

Investigation of the multiple-demand network at multiple spatial scales



Sneha Shashidhara

Clare College

MRC – Cognition and Brain Sciences Unit

This dissertation is submitted for the degree of
Doctor of Philosophy at University of Cambridge

September 2019

Preface

I declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration. Where reference is made to the work of others, the extent to which that work has been used is indicated. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

The study outlined in Chapter 2 have been published in a peer-reviewed journal, and the studies outlined in Chapter 3 and Chapter 4 have been submitted to a preprint server. The references for these publications are as follows:

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*.

Shashidhara, S., Spronkers, F., Erez, Y. (2019). Localizing the ‘multiple-demand’ frontoparietal network in individual subjects. *bioRxiv*, 661934.

Shashidhara, S., Erez, Y. (2019). Reward motivation modulates coding of behaviorally relevant category distinctions across the frontoparietal cortex. *bioRxiv*, 609537.

The study outlined in Chapter 3 was done in collaboration with Spronkers, F., who tested the participants.

The study outlined in Chapter 5 was done in collaboration with Watanabe K., Kadohisa M., and Kusunoki M., who provided the data.

This dissertation does not exceed 60,000 words in length.

Acknowledgment

I would like to thank everybody who made this PhD possible.

My supervisor, John Duncan, for his encouragement and invaluable guidance. In particular those unplanned discussions that made the past four years both productive as well as fun and interesting!

Yaara Erez for navigating me through multiple projects. Her advice has been extremely helpful in all respects. Daniel Mitchell for teaching me and patiently answering all my questions (perhaps the same questions multiple times). The imaging community and admin for making these studies possible.

Everyone else at the CBU whose doors I have knocked for help, which was always so promptly given.

A big thank you to my parents and friends who kept me almost sane during this time. The Gates Trust and Clare College for funding and supporting me through this journey. The people who patiently helped the writing process, thank you. Last but not the least, coffee and cakes that made functioning possible.

Abstract

This dissertation investigates the frontoparietal ‘multiple-demand’ (MD) network that is involved in the processing of diverse cognitive demands. This network is active when the task at hand is made more demanding, in a variety of different tasks including working memory, task switching, inhibition, math, language etc.

While the different MD regions have partly different functions, they are highly interconnected allowing them to function together as a network. The experiment in Chapter 2 looked at the interplay between functional differences as well as co-recruitment within this multiple-demand network. Quantitative differences between regions were more prominent in simple tasks. A strong co-recruitment was seen with increased challenge or incentive.

In Chapter 3, task preferences were studied at the voxel level. MD regions were equally well localised in single-subjects using any of three task demands. Voxels localised by all three tasks also captured the underlying neural representations to a similar level in a separate criterion task.

Chapter 4 investigated if task representations, as measured by multi-voxel patterns, were modified due to external motivation. The effect was limited to the cue phase and did not extend to the stimulus processing phase where the stimulus is integrated with the cue to arrive at the response.

Chapter 5 examined neural representations in frontal and parietal regions more directly through single unit activity and local field potentials (LFPs), during a spatial working memory task. While single neurons showed dynamic coding of target information rather than persistent coding, LFPs held this information constant through time. The impact of reference voltages on LFP data was further investigated.

Together, these results explore the functional differences between and within the MD regions, and provide evidence for flexible task representations at the voxel and neuronal level.

Table of Contents

Preface.....	2
Acknowledgment.....	3
Abstract.....	4
Table of Contents	5
Chapter 1	8
Introduction.....	8
Defining the MD network in the brain	10
Historic perspective of the structure and function of the frontal lobe.....	17
Norman and Shallice model and the implication for functional differences in the frontal lobe	19
Alternate theories of frontal lobe function	22
MD regions form a network: Evidence from connectivity studies.....	28
Function of the MD regions: Evidence from multi-voxel patterns	29
Neuronal functions and mechanisms in the MD regions	35
Précis.....	43
Chapter 2	46
Progressive recruitment of the frontoparietal multiple demand system with increased task complexity, time pressure and reward.....	46
Introduction	46
Methods	49
Participants	49
Task Details.....	49
Pre-Scanning Session	53
Scanning Session.....	53
Data Acquisition.....	53
Eye Tracking	54
fMRI analysis	54
Results.....	56
Behaviour	56
Imaging.....	57
Eye tracking.....	65
Discussion	66
Chapter 3	70

Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the ‘multiple-demand’ frontoparietal network	70
Introduction	70
Methods	74
Participants	74
Experimental paradigm	74
Localiser tasks and criterion task	74
fMRI data acquisition	80
Data Analysis	80
Statistical testing and code	85
Results	86
Behavioural results	86
Whole brain and ROI univariate analysis	87
Comparisons of activity patterns between localisers	89
Subject-specific ROIs and univariate activity in the rule-based criterion task	92
Subject-specific ROIs and rule decoding in the rule-based criterion task	94
Effect of ROI size on decoding results	98
Discussion	99
Chapter 4	104
Reward motivation does not modulate coding of behaviourally relevant category distinctions across the frontoparietal cortex	104
Introduction	104
Materials and Methods	108
Participants	108
Task Design	108
Stimuli	111
Structure and Design	111
Functional Localisers	112
Scanning Session	113
Data Acquisition	113
Data and Statistical Analysis	113
Preprocessing	114
General Linear Model (GLM) for the Main Task	114
GLM for the Functional Localisers	115
Univariate Analysis	115
ROI Definition	115

Voxels selection for MVPA	116
Multivoxel pattern analysis (MVPA)	116
Whole brain searchlight pattern analysis	117
Results	119
Behaviour	119
Activity across the MD network during the cue epoch	120
Univariate activity in the MD network during the stimulus epoch	121
Effect of reward motivation on discrimination of behaviourally relevant category distinctions in the MD network	123
Effects of reward motivation on behaviourally relevant category distinctions in LOC	127
Conflict-contingent vs. visual category effects	129
Discussion	132
Chapter 5	136
A new working memory signal in frontoparietal local field potentials	136
Introduction	136
Methods	139
Subjects	139
Task	139
Recordings	142
Data and Analysis	142
Results	145
Behaviour	145
Pattern of coding in spikes and LFPs	146
Stability of information across time	150
Origin of the memory trace	152
Discussion	154
Discussion	159
Bibliography	165
Appendix	190

Chapter 1

Introduction

How we solve any problem or task is one of the challenging open questions in systems neuroscience. Any everyday activity can be broken down into smaller subparts or steps needed to achieve a goal. For example, if the goal is have a cup of coffee, this can be achieved through 1.) Walking to the kitchen, 2.) Pouring water over the coffee, 3.) Adding milk and sugar etc. There has to be a system in the brain that breaks any problem into such smaller action steps, while keeping the end goal in mind. The key is to complete these subparts in the correct sequence, rapidly transitioning from one to the next. Having a separate system for each problem we encounter is not feasible. We can thus conceive of a system that can tackle any kind of problem, ranging from math to planning a vacation.

This system is involved in ‘cognitive control’ – the process by which behaviour is planned in order to achieve a goal. It is primarily concerned with the structure and requirements of complex behaviour. The subparts of the problem could include different cognitive processes, requiring this system to process multiple types of cognitive demand. It is a general problem solving system involved in all types of executive functions – processes needed for cognitive control such as selecting and monitoring behaviours that lead to attainment of chosen goals. Thus we need a system involved in assembling series of sub-tasks rather than something restricted to specific tasks, like a region only interested in moving images or sounds of certain frequencies.

