistics were varied to create four shape classes: ‘strongly X-shaped’, ‘weakly X-shaped’, ‘weakly non-X-shaped’, ‘strongly non-X-shaped’. For the shape-based task, participants classified the target object as ‘X-shaped’ or ‘non-X-shaped’ (Figure 35, Panel A). Colour was also varied to create four classes: ‘strongly red’, ‘weakly red’, ‘weakly green’, ‘strongly green’. For the colour-based task, participants classified the target object as either ‘reddish’ or ‘greenish’ (Figure 35, Panel A). The four feature values along each dimension meant that for both tasks, stimuli were either far from the decision boundary (e.g., strongly red: ‘easy’ trials), or close to the decision boundary (e.g., weakly red: ‘hard’ trials).

The four colours were defined on a line joining strongly red and strongly green coordinates in $u'v'$ space; weakly red and weakly green were placed at a distance 23% from the midpoint to the applicable endpoint of the line. The projector’s colour profile was calibrated using a Spyder colorimeter and after calibration, colours for this experiment were chosen such that luminance was equivalent across the four colours. Luminance was set to 50cd/m² and RGB and chromaticity coordinates for each colour as follows: strongly red (RGB: 182, 88, 38; $u'v'$: 110.34, 41.379), weakly red (RGB: 158, 105, 34; $u'v'$: 62.069, 48.276), weakly green (RGB: 142, 113, 32; $u'v'$: 33.103, 52.414), strongly green (RGB: 110, 124, 28; $u'v'$: -15.172, 59.31).

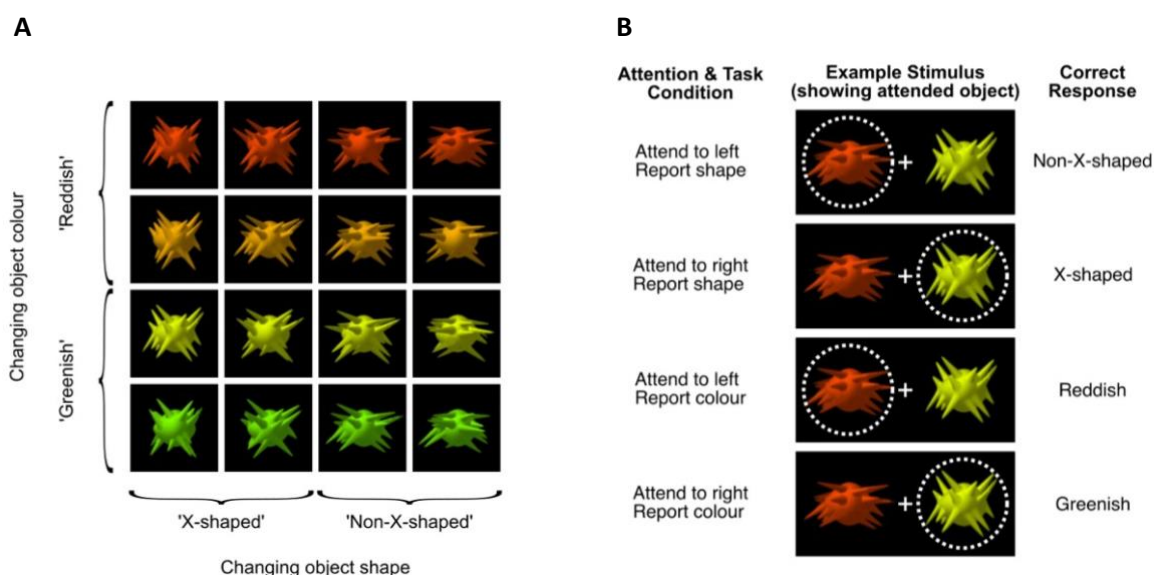


Figure 35. Panel A: Visual stimuli used in the experiment. These stimuli are the same as used in Chapter 3 and are provided again here for ease of reference. Each object varied systematically across two dimensions: colour and shape. In the colour task, participants categorised the attended object as either 'reddish' or 'greenish'. In the shape task, participants categorised the attended object as either 'X-shaped' or 'non-X-shaped', based on the orientations of the object's spikes. Panel B: Visual stimuli as presented in the MEG. Participants were asked to covertly attend to one side of space, and report (via button press) either the shape or colour of the object in the attended space. The dotted circle shows the locus of spatial attention and was not shown to participants.

4.4.4 Selective attention task

The visual attention paradigm used in this study was again based on the one used in Goddard et al. (2022), with three modifications to be more suitable for patients. Although the same task was used in Chapter 3, this study and Goddard et al. (2022) both employ the paradigm in an MEG context. I thus detail the main modifications made relative to Goddard et al. (2022). First, data in this study were acquired across two separate sessions (rather than in a single session). This was done to reduce the demands of the task for patients and older controls (e.g., by reducing task switching as response mappings were kept constant within a single session). Second, patients received more extensive training on this task. As also detailed in Section 4.4.5.2.1, all participants completed a training version of the task immediately prior to commencing the MEG

session, which comprised two blocks each containing 40 trials. However, patients additionally completed further training during their Neuropsychological Session (Session 1), which comprised two full blocks (258 trials per block) of the task. This additional training was provided to ensure adequate task proficiency (see Section 4.4.5.1 for more details) and to mitigate anxiety by increasing task familiarity. Finally, I introduced a flickering white fixation cross to denote the start of a block. This was included to capture participants' attention more instinctively, thereby simplifying the instruction to focus on the centre of the screen.

For reference, the task procedure is described in full. On each trial, two stimuli were presented simultaneously against a black screen, one left and one right of a grey fixation cross, with height and width of 1dva (Figure 35, Panel B). Participants were instructed at the beginning of every block to covertly attend to one side of space and report either the colour ('reddish' or 'greenish') or the shape ('non-X-shaped' or 'X-shaped') of the attended object. Covert attention was ensured using an eye tracker; participants were required to fixate on the fixation cross for a variable duration before the task would proceed (details below). Participants responded by pressing one of two response keys on a button box using the index and middle fingers of their right hand (unless stated otherwise).

Participants performed eight blocks in total across two separate MEG testing sessions (four blocks per session). Each block comprised a unique combination of the four possible spatial and feature attention manipulations (i.e., attend to one of the two locations: left or right, and one of the two features: colour or shape). Response mapping instructions were held constant during each MEG session and was the only element of the instructions to change between the two sessions. The response button mappings were switched between MEG sessions one and two to avoid confounding the response with the stimulus category (Grootswagers et al., 2017).

Each block comprised 258 trials, representing all possible pairs of the 16 objects plus an additional trial at the beginning and end that were later discarded (these trials were randomly chosen repetitions of two trials). Within each block, objects were presented

in a random order so that objects of each shape and colour were equally likely to precede objects of all shapes and colours.

Each block proceeded as follows. A flickering white cross on a black background denoted the start of the block. Each trial then began with the display of a single grey fixation cross (present throughout the entire block) on a black background. The fixation cross was displayed for a minimum of 500ms (maximum of 1000ms), of which at least 300ms (maximum of 800ms) of this time required participants fixate at the central cross, as verified by the eye tracker, before the trial would proceed. Variation in these times was included to reduce the predictability of stimulus onset.

Stimuli then appeared either side of the fixation cross and remained visible for 132ms. After participants made a response, the fixation cross changed colour for 216ms to indicate whether the response was correct (green) or incorrect/missed (blue). This feedback was provided no sooner than 1000ms after stimulus onset; if no response was recorded by 3000ms after stimulus offset, this trial was marked as missed.

The fixation cross then changed back to grey and fixation was then verified for the subsequent trial. Additionally, a calibration dot appeared after every 43 trials to monitor the eye tracker's accuracy; adjustments to the eye tracking calibration were made at this point if necessary (e.g., if the previously established central gaze position was detected to have shifted significantly from the start of the block). Participants received block-wise feedback at the end of the block (average accuracy and reaction time).

4.4.5 Data acquisition

Patients were typically seen on three separate occasions (four, if they required an MRI structural scan), comprising a neuropsychological testing session and two MEG sessions. Controls were typically seen on two separate occasions (three, if they required an MRI structural scan), and completed the neuropsychological tasks before or after their MEG sessions, as time permitted. All participants completed all sessions and MEG

experimental blocks, except for one control subject (sub-59) who, due to technical problems at their second MEG session, only completed five of the eight experimental blocks in total.

Where participants were asked to use a single hand to respond (on all tasks except CamTrip, where both hands were used, see earlier), participants used their dominant, right hand. There were two exceptions to this: one patient (sub-06) completed all tasks with their non-dominant left hand

REDACTION: Personal data removed for confidentiality

reasons.

, and one control (sub-57), a demographic-matched control of sub-06, was asked to use their non-dominant left hand on all tasks accordingly.

4.4.5.1 Neuropsychological session

For patients, the first session represented a neuropsychological testing session typically conducted at the patient's home, involving the administration of the measures detailed previously. Given the demanding nature of the main experimental paradigm, a shorter behavioural version of the task was also administered in this session to assess the patient's suitability for the subsequent MEG sessions. Details of the procedure for the shorter behavioural task are provided below. In contrast, control participants underwent neuropsychological testing either before or after their MEG sessions, without participating in the additional behavioural task.

The shorter behavioural version of the paradigm is largely as described previously, except patients completed only two blocks of the task and it was run on a Dell Latitude laptop. To ensure covert attention, patients' eye movements were tracked with a portable Tobii Pro Nano used for eye tracking. Stimuli were presented using MATLAB R2020b with Psychtoolbox version 3.0.17.

The procedure for running this behavioural version of the task proceeded as follows. First, task instructions were clarified with the patient (displayed through a PowerPoint presentation for consistency). Then, to ensure familiarity with the stimulus categories

(i.e., ‘reddish’ and ‘greenish’ for colour, and ‘X-shaped’ and ‘non-X-shaped’ for shape) and task instructions, patients completed a training session that involved completing a minimum of two training blocks (comprising two different block types such that, across the two blocks, patients practiced attending to both features and both locations), each comprising 30 trials.

Some patients also completed additional training blocks if a slower introduction to the task was thought necessary. Here, patients initially completed the training blocks with the stimuli presented for far longer (e.g., one second). When patients felt comfortable with the task and instructions with the slower version of the task, the stimulus duration was decreased to the appropriate length. Following this and in all other cases, stimuli were presented for 132ms, as per the MEG version of the task. A minimum score of 70% was required for each training block and repeated until this was achieved. Training took approximately 20 minutes.

Following this, patients completed full versions of the blocks on which they had been trained. Each block consisted of 258 trials (as in the MEG version of the task) and were otherwise identical to the training versions. Completing this shorter behavioural version of the task took approximately 40 minutes. Patients were required to score at least 70% on this task to continue in the study. All patients met this minimum requirement and were invited to the subsequent MEG sessions.

4.4.5.2 MEG sessions

To reduce the complexity and the duration of the main experimental task, MEG data were acquired across two separate sessions for all participants. Block order was counterbalanced across participants and arranged such that the attended location and/or attended feature alternated between blocks. Starting location, starting attended feature, and which response mapping participants used in their first session were also counterbalanced across participants. With the exception that neuropsychological tests

were administered to controls before or after these sessions, each MEG session proceeded in a similar fashion for all participants and lasted a maximum of 180 minutes (including set-up time; the MEG data acquisition lasted no longer than 120 minutes). The procedure is described in full.

4.4.5.2.1 *MEG, pre-task training*

First, all participants were given or reminded of the task instructions (displayed again through a PowerPoint presentation for consistency). All participants completed training of the task at a Dell desktop computer fitted with a head rest, situated outside of the MEG room. Stimuli were presented using MATLAB R2017a with Psychtoolbox version 3.0.17, and to ensure covert attention, eye movements were tracked using EyeLink 1000 tracking system. The eye tracker tracked the participant's right eye, unless this was not possible

REDACTION: Medical information removed. Personal data removed for confidentiality reasons.

Each participant completed a minimum of two training blocks (again, the two blocks chosen such that across the training blocks, participants practiced attending to both features and both locations), each comprising 40 trials. If a participant scored below 70% for either the colour or shape task block, they completed additional blocks of that task type until their performance reached 70% or above. Training took approximately 20 minutes.

4.4.5.2.2 *MEG, set-up*

Participants were then set up for the MEG session. Five electrodes were placed on the participant's face to record eye blinks and eye movements (i.e., electrooculogram (EOG): above and below the right eye, to the lateral side of each eye, and on the left jaw (ground)). Five head position indicator (HPI) coils were then placed at roughly evenly dispersed positions on the participant's head (typically placed at edge of hairline on left

and right sides of forehead, behind ears, and on crown of head). Heads were digitised using the 3D FASTRAK Polhemus system. Digitising goggles provided by the system were placed on the participant's head, acting as a reference for the system. The digitising goggles and digitising pen interacted with a receiver on the back of the wooden chair on which participants were seated, allowing for points to be registered by the system. First, anatomic fiducial points (i.e., nasion, left and right preauricular) were digitised, then the HPI coils, and an additional 200 to 300 points evenly distributed over the scalp to capture the participant's head shape and nose. The position of the HPI coils were checked at the end to ensure minimal movement (less than 2mm) during the digitising procedure. After digitisation, digitising goggles were removed, and electrodes were filled with gel.

4.4.5.2.3 *MEG equipment*

MEG data were acquired in a magnetically shielded room using a 306-channel (204 planar gradiometers and 102 magnetometers) MEGIN TRIUX Neo system at the MRC CBU at the University of Cambridge. All data were acquired with a sampling rate of 1000Hz and band-pass filtered online at 0.03 to 330Hz. Head position and EOG activity were continuously tracked using the HPI marker coils and electrodes, respectively. Measurement took place with participants in the upright seated position within the MEG system within a magnetically shielded room.

Stimuli were presented using MATLAB R2020b with Psychtoolbox version 3.0.17 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) and were back-projected using a VPixx Technologies PROPixx projector (60Hz refresh rate) onto a 6100 x 4200mm screen situated approximately 130cm in front of the participant.

To ensure covert attention, participant's eye movements were tracked using an EyeLink CL 5.15 tracker 1000 Plus remote tracking system (1000Hz). As in training, the eye

tracker tracked the participant's right eye, unless this was not possible

REDACTION:

Medical information removed. Personal data removed for confidentiality reasons.

A photodiode marked the onset of each stimulus in the MEG data. It was activated by the presentation of a small white circle on the bottom left-hand side of the screen during each stimulus presentation.

4.4.5.3 MEG data acquisition

Once seated in the MEG room, participants were handed the button box. All participants held the button box with their right hand (except sub-06 and sub-57, both of whom used their left hand, see earlier), and used their index and middle finger to respond. To ensure familiarity with using the button box, participants completed one further training block in the MEG with the button box (10 trials); the block type used in this training block was the same as the first block of the session.

At the start of each block, participants completed an eye tracker calibration and validation task. Task instructions and response mappings for the block were then displayed along with a graphical reminder of the attended feature categories. The experimenter read these instructions to participants and verbally checked comprehension. The block then proceeded as described in Section 4.4.4.

4.4.6 Analyses

4.4.6.1 MEG preprocessing

MEG analysis was conducting using MNE-Python (version 1.2.1, using Python 3.10.6, Gramfort et al., 2013).

MEG data were first processed using the MNE-Python implementation of Maxwell filtering and spatiotemporal signal-space separation (tSSS) (Taulu & Kajola, 2005; Taulu & Simola, 2006). First, I applied the automatic bad channel detection and interpolation routine implemented within Maxwell filtering. I then applied tSSS to suppress environmental noise, compensate for head movements within each block, and to reorient data from both sessions to the participant's head position in a representative reference block (defined as block two of session one). Alignment of both session data to a single reference was considered appropriate as variability across head position measurements taken at the start of each block for a given participant was minimal (typically within 10mm).

Data for each participant were then concatenated across the eight recording blocks. To reduce the influence of eye blinks, I performed an independent component analysis (ICA) using the FastICA algorithm (Hyvärinen, 1999; Hyvärinen & Oja, 2000) implemented in MNE-Python. Before ICA fitting, data were filtered between 1 and 40Hz, scaled to unit variance (using the standard deviation across all channels comprising each channel type) and whitened using principal component analysis (PCA). The first n PCA components to explain the cumulative variance of 95% of the data were then passed to the ICA algorithm. To detect and exclude ICA components that captured eye-blink artifacts, independent components were correlated to epochs centred around detected eye blinks extracted from the vertical EOG channel. Independent components that matched the vertical EOG signal through this procedure, and thus likely reflected an eye-blink artifact, were manually inspected before being marked for exclusion. The computed ICA was then applied to the concatenated raw data, yielding data reconstructed with eye-blink artifacts removed.

For four participants, I was not able to identify independent components which reflected eye-blink artifacts with the above procedure. I re-ran this procedure for these participants, specifying for more PCA components be entered in the ICA decomposition (i.e., n PCA components to explain cumulative variance of 99% of the data) to see whether these artifacts might be captured with more components. I detected and removed eye-blink artifacts in one additional participant. Therefore, eye-blink artifacts were detected and removed from all but three participants (all controls).

A low-pass filter of 66.67Hz was then applied to the concatenated, reconstructed data, followed by a notch filter at 50Hz to correct for line noise. Finally, I extracted epochs across all sensors from -200 to 1500ms relative to the stimulus onset of each trial, and downsampled the data to 200Hz. No baseline correction was applied to these data.

4.4.6.2 MEG rule and stimulus information decoding

I used MVPA to measure the extent to which rule information and information about the attended and unattended stimulus was reflected in brain activity at each timepoint. I performed the decoding analysis using the `mne.decoding` methods of MNE-Python, which wrap the `scikit-learn` package (Pedregosa et al., 2011). Data from both magnetometers and gradiometers was used in the analyses, using the `StandardScaler` function from `sklearn` to normalise each channel within itself across all epochs for a given timepoint.

MVPA was performed at the individual subject level in sensor space. I created several different datasets for the classification analysis. The first comprised all preprocessed data from all sensors (whole brain sensor space). I then selected sensors over my regions of interest: occipital (72 sensors), parietal (78 sensors) and frontal (78 sensors), split into left and right hemispheres. Decoding of rule information was conducted on whole brain sensor space. My primary aim was to explore how parietal lesions affect neural prioritisation of relevant information, with the view to examine how damage might

affect processing of stimulus information elsewhere in the brain (e.g., compensatory mechanisms). To this end, I examined the dynamic coding of attended and unattended stimulus information in whole brain sensor space, and in sensors over left and right occipital, parietal and frontal regions.

For decoding of rule information, I trained separate linear classifiers (Support Vector Machines, SVMs) to discriminate between labelled data from two location rules ('attend left' and 'attend right') and two feature rules ('attend colour' and 'attend shape').

For the decoding of stimulus information, separate classifiers were trained to discriminate between each of the six possible pairings of the four feature values (1v2, 1v3, 1v4, 2v3, 2v4 and 3v4) for shape and colour separately, in each of the four attention conditions, as in our previous work (e.g., Goddard et al., 2022). The four attention conditions were as follows: attended location attended feature (aLaF, e.g., the colour of the attended object in attend colour blocks), attended location unattended feature (aLuF, e.g., the colour of the attended object in attend shape blocks), unattended location attended feature (uLaF, e.g., the colour of the unattended object in attend colour blocks), and unattended location unattended feature (uLuF, e.g., the colour of the unattended object in attend shape blocks).

For rule and stimulus decoding, I created 'pseudotrials' to improve the signal-to-noise ratio for the discrimination of interest (Grootswagers et al., 2017; Guggenmos et al., 2018; Scrivener et al., 2023). That is, for a given classification, all epochs pertaining to the two categories of interest were extracted. With category labels maintained, epochs were randomly averaged in groups of four (a pseudotrial); where the number of epochs was not divisible by four, the remainder was excluded. Classifiers were then trained to distinguish between the pseudotrial data pertaining to the two categories of interest and were tested on held-out pseudotrial data using five-fold stratified cross-validation. Accuracy was quantified using the Area Under the Receiver Operating Characteristic Curve (ROC AUC) scoring metric. The process of generating pseudotrials and training/testing with cross-validation was repeated 100 times. Accuracy for a given pairwise classification was then the average accuracy across the cross-validation folds,

and across the 100 iterations of pseudotrial generation. This process was repeated at each timepoint.

For decoding of stimulus information within a given set of sensor regions, this process was repeated for each pairwise classification comprising a task condition, and then repeated for all four task conditions (i.e., ‘attend left, report colour’; ‘attend left, report shape’; ‘attend right, report colour’; ‘attend right, report shape’). Data for a given task condition came from two blocks, one from each session, each of which had used different response mappings. Accordingly, the classifier was required to generalise across session and response mapping. Decoding accuracy for a given attention condition (i.e., aLaF, aLuF, uLaF, uLuF) was the average accuracy across all comprising classifications at each timepoint (e.g., decoding accuracy for relevant information, aLaF, at a given timepoint was the average decoding accuracy at that timepoint across the six feature-level pairwise classifications, across the four different task conditions). This process yielded four decoding or classification accuracy (ROC AUC) traces per participant (one for each attention condition) for each set of sensors.

Only correct trials were used in all decoding analyses. To mitigate the effects of potential class imbalance that could arise as a result of excluding error trials, I used stratified k-fold cross-validation, which maintains the degree of class imbalance within each fold to mitigate the effects of imbalance, and ROC AUC as the measure of decoding accuracy. These are recommended approaches for dealing with imbalanced data (Thölke et al., 2023), however, it is worth noting that the degree of imbalance seen in my data was not severe and that baseline decoding accuracies for stimulus information hovered at the expected chance level (0.5).

4.4.6.3 Quantification of decoding strength

To facilitate comparison of the strength of neural representations between patients and controls, I quantified the strength of rule or stimulus information coding (i.e., decoding accuracies) for each participant.

To quantify strength of decoding of the two task rules (i.e., location or feature), I summed decoding accuracies (ROC AUC scores) for a given rule within a set time window. This time window was defined from stimulus onset to 600ms post-stimulus onset. Previous work using this task showed the dominance of relevant information over irrelevant from 400ms (Goddard et al., 2022), and minimum average reaction time across participants in the task was 589ms. The window of 0-600ms was chosen to balance inclusion of possible relevant information dominance and minimise reaction time effects. The strength of rule decoding score for a given rule thus represents the area under the curve (AUC) within this time window (henceforth referred to as ‘rule coding AUC’).

4.4.6.3.1 *Attentional prioritisation*

To quantify the strength of attentional prioritisation (i.e., stronger coding of relevant over irrelevant information), I first computed a timecourse reflecting attentional prioritisation for each participant. I derived this measure by first combining the three irrelevant information decoding traces (i.e., decoding traces for aLuF, uLaF, uLuF). This was done by averaging decoding accuracies for the three conditions at each timepoint, yielding an average decoding trace for irrelevant information. The attentional prioritisation decoding trace was then calculated by subtracting the averaged irrelevant decoding value from the decoding accuracy for relevant information (i.e., aLaF) at each timepoint. The attentional prioritisation decoding trace thus represents the degree to which relevant or attended information was more strongly represented than irrelevant

information across time for a given participant. Accordingly, a positive value at any given timepoint represented stronger decoding of relevant over irrelevant information at that timepoint, and zero represents no advantage for the representation of relevant information over irrelevant.

Separate attentional prioritisation timecourses were calculated for the different sensor space regions. Namely, separate attentional prioritisation timecourses were computed for sensors over occipital, parietal and frontal regions in the unaffected (contralesional) hemisphere and affected (ipsilesional) hemisphere. In controls, attentional prioritisation timecourses were calculated for sensors over left and right regions of interest separately and then averaged. To examine whether strength of attentional prioritisation for each patient was stronger in their unaffected relative to affected hemispheres, I also subtracted the affected timecourse from the unaffected timecourse for each sensor space region. Accordingly, positive values at any given timepoint here reflected stronger attentional prioritisation in sensors over the unaffected hemisphere of that region. To facilitate patient comparison with a corresponding hemisphere subtraction in controls, I created two attentional prioritisation difference timecourses for each control: subtracting right from left hemisphere, and vice versa, for each sensor space region.

Finally, I quantified the strength of attentional prioritisation (i.e., an overall attentional prioritisation score) for sensors over each region of interest for each participant. The process of quantification is as described previously for quantification of rule decoding. That is, for each participant, for each set of sensors, I summed the attentional prioritisation scores for each timepoint within an a priori time window (from stimulus onset to 600ms post-stimulus onset). As with rule coding AUC, my overall attentional prioritisation score here represents the AUC within this time window (accordingly, I refer to this as the ‘attentional prioritisation AUC’ for a given set of sensors).

I also quantified the difference in attentional prioritisation strength between the two hemispheres (unaffected minus affected in patients; left minus right and right minus left for controls) for each participant, for each sensor space region. As with the original

attentional prioritisation score, this was derived by summing the difference scores across specified time window (from stimulus onset to 600ms post-stimulus onset).

4.4.6.4 Statistical tests

Heterogeneity in the patient sample and limited sample size precluded the use of standard group-level statistical tests. Accordingly, I took an exploratory, case-series approach to analysis, focussing on individual-level statistics and case-control relationships. All statistical analyses were performed using R (v4.3.2) in RStudio (v2023.12.0).

4.4.6.4.1 *Testing for differences between patients and controls*

To assess if patients differed from our group of controls on neuropsychological measures, behavioural performance (accuracy and reaction time), and our neural data (rule coding and attentional prioritisation), I used a modified paired sample t-test for case-control designs described by Crawford and Howell (1998; discussed further in Crawford et al., 2010; Crawford & Garthwaite, 2012). Similar to a standard t-test, this approach tests whether a patient's score is sufficiently low (or different, in the case of a two-tailed test) to suggest that it does not come from the control group. However, a standard t-test treats control mean and standard deviation (SD) as population parameters, which can inflate Type I error when used to detect patient impairment against a small control sample (Crawford & Garthwaite, 2012; Crawford & Howell, 1998). The Crawford and Howell (1998) modified t-test treats control mean and SD as sample statistics, which is considered more appropriate when comparing a patient to a small normative or control sample, as was the case here. Moreover, this approach has been shown to be more robust to violations of the normality assumption for the control

distribution, which is more likely when assessing smaller cohorts (Crawford et al., 2006).

Here, I used this approach to test whether each patient differed from the control group on a given measure. Result statistics are reported with Z-case-control (z_{cc}) score as an interval estimate of the effect size. To correct for multiple comparisons, I applied a false discovery rate (FDR) correction within each patient across multiple related tests (adjusted p-values are denoted as adj-p, where applicable).

4.4.6.4.2 Testing for relationships between behavioural and neural measures

To test whether differences in attentional prioritisation might relate to behavioural performance on the experimental task, I used an approach that examined whether participants who are more dissimilar in terms of their neural data are also more dissimilar in terms of their behavioural data (Lorenz et al., 2018). First, I computed a dissimilarity matrix for each measure of interest: three neural measures (i.e., attentional prioritisation AUC scores for sensors over occipital, parietal, and frontal regions (averaged over hemispheres)) and two behavioural measures (i.e., accuracy and reaction time). Each symmetric dissimilarity matrix comprised difference scalars, reflecting in each column the difference between each participant's score on a given metric to all other participants' scores (i.e., inter-individual dissimilarity). I then performed six Spearman rank correlations, correlating the upper triangle of each neural dissimilarity matrix with the upper triangle of both behavioural dissimilarity matrices.