The next step would be to isolate such a system using task complexity. Demand of the task determines how hard or how long one needs to work to successfully complete it. When a problem becomes harder, the demand increases, and the system will work harder to solve it. While solving a math problem, $5 + 7$ presented visually, many areas in the brain would be activated including visual areas, problem solving areas, memory areas etc. When solving the more complex problem of $1564 + 5789$, all of the areas involved in solving the easier problem would be active and the problem solving system would be more active. The key aspect of this system is its ability to solve different types of demand, thus in order to isolate it we need multiple such demand increases. Other increases in demand could be memory related: remembering 4 items instead of 2, choice related: choosing an answer from 6 options

versus 2. Learning names of faces is harder if the gender of the face and name don't match (having to overcome/inhibit a dominant association), compared to when they do match. Regions that are more active in the difficult condition of all these tasks form a general problem solving system.

Experiments like this suggest a distributed network in the brain termed the multiple-demand (MD) network. A network consisting of different regions all active during various tasks, begs the question of whether all the regions are doing the same thing. The function of any node or region in the brain essentially depends upon its input, which is dependent on its anatomical location and the regions it is connected to. In a distributed network different nodes could have access to different types of information and have partly different functions. On the other hand, strong interconnections between them would lead to fast information exchange and allow for a network function.

An illustration of the concept of the multiple-demand network is shown in Figure 1.1. The stimuli and problem related information are fed as input to the MD system that uses other resources to help solve it. These resources could include information from memory of having solved a similar problem, or other relevant pieces of knowledge in memory. The network function will depend on the reward to be gained on solving the problem. In other words, reward would determine motivation to solve the task. Similarly, complexity of the task would also control the activity of the network. It helps determine the amount of effort the system needs to put in to solve the task. This thesis seeks to examine how the MD network handles various complex tasks. Its function is investigated at the level of the network, individual nodes, and the scale of individual cells in the brain.

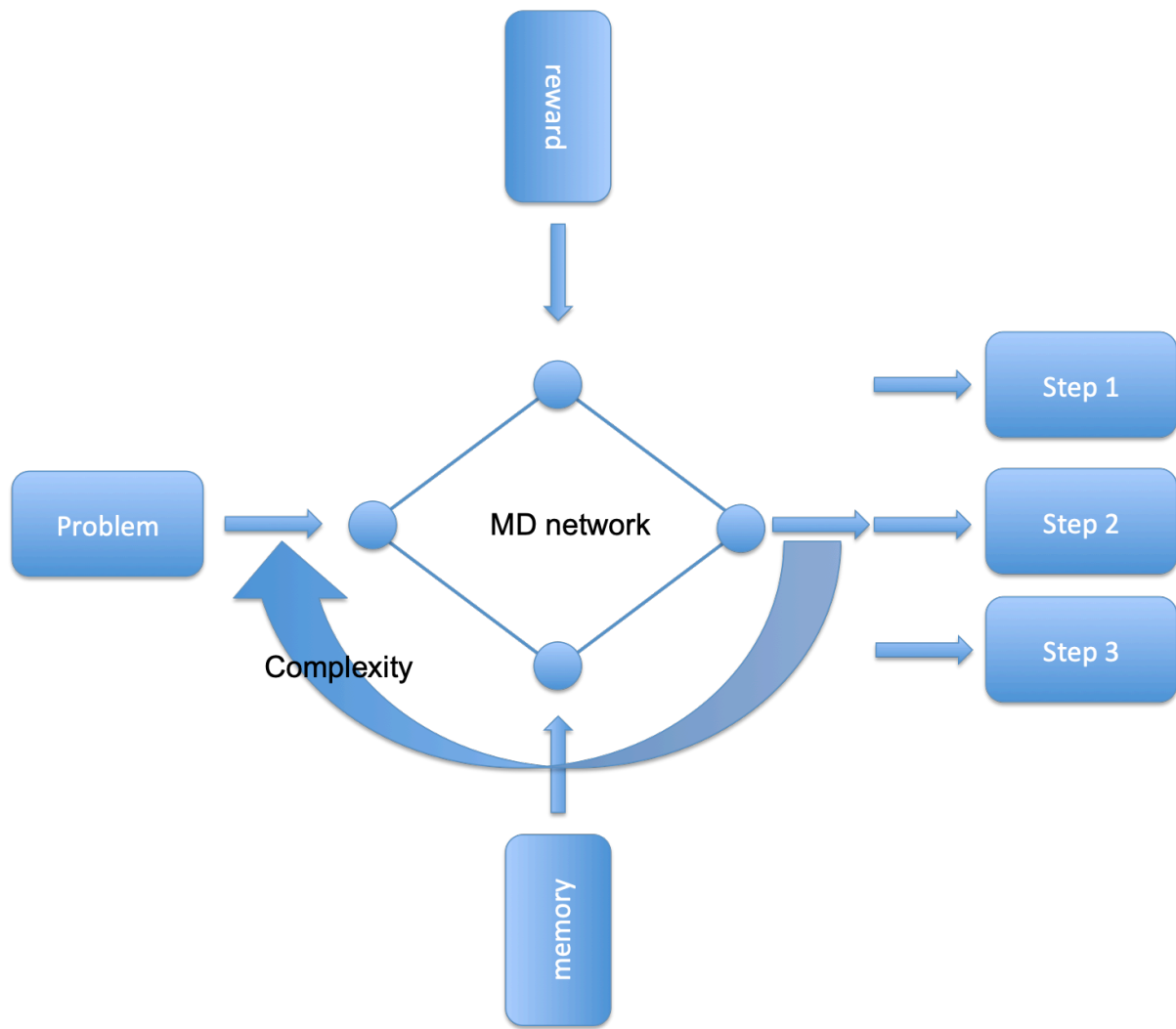


Figure 1.1: An illustration of the multiple-demand network. When presented with a problem, a network of highly interconnected nodes – in conjunction with other systems providing information from memory, estimate of reward to be gained on completing the task etc. – solves the task by breaking it into smaller manageable steps. Depending on the complexity of the problem, the MD network increases its activity or the effort it puts in to solving the problem.

Defining the MD network in the brain

The MD network is a specific and distributed network, comprising regions in the frontal and parietal cortices. The original network consisted of only frontal regions as shown by Duncan and Owen (2000). They compiled results from different studies that investigated the frontal lobe through various tasks. The advent of functional magnetic resonance imaging (fMRI) and

positron emission tomography (PET) made possible the examining of activation in the brain, in a non-invasive manner. The selected 20 such fMRI and PET studies that came under five broadly defined demands: response conflict, novelty, working memory including studies that investigated the effect of load and those studying delay activity, and perceptual difficulty (Duncan and Owen 2000). Response conflict is when the task leads to simultaneous activation of two incompatible responses. This demand included suppression of prepotent or dominant responses through Stroop studies, reversing previously trained stimulus-response mappings, and through incompatible response mappings such as saccade away from target etc. Given the importance of frontal lobe in learning of novel tasks, novelty studies included here compared the initial learning of an unfamiliar cognitive task, with later well-practised or more automatic performance. Working memory is a system that holds information for short periods of time and allows manipulation of such information, necessary for behaviour. The load on this system was investigated using immediate word recall, recognition and n-back studies. It was measured by assessing activation when more items were to be remembered compared to fewer items. Storage of information was studied through brain activity during the delay period. Lastly, the perceptual difficulty studies included degraded stimulus studies and object recognition from differing viewpoints.