For statistical inference of these correlation analyses and correcting for multiple comparisons across three neural measures, I performed permutation testing (10,000 permutations) in MATLAB (vR2021b) where I replicated this procedure but randomly shuffled participants' labels. I then obtained p-values with an alpha-level set at 0.1 for a one-sided test (Groppe et al., 2011). In addition, by using the 'max statistic' method (i.e., in each permutation, I retained only the minimum/maximum correlation value across

sensor regions for reaction time/accuracy respectively, from which the null distribution is generated) (Groppe et al., 2011). Finally, I adjusted the p-values to correct for the three multiple comparisons across sensor regions (for accuracy and reaction time, separately) using family-wise error rate correction (FWER). Correlations were tested with a one-tailed test, as the direction of each of the inter-individual dissimilarity correlations was maintained from the original relationship between the outcome variables. For example, should faster (lower) reaction time be associated with higher attentional prioritisation in sensors over a given region (i.e., a negative correlation), a significant relationship between inter-individual differences in reaction time and inter-individual differences in attentional prioritisation would similarly yield a negative correlation.

This approach of comparing dissimilarity matrices, which captures differences between individuals, has the potential to detect subtle and non-linear relationships between neural and behavioural data that might be missed when directly correlating outcome measures across individuals. Accordingly, it allows for a more statistically powerful, nuanced examination of the relationship between patterns of neural activity and behavioural performance.

4.5 Results

4.5.1 Neuropsychological results

Individual performance on the CamTrip task is shown in Figure 36. As described in Table 8, evidence of a subtle extinction-like spatial bias can be reflected in maintenance scores less than 1 (i.e., poorer performance detecting bilateral relative to unilateral targets, Figure 36, Panel A) or in bias scores deviating from 0.5 (Figure 36, Panel B). Difficulty in exerting top-down attentional control is reflected in higher values of alpha (Figure 36, Panel C). As is evident in Figure 36, patients largely fall within the bounds of control performance on all these measures.

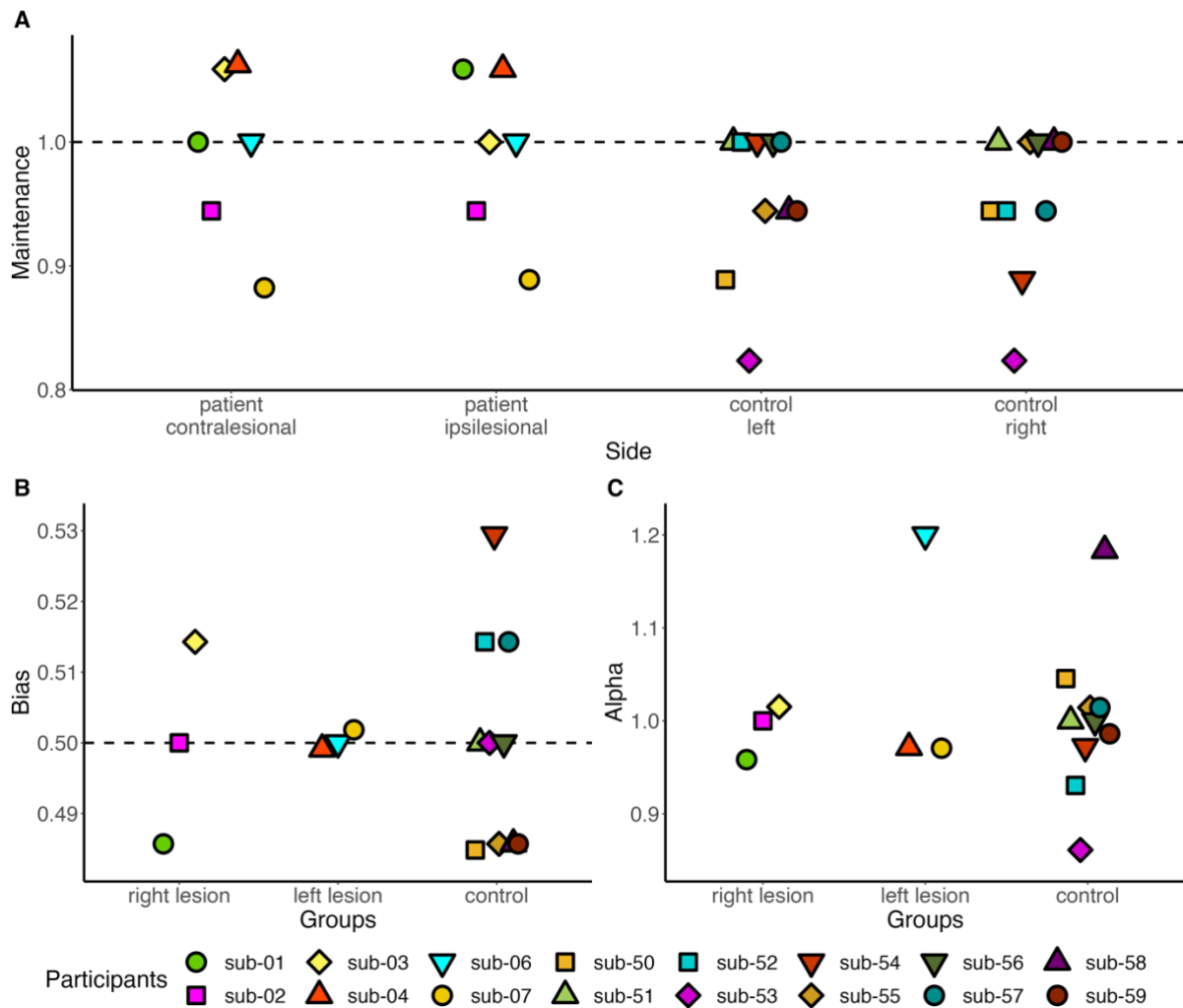


Figure 36. Outcome measures for the CamTrip task, an extinction-like task built in-house (see Peers et al., 2005 for details of an earlier version of this task). Panel A: Maintenance plots show proportion correct in detecting bilaterally presented targets divided by proportion correct in detecting unilaterally presented targets (left or right). Patient maintenance scores shown are bilateral detection relative to unilateral detection on the contralesional side or ipsilesional side of space. For controls, scores shown are bilateral detection relative to unilateral detection on the left side (i.e., left maintenance) or right side (i.e., right maintenance) of space. Scores less than 1 (dotted line) on these plots reflects poorer performance in detecting targets in bilaterally presented stimuli relative to targets presented unilaterally, and thus scores well below 1 may suggest extinction. As shown, patient scores appear consistent for ipsilesional or contralesional maintenance, and patient scores overall do not fall below the range of control performance on this task. Panel B: Bias scores reflect whether a participant show a strong bias to the right (score close to 0), or a strong bias to the left (score close to 1). Dotted line represents no bias to either side. Panel C: Alpha reflects ability to exert top-down control, with higher values indicating poorer top-down control. Each participant's performance is displayed with a unique combination of shape and colour. Participants in different shades of the same colour (e.g., different shades of yellow) are matched with regard to their demographic profiles (i.e., age, gender, education level).

Individual performance on neuropsychological tests of spatial attention and working memory from the OCS-BRIDGE are shown in Figure 37. On measures of spatial attention bias (Hearts Cancellation, SALT and Line Bisection Task, Figure 37, Panels A to D) and working memory (Digit Span, forwards and backwards, Figure 37, Panels E and F, respectively), the range of control performance was large and indeed impaired performance (denoted as falling within the shaded areas of Figure 37, and highlighted with a pale violet outline) was detected for both controls and patients on some measures.

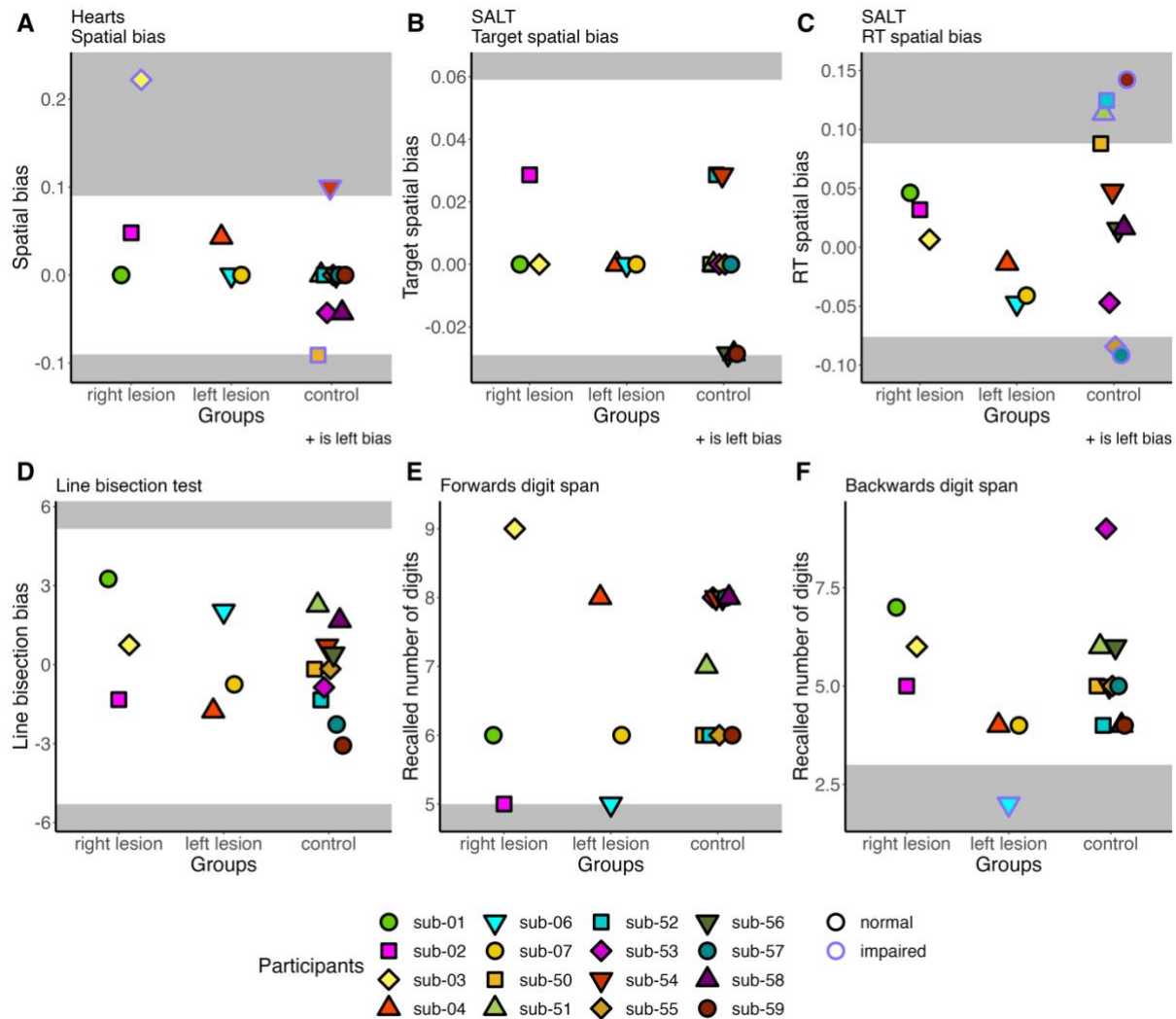


Figure 37. OCS-BRIDGE neuropsychological performance across administered tests: Hearts Cancellation (measure of spatial bias, Panel A), Sustained Attention Lighthouse Test (SALT) (spatial bias measured through target detection, Panel B; spatial bias in reaction times (RT), Panel C), Line Bisection Test (bias when estimating centre of horizontal line, Panel D), Digit Span Forwards (Panel E) and Backwards (Panel F). Patients are shown in leftmost columns of each plot, split by lesion location (right lesion on far left, and left lesion in middle), and controls are shown in rightmost column. Each participant's performance is displayed with a unique combination of shape and colour. Participants in different shades of the same colour (e.g., different shades of yellow) are matched with regard to their demographic profiles (i.e., age, gender, education level). Impairment (shaded area) is defined as performance exceeding two standard deviations from the mean of a reference control sample; the white area of each plot thus represents the range of scores that are 'within normal limits'. Participants falling outside of normal limits according to normative data, and thus show an 'impaired' performance on a given test are highlighted with a pale violet outline.

To test whether patients fell outside the range of control performance on any of the neuropsychological outcome measures used, I compared each patient's performance against the control group using case-control t-tests (Crawford & Howell, 1998), one- or two-tailed as appropriate. After correcting for multiple comparisons within tests for each patient (FDR correction within patient, across 10 outcome measures), patient performance was within control limits on every measure except one: sub-03 showed evidence of left spatial bias on the Hearts Cancellation Task that was outside of control performance (Figure 37, Panels A, shown as yellow diamond; $t(9)=4.49$, $z_{cc}=4.71$, adj- $p=0.015$). This finding of left spatial bias in this patient was inconsistent with their (right-sided) lesion.

Even when examining uncorrected p-values, no evidence of spatial bias was found for the other patients (uncorrected p-values on other spatial bias tasks >0.073). Examination of uncorrected p-values on other measures suggested trends towards poorer top-down control of attention than controls for sub-06 (CamTrip alpha: $t(9)=2.30$, $z_{cc}=2.42$, uncorrected $p=0.023$), and poorer performance on working memory tasks, relative to controls, for both sub-02 and sub-06 (sub-02: forwards digit span, $t(9)=-2.01$, $z_{cc}=-2.11$, uncorrected $p=0.037$; sub-06: forwards digit span, $t(9)=-2.01$, $z_{cc}=-2.11$, uncorrected $p=0.037$; backwards digit span, $t(9)=-2.11$, $z_{cc}=-2.21$, uncorrected $p=0.032$). All other uncorrected p-values were >0.139).

4.5.2 Behavioural results

Patients and controls performed well on the main behavioural task. There was no evidence of a difference in accuracy between patients and controls (t-test between groups: $t(14) = -0.562$; $p=0.583$), however patients were significantly slower than controls ($t(14)=3.67$; $p=0.025$). Mean accuracy and reaction time (for correct trials only) for each individual participant and performance distribution across the groups are shown in Figure 38. Using case-control one-tailed t-tests (Crawford & Howell, 1998), no individual patient was less accurate than controls (p-value range 0.067-0.829). However, four patients were significantly slower than controls (sub-01: $t(9)=1.99$, $z_{cc}=2.09$, $p=0.039$; sub-04: $t(9)=2.49$, $z_{cc}=2.62$, $p=0.017$; sub-06: $t(9)=4.72$, $z_{cc}=4.95$, $p=0.001$; and sub-07: $t(9)=2.58$, $z_{cc}=2.70$, $p=0.015$).

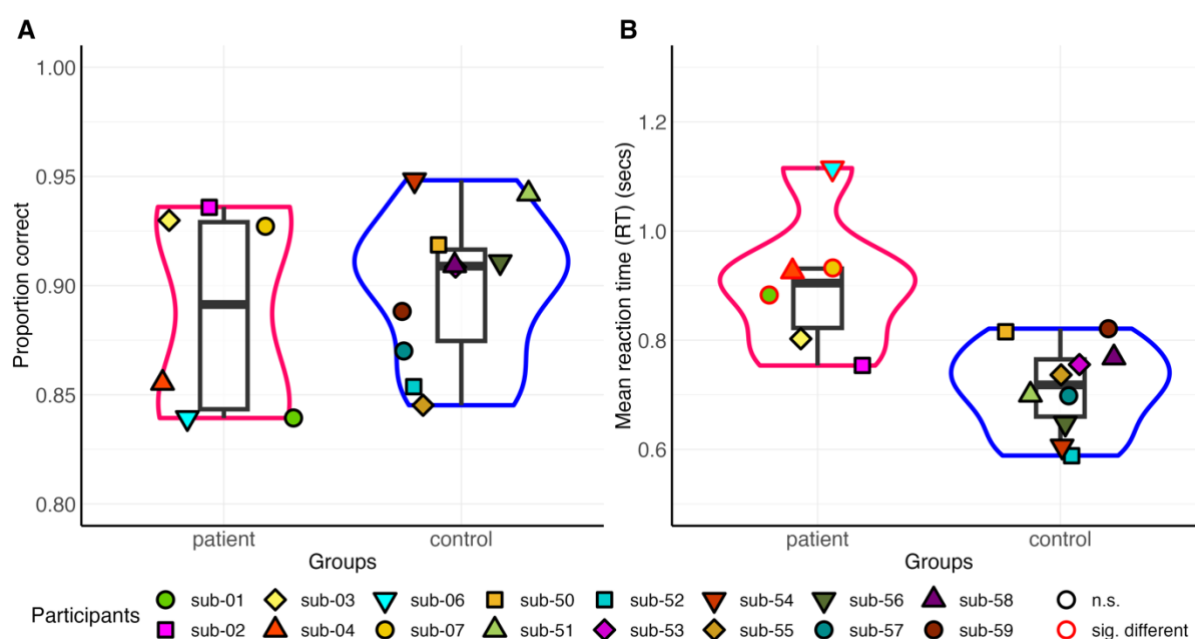


Figure 38. Behavioural performance for individual patients and controls. Panel A shows mean accuracy (proportion correct). Panel B shows mean reaction time (RT, in seconds; correct trials only). Each participant's performance is displayed with a unique combination of shape and colour. Participants in different shades of the same colour (e.g., different shades of yellow) are matched with regard to their demographic profiles (i.e., age, gender, education level). Violins show distribution for patients (pink) and controls (blue). Individual patients who showed significantly poorer performance relative to controls are shown with red outline, denoted as 'sig. different' (= significantly different); 'n.s.' (= not significantly different), black outline, denotes patient not found to significantly differ from controls.

Figure 39 shows behavioural performance (accuracy in Panel A and reaction time for correct trials in Panel B), split by the attended visual field (i.e., for the ipsilesional and contralesional visual fields for patients, left and right visual fields for controls).

I had anticipated patients might show some evidence of spatial bias or inattention on this task, for instance, poorer performance (i.e., lower accuracy or slower reaction times) when attending to their contralesional visual field compared to their ipsilesional visual field. Case-control one-tailed t-tests comparing patient performance when attending to their ipsilesional or contralesional visual field against the same visual field in controls showed little evidence suggestive of spatial bias on behavioural measures of this task. Only one patient (sub-01) showed significantly reduced accuracy relative to controls when attending to their contralesional visual field ($t(9)=-2.40$, $z_{cc}=2.52$, $\text{adj-}p=0.040$; $\text{adj-}p$ values for all other patients and tests >0.077). Moreover, this seemed to indicate spatial bias, with sub-01 demonstrating a significantly greater difference in accuracy between the two hemifields compared to the equivalent difference in controls ($t(9)=2.15$, $z_{cc}=2.25$, $p=0.030$, indicated as red line in Figure 39, Panel A; p -values for all other patients >0.113).

Three of the patients previously identified as significantly slower than controls on this task were again identified as being slower than controls when the hemifields were considered separately. For all three patients, significantly slower reaction times were observed in both the ipsilesional and contralesional visual field (sub-04: ipsilesional visual field $t(9)=2.53$, $z_{cc}=2.66$, $\text{adj-}p=0.020$; contralesional visual field $t(9)=2.41$, $z_{cc}=2.53$, $\text{adj-}p=0.020$; sub-06: ipsilesional visual field $t(9)=5.14$, $z_{cc}=5.39$, $\text{adj-}p<0.001$; contralesional visual field $t(9)=4.26$, $z_{cc}=4.46$, $\text{adj-}p=0.001$; sub-07: ipsilesional visual field $t(9)=2.33$, $z_{cc}=2.44$, $\text{adj-}p=0.022$; contralesional visual field $t(9)=2.74$, $z_{cc}=2.87$, $\text{adj-}p=0.022$). Only for sub-07 was the difference in reaction time between the two hemifields significantly greater than the equivalent difference in controls, suggestive of spatial bias ($t(9)=-2.65$, $z_{cc}=-2.78$, $p=0.013$, indicated as red line in Figure 39, Panel B; p -values for all other patients >0.226).

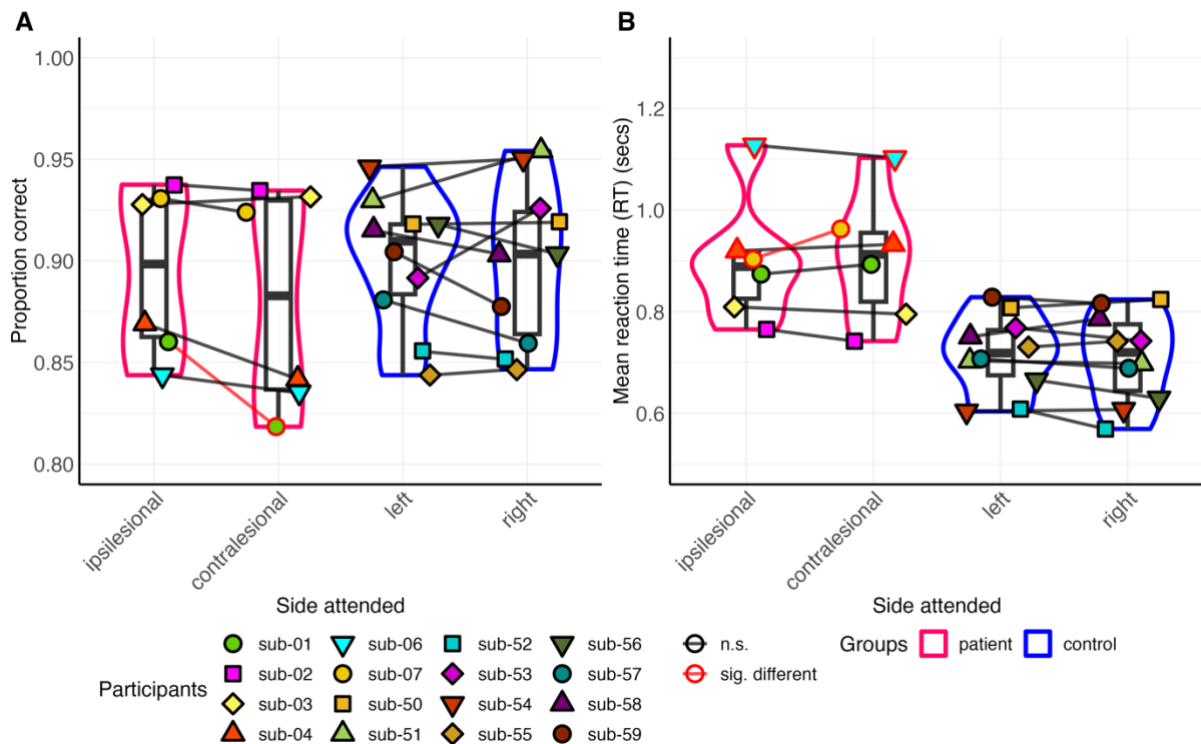


Figure 39. Behavioural performance for individual patients and controls, split by side attended. Panel A shows mean accuracy (proportion correct). Panel B shows mean reaction time (RT, in seconds; correct trials only). For patients, average performance across trials is shown when attention was directed to the ipsilesional visual field or contralesional visual field. For controls, average performance across trials is shown when attention was directed to the left or right visual fields. Each participant's performance is displayed with a unique combination of shape and colour. Participants in different shades of the same colour (e.g., different shades of yellow) are matched with regard to their demographic profiles (i.e., age, gender, education level). Violins show distribution for patients (pink) and controls (blue). Individual patients who showed significantly poorer performance relative to controls are shown with red outline, denoted as 'sig. different' (= significantly different); 'n.s.' (= not significantly different), black outline, denotes patient not found to significantly differ from controls (after multiple correction, where appropriate). Poorer performance (i.e., lower accuracy and higher/slower reaction times) is expected when attention is directed to the contralesional visual field, relative to the ipsilesional visual field, in patients demonstrating spatial bias. Patients who showed significantly poorer performance when attending to their contralesional versus ipsilesional visual field, relative to the equivalent difference in controls, are shown with a red connecting line (i.e., sub-01 in Panel A, and sub-07 in Panel B).

I also explored whether subtle differences in attending to the two sides of space might be more pronounced for difficult compared to easy trials. Task difficulty was defined by the proximity of the attended feature dimension of the target to the decision boundary, with 'easy' trials involving categorisations far from the decision boundary and 'hard' trials involving categorisations close to the decision boundary (see Figure 35). This

breakdown of behavioural results by side attended and difficulty is shown in Supplementary Figure S 8. Participants were generally more accurate and faster with 'easy' categorisations compared to 'hard' ones.

4.5.3 Decoding results

4.5.3.1 Rule decoding

I first examined the dynamics with which patients and controls coded rule information from whole brain neural data. The two main rule components of the task were decoded separately: the location of the target object (left or right), and the feature of the target object (colour or shape). Decoding of location and feature rule information for individual patients and the control group is shown in Figure 40 (Panel A and B, respectively). Individual rule decoding plots for all participants are shown in Supplementary (for location rule in Figure S 9 and feature rule in Figure S 10).

As rule information (i.e., attend left or right, attend colour or shape) was blocked and thus known before each trial, I expected that this information would be present throughout the trial, including prior to stimulus onset (Barnes et al., 2022; Goddard et al., 2022; Hebart et al., 2018). In line with this, all participants showed rule information decoding prior to stimulus onset. Rule decoding further increased after stimulus display, once the rule could be put into action.

To explore whether patients showed weaker rule decoding relative to controls, I compared each patient's rule AUC scores to that of the control group for the location rule and feature rule separately. Rule AUC scores reflect the strength of rule decoding within the first 600ms from stimulus onset (i.e., the prescribed window of interest). Rule AUC scores for each participant are shown in the insets of Figure 40, for location (Panel A) and feature (Panel B). Case-control t-tests showed that sub-06 exhibited weaker location and feature rule decoding relative to controls (location: $t(9)=-2.39$, $z_{cc}=-2.50$, adj-p=0.02; feature: $t(9)=-3.40$, $z_{cc}=-3.57$, adj-p=0.01). In addition, sub-03 also

showed weaker feature rule decoding relative to controls ($t(9)=-2.47$, $z_{cc}=-2.59$, adj- $p=0.04$). No other patients showed weaker rule decoding, for location or feature, relative to controls (all other adj- p s >0.05).