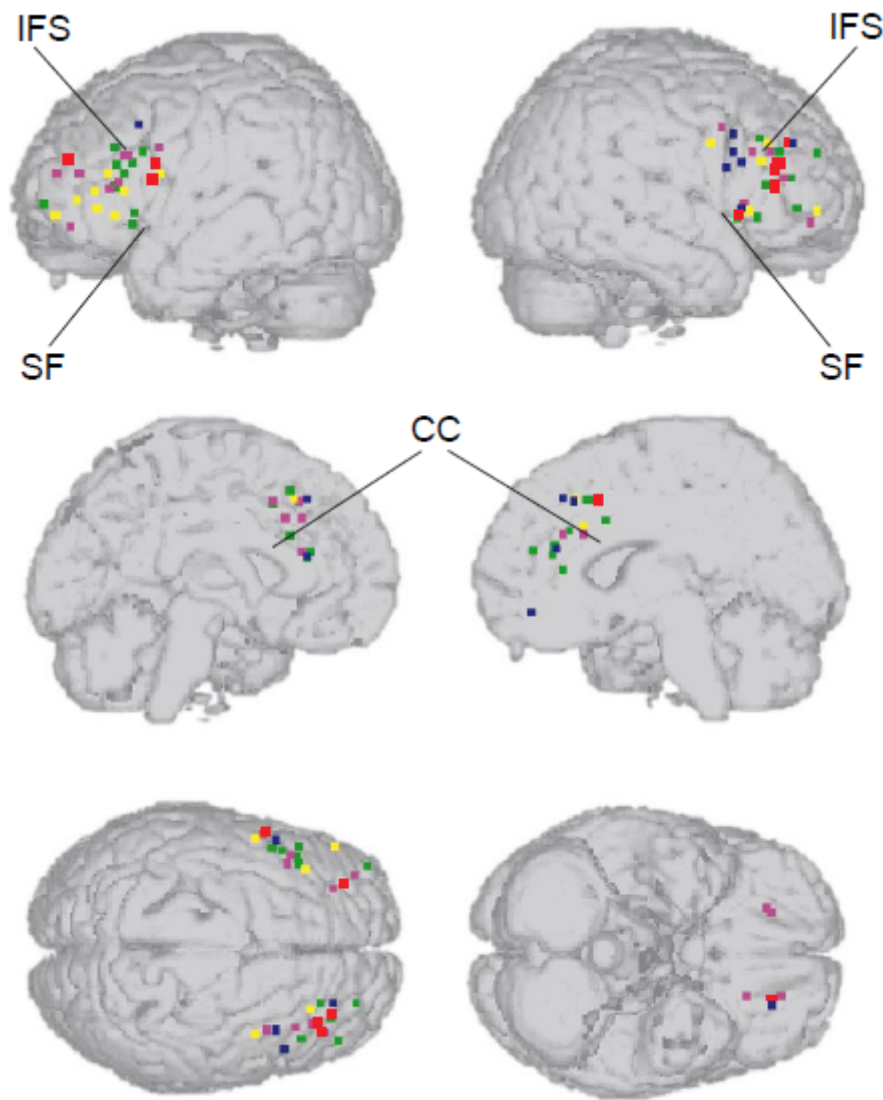


Figure 1.2: Clustering of activations from 5 different cognitive demands. Taken from Duncan and Owen (2000). Activations from the different cognitive demands are shown in different colours, response conflict (green), task novelty (pink), working memory load (yellow), working memory delay (red) and perceptual difficulty (blue). Abbreviations: CC, corpus callosum; IFS, inferior frontal sulcus; SF, Sylvian fissure.

These 20 studies used different tasks and task-designs, different stimuli, and included both PET and fMRI experiments. Clustering the activations from these diverse tasks showed a set of common regions in the frontal lobe (Figure 1.2). This provided the first definition of the MD network – a common set of areas in the frontal lobe – involved in different cognitive tasks and demands, stimulus modalities etc. Duncan (2006) first introduced IPS as a part of

the MD network (Figure 1.3), based on the same 20 studies of Duncan and Owen (2000). The MD network became a common set of regions across the frontal and parietal cortices involved in diverse tasks and behaviour. While there have been many studies focused on understanding the function of one of these regions such as parietal cortex (Simon et al. 2002), medial prefrontal cortex (Rushworth et al. 2004), and anterior prefrontal cortex (Höller-Wallscheid et al. 2017) etc., many of the same regions were indeed recruited by different demands, suggesting a specific frontoparietal network solving diverse cognitive problems. How these mid-dorsolateral, mid-ventrolateral, anterior cingulate and insular, and intraparietal regions work together to meet such diverse challenges remained a challenging question.