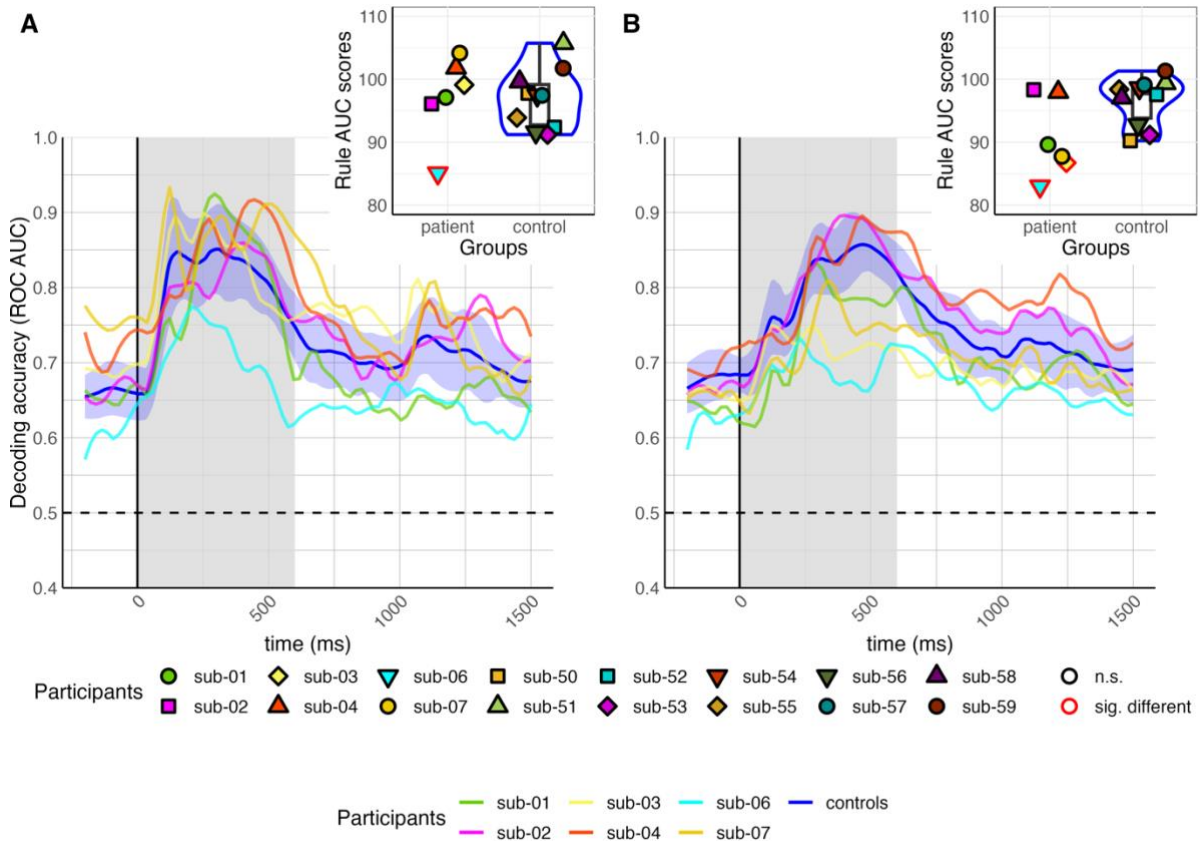


Figure 40. Rule decoding traces for location rule (Panel A) and feature rule (Panel B) for each patient (different colours) against the averaged control group (shaded blue area shows standard deviation around the control average). Plots show decoding accuracy (ROC AUC) from -200 to 1500ms, relative to stimulus onset (shown with vertical black line at 0) across all sensors (whole brain). Plots are smoothed using loess method. Chance is 0.5 and is indicated by the dotted horizontal black line. Shaded area in grey (0-600ms) reflects the area of decoding accuracies that are summed for the purpose of calculating rule decoding AUC scores. Inset shows individual rule decoding AUC scores, with distribution for control group shown with blue violin. As in previous plots, participant's performance is displayed with a unique combination of shape and colour. Participants in different shades of the same colour (e.g., different shades of yellow) are matched with regard to their demographic profiles (i.e., age, gender, education level). Individual patients who showed significantly poorer performance relative to controls are shown with red outline, denoted as 'sig. different' (= significantly different); 'n.s.' (= not significantly different), black outline, denotes patient not found to significantly differ from controls. See Supplementary Materials for individual control decoding traces for location rule (Figure S 9), and feature rule (Figure S 10).

4.5.3.2 Stimulus decoding

4.5.3.2.1 *Whole brain sensor space, each attention condition*

To explore how parietal lesions might affect the neural representation of attended and unattended stimulus information, I first examined individual-level decoding traces of this information across all sensors (whole brain). In a previous MEG study using this task (Goddard et al., 2022), all stimulus information (attended and unattended) was briefly decodable after stimulus onset, resolving to a selective and sustained representation of task-relevant information from approximately 400ms. While this previous work reported group-level decoding timecourses, I hoped to see a similar pattern of results at an individual level in my data. I explored this visually, shown in Figure 4I. While apparent to varying degrees, a selective representation of task-relevant information (shown in red) is clearly detectable on visual inspection for all participants bar one possible exception, a focal lesion patient: sub-06. This patient does not appear to show a clear boost in coding of relevant information, and indeed decoding of all attention conditions appears mostly at chance across time. However, there is a perceptible selective increase in coding of relevant information around their average response time (1115ms).

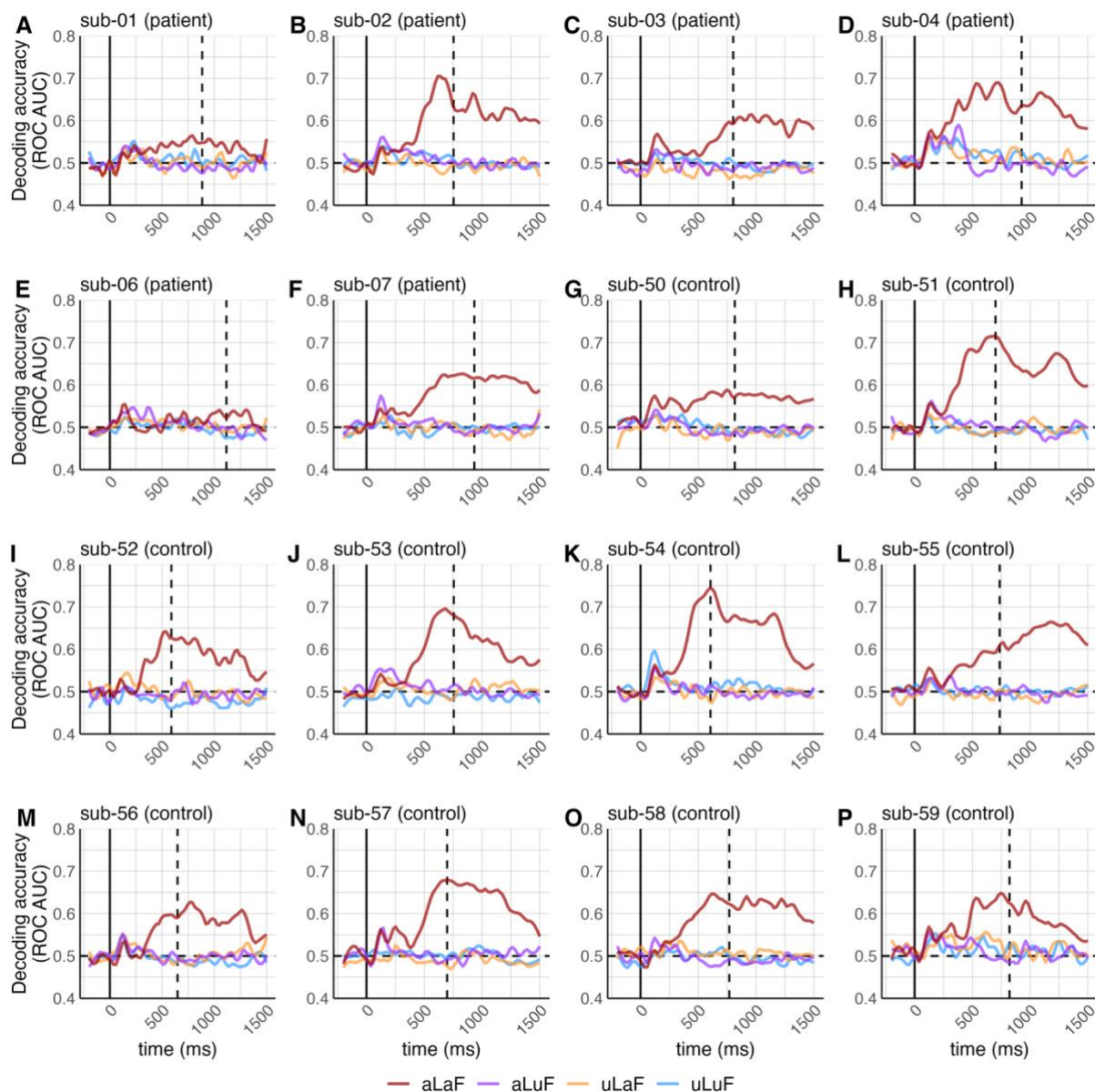


Figure 41. Stimulus decoding across time, for each patient (first six panels) and each control (remaining panels). Each coloured line represents average decoding accuracy (ROC AUC) for pairwise comparisons comprising each of the four attention conditions from -200 to 1500ms, relative to stimulus onset (shown with vertical solid black line at 0) across all sensors (whole brain). The four attention conditions and representative colours are as follows: red represents decoding of relevant information (attended feature information at the attended location [aLaF]), other colours represent irrelevant information decoding (purple = unattended feature information at the attended location [aLuF], yellow = attended feature information at the unattended location [uLaF], blue = unattended feature information at the unattended location [uLuF]). Plots are smoothed using loess method. Each individual's average reaction time on the task is shown as a vertical dashed line. Chance is 0.5 and is indicated by the dotted horizontal black line. Stimulus decoding of relevant information is visually evident for all participants but is markedly reduced for sub-06.

4.5.3.2.2 *Attentional prioritisation*

I then formalised the degree of attentional prioritisation for each individual by subtracting decoding of irrelevant from relevant stimulus information (Figure 42). I explored this in sensors over occipital, parietal and frontal regions of the unaffected and affected hemispheres separately. Figure 42 shows individual attentional prioritisation timecourses for sensors over the unaffected and affected hemisphere for patients against the average of both hemispheres in controls (Figure 42, two leftmost columns). Positive values reflect the degree to which relevant information was more strongly represented than irrelevant information for a given timepoint, with attentional prioritisation scores at zero representing no advantage for the representation of relevant information over irrelevant. Attentional prioritisation timecourses for each individual participant including controls are shown in Supplementary Figure S 11.

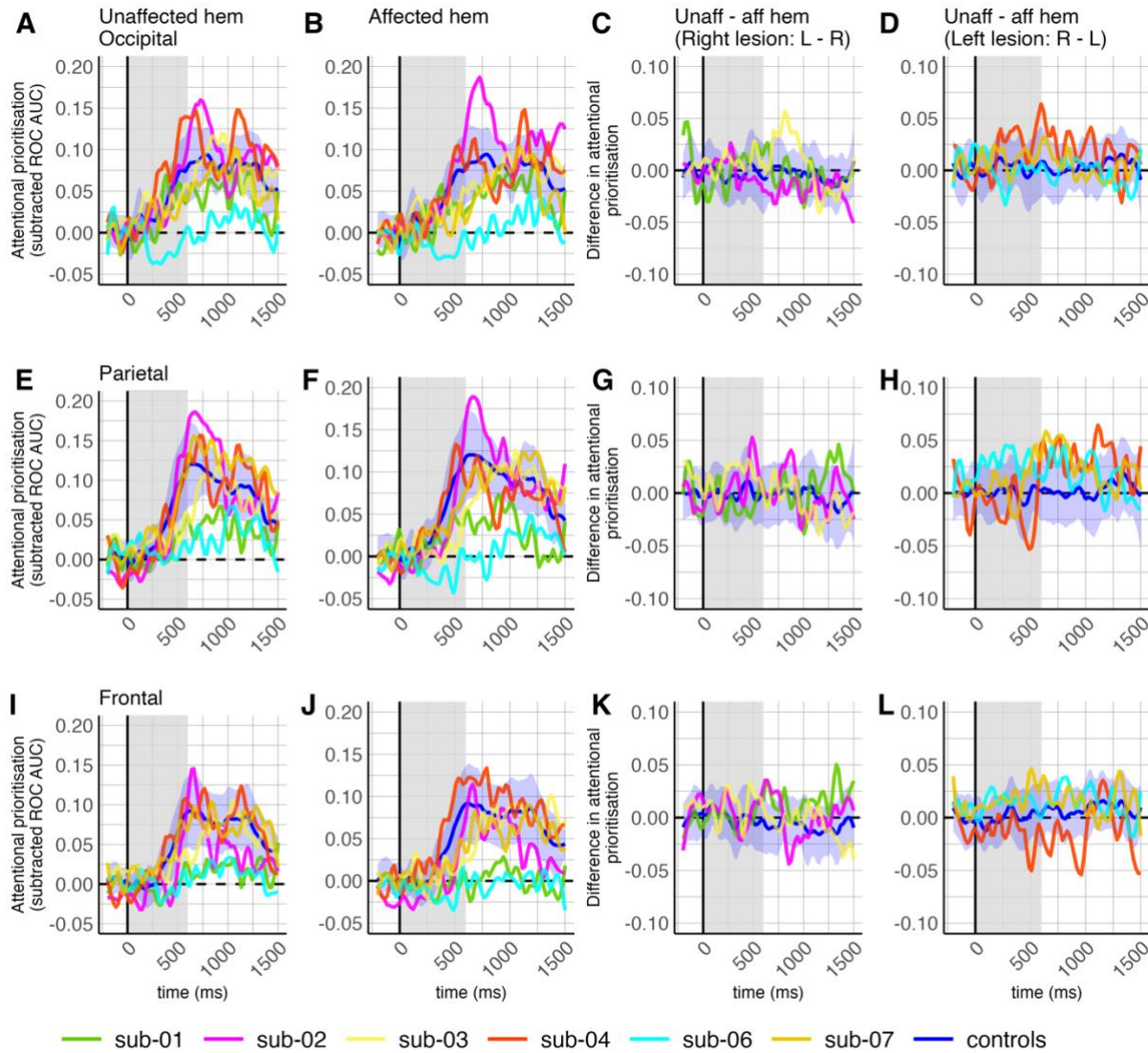


Figure 42. Attention prioritisation timecourses for sensors over occipital (top panels, Panels A- D), parietal (middle panels, Panels E-H) and frontal (bottom panels, Panels I-L) regions. Attentional prioritisation reflects the relative strength of relevant over irrelevant information representation across time, from -200ms to 1500ms relative to stimulus onset (vertical black line at 0ms), with values above zero indicating stronger representation of relevant information. Accordingly, attentional prioritisation at zero (dotted horizontal black line) reflects no relative representation of relevant information over irrelevant information. Each patient's attentional prioritisation timecourse is shown for sensors over their unaffected (first column, Panels A, E, I) or affected (second column, Panels B, F, J) hemisphere separately; for control group (shown in blue, with shaded blue area showing standard deviation around control average), attention prioritisation is averaged over both hemispheres and thus is the same for unaffected and affected plots of a given set of sensor regions. The difference in attentional prioritisation between hemispheres is shown against equivalent subtraction in controls (third and fourth columns). That is, for right lesion patients (third column, Panels C, G, K), the unaffected minus affected hemisphere subtraction is shown against left minus right hemisphere in the control group; for left lesion patients (fourth column, Panels D, H, L), the hemisphere subtraction is shown against right minus left hemisphere in the control group. Shaded area in grey (0-600ms) reflects time window used to calculate overall attentional prioritisation score or AUC. Plots are smoothed using loess method. Each individual's timecourse is shown in Supplementary Figure S II. Hem = hemisphere, unaff = unaffected (i.e., contralesional hemisphere), aff = affected (i.e., ipsilesional hemisphere).

The control group (blue line) showed attentional prioritisation in all ROIs, peaking around 630ms in frontal regions (Figure 42, Panel I), 700ms in parietal regions (Figure 42, Panel E) and 715ms in occipital regions (Figure 42, Panel A). The prioritisation was sustained in frontal and occipital regions, and fell gradually in parietal regions, until around 1250ms after which it declined more rapidly. Individual patient attentional prioritisation traces are noisier than the control average and thus trends are more difficult to visually discern. Most patients showed a degree of attentional prioritisation that roughly matched controls, for sensors over both their unaffected and affected hemispheres. The exceptions were sub-06 (cyan), for whom attentional prioritisation hovers around zero, particularly for the affected hemisphere, and sub-01, for whom attentional prioritisation hovers around zero particularly for frontal regions of the affected hemisphere.

Next, for each patient, I compared the strength of attentional prioritisation between unaffected and affected hemispheres (Figure 42, shown separately for right lesion patients in third column, and left lesion patients in fourth column). Here, positive values for patients reflect stronger attentional prioritisation in the unaffected relative to the affected hemisphere; for controls, positive values indicate stronger attentional prioritisation in the hemisphere that corresponded to the unaffected hemisphere for the given set of patients (e.g., left minus right for comparison with right lesion patients). I had anticipated that patients might show stronger attentional prioritisation in sensors over unaffected regions. This was visually apparent for all three left lesion patients for sensors over parietal regions (where lesions were located) (Figure 42, Panel H, sub-04 (orange) sub-06 (cyan), and sub-07 (mustard)). However, for sub-04 and sub-07 this difference occurred after my a priori time window of interest (0-600ms). Sub-04 (orange) also showed visually stronger attentional prioritisation in unaffected relative to affected occipital regions (Figure 42, Panel D). Of the right lesion patients, only sub-03 (yellow) showed a visibly noticeable difference between hemispheres, with attentional prioritisation peaking earlier in unaffected occipital relative to affected occipital regions, resulting in a difference in prioritisation between the hemispheres, again occurring after the time window of interest (peaking around 800ms, Figure 42,

Panel C). As expected, controls did not show any clear differences between hemispheres (i.e., blue line appears flat in two rightmost columns of Figure 42).

To formally test these differences, I used the attentional prioritisation AUC score which summed attentional prioritisation values over the first 600ms (Figure 43). First, I tested whether attentional prioritisation scores were lower in patients than controls, for sensors over affected or unaffected occipital, parietal and frontal regions, separately.

After correcting for multiple comparisons, case-control tests revealed that only one patient showed significantly reduced attentional prioritisation scores relative to controls in the first 600ms. Sub-06 (cyan upside down triangle, Figure 43, left panel) exhibited reduced attentional prioritisation relative to the control group in parietal and frontal regions on the affected side ($t(9)=-3.39$, $z_{cc}=-3.55$, $\text{adj-}p=0.012$; $t(9)=-2.57$, $z_{cc}=-2.70$, $\text{adj-}p=0.028$, respectively), and occipital regions in both hemispheres (affected: $t(9)=-4.27$, $z_{cc}=-4.48$, $\text{adj-}p=0.009$; unaffected: $t(9)=-4.27$, $z_{cc}=-4.48$, $\text{adj-}p=0.009$). No other patients showed reduced attentional prioritisation relative to controls after correction for multiple comparisons (all other $\text{adj-}ps > 0.10$).

I had also suspected that sub-01 might show reduced attentional prioritisation in frontal regions, especially on the affected side, based on visual inspection of Figure 42 (Panel J). Uncorrected case-control tests suggested sub-01 showed reduced attentional prioritisation for sensors over affected frontal regions ($t(9)=-1.96$, $z_{cc}=-2.05$, uncorrected $p=0.041$, Figure 43, Panel E, green circle). In addition, uncorrected case-control tests suggested sub-07 showed reduced attentional prioritisation for sensors over affected occipital regions in this time window ($t(9)=-1.93$, $z_{cc}=-2.03$, uncorrected $p=0.043$, Figure 43, Panel A, mustard circle), reflecting an increased latency of attentional prioritisation for this patient (Figure 42, Panel B).

I also tested whether patients exhibited a greater difference in the strength of attentional prioritisation between hemispheres, relative to controls, in the first 600ms. Specifically, I tested whether patients showed stronger attentional prioritisation in sensors over their unaffected, relative to affected, hemisphere for occipital, parietal and frontal regions, compared to the corresponding subtraction in the control group (i.e.,

the difference plots shown in rightmost columns of Figure 42). The computed difference attentional prioritisation scores are shown for each participant, for sensors over each region of interest, in Figure 43 (right column). Visual inspection of Figure 43 suggests patients broadly show stronger attentional prioritisation in sensors over unaffected regions (i.e., most patients show positive difference values). When tested against the relevant control comparison, only sub-06 showed significant asymmetry in strength of attentional prioritisation between hemispheres. Sub-06 showed significantly stronger attentional prioritisation in parietal regions of their unaffected relative to affected hemisphere, compared to controls ($t(9)=2.56$, $z_{cc}=2.68$, $\text{adj-}p=0.028$). In this early time window, this was mainly driven by below zero attentional prioritisation (i.e., prioritisation of the irrelevant information) in sensors over the affected parietal lobe (see Figure 42, Panels E and F).

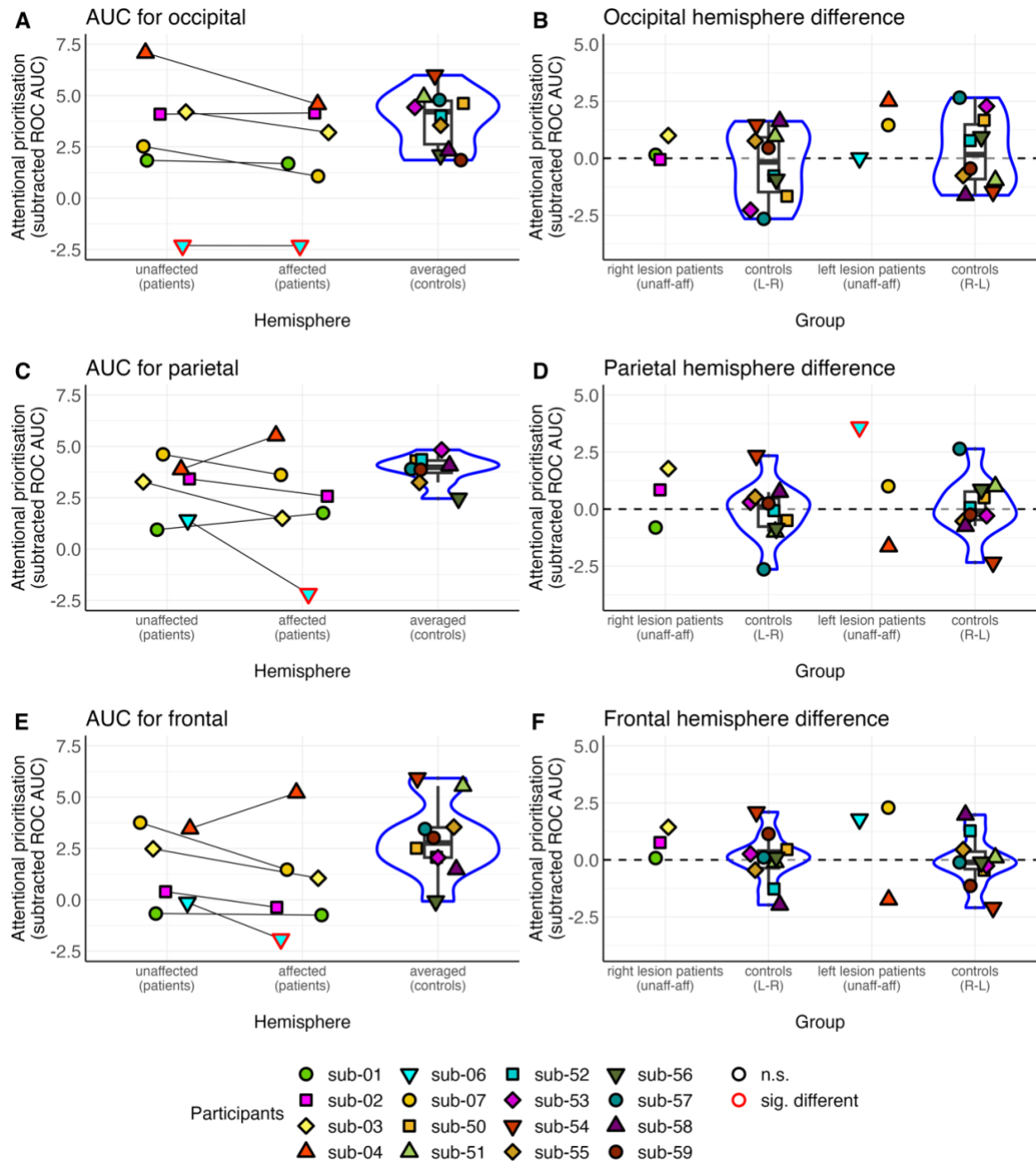


Figure 43. Attentional prioritisation AUC scores, shown separately for each patient and for the control group (blue violin), for sensors over occipital (top panels, Panels A and B), parietal (middle panels, Panels C and D), and frontal sensors (bottom panels, Panels E and F). Left panels (Panels A, C, and E) show attentional prioritisation scores for sensors over a given region separated into patient's unaffected and affected hemispheres, and averaged over both hemispheres in controls. Right panels (Panels B, D, and F) show the difference in attentional prioritisation AUC scores. For patients (shown separately by lesion side), the difference is between AUC scores for sensors over their unaffected and affected hemisphere for a given region. Accordingly, positive values reflect stronger prioritisation in sensors over unaffected, relative to affected, hemisphere of that region. For controls, the subtraction was done in both directions, and patients were compared to the appropriate comparison. The appropriate control hemisphere subtraction for a given patient group is shown right of the patient difference AUC scores. Zero values in the difference plots (horizontal dotted line) reflect no difference between hemispheres. As in previous plots, each participant's score is displayed with a unique combination of shape and colour. Individual patients with significantly different (i.e., 'sig. different') AUC scores, relative to controls after multiple correction, are shown with red outline; patients not significantly differ from controls (i.e., 'n.s.') shown with black outline.

4.5.3.2.3 *Relationship between strength of prioritisation and behavioural performance*

Finally, I asked whether neural prioritisation of relevant over irrelevant information predicted behavioural performance. To explore this, I asked whether inter-individual dissimilarity in attentional prioritisation for each region (i.e., attentional prioritisation AUC over the first 600ms for sensors over occipital, parietal and frontal regions) correlated with inter-individual dissimilarity on behavioural measures (i.e., accuracy and reaction time). These relationships are shown in Figure 44. There was a trend towards positive relationships between AUC differences and accuracy (people who were more dissimilar in attentional prioritisation tend to be more dissimilar in accuracy, with higher AUC scores predicting higher accuracy). However, after correcting for multiple comparisons, the permutation test showed that this relationship was not significantly greater than chance in any region (Spearman R: occipital $R=0.34$, adj- $p=0.275$; parietal $R=0.52$, adj- $p=0.057$; frontal $R=0.30$, adj- $p=0.363$, Figure 44, Panels A-C). For reaction time, there was a significant correlation between inter-individual dissimilarity in reaction time and inter-individual dissimilarity in neural prioritisation for sensors over occipital lobe ($R=-0.58$, adj- $p=0.030$, Figure 44, Panel D), but not sensors over parietal or frontal regions (parietal $R=-0.47$, adj- $p=0.090$; frontal $R=-0.28$, adj- $p=0.337$, Figure 44, Panels E and F). This significant negative correlation indicates that participants who are more dissimilar in attentional prioritisation scores for sensors over occipital lobe are also more dissimilar in reaction time, with stronger attentional prioritisation associated with faster behavioural performance and weaker attentional prioritisation associated with slower behavioural performance.

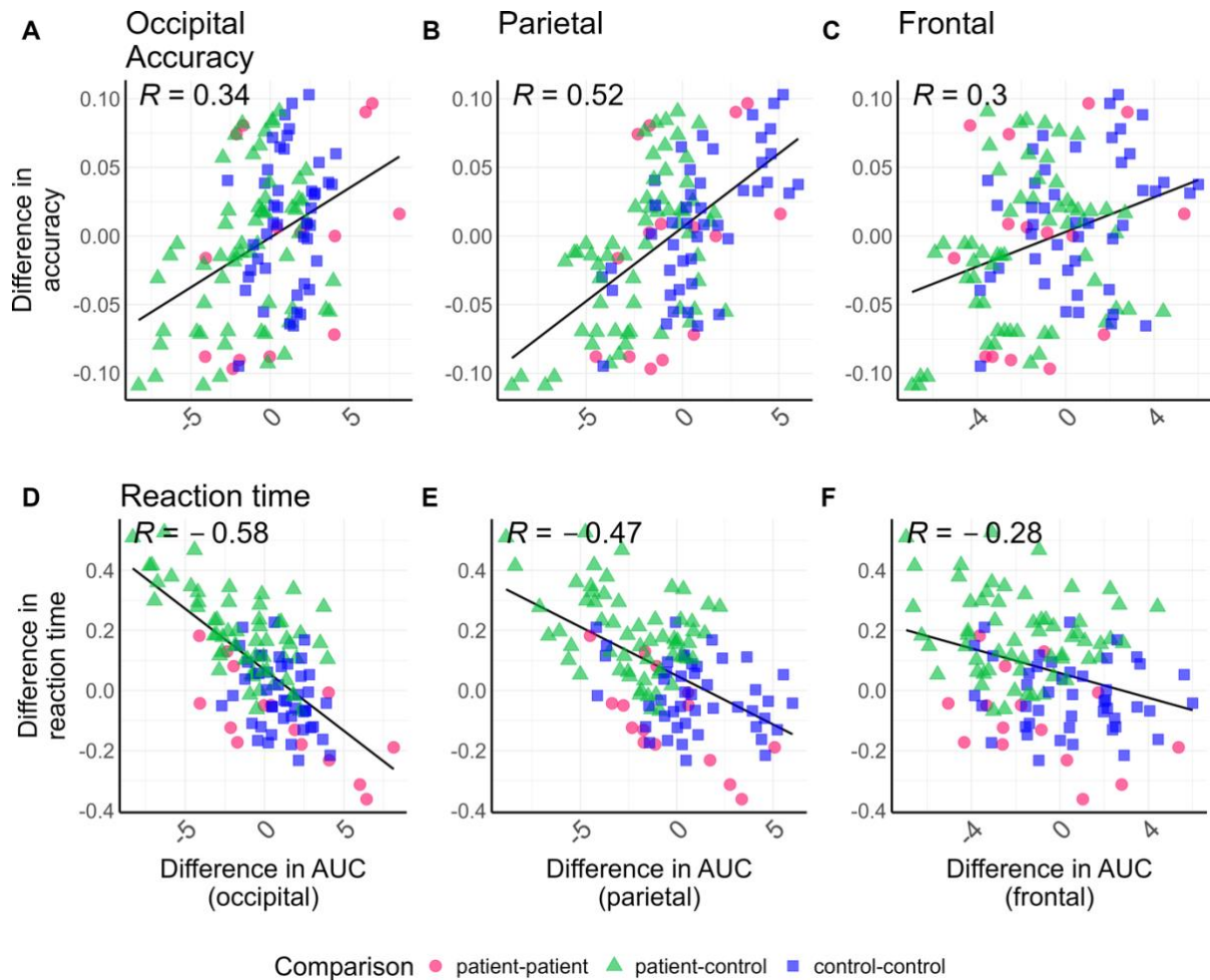


Figure 44. Scatterplots showing relationship between inter-individual dissimilarity in neural measures (i.e., attentional prioritisation values for each sensor space region: sensors over occipital, leftmost panels, Panels A and D; parietal, middle panels, Panels B and E; and frontal, rightmost panels, Panels C and F) and inter-individual dissimilarity in behavioural measures (i.e., accuracy, top panels, Panels A to C; and reaction time (RT) bottom panels, Panels D to F). For interest, for each measure, differences between two patients are shown in pink; differences between a patient and a control participant in green; differences between two controls in blue. Black lines show linear relationship between these dissimilarity profiles. R values shown are Spearman correlation coefficients. Inter-individual differences in attentional prioritisation (AUC) for sensors over occipital regions were significantly correlated with inter-individual differences in RT (Panel D). No other correlations were significant after multiple comparison correction (FWER across three neural measures).

4.6 Discussion

In this study, I used MEG with MVPA to explore how focal parietal lesions affect spatial and feature attention processes. Specifically, I examined how parietal damage in six patients affects the neural prioritisation of relevant information at an individual level, and whether individual differences in the neural prioritisation of relevant information relate to behavioural performance. This study was a proof-of-concept examination of the effects of parietal damage on attentional processes. With so few patients (and particularly given the variation in lesion and patient characteristics therein), and with only 10 controls against which to compare patient performance, the conclusions I can draw from this study are limited.

This caveat notwithstanding, I provide preliminary evidence that patients with parietal lesions can show subtle differences in the strength and dynamics of attentional prioritisation, despite high behavioural accuracy. I focus on four observations. First, there was limited evidence of attentional impairments detected in these patients on neuropsychological measures and on the behavioural task. Notably, patients were able to successfully complete the main experimental task to a high level of accuracy (at a similar level to controls) but were slower than control participants. Second, individual differences in reaction time were correlated with differences in strength of attentional prioritisation in sensors over occipital regions, providing some evidence of a relationship between neural and behavioural measures. Third, one patient, sub-06 showed evidence of reduced attentional prioritisation, against a background of good (albeit slower) behavioural performance. Finally, I consider the observation that, overall, there was limited evidence of altered neural attentional processing for most patients.

4.6.1 Neuropsychological and behavioural performance shows limited evidence for attentional impairments in patients

Parietal lesions are typically associated with spatial attention deficits, most commonly following right hemisphere lesions (e.g., Mort et al., 2003), although spatial biases have also been observed in patients with left hemisphere lesions (e.g., Peers et al., 2005). In contrast, the parietal patients in this study did not demonstrate clear evidence of spatial bias on the neuropsychological measures administered. Indeed, on these measures, only one patient (sub-03) demonstrated unilateral (left) bias that differed from controls. This observation was broadly inconsistent with their right hemisphere lesion (although, right-sided inattention can follow right hemisphere lesions, e.g., Moore et al., 2023; Ten Brink et al., 2017) and was only observed on one of three neuropsychological measures assessing spatial bias (Hearts Cancellation).

That spatial bias was not detected in other patients, or for sub-03 on other spatial bias tasks, is not wholly unexpected. Damage to parietal lobe does not necessarily result in spatial attention deficits (e.g., Karnath et al., 2011; Molenberghs et al., 2012; Moore et al., 2023). Moreover, all patients in this study were assessed at the chronic stage, and none reported observable spatial attention deficits in their everyday life. One further caveat is that the tests comprising the OCS-BRIDGE neuropsychological battery primarily function as screening tools rather than as comprehensive assessments of cognitive ability. However, the most sensitive spatial attention measure used in this study was the CamTrip, and on this task no patient showed a spatial bias outside the range of controls.

Interestingly, while sub-03 showed some evidence of left bias on neuropsychological testing, this was not reflected in their behavioural performance on the main experimental task, with high accuracy and fast reaction times in keeping with control performance when attending to either side of space. In terms of neural data, while sub-03 showed reduced feature rule decoding relative to controls, they showed location rule decoding and strength of neural attentional prioritisation in keeping with that seen in

controls, across both hemispheres. While spatial bias was not directly tested in the neural data, the only possible difference I observed was visibly faster and higher attentional prioritisation in this patient's unaffected occipital regions relative to their affected occipital region, but this was after the time window of interest that I analysed formally. Therefore, there were no clear indicators, for this individual, that the potential spatial bias revealed in the Hearts Cancellation test affected performance on the other neuropsychological tests, the main task, or the neural data. Conversely, sub-06, who demonstrated both reduced location and feature rule decoding, and reduced strength of attentional prioritisation relative to controls, did not differ from controls on neuropsychological measures.

On the main experimental task, there was evidence for spatial bias or inattention in accuracy for one patient, and in reaction time for a different patient. While sub-01's accuracy when attending to their ipsilesional side was equivalent to control performance for that side, their accuracy for targets on their contralesional side was significantly lower than control performance. Moreover, this difference in accuracy between the two hemifields, favouring the ipsilesional over the contralesional visual field, was significantly greater than the equivalent difference in controls. This suggests that sub-01 exhibited some inattention to their contralesional side on this task. However, there was no evidence of spatial bias in their reaction time or on neuropsychological measures. A different patient, sub-07, showed evidence of spatial bias in their reaction times. For sub-07, the difference in reaction times between the two hemifields was significantly greater than the equivalent difference in controls, favouring the ipsilesional over the contralesional visual field, and they were slower than controls for both hemispheres. However, there was no evidence of equivalent bias in their accuracy on the main task or on neuropsychological measures.

Overall, then, there were hints of lateralised spatial inattention in behaviour for three of the six patients, but the pattern was not consistent across measures for any patient. As a cohort, patient performance on neuropsychological measures also did not suggest substantial difficulties with attention or working memory. This was perhaps reflected in the good behavioural accuracy observed in the experimental task and reasonable

neural attentional prioritisation seen for most patients (see later). Indeed, despite initial concerns that patients might find the task too demanding, all participants were able to complete the task to a high level of accuracy, in keeping with the age-matched controls tested here, and also broadly in keeping with the performance of younger controls (as in Chapter 3, and Goddard et al., 2022). Patients were, however, slower than controls at the group level, and four patients were significantly slower than controls individually: sub-01, sub-04, sub-06, and sub-07.

In summary, behavioural performance indicated patients were able to do the task as well as controls, albeit slower in some cases. Moreover, taken together with neuropsychological test performance, the evidence for attentional deficits was mixed, with no clear or consistent indication of attentional deficits or spatial bias in any individual patient across the various measures.

4.6.2 Preliminary evidence that reaction time on the main task relates to strength of neural attentional prioritisation

One aim of this study was to examine whether the strength of attentional prioritisation related to behaviour. To address this, I asked whether dissimilarity between participants in their behavioural performance (i.e., accuracy and reaction time) correlated with dissimilarity on neural measures (i.e., attentional prioritisation across the three sensor space regions). I found that inter-individual differences in reaction time were significantly correlated with differences in attentional prioritisation in occipital regions. This indicated that participants who were more different in their reaction time were also more different in their ability to neurally prioritise relevant information, as detected in sensors over occipital regions. Moreover, the negative correlation implied that faster reaction times were associated with stronger attentional prioritisation in these regions (and vice versa). This thus provides preliminary evidence linking strength

of neural prioritisation in the first 600ms, in occipital regions, to behavioural performance.

A growing body of MEG and electroencephalography (EEG) work shows attentional prioritisation effects within this time window, with several studies showing dominance of task-relevant coding emerging as early as 200ms and sustained through time (Barnes et al., 2022; Battistoni et al., 2020; Goddard et al., 2022; Grootswagers et al., 2021; Moerel et al., 2021, 2022; Noah et al., 2023). The preliminary finding in this study adds to this literature by suggesting that early prioritisation of task-relevant information in occipital regions may have a causal relationship with behaviour. This, moreover, accords with other work that has also demonstrated a link between occipital codes and behaviour. For instance, task-related representations have been shown to be stronger on correct compared to incorrect trials in parts of occipital cortex (Williams et al., 2007), and distortions in the fidelity of working memory representations in early visual cortex have been shown to predict memory errors (Hallenbeck et al., 2021).

I did not find significant relationships between dissimilarity in reaction time and strength of attentional prioritisation for other regions, nor between dissimilarity in accuracy and attentional prioritisation in any region. The lack of observed relationships between other behavioural and neural measures is perhaps surprising. Given the purported critical role of frontoparietal regions in selective attention processes, one might predict the strength of attentional prioritisation in these regions to show a close link to behaviour (although see Zheng et al. (2024) and Chapter 5 for discussion on the difficulties of relating neural codes to behaviour). However, the relationships between behavioural and neural measures were consistently in the expected direction across all sensor space regions (i.e., positive between dissimilarity in accuracy and neural measures, and negative for dissimilarity in reaction time and neural measures). Given the small sample size, these results warrant further study in a larger cohort. In addition, it is worth noting the approach taken here considered strength of attentional prioritisation over the first 600ms only. Thus, reduced attentional prioritisation scores may reflect latency in the prioritisation of task-relevant information and/or overall reductions or inefficiencies in prioritisation. This approach is insensitive to neural

prioritisation that occurs later in the timecourse. It would be interesting in the future to explore the rest of the timecourse in more detail, for example by calculating prioritisation scores within a time window before the response or by focussing on a time window over peak attentional prioritisation in controls. This would provide more information on how the dynamics of information coding relates to behavioural performance.

4.6.3 Changes in attentional prioritisation dynamics detected in one patient

Of the six patients who participated in this study, the results of sub-06 warrant further consideration. Sub-06 was the only patient that reliably showed reduced rule decoding and attentional prioritisation, relative to controls. This patient showed significantly reduced attentional prioritisation in sensors over the entire affected hemisphere as well as in sensors over unaffected occipital regions, in some cases appearing, if anything, to prioritise the incorrect information at early timepoints. In addition, while prioritisation in their unaffected parietal and frontal regions were not significantly reduced relative to controls, it was amongst the weakest seen across all controls and patients, particularly in the time window that I assessed formally. These findings point to a broad reduction in attentional prioritisation following their damage, consistent with their reduced coding of the task rules and with the hypothesis that damage to the MD system may lead to a reduction in attentional prioritisation throughout the brain. However, this patient's lesion was the most extensive of the cohort, extending into frontal lobe. It is thus unclear which characteristics of the patient's lesion (e.g., its large extent or extension into frontal lobe) could account for the observation of altered neural processing in this patient. A more nuanced examination, ideally at the source level, in future work could help to elucidate whether and how damage to the MD system affects the processing of task-related information in both affected and unaffected hemispheres.

This patient also showed stronger attentional prioritisation in sensors over unaffected relative to affected parietal regions, relative to the corresponding subtraction in controls. It is difficult to completely disentangle whether this reflects compensation in the unaffected side (particularly given the, presumably, poorer input from occipital lobe), or just more extreme difficulty over the affected side. Since prioritisation in sensors over the unaffected parietal lobe was not significantly different from controls, perhaps the difference primarily reflects lower prioritisation in the affected hemisphere, rather than any compensatory increase in prioritisation in the unaffected hemisphere (see Figure 43).

It is important to note that these findings of reduced attentional prioritisation occurred against a backdrop of little suggestion of attentional difficulties on neuropsychological measures (note, however, sub-06 showed slightly poorer top-down control relative to the control group on the CamTrip task (albeit not significant after multiple correction), possibly indicating poorer ability to inhibit response to irrelevant information). In addition, this patient's behavioural performance was not significantly lower than controls for either ipsilesional or contralesional hemifields. It is interesting that reduced attentional prioritisation in this individual patient was not reflected in these measures. However, numerically, sub-06's accuracy was among the lowest of all participants (along with sub-01, accuracy overall was 83.9%). Since neural measures were calculated on the basis of correct trials only, fewer trials contributed to calculating this patient's attentional prioritisation timecourse compared to other participants, potentially leading to a more variable timecourse. Indeed, sub-06 and sub-01 (i.e., the two participants with lowest accuracy) show visually different task-related decoding traces. However, the difference in behavioural accuracy between sub-06 (and sub-01) and some controls was minimal, and thus it is unlikely that differences in attentional prioritisation were due to trial count alone.

However, sub-06 showed slower behavioural performance for items across the visual field. This may have been related to their significantly reduced attentional prioritisation in the first 600ms, which in this case appeared to reflect both an overall reduction in attentional prioritisation and an increase in the *latency* of attentional prioritisation.

Indeed, visual inspection of this patient's attentional prioritisation traces (see Figure 42) suggest that they did eventually show modest attentional prioritisation but that the onset of this was extremely late at around 1000ms (compared to around 300ms for the control group). This could explain their long reaction times. In addition, their attentional prioritisation profile, even at this late stage, was far less pronounced than other participants, perhaps reflecting their relatively poor behavioural accuracy while at the same time providing a neural basis for their being able to complete the task at all. Future work could explore this further by examining the whole timecourse.

One other consideration is that this patient was the only patient to perform the task with their non-dominant hand REDACTION: Medical information removed. Personal data removed for confidentiality reasons. It is not clear whether this influenced this patient's results. Sub-57 (a demographic-matched control of sub-06) also completed the task with their non-dominant hand, with little obvious consequence (on the basis of visual inspection) for their behavioural performance or coding of task-relevant information (Figure 41).

4.6.4 Limited evidence of altered neural processing in most patients

Parietal lobe is implicated in selective attention (Behrmann et al., 2004; Duncan, 2013; Duncan et al., 2020), particularly spatial attention (Husain & Nachev, 2007), and thus I had anticipated that damage to these regions would likely be reflected in altered neural processing of task-related information during the experimental task. While subtle differences in neural prioritisation of relevant information were detected, particularly for sub-06 as detailed previously, overall, the results did not point to altered neural processing for most patients. For instance, in most patients, strength of location and feature rule information coding within the first 600ms was similar to that seen in controls. Only two patients showed comparatively weaker rule coding within this window: sub-06 showed weaker coding for both rules, while sub-03 showed weaker feature rule coding. While not formally tested, even these patients appeared to show

evidence of coding of rule information: decoding traces for all patients appeared well above chance and followed a similar pattern to controls.

In addition, all patients except sub-06 showed attentional prioritisation within the first 600ms at a strength similar to controls. Indeed, visual inspection of stimulus decoding traces (Figure 41) and attentional prioritisation timecourses (Figure 42) indicates that only sub-06, and potentially sub-01, showed signs of altered neural attentional processing. While not formally tested, it is interesting to note that the other four patients appeared to show clear above-chance coding of relevant information, and these decoding traces visually resembled that of controls (i.e., initial representation of all information, with only a sustained selective representation of relevant information maintained across time).

There are several possible interpretations for the lack of clear neural effects for most patients. First, the lack of effects could reflect limitations of the method. For instance, my measure of attentional prioritisation was reasonably coarse, quantifying relative coding of relevant information across a set time window that was linked to stimulus onset. This window was chosen on the basis of previous work suggesting attentional prioritisation effects occur within this window (e.g., Goddard et al., 2022), and to avoid any possible contamination from reaction time effects. However, for patients and controls, selective task-relevant coding and peak attentional prioritisation seems to occur after this time window. Thus, this approach may have underestimated the true extent of neural differences between patients and controls. Indeed, reaction time varied across participants and there is a hint, in visual inspection of the stimulus decoding traces (Figure 41), that reaction times follow a period of selection prioritisation of relevant information. Thus, reduced attentional prioritisation may reflect an overall reduction in this measure, or increased latency in prioritising relevant information. It would be interesting for future work to consider realigning the data to response times and/or quantifying attentional prioritisation over a set time window prior to the time of each participant's response.

Relatedly, my approach was reasonably coarse in terms of being able to detect other types of neural differences. For instance, I primarily focussed on identifying weaker

coding or reduced attentional prioritisation relative to controls. However, neural inefficiencies or alterations in neural processing could also be reflected in a slower attentional prioritisation timecourse. While a slower attentional prioritisation timecourse might be reflected in a lower attentional prioritisation score, I did not directly test whether patients showed latency in attentional prioritisation. There is some suggestion in the data that patients may have exhibited slower prioritisation than controls. For instance, for the affected hemisphere, sub-01, sub-03, sub-06 and sub-07 show hints of a later peak in attentional prioritisation than controls. It would be interesting to explore this with a more nuanced measure in future work by, for instance, comparing time to peak between patients and controls.

It is also worth considering whether the lack of significantly altered neural effects in this study may reflect something about the patients I recruited. All patients were chronic lesion patients, but pronounced spatial inattention (e.g., neglect) is typically associated with the acute stage post-insult (Karnath et al., 2011). I had hoped that this task would be demanding enough to detect subtle spatial attention deficits, however, it is possible that most patients in this study either never had spatial attention deficits or had recovered sufficiently from their injury (e.g., due to neural reorganisation and/or recovery of functional network connectivity). Future work could benefit from using additional analyses to test for compensatory mechanisms (e.g., by examining whether increased activity in the unaffected regions carries additional task information that enhances task performance, as in Knights et al., (2021)). Alternatively, one possible interpretation is that these results indicate that parietal regions are not critical for selective attention, contrary to the attentional episodes framework (Duncan, 2013; Duncan et al., 2020). Yet given the limitations of the method, the small control sample size, and the mixed results across individuals, it is unclear whether the observed lack of neural effects in most patients reflects a true absence of altered attentional processing.

4.6.5 Additional limitations and future directions

The conclusions presented in this work are mostly speculative as they depend strongly on what was observed in the control sample, which was small. Moreover, the strength of the conclusions one can draw from focal lesion patients are complicated and tempered by heterogeneity in lesion and patient characteristics (see Chapter 1). While these limitations are characteristic of focal lesion studies and not restricted to the current work, it is important to note that they are arguably more pronounced here as I examine profiles of individual patients against a small group of control participants. As highlighted recently by Marek et al. (2022), large sample sizes are needed to reliably detect the subtle and often smaller-than-expected effect sizes typical of studies exploring brain-behaviour relationships. Future work with a larger sample (of controls and/or patients) would permit more powerful group-level analyses across groups of patients (e.g., left lesion and right lesion, MD lesion and non-MD lesion), and more reliable comparisons of individual patient profiles.

Another potential limitation to consider is the implicit assumption that differences in the strength of attentional prioritisation are meaningful. Most decoding work compares decoding accuracy against chance, with the inference being that if two categories can be discriminated at an accuracy greater than chance, information about those categories exists in the neural data. In this work, I go a step further, and instead aim to interpret difference in decoding accuracies across individuals. Yet inferring relative representational strength from differences in decoding accuracy can be problematic in certain circumstances. For example, when comparing different regions of the brain, differences in decoding strength may stem from extraneous factors (e.g., differences in vasculature in fMRI, distance from sensors and folding in MEG). Similarly, variation in decoding strength across individuals may reflect idiosyncratic anatomical or physiological characteristics (e.g., head shape, cortical folding), which will tend to obscure true relationships between neural decoding and individual differences in behaviour (Zheng et al., 2024). Here I attempt to account for this by comparing individuals on the basis of a difference measure of attentional prioritisation (i.e.,

strength of relevant over irrelevant information coding). This aimed to make for a fairer comparison between individuals by considering an individual's decoding strength in the context of their baseline performance. However, given that the baseline (i.e., irrelevant information coding) was largely at chance, this would have had minimal impact. Thus, the finding that attentional prioritisation was related to reaction time is important for the interpretation that individual differences in decoding strength are meaningful. More work is needed to understand the neural generators of differences in decoding accuracies between conditions, regions and individuals. In addition, further work is required to elucidate whether the relationship between inter-individual differences on neural and behavioural measures found in this work holds in larger cohorts; this will help to clarify the significance of differences in representational strength.

I note also that this work was done on sensor space. It was unclear how brain damage and factors relating to this (e.g., craniotomy) might affect source space reconstruction, especially in the context of MVPA, and there is limited precedent in the field to vouch for the validity of such an approach. Accordingly, as a proof-of-concept study, I decided to restrict the analysis to sensor space. Naturally, this affects the spatial specificity of my conclusions and accordingly I have endeavoured to be cautious in spatial interpretations. However, it would be interesting in the future to carry out source-space estimation in order to consider a more spatially nuanced account of how parietal damage affects attentional prioritisation and/or whether there are any detectable compensatory mechanisms. The findings of the current work can inform future studies which explore these complexities.

This study was conducted using correct trials only and thus an acceptable level of accuracy was required to be included in the study. I did not exclude any patient or control on the basis of behavioural performance, but I did consider likely suitability in this regard during recruitment. Accordingly, it was perhaps unlikely that the patients in this study would show vastly different neural profiles to controls, especially at the coarse sensor-space level. It would be interesting to examine how neural representations change following errors on this task (e.g., as in A. K. Robinson et al., 2022), or in patients with poorer accuracy. However, to look at representational content using linear

classifiers, a large number of trials (of correct trials, as is the case here, or a high number of errors if studying error trials) is needed. An interesting future avenue could be to collect multiple sessions of data from a single or small set of patients who perform poorly on this task. This would permit an exploration into representational dynamics when behaviour is more affected.

Finally, I note that in this study I examined differences in strength of neural attentional prioritisation in sensors over affected and unaffected occipital, parietal and frontal regions, rather than differences in attentional prioritisation for stimuli on the contralesional and ipsilesional side of space. This was motivated by the aim of exploring how parietal damage affects neural attentional prioritisation processes across the brain, and the observation of minimal behavioural attentional bias in these patients. Nonetheless in future work it would be interesting to examine the neural data for evidence of attentional bias. For instance, attentional bias could manifest as stronger attentional prioritisation when attending to the ipsilesional hemifield, relative to the contralesional hemifield. In addition, previous work has indicated patients with spatial bias may still demonstrate attentional deficits when attending to their ipsilesional visual field. For example, Duncan et al. (1999) found bilateral reductions in visual processing in left-sided neglect, and Snow and Mattingley (2006) found impairments in the inhibition of task-irrelevant information in ipsilesional visual field. Accordingly, it would be interesting to explore what task-related information is represented when attending to contralesional and ipsilesional hemifields. Having now established that the approach taken in this work can explore individual-level attentional prioritisation processes following brain damage, studying how representational content changes with respect to spatial bias would be an exciting next step.

4.6.6 Conclusions

This work served as a proof-of-concept exploration of how parietal damage might affect neural attentional processes, and how these changes might relate to behaviour. Despite varying degrees of parietal damage, all patients were able to complete the task as successfully, albeit in some cases more slowly, than controls. Against this background of high behavioural accuracy, I provide early evidence that parietal damage can affect coding of rule information and the relative prioritisation of relevant information in the brain. Moreover, this work provides preliminary evidence that neural attentional prioritisation relates to behaviour, with individual differences in strength of attentional prioritisation (sensors over occipital regions) related to individual differences in behavioural reaction times. Overall, this study highlights the nuanced effects chronic parietal damage can have on the representation of stimulus information during selective attention and underscores the potential of MVPA to probe subtle neural consequences that can follow brain injury.

4.7 Afterword

This study explored how the neural representation of relevant over irrelevant information (i.e., attentional prioritisation), and the timecourse over which this information is prioritised, changes following chronic parietal damage in six patients. The work from Chapter 3 and Goddard et al. (2022) used the same spatial and feature attention paradigm featured in this chapter and provided a spatial and temporal account of where and when relevant information is neurally prioritised in young, healthy adults. Namely, these studies showed that, following an initial period where all stimulus information is represented, relevant information is ultimately prioritised in the brain.

It was unclear when I started this study whether I would be able to look for this pattern at an individual level, particularly in individual patients. Yet I have shown that attentional prioritisation was visually detectable at an individual level, and moreover, that nearly all parietal patients showed strength of attentional prioritisation at a similar level to controls. In the context of the other uncertainties I detailed in the Foreword to this chapter, this alone was an exciting result. I also have been able to demonstrate that reaction time performance on the main experimental task has some relationship to my neural measure, providing preliminary evidence that my method of examining neural representations might be capturing something meaningful about behaviour. How parietal damage affects neural attentional prioritisation dynamics is still an open question, however. I have been able to show some evidence of a reduction in attentional prioritisation for one patient that may, in part, be linked to their slower reaction time. However, my attempt to probe individual-level changes following parietal damage was coarse, constrained by variation in patient, lesion and recovery characteristics and the small number of controls I had in this study.