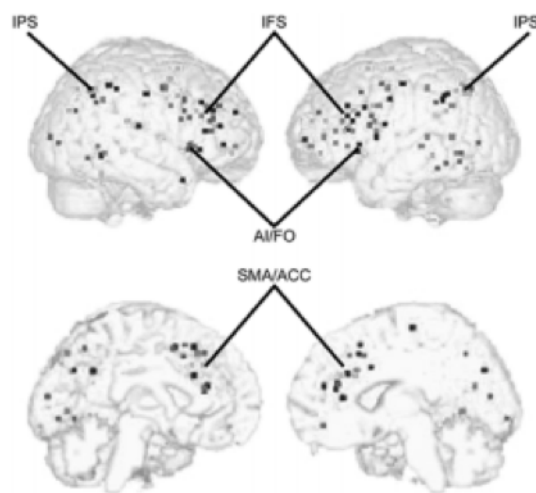


Figure 1.3. MD network including IPS. Taken from Duncan (2006). Abbreviations: IFS, inferior frontal sulcus; IPS, intraparietal sulcus; AI, anterior insula; FO, frontal operculum; SMA, supplementary motor area; ACC, anterior cingulate cortex.

Fedorenko et al. (2013) made a systematic attempt at localising the task-based multiple-demand network in fMRI ($n = 48$). To improve spatial localisation, they used activations from single subjects doing diverse tasks in an MRI scanner. The study included seven different tasks (Figure 1.4) and thresholded the mean activation of Hard versus Easy conditions across all the tasks. The tasks were the following: A sentence task included reading non-words (Hard) versus sentences (Easy). A math task required adding two to four numbers (Easy) versus adding six to eight numbers (Hard). In spatial/verbal working memory tasks, participants remembered four (Easy) and eight (Hard) locations/digit names. In

numerical/verbal multisource interference tasks, participants saw three digits/words and responded with the identity of the non-repeated digit/word. In the easy condition, the position of the non-repeated digit/word corresponded to the position of the response button, and the distractors were not possible responses (e.g., 100; left none none); in the hard condition, the position of the target digit/word did not match the position of the button, and the other digits were possible responses (e.g., 212, middle left middle). In Stroop, participants saw a word and overtly named the colour of the word's font. In the Easy condition the words were non-colour adjectives, and in the Hard condition the words were colour adjectives. In half of the trials in a given block, the font colour did not match the colour of the word.