I have presented a first account of how this approach could be used to probe neural changes following focal, chronic lesions. It by no means represents a comprehensive account of how parietal damage can affect neural attentional prioritisation dynamics.

Indeed, many interesting questions remain unanswered. For instance, all the patients were able to complete the demanding spatial and feature attention task with high behavioural accuracy, despite, in some cases, rather large lesions – how did this transpire in the brain? Does slower reaction time in some cases reflect that task information was processed through a circuitous route? Can we see evidence of compensatory processes by examining the whole timecourse of information coding and the exchange of these codes between brain regions?

I have now established that we can decode meaningfully from individual older controls and focal lesion patients, and that this approach of examining attentional prioritisation might allow us to detect subtle changes in neural codes. Accordingly, I hope we may be able to answer these questions in future studies. Indeed, I consider a real strength of this study to lie in the potential avenues it now opens for future research.

4.8 Supplementary Materials for Chapter 4

4.8.1 Additional behavioural plots

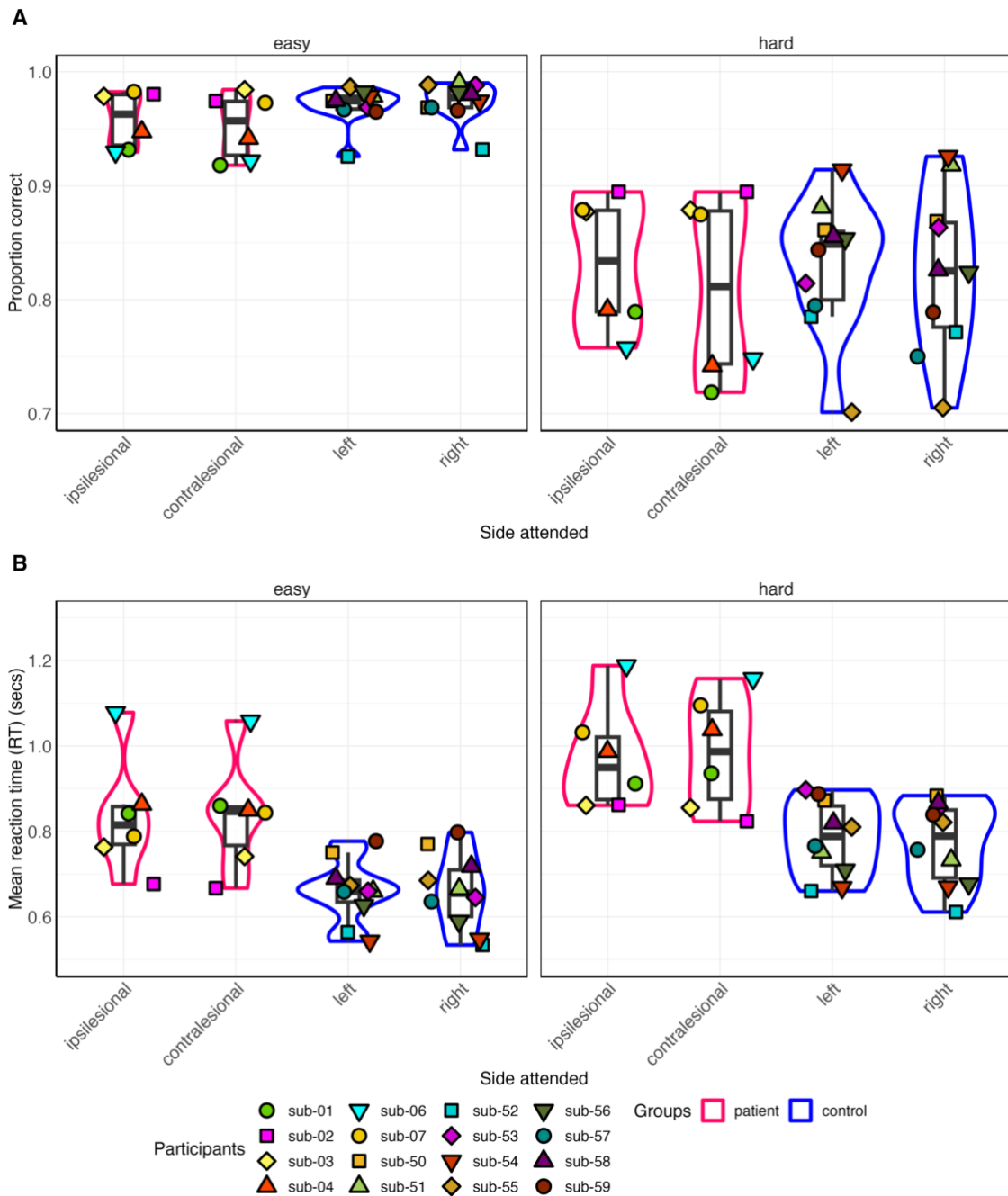


Figure S 8. Behavioural performance for accuracy (Panel A) and reaction time (Panel B) for patients (pink) and controls (blue). Performance metrics split by side attended (ipsilesional and contralesional side for patients; left and right for controls), and task difficulty, and thus are averaged over feature attended (e.g., attend colour, attend shape). Task difficulty is defined by proximity of the attended feature dimension of the target to the decision boundary, where ‘easy’ trials involved a categorisation far from the decision boundary (i.e., feature steps 1 and 4) and ‘hard’ trials involved a categorisation close to the decision boundary (i.e., feature steps 2 and 3, see Figure 35). As anticipated, visual inspection of the above suggests participants (patients and controls) were more accurate and faster when making easy relative to hard categorisations.

4.8.2 Individual rule decoding plots

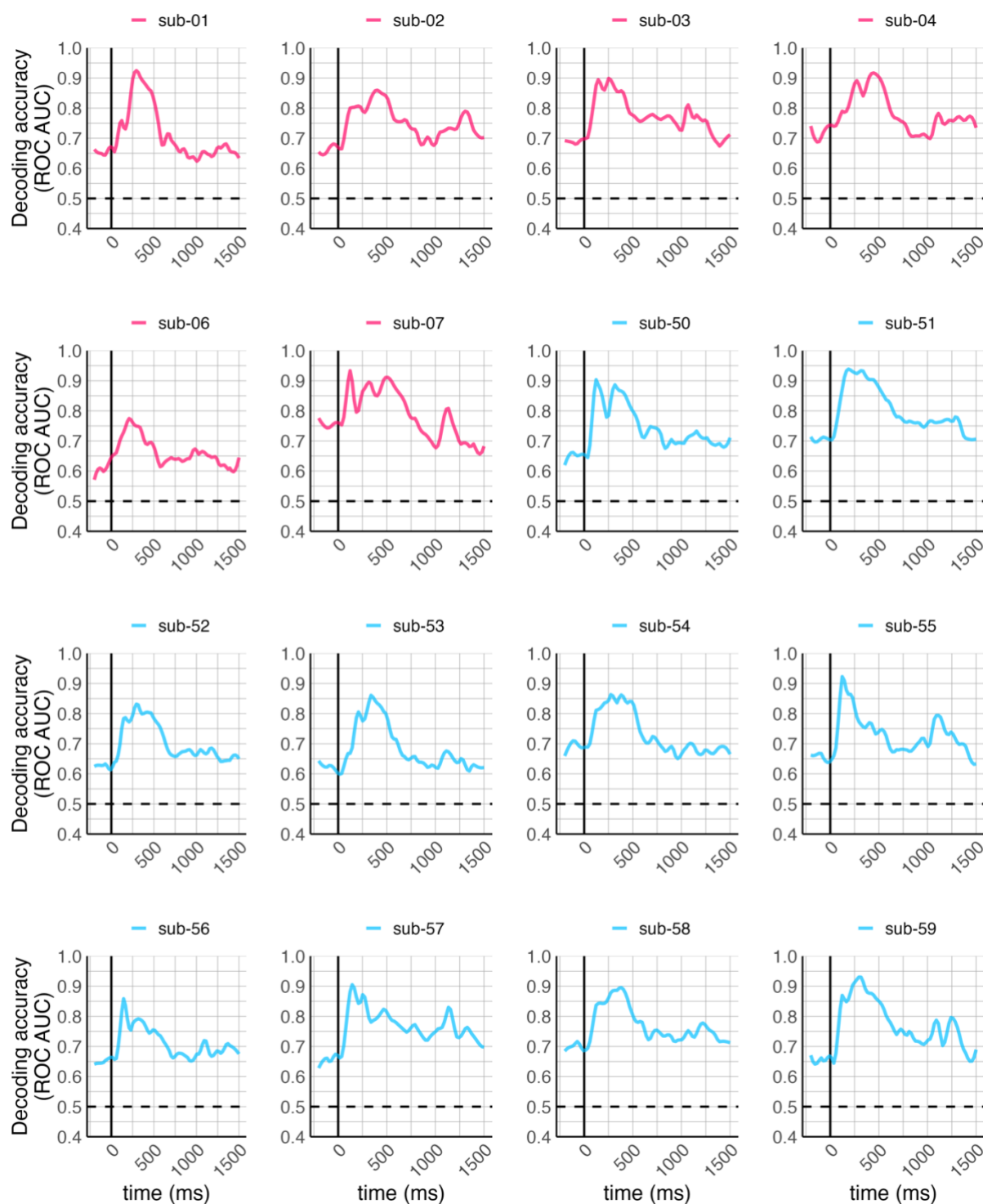


Figure S9. Rule decoding for target location for each patient (pink lines) and each control (blue lines). Each plot shows decoding accuracy (ROC AUC) for classification of attend left from attend right instruction from -200 to 1500ms, relative to stimulus onset (shown with vertical black line at 0) across all sensors (whole brain). Chance is 0.5 and is indicated by the dotted horizontal black line. Note – decoding accuracy is above chance prior to stimulus onset for every participant. This is expected as task instructions were blocked. Plots are smoothed using loess method.

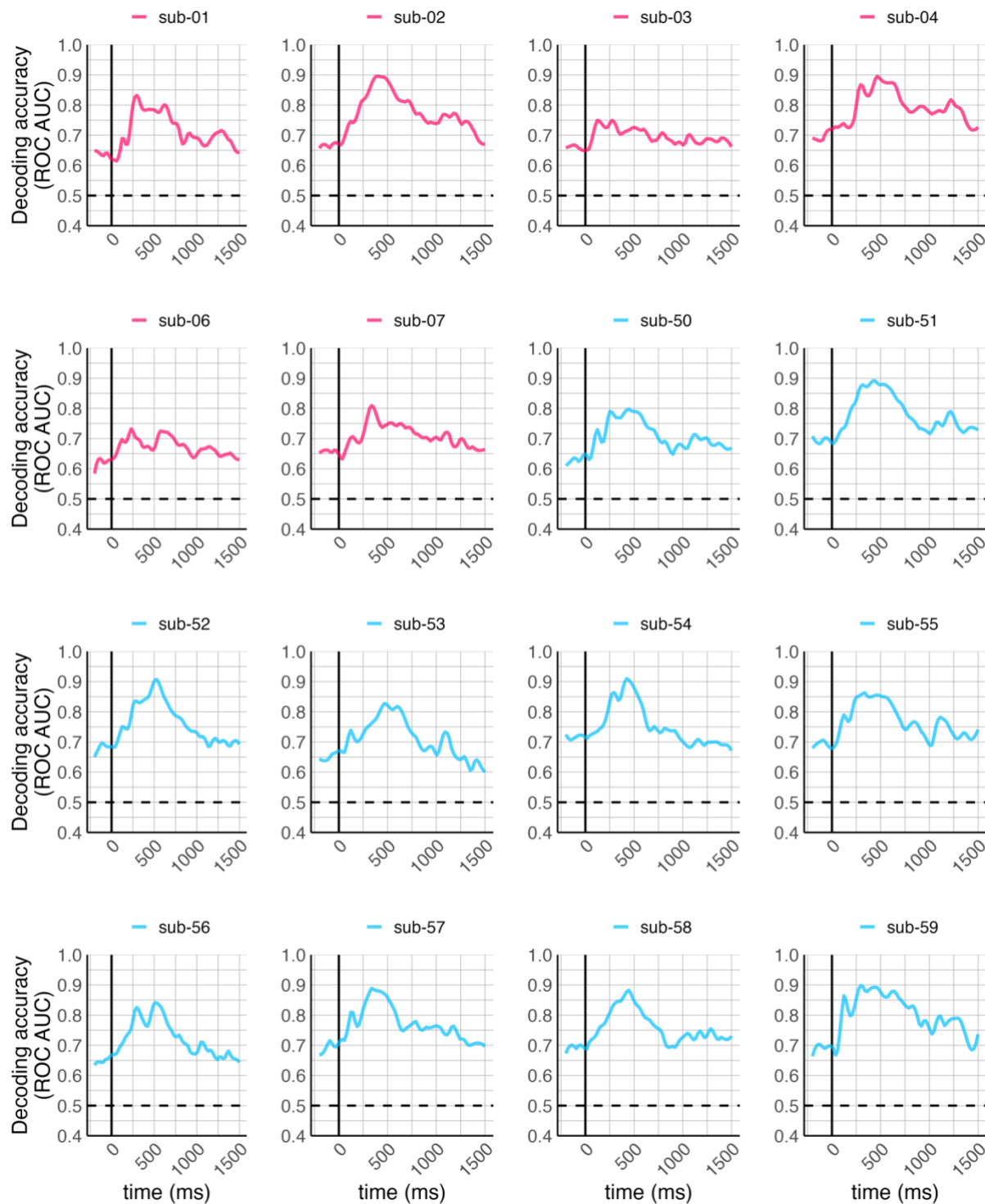


Figure S10. Rule decoding for target feature for each patient (pink lines) and each control (blue lines). Each plot shows decoding accuracy (ROC AUC) for classification of attend colour from attend shape instruction from -200 to 1500ms, relative to stimulus onset (shown with vertical black line at 0) across all sensors (whole brain). Chance is 0.5 and is indicated by the dotted horizontal black line. Note – decoding accuracy is above chance prior to stimulus onset for every participant. This is expected as task instructions were blocked. Plots are smoothed using loess method.

4.8.3 Individual attention prioritisation timecourses

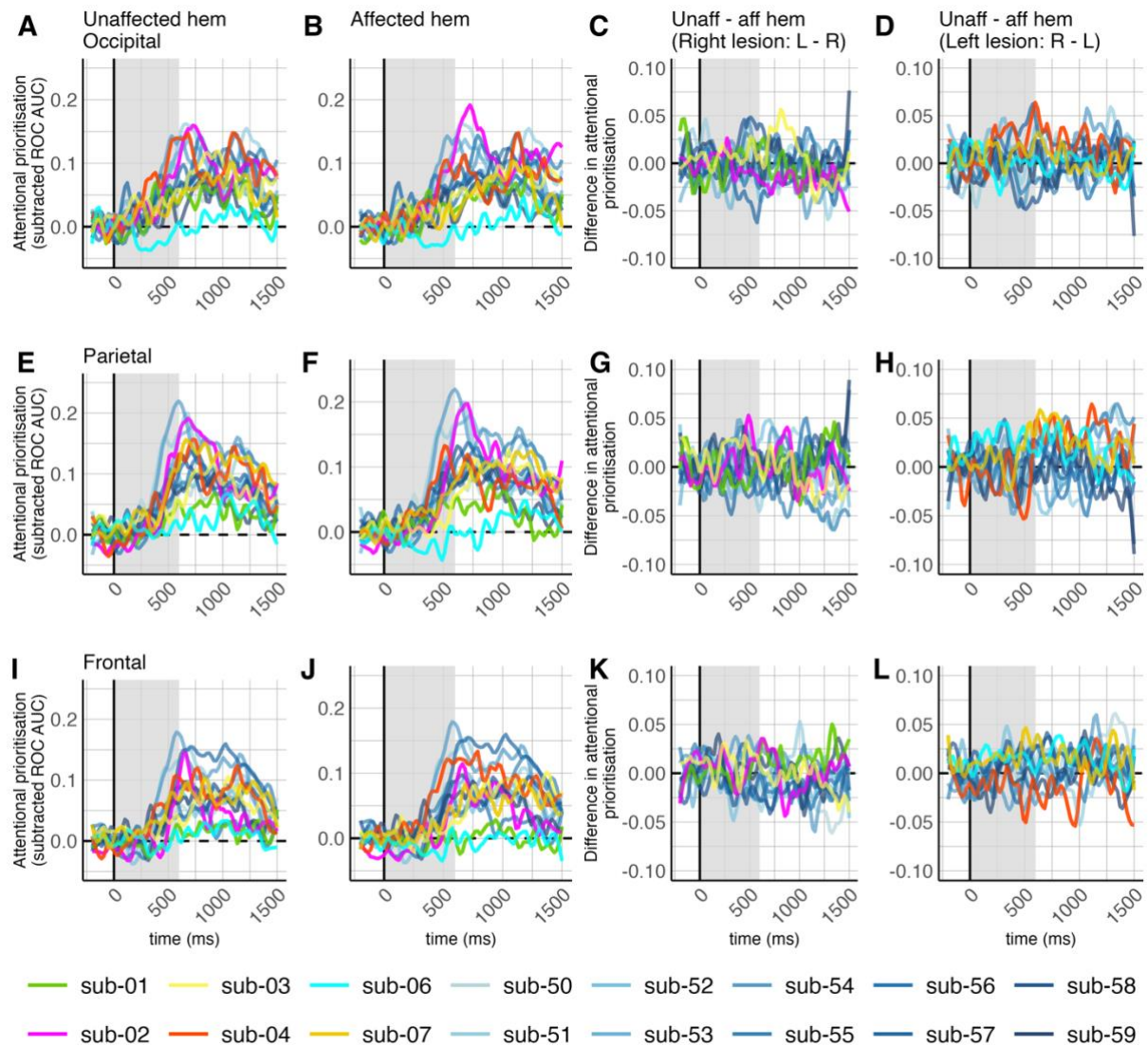


Figure S II. Individual attention prioritisation timecourses (c.f. Figure 42) for sensors over occipital (top panels, Panels A to D), parietal (middle panels, Panels E to H) and frontal (bottom panels, Panels I to L) regions. Attentional prioritisation reflects the relative strength of relevant over irrelevant information representation across time, from -200ms to 1500ms relative to stimulus onset (shown with vertical black line at 0ms), with values above zero indicating stronger representation of relevant information. Accordingly, attentional prioritisation at zero (shown as the dotted horizontal black line) reflects no relative representation of relevant information over irrelevant information. Each patient's attentional prioritisation timecourse is shown for sensors over their unaffected (first column, Panels A, E, I) or affected (second column, Panels B, F, J) hemisphere separately; for controls (shown in shades of blue), attention prioritisation is averaged over both hemispheres and thus is the same for the unaffected and affected plots of a given set of sensor regions. The difference in attentional prioritisation between unaffected and affected hemispheres is shown against the relevant subtraction in controls (third and fourth columns). That is, for right lesion patients (third column, Panels C, G, K), the unaffected minus affected hemisphere subtraction for these three patients is shown against left minus right hemisphere in controls; for left lesion patients (fourth column, Panels D, H, L), the hemisphere subtraction for these three patients is shown against right minus left hemisphere in controls. Shaded area in grey (0-600ms) reflects the time window used to calculate overall attentional prioritisation score or AUC. Plots are smoothed using loess method.

Chapter 5 |

General discussion

The work reported in this thesis investigated how a network of frontoparietal regions, the multiple-demand (MD) system, gives rise to complex human cognition and behaviour, with a particular focus on its role in fluid intelligence and attention. I sought to address the following research questions:

Chapter 2	Which brain networks are critical to fluid intelligence?
Chapter 3	What task information is prioritised in the MD system when spatial and feature attention are used together?
Chapter 4	How is the representation and processing of task-related information affected following damage to the MD system (specifically parietal lobe)?

To address these questions, I conducted three studies using a range of different methodological approaches, including focal lesion models; quantification of grey matter (GM) resting-state network damage and white matter (WM) network disconnection; multivariate pattern analysis (MVPA) with functional magnetic resonance imaging

(fMRI) and magnetoencephalography (MEG) data; and MEG-fMRI fusion. In the following section, I will summarise the findings of these studies. I will then discuss the implications of these results for our broader understanding of how frontoparietal regions support intelligent human behaviour. Although I have considered specific limitations for each study within their respective chapters, in this chapter, I will discuss some general limitations of this work and of the methods used, along with the challenges these pose for interpretation. I will also mention some exciting possible avenues for future research that stem from the current work before concluding the thesis.

5.1 Summary of findings

5.1.1 Chapter 2: Neuropsychological evidence for the role of frontoparietal cortex in fluid intelligence

In this chapter, I used a focal lesion approach to examine the causal contribution of different resting-state networks to fluid intelligence. I drew on a large cohort of focal lesion patients ($N=174$) to relate the extent of GM and WM damage to seven canonical resting-state networks (Yeo et al., 2011) to fluid intelligence deficit. I showed that damage to frontoparietal network (FPN), a resting-state network closely related to the MD system, predicted fluid intelligence deficit over and above any contributions from other networks. However, this result could not be separated from total lesion volume, limiting the strength of conclusions that could be drawn. Motivated by previous work highlighting the importance of WM connections in facilitating intelligence, I also tested whether my two approaches for quantifying network damage implicated different networks in fluid intelligence. They did not; both approaches implicated similar networks and were only significant when considering GM network damage. When only GM damage was considered, damage to both FPN and dorsal attention network (DAN) predicted fluid intelligence deficits over contributions from other networks, but not over total lesion volume. These findings suggest that, of all resting-state networks, FPN may be particularly important in supporting fluid intelligence. These findings also highlight some of the challenges faced when using lesion data to establish clear brain-function relationships.

5.1.2 Chapter 3: Spatial and feature-selective attention interact to drive selective coding in frontoparietal cortex

Next, I used fMRI and MVPA to consider the role of the MD system in selective attention. I scanned healthy adults ($N=30$) while they completed a demanding spatial (attend left or right object) and feature (attend colour or shape) attention task in the fMRI scanner. I examined what task information is prioritised in the MD system when these two types of selective attention are deployed. I showed that MD regions selectively coded behaviourally relevant information at the intersection of the two attention types, emphasising that MD codes are tightly linked to behaviour. This selectivity was also seen across the brain, although it was strongest in MD regions. I also showed that MD codes reflected both task difficulty and stimulus categorisation decisions. Finally, I related my results to extant MEG data using MEG-fMRI fusion to estimate when the attention effects arose in the brain. However, I was unable to determine whether the selective processing of relevant information arose in the MD system before or after the visual system.

5.1.3 Chapter 4: How do focal brain lesions affect information coding during selective attention?

The selective coding of behaviourally relevant information in the MD system, as seen in Chapter 3, is proposed to be critical for driving brain-wide focus on this information, thereby facilitating selective attention and goal-directed behaviour. Accordingly, damage to this system should affect neural prioritisation of relevant information in the brain. In Chapter 4, I provided a preliminary test of this hypothesis using a novel combination of focal lesion, MEG, and MVPA methods to explore how damage to parietal lobe affects the prioritisation of relevant information during selective attention.

I used MEG to record the neural activity of patients with focal parietal lesions ($N=6$) and controls who were matched to patients on the basis of age, gender and education ($N=10$) while they completed the demanding spatial and feature attention task used in Chapter 3. Relative to controls, I showed that one of the six parietal patients demonstrated subtle reductions in the prioritisation of relevant over irrelevant information (i.e., attentional prioritisation), despite high behavioural accuracy. This result provides early evidence that parietal damage may affect the prioritisation of attended information in the brain. Moreover, inter-individual differences in reaction time related to inter-individual differences in attentional prioritisation in sensors over occipital lobe, providing preliminary evidence that the approach I took to quantify the strength of task-related representations captures something meaningful about behaviour.

5.2 Implications

5.2.1 Implications of this work for understanding the role of the multiple-demand (MD) system in complex human behaviour

This work stems from a prominent hypothesis that proposes a core function of the multiple-demand (MD) system is attentional integration: assembling a structured mental control program through segmentation and integration (Duncan, 2013; Duncan et al., 2020). That is, complex thoughts and behaviours arise from segmenting a task into a sequence of simpler parts, referred to as attentional episodes (Duncan, 2013). Organised behaviour relies on integrating the elements required for each episode into an overall mental control program, with each part focussed on in turn. This proposal, therefore, posits that flexible selection and integration of information from the world and our memories may be one critical component underpinning fluid intelligence, as tested through novel problem-solving. Accordingly, under this framework, fluid intelligence and selective attention are fundamentally linked, with both subserved by the MD system. While I outlined the current evidence that speaks in favour of this proposal in Chapter 1, I also highlighted that there are still gaps in our understanding of how the MD system supports these critical components of intelligent human behaviour. Now, I will consider what the work in the preceding three chapters has added to this story.

Chapter 2 used neuropsychology to examine whether frontoparietal regions have a critical role in supporting fluid intelligence. While previous neuropsychological studies had implicated MD regions in fluid intelligence (e.g., Woolgar et al., 2010, 2018), various other studies implicated different networks (e.g., Cipolotti et al., 2023; Dubois et al., 2018; Santarnecchi et al., 2017). In the context of small patient cohorts and strongly hypothesis-driven comparisons (e.g., comparing MD damage to non-MD damage, or voxel-based methods) the field was missing a data-driven comparison of the contribution of different networks. To address this, I drew on a large cohort of focal lesion patients and examined the relationship between grey matter (GM) and white

matter (WM) network damage to each of seven canonical resting-state networks (Yeo et al., 2011) and fluid intelligence deficit. I observed that selective damage to frontoparietal network (FPN) – a network closely linked to the activation-defined MD system – predicted fluid intelligence deficit over contributions from other networks. Broadly speaking, these findings accord with previous work that has highlighted the role of frontoparietal regions in supporting fluid intelligence (e.g., Barbey et al., 2014; Duncan et al., 1995; Finn et al., 2015; Gläscher et al., 2009, 2010; Santarnecchi et al., 2017; Woolgar et al., 2010, 2018) and adds to this knowledge base by specifying that fluid intelligence may depend on FPN over and above other networks, including other networks which are distributed across frontal and parietal lobes. For example, default mode network (DMN) was not implicated in supporting fluid intelligence, once the effect of FPN damage was accounted for, despite previous work suggesting that it may play an important role in supporting fluid intelligence (e.g., Dubois et al., 2018; Finn et al., 2015; Sripatha et al., 2019). Nonetheless, future work with patients exhibiting more widely distributed lesions is needed to separate the effect of FPN damage from the general effect of total lesion volume.

The extent to which we can draw conclusions about the (activation-defined) MD network from the findings of Chapter 2 depends on the extent to which the (resting-state derived) FPN overlaps with the MD network. This is challenging to specify definitively, because both resting-state parcellations and activation-based definitions of these networks vary between studies (e.g., granularity or number of resting-state networks delineated in a parcellation, which tasks are used to define MD regions). However, taking two recent definitions, Assem, Glasser, et al. (2020) compared the spatial similarity between the MD network (defined in surface space from overlap of three functional tasks in Human Connectome Project (HCP) data) and a surface space resting-state network parcellation, the Cole-Anticevic Brain Network Parcellation (CAB-NP; Ji et al., 2019). They found that the MD system showed extensive overlap with FPN, but also some overlap with other resting-state networks, including cingulo-opercular network, dorsal attention network (DAN) and DMN (See Figure 2 from Chapter 1). Critically, core MD regions (the most domain-general part of the MD system) fell entirely within FPN (although, core MD regions did not fully account for

FPN). This suggests that FPN is a reasonable approximation that captures much, but perhaps not all, of the MD system.