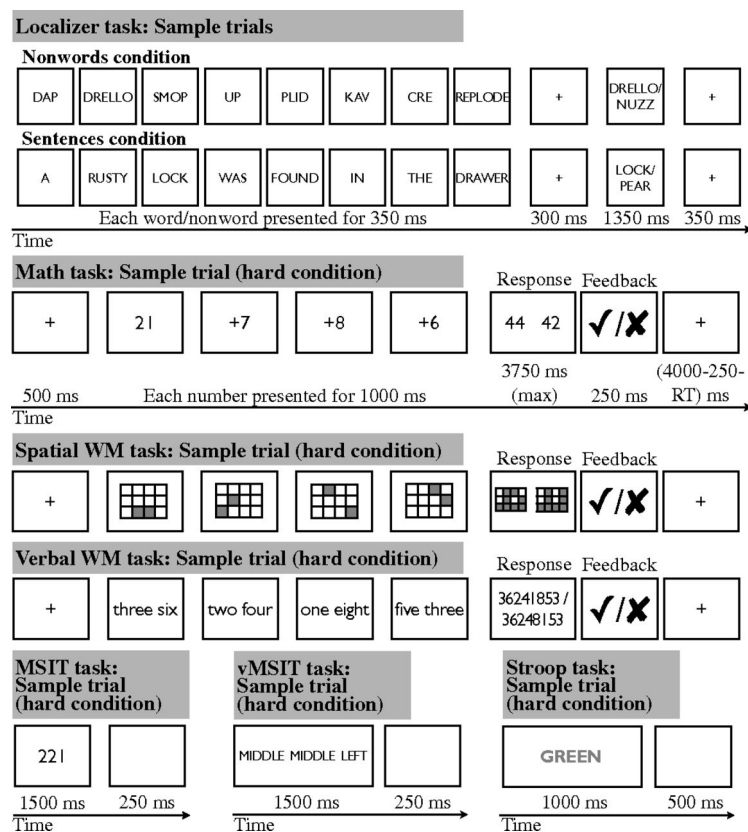


Figure 1.4: Diverse tasks used to localise the MD network. Taken from Fedorenko et al. (2013). Sample trials for the seven tasks are shown. The sentence task, also used as the localiser involved reading non-words contrasted with reading words. In the math task, participants added a smaller (Easy) or larger number of addends (Hard). In the spatial/verbal working memory tasks, participants remembered four (Easy) or eight (Hard) locations/digit names. In the multisource interference task, participants saw triplets of digits and had to press a button corresponding to the identity of the non-repeated digit (here 1). In the Easy

condition, the target digit and the response were the same, and the distractors could not be possible responses, while in the hard condition, the target digit and the response were different, and the distractors were possible responses. The verbal multisource interference task was a variant with words (left, middle and right) used instead of digits. In the Stroop task, participants overtly named the font colour of the words. In the Easy condition, the words were non-colour adjectives, and in the Hard they were colour adjectives whose font colour did not match in half the trials.

The tasks varied along several dimensions. First, the tasks differed in the kinds of stimuli they involved: the sentence, verbal working memory, verbal multisource interference tasks, and Stroop used word stimuli and verbal representations. Math and numerical multisource interference task used Arabic numerals, and the spatial working memory task used spatial locations. The tasks also differed in the kinds of cognitive processes they used. The sentence task required participants to read sequences of non-words or sentences and to respond to a memory probe at the end of each; working memory tasks required keeping sets of digits or spatial locations in memory for a brief period; the math task required manipulating representations and storing/updating intermediate results; and multisource interference tasks and Stroop required inhibiting a prepotent response. The common voxels activated by all the tasks were thresholded and divided to form seven regions of the multiple-demand network (Figure 1.5).

The MD voxels involved in all these diverse tasks. As they were isolated in the Hard versus Easy contrasts for each of the tasks, they increased their activation with an increase in demand. Definition of Hard or increase in demand was task dependent and was not always parametrically modulated. However, overlapping voxels or regions were activated in all these different types of increases in complexity. Functional profiles of these regions showed reliable Hard > Easy effects for all (or six of seven) tasks. The study also showed that other regions, including language regions and a bilateral temporal pole region, did not show this MD profile of activation during all types of cognitive demands. While the aspect of functional differences was not systematically analysed, the activation profiles of different MD regions showed different task preferences.

Regions included in the MD network are the posterior, middle and anterior parts of the middle frontal gyrus (pMFG, mMFG and aMFG), the posterior part of the dorsolateral frontal

cortex (pdLFC), the intra-parietal sulcus (IPS) and surrounding regions in the parietal lobe, the anterior insula and the adjoining frontal operculum (AI/FO), and the pre-supplementary motor area and the anterior cingulate cortex along the media surface (preSMA/ACC). These bilateral masks were used as MD regions of interest (ROIs) in the studies presented in this thesis, as well as in numerous other studies.