In my work, I used the volumetric resting-state parcellation provided by Yeo et al. (2011). A recent study constrained their analysis of the MD system to just the regions that fell within Yeo et al.'s (2011) FPN (Shashidhara et al., 2019), and reported how this affected the MD definition taken from Fedorenko et al., 2013. Figure 45 shows the full MD network (Panel A, MD template from Fedorenko et al., 2013) and the 'reduced' MD network, where only parts of the network that shared overlap with Yeo et al.'s (2011) FPN definition are shown (Panel B). As is clear in Figure 45, much, but not all, of the MD system falls within FPN.

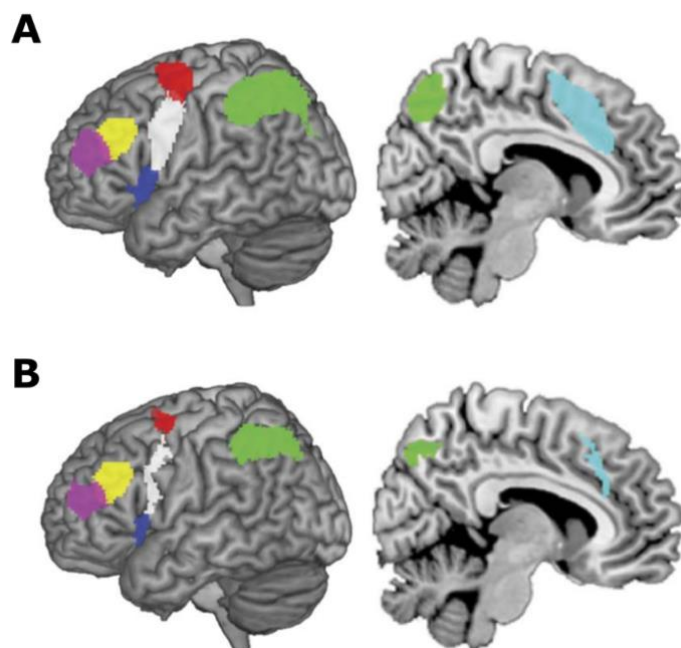


Figure 45. A visual comparison of the full MD network (Fedorenko et al., 2013), shown in Panel A, against a 'reduced' MD network, formed by the intersection of MD regions with the frontoparietal network definition from Yeo et al. (2011), shown in Panel B. The MD system shows substantial, but not complete, overlap with the frontoparietal network. Figure adapted with permission from Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *Journal of Cognitive Neuroscience*, 31(11), 1617–1630 ©2019 Massachusetts Institute of Technology. All rights reserved.

However, Shashidhara et al., (2019) did not examine which other resting-state networks overlapped with the MD system. To provide some indication of this, Figure 46 shows the most recent volumetric activation-based MD region template (Woolgar et al., 2018) overlaid on the seven-network Yeo et al. (2011) parcellation. Here, MD regions (defined using data from 63 controls) predominantly overlap with FPN and DAN. While comparison of these definitions again suggests some correspondence between FPN and the MD system, MD regions also clearly overlap with DAN (particularly in parietal lobe), and, conversely, much of FPN falls outside of the MD network. Interestingly, the results from Chapter 2 also hinted at the possible contribution of DAN to fluid intelligence.

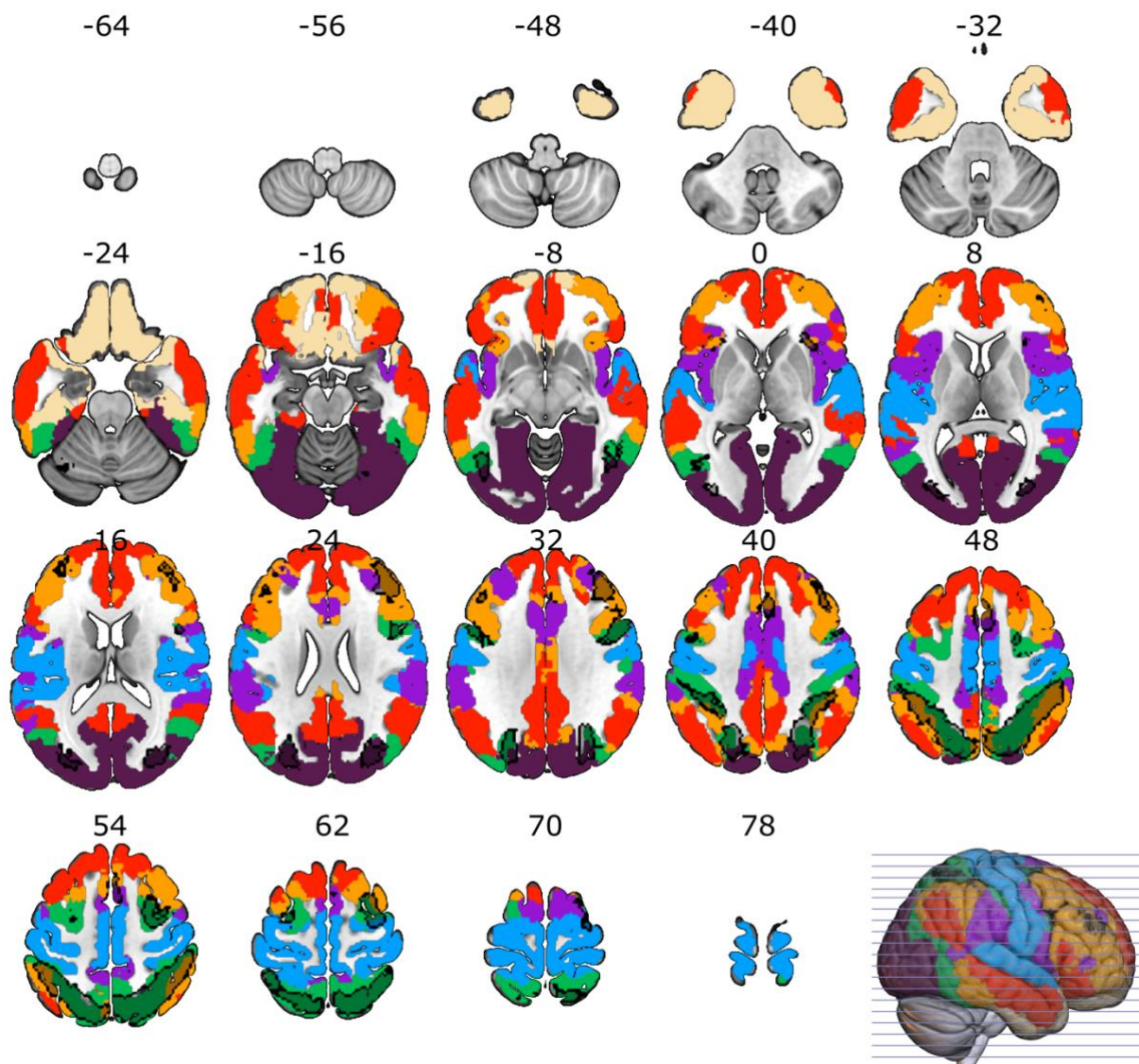


Figure 46. MD regions, from a probabilistic fMRI activation map (minimum threshold = 0.15, meaning that at least 15% of participants showed significant activation in this voxel) based on 63 healthy participants (Woolgar et al., 2018), are shown as shaded areas outlined in black, overlaid on the seven resting-state network parcellation from Yeo, et al. (2011). Colours for Yeo, et al. (2011) parcellation are the same as in Chapter 2 and are as follows: orange = frontoparietal network, green = dorsal attention network, dark purple = visual network, blue = somatomotor network, violet = ventral attention network, cream = limbic network, red = default network. MD regions primarily overlap with frontoparietal network and, especially in parietal lobe, the dorsal attention network.

The motivation for using resting-state networks (namely, Yeo et al., 2011) in this study, rather than an activation-based definition of the MD network, was to be able to contrast contributions from multiple canonical networks comprising a parcellation of the whole cortex. While this approach allowed a more data-driven test of which networks subserve fluid intelligence, it precluded a more spatially sensitive test of whether MD damage specifically predicts fluid intelligence deficit. Nonetheless, to the extent that MD regions overlap with FPN (and perhaps DAN), the data support the idea that the MD system has a key role in fluid intelligence.

Chapters 3 and 4 examined the role of MD regions in supporting selective attention processes. In both chapters, I examined neural task-related representations when two types of selective attention are deployed: spatial and feature attention. For this I used a demanding visual attention task (Goddard et al., 2022) that manipulated spatial attention (attend to left or right object) and feature attention (attend to colour or shape of the target object).

These types of attention are frequently studied separately, with MD regions showing preferential representation of task-relevant information in line with the proposal that these regions facilitate goal-directed behaviour. For instance, when attention is directed to an object in a particular location, MD regions show stronger coding of the attended object over the unattended object (Woolgar, Williams, et al., 2015), and stronger coding of the attended feature dimension over the unattended feature dimension when attention is directed to a single feature of a single object (Jackson et al., 2017). However, these types of attention are often used together and, to date, it has been unclear what task information is prioritised in MD regions when attention is simultaneously directed to both spatial and feature information.

I sought to address this point in Chapter 3 by scanning healthy adults with fMRI while they completed a demanding spatial and feature attention task. Based on previous theories, I had anticipated that the attention effect reflected in MD codes could be additive (i.e., show graded enhancement of all attended information) or multiplicative (i.e., selective enhancement of only information at the intersection of spatial and feature attention demands). I found evidence strongly in favour of the multiplicative account:

the MD system was highly specific in its representation of task-related information, representing only information about the attended feature of the attended object. This finding aligns well with previous work demonstrating the dominance of task-relevant information in frontoparietal regions (Jackson et al., 2017; Woolgar, Afshar, et al., 2015; Woolgar et al., 2011; Woolgar, Williams, et al., 2015) and extends this work by specifying the highly selective coding scheme employed when attention is applied in both physical and feature space.

While selective coding of task-relevant information was strongest in MD regions, a similar effect was seen in visual cortex and across the brain more broadly. This could reflect top-down feedback from the MD system, driving system-wide prioritisation of the attended information (e.g., Desimone & Duncan, 1995), although the poor temporal resolution of the blood-oxygen-level-dependent (BOLD) response, and the highly correlated responses between regions in the MEG-fMRI fusion analysis, made it difficult to ascertain the temporal order in which this arose across regions.

In line with other work (e.g., Woolgar et al., 2011, 2015), I expected that difficulty would be reflected in MD coding of attended information, with stimuli close to the decision boundary represented more strongly than stimuli far from the decision boundary. Instead, I revealed that MD codes were organised in a more complex fashion, with strength of representations influenced by two factors: the difficulty of the decision and the stimulus dimension. This result is consistent with other research highlighting high-dimensionality in prefrontal patterns that are related to behaviour (Badre et al., 2021; Bartolo et al., 2020; Rigotti et al., 2013).

Overall, the work in Chapter 3 feeds into a growing evidence base that MD codes closely reflect what is needed for behaviour. Collectively, these findings support a key prediction of the attentional integration proposal: in order for the MD system to support complex behaviour, this system should selectively code information relevant to a given task.

Yet Chapter 3 was unable to ascertain whether preferential coding of task-relevant information in MD regions facilitates cognitive control through driving brain-wide

focus on this information. The work in Chapter 4 represents a first step to being able to address such questions using a novel approach that uses insight from neuroimaging and multivariate pattern analysis (MVPA) with focal lesion patients to address questions of causality. Here, I examined how damage to parietal lobe affected the prioritisation of relevant information in individual focal lesion patients using MEG, finding reduced preferential coding of relevant information in one of six patients in the context of good behavioural accuracy but slower reaction times. Despite large damage in some cases, the remaining five patients evinced neural evidence of attentional prioritisation within a prescribed time window to a level similar to controls. Variability in patient, lesion, and recovery characteristics, as well as the small control sample used as a reference to identify differences in neural processes, limit the strength of the conclusions that can be drawn from this work; however, these results provide some evidence that parietal damage can affect the neural prioritisation of relevant information.

In summary, the results from this body of work broadly accord with the prevailing account of how frontoparietal regions, or the MD system more specifically, support intelligent human behaviour. Yet future work is needed to establish a causal role for the MD system in supporting fluid intelligence and selective attention. The new approach I introduced in Chapter 4, whereby the impact of brain damage is examined in terms of its effects on neural representational content, offers an exciting new means to address the need to establish causality.

5.2.2 Methodological implications

In this thesis, I have taken two different approaches to examine how focal damage relates to changes in cognitive processes. In Chapter 2, I related the extent of GM and WM structural network damage to fluid intelligence deficit to identify key networks underpinning this process. In Chapter 4, I examined how focal parietal damage affects the prioritisation of relevant information across the brain. The findings from these

studies, particularly regarding what we can learn from the methodological approaches taken, have implications for future work. I will consider these implications now.

To my knowledge, no previous study has examined the GM and WM contributions of canonical resting-state networks to fluid intelligence in a neuropsychological context. Yet as focal lesions can disrupt functional networks, focal damage should be considered in the context of brain-wide networks (Gratton et al., 2012; Guggisberg et al., 2019; Woolgar et al., 2013). Additionally, focal damage can cause behavioural changes due to both structural damage at the lesion site and network disconnection (Carrera & Tononi, 2014), emphasising the importance of considering both GM and WM contributions. Yet previous neuropsychological work has largely focussed on relating fluid intelligence deficits to individual brain voxels (e.g., voxel-based methods), with an emphasis on GM contributions, precluding examination of the contribution of brain networks to cognition.

The work in Chapter 2 thus represents an extension of previous work by considering how focal lesion damage relates to fluid intelligence deficits in the context of disruption to resting-state networks. I quantified network damage in two ways, permitting a comparison of the sensitivity of these approaches: volume of GM network damage and estimation of structural WM within-network tract damage using a large ($N=1042$) sample of healthy WM tractograms as reference. Despite a large patient sample size ($N=174$) and robust estimation of WM network damage, I found these two approaches to estimating network damage led to similar predictions and were highly correlated, suggesting they explained similar variance in fluid intelligence. One implication of this may be that examining the extent of within-network disconnection caused by damage to structural WM may not provide much more information than examining the impact of GM structural damage alone. Of course, there is still the possibility that alternative approaches for examining network dysfunction might be more sensitive or capture other sources of individual differences, and accordingly offer clearer insights into brain-function relationships. For example, new insights might be possible by considering how WM tract damage affects communication between networks (i.e., inter-network disconnection). This could be particularly important to consider for systems like the

MD network which may derive their central role in domain-general cognition through broad and flexible connections with a range of more domain-specific networks (Cole et al., 2013; Duncan et al., 2020).

In addition, neural reorganisation can occur during and following recovery from focal damage (Grefkes & Fink, 2020) and lead to changes in local and network function, complicating inferences drawn from studies that focus solely on relating structural damage to cognitive changes, as in Chapter 2. I sought to address this point in Chapter 4 by considering how focal parietal damage affected the dynamics of functional neural representation in MEG. As far as I am aware, no previous study has explored how information coding, or the strength of information codes, change following brain damage. With such a small sample of patients and controls, I could not make strong claims to this point. However, my study provided preliminary evidence that parietal damage can lead to subtle changes in the dynamics of attentional prioritisation that could be detected at an individual level using my approach. Moreover, I was able to demonstrate that the strength of attentional prioritisation could differ between sensors over the affected and unaffected hemispheres of an individual. This provides early evidence that this approach might be able to track altered or compensatory neural processing within a single subject. Having now established the feasibility of this approach for exploring individual-level attentional prioritisation following brain damage, exciting opportunities open up to understand how focal damage affects cognitive dynamics through neural reorganisation and compensation.

5.3 Limitations and Future Directions

5.3.1 How are fluid intelligence and selective attention linked?

A core theme of this work is that fluid intelligence and selective attention processes are linked through the process of assembling structured mental programs for behaviour, subtended by a core system that is recruited across multiple demands (Duncan, 2013; Duncan et al., 2020). While my observations align well with the proposal that the MD network is involved in both fluid intelligence and selective attention, I do not directly test how these two constructs relate. My findings sit within a growing body of neuropsychological and neuroimaging evidence that also implicate MD regions in fluid intelligence (e.g., Assem, Blank, et al., 2020; P.-Y. Chen et al., 2020; Woolgar et al., 2010, 2018) and selective attention (e.g., Jackson et al., 2017, 2021; Woolgar, Williams, et al., 2015). Similarly, the studies in this thesis explore the neural mechanisms of fluid intelligence and selective attention independently, with the MD system emerging as a common core. Thus, a limitation of my work is that the conclusion that frontoparietal regions subtend both processes is inferred, rather than directly tested. To bring the pieces of this story together, we need to understand how selective attention and the processes underpinning fluid intelligence relate and work together to support complex, flexible behaviour. In future work, the approaches I have used in this thesis could be extended to bridge this gap. For instance, the approach taken in Chapter 4, combining lesion patients and neuroimaging to assess information processing during selective attention, could be extended to study the neural correlates of fluid intelligence deficit after MD damage. For example, it would be interesting to test whether changes in the dynamics or extent of neural prioritisation of relevant information predict fluid intelligence deficits, and whether and how this varies with lesion size and location. This could lead to important insights into the interplay between fluid intelligence and selective attention in supporting adaptive, intelligent behaviour.

5.3.2 Methodological limitations

Recognising the limitations inherent to any single method, I drew on a range of methodological approaches to examine how the MD system supports intelligent behaviour. In Chapter 2, I used focal lesion models to investigate the relationship between GM and WM network damage and fluid intelligence loss. In Chapter 3, I used fMRI to address a spatial question, examining what task-related information is held in MD regions when attention is directed to spatial and feature information. In Chapter 4, I combined focal lesion models with MEG to explore how brain damage affects the dynamics of neural attentional processes. In both Chapter 3 and 4, I use MVPA to assess informational content. Here, I will discuss some of the broader limitations of each of these methods and consider how I have used different methods in this thesis help to address some of these shortcomings.

The work in Chapter 2 represents an important extension on previous focal lesion research into the neural mechanisms of fluid intelligence. As cognitive and behavioural changes can result from structural damage at the lesion site and from damage to connected regions distal to the lesion (Carrera & Tononi, 2014), I considered contributions of both GM network damage and structural network disconnection caused by damage to WM tracts. Yet without functional data, we cannot assess the impact that neural reorganisation or changes to functional network dynamics have on fluid intelligence. Thus, to have a better understanding of whether the MD system critically underpins fluid intelligence, future work would ideally individually localise this system and relate functional network alterations to cognitive changes. While Chapter 4 does not directly tackle this, it takes us a step closer by combining focal lesion methods with MEG and MVPA to examine how focal brain damage affects dynamic attentional processes, thereby shedding light on how focal damage might affect functional processes.

In Chapter 3, I utilised the high spatial resolution of fMRI along with MVPA methods to address a spatial question, finding MD regions selectively code task-relevant information, with a similar but somewhat weaker effect seen across the cortex. Given

the poor temporal resolution of fMRI, however, I was unable to derive a timecourse for when attention effects arose in the brain. To overcome this limitation, I related my fMRI results to extant MEG data using MEG-fMRI fusion in an attempt to derive a spatially and temporally coherent account of when these attention effects arose in the brain. Yet as the attention effect was so large and widespread, I was not able to determine whether the selective processing of relevant information arose in the MD system before or after the visual system. Thus, regarding whether MD codes affect representations elsewhere in the brain, in line with the biased competition hypothesis (Desimone & Duncan, 1995), the results from Chapter 3 are inconclusive. Accordingly, in Chapter 4, I turned to MEG to examine how damage to part of the MD system might affect the prioritisation of relevant information elsewhere in the brain, moving us closer to being able to establish a causal functional account of the MD system.

In Chapter 4, I sought to combine the advantages of focal lesion methods (namely, causal inference) with the high temporal resolution of MEG and used MVPA to track how brain damage affects the dynamics of attentional prioritisation. An advantage of this approach is that it offers a framework to investigate the dynamics of how the MD system causally supports goal-directed information processing across the brain. In the case of Chapter 4, constrained by the small sample size, I was unable to definitively test whether the MD system has a causal role in driving task-relevant information coding throughout the brain. However, the results from this proof-of-concept exploration indicate that parietal damage may be associated with subtle changes to attentional prioritisation that can be detected using this approach. This opens the possibility for future work to track whether the extent of damage to the MD system affects the strength and exchange of relevant information throughout the brain. For instance, with larger patient and control cohorts, one possible approach could be to localise the MD system within each individual (e.g., using functional localisers, or by transforming a template activation-based definition) and then relate the extent of disruption to this system to changes in dynamic attentional prioritisation processes.

Another potential application of the approach in Chapter 4 could be to examine compensatory dynamics. Despite parietal damage (in some cases extensive), all six

patients completed the demanding selective attention task with accuracy similar to that of controls. Given the purported role of parietal lobe in supporting selective attention processes (Behrmann et al., 2004; Duncan, 2013; Duncan et al., 2020), it is interesting to consider whether compensatory processes might be at play. However, discerning whether an observed result reflects compensation can be challenging. In the context of Chapter 4, for instance, stronger attentional prioritisation in a patient's unaffected relative to affected hemisphere, compared to the equivalent difference in controls, may indicate a compensatory mechanism or may reflect a circumscribed effect of damage over the affected side. Although a coarse comparison in sensor space with a small control group failed to reveal evidence of compensatory mechanisms in Chapter 4, MVPA methods have elsewhere been shown to hold promise for examining compensatory processes. A recent study demonstrated that MVPA could differentiate between compensation and general inefficiencies in brain activity in a large cohort ($N=81$) by assessing whether increased activity in regions not typically recruited in a task carried information that enhanced task performance (Knights et al., 2021). Accordingly, with a larger sample size, possibly combined with greater spatial resolution by examining these dynamics in source space, it may be possible to apply this approach to test for evidence of compensatory mechanisms in these patients.

5.3.3 What do representational codes 'mean'?

A final set of considerations involves the limits on inference from MVPA. My work has relied heavily on these methods to examine what information is coded in the brain. As discussed in Chapter 1, if different events reliably elicit different patterns of neural activity, information coding is inferred. Moreover, if information is decodable in a particular brain region, it is taken as evidence that this region is involved in processing that information. This approach is widely used to probe the representational content of brain regions (e.g., Grootswagers et al., 2017; Haynes & Rees, 2006; Kriegeskorte et al., 2008; Mur et al., 2009) and to track the exchange of information throughout the brain (e.g., Goddard et al., 2022). A key assumption underlies the use of such techniques to

infer neural mechanisms: that the representational codes detected by MVPA methods reflect information that is processed and used by the brain. Yet the validity of this assumption has been questioned, with a primary concern that successful decoding does not necessarily imply that the decoded information is actually utilised by the brain in meaningful ways (de-Wit et al., 2016; Ritchie et al., 2019). One way to address these concerns is to link aspects of neural representations with behaviour. In Chapter 4, I examined whether differences in neural representational strength related to behaviour in a neuropsychological context. Comparing focal lesion patients to controls may reveal larger and more meaningful individual differences than those found in a sample of only healthy controls, offering a promising context for exploring these relationships. Across all patients and controls, I explored whether inter-individual differences in neural attentional prioritisation (i.e., relative prioritisation of relevant over irrelevant information in a prescribed time window) related to inter-individual differences in behavioural performance (accuracy and reaction time). Thus, this approach probed whether participants with similar attentional prioritisation scores were also similar to one another in accuracy or reaction time. Indeed, I found evidence of a significant relationship between inter-individual differences in attentional prioritisation for sensors over occipital lobe and inter-individual differences in reaction time. This result suggests that differences in strength of neural codes may be meaningful for behaviour.

However, it is important to note that the evidence for a direct link between neural representations and behaviour has been mixed in the literature (see Zheng et al., 2024 for a recent overview). In particular, individual differences in aspects of neural representation, such as decoding accuracy, may reflect extraneous factors (e.g., head shape and cortical folding) rather than meaningful differences related to behaviour (Zheng et al., 2024). This is the case in healthy individuals and might be exaggerated in patients due to heterogeneity induced by differences in lesion characteristics. Accordingly, I used a difference measure to compare patients against controls (i.e., strength of relevant minus irrelevant information coding) in Chapter 4. However, this approach may also be limited as irrelevant information coding (at least, in this study) was close to chance. Future work with larger cohorts is needed to clarify whether the

results from Chapter 4 are robust, and to further elucidate the significance of individual differences in decoding strength.

In summary, the question of whether representational codes and the strength of information decoding reflect meaningful aspects of brain information processing remains open and significant for interpreting this body of work. However, the preliminary finding that dissimilarity in neural attentional prioritisation relates to dissimilarity in behavioural performance in a neuropsychological context suggests promising avenues for future work. It would be interesting to consider whether a similar dissimilarity approach to the one taken in Chapter 4 could also be applied to explore the link between neural codes and behaviour in healthy individuals.

5.4 Conclusions

This thesis sought to elucidate the role of a network of predominantly frontoparietal regions, the MD system, in supporting intelligent human behaviour. Using insights from lesion methodology and both structural and functional neuroimaging methods, I have demonstrated that these regions are involved in two critical aspects underpinning our capacity for cognitive control and complex behaviour: fluid intelligence and selective attention. I have shown that damage to FPN is linked to deficits in fluid intelligence, suggesting this network plays a critical role in supporting higher cognitive functions. I have demonstrated that the MD system selectively prioritises the information that is most important for behaviour, underscoring its role in goal-directed cognition. Finally, I have provided preliminary evidence that parietal lobe damage affects neural attentional processes, demonstrating that the combination of neuropsychological, neuroimaging, and MVPA methods holds potential to examine how brain damage can affect the dynamics of neural information processing in the brain. Taken together, these studies add to a growing body of work that suggests MD regions play an important role in processes fundamental to intelligent human behaviour. They also highlight the unique contributions lesion methodology, neuroimaging, and their combination have in advancing our understanding of the neural mechanisms underpinning human cognition.

References

- Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test Performance as a Measure of Frontal Lobe Damage. *Journal of Clinical and Experimental Neuropsychology*, 13(6), 909–922. <https://doi.org/10.1080/01688639108405107>
- Anderson, S. W., Damasio, H., & Tranel, D. (1990). Neuropsychological Impairments Associated With Lesions Caused by Tumor or Stroke. *Archives of Neurology*, 47(4), 397–405. <https://doi.org/10.1001/archneur.1990.00530040039017>
- Aoi, M. C., Mante, V., & Pillow, J. W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nature Neuroscience*, 23(11), 1410–1420. <https://doi.org/10.1038/s41593-020-0696-5>
- Arazi, A., Yeshurun, Y., & Dinstein, I. (2019). Neural Variability Is Quenched by Attention. *Journal of Neuroscience*, 39(30), 5975–5985. <https://doi.org/10.1523/JNEUROSCI.0355-19.2019>
- Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural Activity in the Primate Prefrontal Cortex during Associative Learning. *Neuron*, 21(6), 1399–1407. [https://doi.org/10.1016/S0896-6273\(00\)80658-3](https://doi.org/10.1016/S0896-6273(00)80658-3)

References

- Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-Specific Neural Activity in the Primate Prefrontal Cortex. *Journal of Neurophysiology*, 84(1), 451–459. <https://doi.org/10.1152/jn.2000.84.1.451>
- Assem, M., Blank, I. A., Mineroff, Z., Ademoğlu, A., & Fedorenko, E. (2020). Activity in the fronto-parietal multiple-demand network is robustly associated with individual differences in working memory and fluid intelligence. *Cortex*, 131, 1–16. <https://doi.org/10.1016/J.CORTEX.2020.06.013>
- Assem, M., Glasser, M. F., Van Essen, D. C., & Duncan, J. (2020). A Domain-General Cognitive Core Defined in Multimodally Parcellated Human Cortex. *Cerebral Cortex*, 30(8), 4361–4380. <https://doi.org/10.1093/cercor/bhaa023>
- Badre, D., Bhandari, A., Keglovits, H., & Kikumoto, A. (2021). The dimensionality of neural representations for control. *Current Opinion in Behavioral Sciences*, 38, 20–28. <https://doi.org/10.1016/j.cobeha.2020.07.002>
- Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, 344(6182), 424–427. <https://doi.org/10.1126/science.1247003>
- Barbey, A. K., Colom, R., Paul, E. J., & Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Structure and Function*, 219(2), 485–494. <https://doi.org/10.1007/s00429-013-0512-z>
- Barbosa, J., Lozano-Soldevilla, D., & Compte, A. (2021). Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. *PLOS Biology*, 19(10), e3001436. <https://doi.org/10.1371/journal.pbio.3001436>
- Barnes, L., Goddard, E., & Woolgar, A. (2022). Neural Coding of Visual Objects Rapidly Reconfigures to Reflect Subtrial Shifts in Attentional Focus. *Journal of Cognitive Neuroscience*, 34(5), 806–822. https://doi.org/10.1162/jocn_a_01832

- Bartholomew, D. J., Allerhand, M., & Deary, I. J. (2013). Measuring mental capacity: Thomson's Bonds model and Spearman's g-model compared. *Intelligence*, 41(4), 222–233. <https://doi.org/10.1016/j.intell.2013.03.007>
- Bartholomew, D. J., Deary, I. J., & Lawn, M. (2009). A new lease of life for Thomson's bonds model of intelligence. *Psychological Review*, 116(3), 567–579. <https://doi.org/10.1037/a0016262>
- Bartolo, R., Saunders, R. C., Mitz, A. R., & Averbach, B. B. (2020). Dimensionality, information and learning in prefrontal cortex. *PLOS Computational Biology*, 16(4), e1007514. <https://doi.org/10.1371/journal.pcbi.1007514>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2020). The time course of spatial attention during naturalistic visual search. *Cortex*, 122, 225–234. <https://doi.org/10.1016/j.cortex.2018.11.018>
- Baylis, G. C., Driver, J., & Rafal, R. D. (1993). Visual Extinction and Stimulus Repetition. *Journal of Cognitive Neuroscience*, 5(4), 453–466. <https://doi.org/10.1162/jocn.1993.5.4.453>
- Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology*, 14(2), 212–217. <https://doi.org/10.1016/j.conb.2004.03.012>
- Bertolero, M. A., Yeo, B. T. T., Bassett, D. S., & D'Esposito, M. (2018). A mechanistic model of connector hubs, modularity and cognition. *Nature Human Behaviour*, 2(10), 765–777. <https://doi.org/10.1038/s41562-018-0420-6>

References

- Bichot, N. P., Schall, J. D., & Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, 381(6584), 697–699. <https://doi.org/10.1038/381697a0>
- Bowerman, B. L., & O'Connell, R. T. (1990). *Linear statistical models: An applied approach* (2nd ed.). Duxbury.
- Brawn, P. T., & Snowden, R. J. (2000). Attention to overlapping objects: Detection and discrimination of luminance changes. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 342–358. <https://doi.org/10.1037/0096-1523.26.1.342>
- Brett, M., Leff, A. P., Rorden, C., & Ashburner, J. (2001). Spatial Normalization of Brain Images with Focal Lesions Using Cost Function Masking. *NeuroImage*, 14(2), 486–500. <https://doi.org/10.1006/nimg.2001.0845>
- Bright, P., Jaldow, E., & Kopelman, M. D. (2002). The National Adult Reading Test as a measure of premorbid intelligence: A comparison with estimates derived from demographic variables. *Journal of the International Neuropsychological Society*, 8(6), 847–854. <https://doi.org/10.1017/S1355617702860131>
- Burrows, B. E., & Moore, T. (2009). Influence and Limitations of Popout in the Selection of Salient Visual Stimuli by Area V4 Neurons. *Journal of Neuroscience*, 29(48), 15169–15177. <https://doi.org/10.1523/JNEUROSCI.3710-09.2009>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–1525. <https://doi.org/10.1016/J.VISRES.2011.04.012>
- Carrera, E., & Tononi, G. (2014). Diaschisis: Past, present, future. *Brain*, 137(9), 2408–2422. <https://doi.org/10.1093/brain/awu101>
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Houghton Mifflin.

- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 27:1-27:27. <https://doi.org/10.1145/1961189.1961199>
- Chatterjee, A., & Coslett, H. B. (2003). Neglect: Cognitive neuropsychological issues. In T. E. Feinberg & M. J. Farah (Eds.), *Behavioural neurology and neuropsychology* (2nd Edition). McGraw-Hill.
- Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, 2(7), 671–676. <https://doi.org/10.1038/10230>
- Chen, J., Scotti, P. S., Dowd, E. W., & Golomb, J. D. (2021). Neural Representations of Task-relevant and Task-irrelevant Features of Attended Objects. *bioRxiv*, 2021.05.21.445168. <https://doi.org/10.1101/2021.05.21.445168>
- Chen, P.-Y., Chen, C.-L., Hsu, Y.-C., & Tseng, W.-Y. I. (2020). Fluid intelligence is associated with cortical volume and white matter tract integrity within multiple-demand system across adult lifespan. *NeuroImage*, 212, 116576. <https://doi.org/10.1016/j.neuroimage.2020.116576>
- Chen, X., Hoffmann, K.-P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology*, 107(5), 1530–1543.
- Cichy, R. M., Kriegeskorte, N., Jozwik, K. M., van den Bosch, J. J. F., & Charest, I. (2019). The spatiotemporal neural dynamics underlying perceived similarity for real-world objects. *Neuroimage*, 194, 12–24. <https://doi.org/10.1016/j.neuroimage.2019.03.031>
- Cichy, R. M., & Oliva, A. (2020). A M/EEG-fMRI Fusion Primer: Resolving Human Brain Responses in Space and Time. *Neuron*, 107(5), 772–781. <https://doi.org/10.1016/j.neuron.2020.07.001>

References

- Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3), 455–462. <https://doi.org/10.1038/nn.3635>
- Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. *Cerebral Cortex*, 26(8), 3563–3579. <https://doi.org/10.1093/cercor/bhw135>
- Cipolotti, L., Healy, C., Chan, E., Bolsover, F., Lecce, F., White, M., Spanò, B., Shallice, T., & Bozzali, M. (2015). The impact of different aetiologies on the cognitive performance of frontal patients. *Neuropsychologia*, 68, 21–30. <https://doi.org/10.1016/j.neuropsychologia.2014.12.025>
- Cipolotti, L., Ruffle, J. K., Mole, J., Xu, T., Hyare, H., Shallice, T., Chan, E., & Nachev, P. (2023). Graph lesion-deficit mapping of fluid intelligence. *Brain*, 146(1), 167–181. <https://doi.org/10.1093/brain/awac304>
- Cocuzza, C. V., Ito, T., Schultz, D., Bassett, D. S., & Cole, M. W. (2020). Flexible Coordinator and Switcher Hubs for Adaptive Task Control. *Journal of Neuroscience*, 40(36), 6949–6968. <https://doi.org/10.1523/JNEUROSCI.2559-19.2020>
- Cohen, A. O., & Casey, B. J. (2017). The Neurobiology of Adolescent Self-Control. In T. Egner (Ed.), *The Wiley Handbook of Cognitive Control* (pp. 455–475). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118920497.ch26>
- Cohen, J. D. (2017). Cognitive Control: Core Constructs and Current Considerations. In T. Egner (Ed.), *The Wiley Handbook of Cognitive Control* (pp. 1–28). John Wiley & Son. <https://doi.org/10.1002/9781118920497.ch1>
- Cohen, J. R., Gallen, C. L., Jacobs, E. G., Lee, T. G., & D’Esposito, M. (2014). Quantifying the Reconfiguration of Intrinsic Networks during Working Memory. *PLOS ONE*, 9(9), e106636. <https://doi.org/10.1371/journal.pone.0106636>

- Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12), 1594–1600. <https://doi.org/10.1038/nn.2439>
- Cole, M. W., Pathak, S., & Schneider, W. (2010). Identifying the brain's most globally connected regions. *NeuroImage*, 49(4), 3132–3148. <https://doi.org/10.1016/j.neuroimage.2009.11.001>
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16(9), 1348–1355. <https://doi.org/10.1038/nn.3470>
- Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*, 37(1), 343–360. <https://doi.org/10.1016/J.NEUROIMAGE.2007.03.071>
- Colom, R., & Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, 35(3), 243–251. <https://doi.org/10.1016/j.intell.2006.07.008>
- Colom, R., Haier, R. J., Head, K., Álvarez-Linera, J., Quiroga, M. Á., Shih, P. C., & Jung, R. E. (2009). Gray matter correlates of fluid, crystallized, and spatial intelligence: Testing the P-FIT model. *Intelligence*, 37(2), 124–135. <https://doi.org/10.1016/j.intell.2008.07.007>
- Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, 12(4), 489–501. <https://doi.org/10.31887/DCNS.2010.12.4/rcolom>
- Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., & Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence*, 76, 101376. <https://doi.org/10.1016/j.intell.2019.101376>

References

- Crawford, J. R., & Garthwaite, P. H. (2012). Single-case research in neuropsychology: A comparison of five forms of t-test for comparing a case to controls. *Cortex*, 48(8), 1009–1016. <https://doi.org/10.1016/j.cortex.2011.06.021>
- Crawford, J. R., Garthwaite, P. H., Azzalini, A., Howell, D. C., & Laws, K. R. (2006). Testing for a deficit in single-case studies: Effects of departures from normality. *Neuropsychologia*, 44(4), 666–677. <https://doi.org/10.1016/j.neuropsychologia.2005.06.001>
- Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, 27(3), 245–260. <https://doi.org/10.1080/02643294.2010.513967>
- Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, 12(4), 482–486.
- Crawford, J. R., Parker, D. M., Stewart, L. E., Besson, J. A. O., & De Lacey, G. (1989). Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *British Journal of Clinical Psychology*, 28(3), 267–273. <https://doi.org/10.1111/j.2044-8260.1989.tb01376.x>
- Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of multiple, independent categories in the primate prefrontal cortex. *Neuron*, 66(5), 796–807.
- Cumming, T. B., Packer, M., Kramer, S. F., & English, C. (2016). The prevalence of fatigue after stroke: A systematic review and meta-analysis. *International Journal of Stroke*, 11(9), 968–977. <https://doi.org/10.1177/1747493016669861>
- de-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic Bulletin & Review*, 23(5), 1415–1428. <https://doi.org/10.3758/s13423-016-1002-0>

- Deary, I. J., Cox, S. R., & Ritchie, S. J. (2016). Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish g. *Psychological Inquiry*, 27(3), 192–199. <https://doi.org/10.1080/1047840X.2016.1186525>
- Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, 17(2), 255–264. <https://doi.org/10.1037/0894-4105.17.2.255>
- Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W.-L., & Humphreys, G. W. (2015). The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychological Assessment*, 27(3), 883–894. <https://doi.org/10.1037/pas0000082>
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>
- Di Russo, F., Martínez, A., & Hillyard, S. A. (2003). Source Analysis of Event-related Cortical Activity during Visuo-spatial Attention. *Cerebral Cortex*, 13(5), 486–499. <https://doi.org/10.1093/cercor/13.5.486>
- Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12(3), 99–105. <https://doi.org/10.1016/j.tics.2008.01.001>
- Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 104(26), 11073–11078. <https://doi.org/10.1073/pnas.0704320104>

References

- Dosenbach, N. U. F., Raichle, M., & Gordon, E. M. (2024). *The brain's cingulo-opercular action-mode network*. OSF. <https://doi.org/10.31234/osf.io/2vt79>
- Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., Burgund, E. D., Grimes, A. L., Schlaggar, B. L., & Petersen, S. E. (2006). A core system for the implementation of task sets. *Neuron*, 50(5), 799–812. <https://doi.org/10.1016/j.neuron.2006.04.031>
- Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, 1(1), 17–22. <https://doi.org/10.1038/217>
- Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170284. <https://doi.org/10.1098/rstb.2017.0284>
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology. General*, 113(4), 501–517.
- Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews. Neuroscience*, 2(11), 820–829. <https://doi.org/10.1038/35097575>
- Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4), 172–179. <https://doi.org/10.1016/j.tics.2010.01.004>
- Duncan, J. (2013). The structure of cognition: Attentional episodes in mind and brain. *Neuron*, 80(1), 35–50. <https://doi.org/10.1016/j.neuron.2013.09.015>
- Duncan, J., Assem, M., & Shashidhara, S. (2020). Integrated Intelligence from Distributed Brain Activity. *Trends in Cognitive Sciences*, 24(10), 838–852. <https://doi.org/10.1016/j.tics.2020.06.012>
- Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., & Shibuya, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental*

- Psychology: General*, 128(4), 450–478. <https://doi.org/10.1037/0096-3445.128.4.450>
- Duncan, J., Burgess, P., & Emslie, H. (1995). Fluid intelligence after frontal lobe lesions. *Neuropsychologia*, 33(3), 261–268. [https://doi.org/10.1016/0028-3932\(94\)00124-8](https://doi.org/10.1016/0028-3932(94)00124-8)
- Duncan, J., Chylinski, D., Mitchell, D. J., & Bhandari, A. (2017). Complexity and compositionality in fluid intelligence. *Proceedings of the National Academy of Sciences*, 114(20), 5295–5299. <https://doi.org/10.1073/pnas.1621147114>
- Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behavior. *Cognitive Psychology*, 30(3), 257–303. <https://doi.org/10.1006/cogp.1996.0008>
- Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23(10), 475–483. [https://doi.org/10.1016/S0166-2236\(00\)01633-7](https://doi.org/10.1016/S0166-2236(00)01633-7)
- Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., Bishop, S., & Nimmo-Smith, I. (2008). Goal Neglect and Spearman's g: Competing Parts of a Complex Task. *Journal of Experimental Psychology: General*, 137, 131–148. <https://doi.org/10.1037/0096-3445.137.1.131>
- Duncan, J., Seitz, R. J., Kolodny, J., Bor, D., Herzog, H., Ahmed, A., Newell, F. N., & Emslie, H. (2000). A Neural Basis for General Intelligence. *Science*, 289(5478), 457–460. <https://doi.org/10.1126/science.289.5478.457>
- Egley, 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>

References

- Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *Journal of Neuroscience*, 35(36), 12383–12393. <https://doi.org/10.1523/JNEUROSCI.1134-15.2015>
- Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience*, 5(7), 671–676. <https://doi.org/10.1038/nn874>
- Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41), 16616–16621. <https://doi.org/10.1073/pnas.1315235110>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications Ltd.
- Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), Article 11. <https://doi.org/10.1038/nn.4135>
- Fischer-Baum, S., Jang, A., & Kajander, D. (2017). The Cognitive Neuroplasticity of Reading Recovery following Chronic Stroke: A Representational Similarity Analysis Approach. *Neural Plasticity*, 2017. <https://doi.org/10.1155/2017/2761913>
- Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, 102(27), 9673–9678. <https://doi.org/10.1073/pnas.0504136102>
- Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, 443(7107), 85–88. <https://doi.org/10.1038/nature05078>

- 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>
- Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, 47(1), 72–89. <https://doi.org/10.1038/s41386-021-01132-0>
- Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66–74. <https://doi.org/10.1016/j.conb.2016.01.010>
- Gläscher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., & Adolphs, R. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10), 4705–4709. <https://doi.org/10.1073/PNAS.0910397107>
- Gläscher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, A., Grabowski, T., Damasio, H., & Adolphs, R. (2009). Lesion Mapping of Cognitive Abilities Linked to Intelligence. *Neuron*, 61(5), 681–691. <https://doi.org/10.1016/j.neuron.2009.01.026>
- Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows. *Neuroimage*, 128, 385–397. <https://doi.org/10.1016/j.neuroimage.2016.01.006>
- Goddard, E., Carlson, T. A., & Woolgar, A. (2022). Spatial and Feature-selective Attention Have Distinct, Interacting Effects on Population-level Tuning. *Journal of Cognitive Neuroscience*, 34(2), 290–312. https://doi.org/10.1162/JOCN_A_01796
- González-García, C., Formica, S., Wisniewski, D., & Brass, M. (2021). Frontoparietal action-oriented codes support novel instruction implementation. *NeuroImage*, 226, 117608. <https://doi.org/10.1016/j.neuroimage.2020.117608>

References

- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132. [https://doi.org/10.1016/S0160-2896\(97\)90014-3](https://doi.org/10.1016/S0160-2896(97)90014-3)
- Grafman, J., Jonas, B., & Salazar, A. (1990). Wisconsin Card Sorting Test performance based on location and size of neuroanatomical lesion in Vietnam veterans with penetrating head injury. *Perceptual and Motor Skills*, 71(3), 1120–1122.
- Gramfort, A., Luessi, M., Larson, E., Engemann, D., 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. <https://www.frontiersin.org/articles/10.3389/fnins.2013.00267>
- Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38(4), 404.
- Gratton, C., Nomura, E. M., Pérez, F., & D'Esposito, M. (2012). Focal Brain Lesions to Critical Locations Cause Widespread Disruption of the Modular Organization of the Brain. *Journal of Cognitive Neuroscience*, 24(6), 1275–1285. https://doi.org/10.1162/jocn_a_00222
- Gratton, C., Sun, H., & Petersen, S. E. (2018). Control networks and hubs. *Psychophysiology*, 55(3), e13032. <https://doi.org/10.1111/psyp.13032>
- Grefkes, C., & Fink, G. R. (2020). Recovery from stroke: Current concepts and future perspectives. *Neurological Research and Practice*, 2(1), 17. <https://doi.org/10.1186/s42466-020-00060-6>
- Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, 43(3), 509–520. <https://doi.org/10.1016/j.neuroimage.2008.07.065>
- Grootswagers, T., Robinson, A. K., Shatek, S. M., & Carlson, T. A. (2021). The neural dynamics underlying prioritisation of task-relevant information. *Neurons*,

- Behavior, Data Analysis, and Theory*, 5(1), 1–17.
<https://doi.org/10.51628/001c.21174>
- 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
- Groppe, D., Urbach, T., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711–1725.
- Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, 173, 434–447. <https://doi.org/10.1016/J.NEUROIMAGE.2018.02.044>
- Guggisberg, A. G., Koch, P. J., Hummel, F. C., & Buetefisch, C. M. (2019). Brain networks and their relevance for stroke rehabilitation. *Clinical Neurophysiology*, 130(7), 1098–1124. <https://doi.org/10.1016/j.clinph.2019.04.004>
- Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, 11(6), 236–242. <https://doi.org/10.1016/j.tics.2007.04.001>
- Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., & Curtis, C. E. (2021). Working memory representations in visual cortex mediate distraction effects. *Nature Communications*, 12(1), 4714. <https://doi.org/10.1038/s41467-021-24973-1>
- Halligan, P. W., & Marshall, J. C. (1998). Neglect of Awareness. *Consciousness and Cognition*, 7(3), 356–380. <https://doi.org/10.1006/ccog.1998.0362>

References

- Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating Human Intelligence. *Neuron*, 76(6), 1225–1237. <https://doi.org/10.1016/j.neuron.2012.06.022>
- Hauk, O., Stenroos, M., & Treder, M. S. (2022). Towards an objective evaluation of EEG/MEG source estimation methods – The linear approach. *NeuroImage*, 255, 119177. <https://doi.org/10.1016/j.neuroimage.2022.119177>
- Hauk, O., Wakeman, D. G., & Henson, R. (2011). Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *NeuroImage*, 54(3), 1966–1974. <https://doi.org/10.1016/j.neuroimage.2010.09.053>
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, 293(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>
- Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7(7), 523–534. <https://doi.org/10.1038/nrn1931>
- Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, 180, 4–18. <https://doi.org/10.1016/j.neuroimage.2017.08.005>
- Hebart, M. N., Bankson, B. B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *eLife*, 7, e32816. <https://doi.org/10.7554/eLife.32816>
- Hebart, M. N., Görden, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, 8. <https://www.frontiersin.org/articles/10.3389/fninf.2014.00088>

- Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Münte, T. F., Gös, A., Scherg, M., Johannes, S., Hundeshagen, H., Gazzaniga, M. S., & Hillyard, S. A. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, 372(6506), 543–546. <https://doi.org/10.1038/372543a0>
- Henson, R. (2007). Efficient Experimental Design for fMRI. In *Statistical Parametric Mapping* (pp. 193–210). Elsevier. <https://doi.org/10.1016/B978-012372560-8/50015-2>
- Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, 36(2), 185–198. <https://doi.org/10.3758/BF03202679>
- Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57(5), 253–270. <https://doi.org/10.1037/h0023816>
- Huepe, D., Roca, M., Salas, N., Canales-Johnson, A., Rivera-Rei, Á. A., Zamorano, L., Concepción, A., Manes, F., & Ibañez, A. (2011). Fluid Intelligence and Psychosocial Outcome: From Logical Problem Solving to Social Adaptation. *PLOS ONE*, 6(9), e24858. <https://doi.org/10.1371/journal.pone.0024858>
- Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends in Cognitive Sciences*, 11(1), 30–36. <https://doi.org/10.1016/j.tics.2006.10.011>
- 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>
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5), 411–430.

References

- Institute for Personality and Ability Testing. (1973). *Measuring Intelligence with the Culture Fair Tests*. Institute for Personality and Ability Testing.
- Ito, T., Kulkarni, K. R., Schultz, D. H., Mill, R. D., Chen, R. H., Solomyak, L. I., & Cole, M. W. (2017). Cognitive task information is transferred between brain regions via resting-state network topology. *Nature Communications*, 8(1), 1027. <https://doi.org/10.1038/s41467-017-01000-w>
- Jackson, J. B., Feradoes, E., Rich, A. N., Lindner, M., & Woolgar, A. (2021). Concurrent neuroimaging and neurostimulation reveals a causal role for dlPFC in coding of task-relevant information. *Communications Biology*, 4(1), 1–16. <https://doi.org/10.1038/s42003-021-02109-x>
- Jackson, J. B., Rich, A. N., Williams, M. A., & Woolgar, A. (2017). Feature-selective Attention in Frontoparietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *Journal of Cognitive Neuroscience*, 29(2), 310–321. https://doi.org/10.1162/jocn_a_01039
- Jackson, J. B., & Woolgar, A. (2018). Adaptive coding in the human brain: Distinct object features are encoded by overlapping voxels in frontoparietal cortex. *Cortex*, 108, 25–34. <https://doi.org/10.1016/j.cortex.2018.07.006>
- Jeffreys, H. (1998). *The theory of probability*. OuP Oxford.
- Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience*, 31(22), 8210–8219. <https://doi.org/10.1523/JNEUROSCI.6153-09.2011>
- Ji, J. L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., & Cole, M. W. (2019). Mapping the human brain's cortical-subcortical functional network organization. *NeuroImage*, 185, 35–57. <https://doi.org/10.1016/j.neuroimage.2018.10.006>

- Jiang, J., Summerfield, C., & Egner, T. (2016). Visual Prediction Error Spreads Across Object Features in Human Visual Cortex. *Journal of Neuroscience*, 36(50), 12746–12763. <https://doi.org/10.1523/JNEUROSCI.1546-16.2016>
- Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, 30(2), 135–154. <https://doi.org/10.1017/S0140525X07001185>
- Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., Kusunoki, M., & Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron*, 80(1), 235–246. <https://doi.org/10.1016/j.neuron.2013.07.041>
- Kanashiro, T., Ocker, G. K., Cohen, M. R., & Doiron, B. (2017). Attentional modulation of neuronal variability in circuit models of cortex. *eLife*, 6, e23978. <https://doi.org/10.7554/eLife.23978>
- Karnath, H.-O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, 411(6840), 950–953. <https://doi.org/10.1038/35082075>
- Karnath, H.-O., Rennig, J., Johannsen, L., & Rorden, C. (2011). The anatomy underlying acute versus chronic spatial neglect: A longitudinal study. *Brain*, 134(3), 903–912.
- Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron*, 22(4), 751–761. [https://doi.org/10.1016/S0896-6273\(00\)80734-5](https://doi.org/10.1016/S0896-6273(00)80734-5)
- Keller, A. S., Jagadeesh, A. V., Bugatus, L., Williams, L. M., & Grill-Spector, K. (2022). Attention enhances category representations across the brain with strengthened residual correlations to ventral temporal cortex. *NeuroImage*, 249, 118900. <https://doi.org/10.1016/j.neuroimage.2022.118900>

References

- Kievit, R. A., Davis, S. W., Griffiths, J., Correia, M. M., Cam-CAN, & Henson, R. N. (2016). A watershed model of individual differences in fluid intelligence. *Neuropsychologia*, *91*, 186–198. <https://doi.org/10.1016/j.neuropsychologia.2016.08.008>
- Kinsbourne, M. (1993). Orientational bias model of unilateral neglect: Evidence from attentional gradients within hemispace. In I. H. Robertson & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies* (pp. 63–86). Lawrence Erlbaum Associates Ltd. Publishers.
- Kleinman, J. T., Newhart, M., Davis, C., Heidler-Gary, J., Gottesman, R. F., & Hillis, A. E. (2007). Right hemispatial neglect: Frequency and characterization following acute left hemisphere stroke. *Brain and Cognition*, *64*(1), 50–59. <https://doi.org/10.1016/j.bandc.2006.10.005>
- Knights, E., Morcom, A. M., & Henson, R. N. (2021). Does Hemispheric Asymmetry Reduction in Older Adults in Motor Cortex Reflect Compensation? *Journal of Neuroscience*, *41*(45), 9361–9373. <https://doi.org/10.1523/JNEUROSCI.1111-21.2021>
- Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, *27*(3), 151–177. <https://doi.org/10.1080/1047840X.2016.1153946>
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868. <https://doi.org/10.1073/pnas.0600244103>
- Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. <https://doi.org/10.1016/j.tics.2013.06.007>

- Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2. <https://doi.org/10.3389/neuro.06.004.2008>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14(4), 389–433. [https://doi.org/10.1016/S0160-2896\(05\)80012-1](https://doi.org/10.1016/S0160-2896(05)80012-1)
- Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82. <https://doi.org/10.1016/j.tics.2004.12.004>
- Lee, Y. S., Zreik, J. T., & Hamilton, R. H. (2017). Patterns of neural activity predict picture-naming performance of a patient with chronic aphasia. *Neuropsychologia*, 94(February 2016), 52–60. <https://doi.org/10.1016/j.neuropsychologia.2016.11.010>
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). Oxford University Press.
- Lhermitte, F. (1983). ‘Utilization behaviour’ and its relation to lesions of the frontal lobes. *Brain*, 106(2), 237–255. <https://doi.org/10.1093/brain/106.2.237>
- Lhermitte, F., Pillon, B., & Serdaru, M. (1986). Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: A neuropsychological study of 75 patients. *Annals of Neurology*, 19(4), 326–334. <https://doi.org/10.1002/ana.410190404>
- Li, R., Perrachione, T. K., Tourville, J. A., & Kiran, S. (2021). Representation of semantic typicality in brain activation in healthy adults and individuals with aphasia: A

References

- multi-voxel pattern analysis. *Neuropsychologia*, 158, 107893.
<https://doi.org/10.1016/j.neuropsychologia.2021.107893>
- Li, Y., Wang, F., Chen, Y., Cichocki, A., & Sejnowski, T. (2018). The Effects of Audiovisual Inputs on Solving the Cocktail Party Problem in the Human Brain: An fMRI Study. *Cerebral Cortex*, 28(10), 3623–3637.
<https://doi.org/10.1093/cercor/bhx235>
- Liu, T. (2019). Feature-based attention: Effects and control. *Current Opinion in Psychology*, 29, 187–192. <https://doi.org/10.1016/j.copsyc.2019.03.013>
- Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical Mechanisms of Feature-based Attentional Control. *Cerebral Cortex*, 13(12), 1334–1343.
<https://doi.org/10.1093/cercor/bhg080>
- Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, 47(1), 108–113.
<https://doi.org/10.1016/j.visres.2006.09.017>
- Lorenz, R., Violante, I. R., Monti, R. P., Montana, G., Hampshire, A., & Leech, R. (2018). Dissociating frontoparietal brain networks with neuroadaptive Bayesian optimization. *Nature Communications*, 9(1), 1227.
<https://doi.org/10.1038/s41467-018-03657-3>
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77(1), 24–42.
<https://doi.org/10.1152/JN.1997.77.1.24/ASSET/IMAGES/LARGE/JNP.JA28F16.JPG>
EG
- Luo, T. Z., & Maunsell, J. H. R. (2018). Attentional Changes in Either Criterion or Sensitivity Are Associated with Robust Modulations in Lateral Prefrontal Cortex. *Neuron*, 97(6), 1382–1393.e7. <https://doi.org/10.1016/j.neuron.2018.02.007>

- Mandonnet, E., Vincent, M., Valero-Cabré, A., Facque, V., Barberis, M., Bonnetblanc, F., Rheault, F., Volle, E., Descoteaux, M., & Margulies, D. S. (2020). Network-level causal analysis of set-shifting during trail making test part B: A multimodal analysis of a glioma surgery case. *Cortex*, 132, 238–249. <https://doi.org/10.1016/J.CORTEX.2020.08.021>
- Manly, T., & Mattingley, J. B. (2004). Visuospatial and Attentional Disorders. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical Neuropsychology: A Practical Guide to Assessment and Management for Clinicians* (pp. 229–252). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470013338.ch11>
- Manohar, S. G., Zokaei, N., Fallon, S. J., Vogels, T. P., & Husain, M. (2019). Neural mechanisms of attending to items in working memory. *Neuroscience and Biobehavioral Reviews*, 101, 1–12. <https://doi.org/10.1016/j.neubiorev.2019.03.017>
- Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474), 78–84. <https://doi.org/10.1038/nature12742>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), Article 7902. <https://doi.org/10.1038/s41586-022-04492-9>
- Marshall, J. C., & Halligan, P. W. (1995). Seeing the forest but only half the trees? *Nature*, 373(6514), 521–523. <https://doi.org/10.1038/373521a0>
- Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive Filling-in of Visual Surfaces in Parietal Extinction. *Science*, 275(5300), 671–674. <https://doi.org/10.1126/science.275.5300.671>

References

- Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, 29(6), 317–322. <https://doi.org/10.1016/j.tins.2006.04.001>
- McAdams, C. J., & Maunsell, J. H. R. (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>
- McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., Alexander, M., & Kilduff, P. T. (1993). Semantic processing in the neglected visual field: Evidence from a lexical decision task. *Cognitive Neuropsychology*, 10(1), 79–108. <https://doi.org/10.1080/02643299308253457>
- McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron*, 83(1), 202–215. <https://doi.org/10.1016/j.neuron.2014.05.019>
- Menon, V., & D’Esposito, M. (2022). The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology*, 47(1), 90–103. <https://doi.org/10.1038/s41386-021-01152-w>
- Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.1016/j.jml.2020.104092>
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Miller, E. K., & Wallis, J. D. (2009). Executive function and higher-order cognition: Definition and neural substrates. *Encyclopedia of Neuroscience*, 4(99–104).

- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). The integration of plans. In *Plans and the structure of behavior*. Henry Holt and Co. <https://doi.org/10.1037/10039-007>
- Milner, B. (1963). Effects of Different Brain Lesions on Card Sorting: The Role of the Frontal Lobes. *Archives of Neurology*, 9(1), 90–100. <https://doi.org/10.1001/archneur.1963.00460070100010>
- Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2007). Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron*, 55(1), 131–141. <https://doi.org/10.1016/j.neuron.2007.06.018>
- Moerel, D., Grootswagers, T., Robinson, A. K., Shatek, S. M., Woolgar, A., Carlson, T. A., & Rich, A. N. (2022). The time-course of feature-based attention effects dissociated from temporal expectation and target-related processes. *Scientific Reports*, 12(1), 6968. <https://doi.org/10.1038/s41598-022-10687-x>
- Moerel, D., Rich, A. N., & Woolgar, A. (2021). *Selective attention and decision-making have separable neural bases in space and time* [Preprint]. Neuroscience. <https://doi.org/10.1101/2021.02.28.433294>
- Molenberghs, P., Sale, M. V., & Mattingley, J. B. (2012). Is there a critical lesion site for unilateral spatial neglect? A meta-analysis using activation likelihood estimation. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00078>
- Moore, M. J., Hearne, L., Demeyere, N., & Mattingley, J. B. (2023). Comprehensive voxel-wise, tract-based, and network lesion mapping reveals unique architectures of right and left visuospatial neglect. *Brain Structure and Function*, 228(9), 2067–2087. <https://doi.org/10.1007/s00429-023-02702-2>
- 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>

References

- Morey, R., & Rouder, J. (2011). Bayes Factor Approaches for Testing Interval Null Hypotheses. *Psychological Methods*, 16, 406–419. <https://doi.org/10.1037/a0024377>
- Morey, R., & Rouder, J. (2018). *BayesFactor: Computation of Bayes Factors for Common Designs*. <https://CRAN.R-project.org/package=BayesFactor>
- Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., & Husain, M. (2003). The anatomy of visual neglect. *Brain*, 126(9), 1986–1997. <https://doi.org/10.1093/brain/awg200>
- Mountain, M. A., & Snow, W. G. (1993). Wisconsin card sorting test as a measure of frontal pathology: A review. *Clinical Neuropsychologist*, 7(1), 108–118. <https://doi.org/10.1080/13854049308401893>
- Mrah, S., Descoteaux, M., Wager, M., Boré, A., Rheault, F., Thirion, B., & Mandonnet, E. (2022). Network-level prediction of set-shifting deterioration after lower-grade glioma resection. *Journal of Neurosurgery*, 137(5), 1329–1337. <https://doi.org/10.3171/2022.1.JNS212257>
- Muhle-Karbe, P. S., Duncan, J., De Baene, W., Mitchell, D. J., & Brass, M. (2017). Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex. *Cerebral Cortex*, 27(3), 1891–1905. <https://doi.org/10.1093/cercor/bhw032>
- Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, 4(1), 101–109. <https://doi.org/10.1093/scan/nsn044>
- Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, 23(5), 1316–1340. <https://doi.org/10.3758/s13423-015-0982-5>

- Naya, Y., Chen, H., Yang, C., & Suzuki, W. A. (2017). Contributions of primate prefrontal cortex and medial temporal lobe to temporal-order memory. *Proceedings of the National Academy of Sciences*, 114(51), 13555–13560. <https://doi.org/10.1073/pnas.1712711114>
- Nelson, H. E. (1982). *National Adult Reading Test (NART): Test manual*. NFER-Nelson.
- Nelson, H. E., & Willison, J. R. (1991). *The National Adult Reading Test (NART): Test manual* (2nd ed.). NFER-Nelson.
- Niessen, E., Ant, J. M., Bode, S., Saliger, J., Karbe, H., Fink, G. R., Stahl, J., & Weiss, P. H. (2020). Preserved performance monitoring and error detection in left hemisphere stroke. *NeuroImage: Clinical*, 27, 102307. <https://doi.org/10.1016/j.nicl.2020.102307>
- Noah, S., Meyyappan, S., Ding, M., & Mangun, G. R. (2023). Time Courses of Attended and Ignored Object Representations. *Journal of Cognitive Neuroscience*, 35(4), 645–658. https://doi.org/10.1162/jocn_a_01972
- Noudoost, B., Chang, M. H., Steinmetz, N. A., & Moore, T. (2010). Top-down control of visual attention. *Current Opinion in Neurobiology*, 20(2), 183–190. <https://doi.org/10.1016/j.conb.2010.02.003>
- O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, 5(11), 1203–1209. <https://doi.org/10.1038/nn957>
- O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature* 1999 401:6753, 401(6753), 584–587. <https://doi.org/10.1038/44134>
- Olszowy, W., Aston, J., Rua, C., & Williams, G. B. (2019). Accurate autocorrelation modeling substantially improves fMRI reliability. *Nature Communications*, 10(1), 1220. <https://doi.org/10.1038/s41467-019-09230-w>

References

- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.
- Peers, P. V., Ludwig, C. J., Rorden, C., Cusack, R., Bonfiglioli, C., Bundesen, C., Driver, J., Antoun, N., & Duncan, J. (2005). Attentional functions of parietal and frontal cortex. *Cerebral Cortex*, 15(10), 1469–1484.
- Penfield, W., & Evans, J. (1935). The frontal lobe in man: A clinical study of maximum removals. *Brain*, 58(1), 115–133. <https://doi.org/10.1093/brain/58.1.115>
- Penke, L., Maniega, S. M., Bastin, M. E., Valdés Hernández, M. C., Murray, C., Royle, N. A., Starr, J. M., Wardlaw, J. M., & Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry*, 17(10), Article 10. <https://doi.org/10.1038/mp.2012.66>
- Pischedda, D., Görgen, K., Haynes, J.-D., & Reverberi, C. (2017). Neural representations of hierarchical rule sets: The human control system represents rules irrespective of the hierarchical level to which they belong. *The Journal of Neuroscience*, 37(50), 12281–12296. <https://doi.org/10.1523/JNEUROSCI.3088-16.2017>
- Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25. <https://doi.org/10.1080/00335558008248231>
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160–174. <https://doi.org/10.1037/0096-3445.109.2.160>
- Prabhakaran, V., Smith, J. A. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1997). Neural Substrates of Fluid Reasoning: An fMRI Study of Neocortical Activation during Performance of the Raven's Progressive Matrices Test. *Cognitive Psychology*, 33(1), 43–63. <https://doi.org/10.1006/cogp.1997.0659>

- Purcell, J. J., Wiley, R. W., & Rapp, B. (2019). Re-learning to be different: Increased neural differentiation supports post-stroke language recovery. *NeuroImage*, 202, 116145. <https://doi.org/10.1016/j.neuroimage.2019.116145>
- Puy, L., Barbay, M., Roussel, M., Canaple, S., Lamy, C., Arnoux, A., Leclercq, C., Mas, J.-L., Tasseel-Ponche, S., Constans, J.-M., Godefroy, O., & null, null. (2018). Neuroimaging Determinants of Poststroke Cognitive Performance. *Stroke*, 49(11), 2666–2673. <https://doi.org/10.1161/STROKEAHA.118.021981>
- 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>
- Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of What and Where in the Primate Prefrontal Cortex. *Science*, 276(5313), 821–824. <https://doi.org/10.1126/science.276.5313.821>
- Raven, J. C., Court, J. H., & Raven, J. (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. H. K. Lewis.
- Reber, J., Hwang, K., Bowren, M., Bruss, J., Mukherjee, P., Tranel, D., & Boes, A. D. (2021). Cognitive impairment after focal brain lesions is better predicted by damage to structural than functional network hubs. *Proceedings of the National Academy of Sciences*, 118(19), e2018784118. <https://doi.org/10.1073/pnas.2018784118>
- Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79(4), 518–524. <https://doi.org/10.1037/0021-9010.79.4.518>
- 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>

References

- Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585–590.
- Ritchie, J. B., & Carlson, T. A. (2016). Neural Decoding and “Inner” Psychophysics: A Distance-to-Bound Approach for Linking Mind, Brain, and Behavior. *Frontiers in Neuroscience*, 10. <https://www.frontiersin.org/articles/10.3389/fnins.2016.00190>
- Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *British Journal for the Philosophy of Science*, 70(2), 581–607. <https://doi.org/10.1093/bjps/axx023>
- Robinson, A. K., Rich, A. N., & Woolgar, A. (2022). Linking the Brain with Behavior: The Neural Dynamics of Success and Failure in Goal-directed Behavior. *Journal of Cognitive Neuroscience*, 34(4), 639–654. https://doi.org/10.1162/jocn_a_01818
- Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of Consulting and Clinical Psychology*, 48(5), 605–614. <https://doi.org/10.1037/0022-006X.48.5.605>
- Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., Manes, F., & Duncan, J. (2010). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, 133(1), 234–247. <https://doi.org/10.1093/BRAIN/AWP269>
- Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, 12(4), 191–200. <https://doi.org/10.1155/2000/421719>
- 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](https://doi.org/10.1016/0042-6989(94)00156-G)

- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Sàenz, M., Buraças, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience* 2002 5:7, 5(7), 631–632. <https://doi.org/10.1038/nn876>
- Sàenz, M., Buraças, 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](https://doi.org/10.1016/S0042-6989(02)00595-3)
- Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research*, 97(3), 423–436. <https://doi.org/10.1007/BF00241536>
- Salvalaggio, A., De Filippo De Grazia, M., Zorzi, M., Thiebaut de Schotten, M., & Corbetta, M. (2020). Post-stroke deficit prediction from lesion and indirect structural and functional disconnection. *Brain : A Journal of Neurology*, 143(7), 2173–2188. <https://doi.org/10.1093/brain/awaa156>
- Santaracchi, E., Emmendorfer, A., & Pascual-Leone, A. (2017). Dissecting the parieto-frontal correlates of fluid intelligence: A comprehensive ALE meta-analysis study. *Intelligence*, 63, 9–28. <https://doi.org/10.1016/j.intell.2017.04.008>
- Sarvas, J. (1987). Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Physics in Medicine & Biology*, 32(1), 11. <https://doi.org/10.1088/0031-9155/32/1/004>
- Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience*, 17(4), Article 4. <https://doi.org/10.1038/nn.3656>

References

- Schoenfeld, M., Hopf, J.-M., Martinez, A., Mai, H., Sattler, C., Gasde, A., Heinze, H.-J., & Hillyard, S. (2007). Spatio-temporal Analysis of Feature-Based Attention. *Cerebral Cortex*, 17(10), 2468–2477. <https://doi.org/10.1093/cercor/bhl154>
- Schubert, A.-L., Hagemann, D., & Frischkorn, G. T. (2017). Is general intelligence little more than the speed of higher-order processing? *Journal of Experimental Psychology. General*, 146(10), 1498–1512. <https://doi.org/10.1037/xge0000325>
- Schultz, D. H., Ito, T., & Cole, M. W. (2022). Global connectivity fingerprints predict the domain generality of multiple-demand regions. *Cerebral Cortex*, 32(20), 4464–4479. <https://doi.org/10.1093/cercor/bhab495>
- Scrivener, C. L., Grootswagers, T., & Woolgar, A. (2023). *Optimising analysis choices for multivariate decoding: Creating pseudotrials using trial averaging and resampling* (p. 2023.10.04.560678). bioRxiv. <https://doi.org/10.1101/2023.10.04.560678>
- Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *Journal of Neuroscience*, 27(9), 2349–2356. <https://doi.org/10.1523/JNEUROSCI.5587-06.2007>
- Selemon, L. D., & Goldman-Rakic, P. S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: Evidence for a distributed neural network subserving spatially guided behavior. *Journal of Neuroscience*, 8(11), 4049–4068. <https://doi.org/10.1523/JNEUROSCI.08-11-04049.1988>
- Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, 114(2), 727–741. <https://doi.org/10.1093/brain/114.2.727>
- Shashidhara, S., Mitchell, D. J., Erez, Y., & Duncan, J. (2019). Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity,

- Time Pressure, and Reward. *Journal of Cognitive Neuroscience*, 31(11), 1617–1630. https://doi.org/10.1162/jocn_a_01440
- Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, 44(3), 535–551. <https://doi.org/10.1016/j.paid.2007.09.015>
- Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex*, 22(1), 158–165. <https://doi.org/10.1093/cercor/bhr099>
- Siegel, J. S., Ramsey, L. E., Snyder, A. Z., Metcalfe, N. V., Chacko, R. V., Weinberger, K., Baldassarre, A., Hacker, C. D., Shulman, G. L., & Corbetta, M. (2016). Disruptions of network connectivity predict impairment in multiple behavioral domains after stroke. *Proceedings of the National Academy of Sciences*, 113(30), E4367–E4376. <https://doi.org/10.1073/pnas.1521083113>
- Siegel, J. S., Seitzman, B. A., Ramsey, L. E., Ortega, M., Gordon, E. M., Dosenbach, N. U. F., Petersen, S. E., Shulman, G. L., & Corbetta, M. (2018). Re-emergence of modular brain networks in stroke recovery. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 101, 44–59. <https://doi.org/10.1016/j.cortex.2017.12.019>
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98. <https://doi.org/10.1016/j.neuroimage.2008.03.061>
- Snow, J. C., & Mattingley, J. B. (2006). Goal-driven selective attention in patients with right hemisphere lesions: How intact is the ipsilesional field? *Brain: A Journal of Neurology*, 129(Pt 1), 168–181. <https://doi.org/10.1093/brain/awh690>

References

- Sobrado, A., Palenciano, A. F., González-García, C., & Ruz, M. (2022). The effect of task demands on the neural patterns generated by novel instruction encoding. *Cortex*, *149*, 59–72. <https://doi.org/10.1016/j.cortex.2022.01.010>
- Soreq, E., Leech, R., & Hampshire, A. (2019). Dynamic network coding of working-memory domains and working-memory processes. *Nature Communications*, *10*(1), 936. <https://doi.org/10.1038/s41467-019-08840-8>
- Spearman, C. (1904). 'General Intelligence', Objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. <https://doi.org/10.2307/1412107>
- Spearman, C. (1927). *The abilities of man* (pp. xxiii, 415). Macmillan.
- Sperber, C. (2022). The strange role of brain lesion size in cognitive neuropsychology. *Cortex*, *146*, 216–226. <https://doi.org/10.1016/j.cortex.2021.11.005>
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased Attention Enhances Both Behavioral and Neuronal Performance. *Science*, *240*(4850), 338–340. <https://doi.org/10.1126/science.3353728>
- Sprague, T. C., Itthipuripat, S., Vo, V. A., & Serences, J. T. (2018). Dissociable signatures of visual salience and behavioral relevance across attentional priority maps in human cortex. *Journal of Neurophysiology*, *119*(6), 2153–2165. <https://doi.org/10.1152/jn.00059.2018>
- Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience*, *16*(12), 1879–1887. <https://doi.org/10.1038/nn.3574>
- Sripada, C., Angstadt, M., Rutherford, S., & Taxali, A. (2019). *Brain Network Mechanisms of General Intelligence* (p. 657205). bioRxiv. <https://doi.org/10.1101/657205>
- 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>

- Stone, S., Halligan, P., & Greenwood, R. (1993). The incidence of neglect phenomena and related disorders in patients with an acute right or left hemisphere stroke. *Age and Ageing*, 22(1), 46–52.
- Stuss, D. T., Levine, B., Alexander, M. P., Hong, J., Palumbo, C., Hamer, L., Murphy, K. J., & Izukawa, D. (2000). Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: Effects of lesion location and test structure on separable cognitive processes. *Neuropsychologia*, 38(4), 388–402. [https://doi.org/10.1016/S0028-3932\(99\)00093-7](https://doi.org/10.1016/S0028-3932(99)00093-7)
- Taulu, S., & Kajola, M. (2005). Presentation of electromagnetic multichannel data: The signal space separation method. *Journal of Applied Physics*, 97(12), 124905. <https://doi.org/10.1063/1.1935742>
- Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine & Biology*, 51(7), 1759. <https://doi.org/10.1088/0031-9155/51/7/008>
- Teichmann, L., Moerel, D., Baker, C., & Grootswagers, T. (2021). An empirically-driven guide on using Bayes Factors for M/EEG decoding. *bioRxiv*, 2021.06.23.449663. <https://doi.org/10.1101/2021.06.23.449663>
- Ten Brink, A. F., Verwer, J. H., Biesbroek, J. M., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2017). Differences between left- and right-sided neglect revisited: A large cohort study across multiple domains. *Journal of Clinical and Experimental Neuropsychology*, 39(7), 707–723. <https://doi.org/10.1080/13803395.2016.1262333>
- Thiebaut de Schotten, M., Foulon, C., & Nachev, P. (2020). Brain disconnections link structural connectivity with function and behaviour. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-18920-9>
- Thiele, A., & Bellgrove, M. A. (2018). Neuromodulation of Attention. *Neuron*, 97(4), 769–785. <https://doi.org/10.1016/j.neuron.2018.01.008>

References

- Thiele, A., Brandt, C., Dasilva, M., Gotthardt, S., Chicharro, D., Panzeri, S., & Distler, C. (2016). Attention Induced Gain Stabilization in Broad and Narrow-Spiking Cells in the Frontal Eye-Field of Macaque Monkeys. *Journal of Neuroscience*, 36(29), 7601–7612. <https://doi.org/10.1523/JNEUROSCI.0872-16.2016>
- Thölke, P., Mantilla-Ramos, Y.-J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O’Byrne, J., & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, 277, 120253. <https://doi.org/10.1016/j.neuroimage.2023.120253>
- Thomson, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8, 271–281.
- Thye, M., & Mirman, D. (2018). Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. *NeuroImage: Clinical*, 20, 1129–1138. <https://doi.org/10.1016/j.nicl.2018.10.017>
- Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage*, 77, 157–165. <https://doi.org/10.1016/j.neuroimage.2013.03.039>
- Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The Retinotopy of Visual Spatial Attention. *Neuron*, 21(6), 1409–1422. [https://doi.org/10.1016/S0896-6273\(00\)80659-5](https://doi.org/10.1016/S0896-6273(00)80659-5)
- Toth, L. J., & Assad, J. A. (2002). Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature*, 415(6868), 165–168. <https://doi.org/10.1038/415165a>
- 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>

- Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., & Fellows, L. K. (2019). Lesion Studies in Contemporary Neuroscience. *Trends in Cognitive Sciences*, 23(8), 653–671. <https://doi.org/10.1016/J.TICS.2019.05.009>
- Vallar, G. (1993). The anatomical basis of spatial hemineglect in humans. In I. H. Robertson & J. C. Marshall (Eds.), *Unilateral neglect: Clinical and experimental studies* (pp. 27–59). Lawrence Erlbaum Associates Ltd. Publishers.
- Vallar, G., & Calzolari, E. (2018). Unilateral spatial neglect after posterior parietal damage. In G. Vallar & H. B. Coslett (Eds.), *Handbook of Clinical Neurology* (Vol. 151, pp. 287–312). Elsevier. <https://doi.org/10.1016/B978-0-444-63622-5.00014-0>
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Vanduffel, W., Tootell, R. B. H., Schoups, A. A., & Orban, G. A. (2002). The Organization of Orientation Selectivity Throughout Macaque Visual Cortex. *Cerebral Cortex*, 12(6), 647–662. <https://doi.org/10.1093/cercor/12.6.647>
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology*, 100(6), 3328–3342. <https://doi.org/10.1152/jn.90355.2008>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., ... Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>

References

- Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411(6840), 953–956. <https://doi.org/10.1038/35082081>
- Ward, R., Goodrich, S., & Driver, J. (1994). Grouping reduces visual extinction: Neuropsychological evidence for weight-linkage in visual selection. *Visual Cognition*, 1(1), 101–129. <https://doi.org/10.1080/13506289408402295>
- Wen, T., Mitchell, D. J., & Duncan, J. (2018). Response of the multiple-demand network during simple stimulus discriminations. *NeuroImage*, 177, 79–87. <https://doi.org/10.1016/j.neuroimage.2018.05.019>
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on Psychological Science*, 6(3), 291–298. <https://doi.org/10.1177/1745691611406923>
- Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience*, 10(6), 685–686. <https://doi.org/10.1038/nn1900>
- Wisniewski, D., González-García, C., Formica, S., Woolgar, A., & Brass, M. (2023). Adaptive coding of stimulus information in human frontoparietal cortex during visual classification. *NeuroImage*, 274, 120150. <https://doi.org/10.1016/j.neuroimage.2023.120150>
- Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible coding of task rules in frontoparietal cortex: An adaptive system for flexible cognitive control. *Journal of Cognitive Neuroscience*, 27(10), 1895–1911. https://doi.org/10.1162/jocn_a_00827
- Woolgar, A., Bor, D., & Duncan, J. (2013). Global Increase in Task-related Frontoparietal Activity after Focal Frontal Lobe Lesion. *Journal of Cognitive Neuroscience*, 25(9), 1542–1552. https://doi.org/10.1162/jocn_a_00432

- Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour*, 2(3), 200–204. <https://doi.org/10.1038/s41562-017-0282-3>
- Woolgar, A., Golland, P., & Bode, S. (2014). Coping with confounds in multivoxel pattern analysis: What should we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013. *Neuroimage*, 98, 506–512. <https://doi.org/10.1016/j.neuroimage.2014.04.059>
- Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive coding of task-relevant information in human frontoparietal cortex. *The Journal of Neuroscience*, 31(41), 14592–14599. <https://doi.org/10.1523/JNEUROSCI.2616-11.2011>
- Woolgar, A., Jackson, J. B., & Duncan, J. (2016). Coding of visual, auditory, rule, and response information in the brain: 10 years of multivoxel pattern analysis. *Journal of Cognitive Neuroscience*, 28(10), 1433–1454. https://doi.org/10.1162/jocn_a_00981
- Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., Roca, M., Antoun, N., Manes, F., & Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, 107(33), 14899–14902. <https://doi.org/10.1073/PNAS.1007928107>
- Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *Neuroimage*, 109, 429–437. <https://doi.org/10.1016/j.neuroimage.2014.12.083>
- Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology*, 33(3), 292–296. <https://doi.org/10.1037/a0030727>

References

- Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*, *106*, 1125–1165. <https://doi.org/10.1152/jn.00338.2011>
- 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*, 97–107. <https://doi.org/10.1016/j.NEUROIMAGE.2017.05.055>
- Zhang, C. Y., Aflalo, T., Revechkis, B., Rosario, E. R., Ouellette, D., Pouratian, N., & Andersen, R. A. (2017). Partially Mixed Selectivity in Human Posterior Parietal Association Cortex. *Neuron*, *95*(3), 697–708.e4. <https://doi.org/10.1016/j.neuron.2017.06.040>
- Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience*, *12*(1), Article 1. <https://doi.org/10.1038/nn.2223>
- Zheng, Y., Lu, R., & Woolgar, A. (2024). Radical flexibility of neural representation in frontoparietal cortex and the challenge of linking it to behaviour. *Current Opinion in Behavioral Sciences*, *57*, 101392. <https://doi.org/10.1016/j.cobeha.2024.101392>