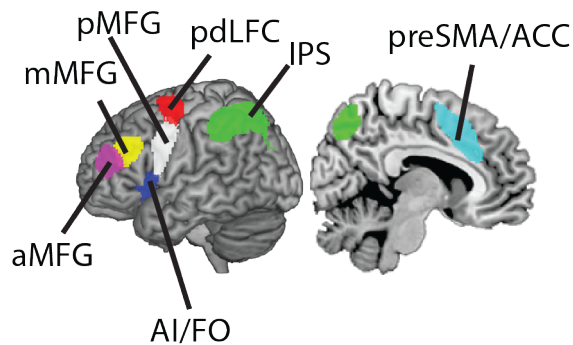


Figure 1.5: The multiple-demand network. The MD template with all the regions labelled, adapted from Fedorenko et al. (2013). While the regions from one hemisphere are shown here, the template is bilateral. Abbreviations: aMFG, anterior middle frontal gyrus; mMFG, mid middle frontal gyrus; pMFG, posterior middle frontal gyrus; pdLFC, posterior dorsal lateral frontal cortex; intraparietal sulcus; AI, anterior insula; FO, frontal operculum; SMA, supplementary motor area; ACC, anterior cingulate cortex.

In a more recent study, Assem et al. (2019) localised the MD network using data from the Human Connectome Project (HCP) with its new parcellation of the cortex into 360 parcels (Glasser et al. 2016). They defined the multiple-demand network as parcels active in a Hard versus Easy contrast of each of three tasks: working memory (2-back versus 0-back), math (math task versus a story task) and relational reasoning (find the dimension of relation between two stimuli and then apply that relation versus apply a known relation). They described a set of 27 parcels active in all three tasks, terming this the extended MD network. 10 of these parcels showed activity exceeding the mean activation of the 27 parcels in at least two out of three tasks, termed the ‘core’ MD network. While the task complexity manipulations were not ideal, this study was unique in identifying the multiple-demand regions in 449 subjects, all performing the same exact tasks. This finding was robust, as indicated by high levels of consistency of activations in the MD network across the subjects. Investigating the functional profiles of these regions, they showed that different parcels have

different relative task preferences. This gives credence to the idea that different MD regions may have access to more than one type of information/ cognitive process leading to partial functional specializations. However, during problem solving, different task contents are integrated and the interconnections between these regions allow for information exchange, enabling the breaking of a complex problem into manageable subparts or attentional episodes (Duncan 2013a).

Historic perspective of the structure and function of the frontal lobe

The frontal lobe, particularly the prefrontal cortex (PFC), is most commonly associated with cognitive control (Miller and Cohen 2001). Many of the structures in the MD network are part of the frontal lobe. In this section I will discuss the definitions of the prefrontal cortex, rooted in its anatomy and anatomical connections, and some seminal lesion studies that gave us some of the first hints of prefrontal function.

PFC is the anterior part of the frontal lobe and is one of the last structures to develop to its current state in the course of evolution. It has increased in size over the course of evolution and in humans occupies as much as 30% of the cerebral cortex. It extends across the lateral, medial and basal surfaces of the brain. This region is often implicated in planning, decision-making, and many other higher-order brain functions (Miller and Cohen 2001). The complex functionality of the region, and the idea that prefrontal cortex is important in making humans unique (Deacon 1997), led researchers to structurally investigate this region, and make comparisons between humans and other primate species.

Brodmann's monograph published in 1909 was most important in characterizing the prefrontal cortex. He speaks about the prefrontal and the parietal cortex: 'For my part I think that histology brings proof of what was previously only 'surmise' that it is the very last pallium to appear in the progress of phylogenesis. I would submit that it is a part with a future, but that at present its evolution is incomplete.' (Campbell 1904, pp. 655–56).

Brodmann defined the prefrontal cortex as a granular frontal region, which is present only in primate species (Carlén 2017). This is one of the most widely accepted definitions of prefrontal cortex today. Using non-human primate brains' volume of the prefrontal cortex, volume in humans was predicted through regression. These predictions usually turned out to be lower than the actual volume (Passingham 1973), suggesting an abnormally large prefrontal cortex in humans. Other studies have identified more specific parts of prefrontal

cortex that have been particularly enlarged in humans (Semendeferi et al. 2001; Petrides 2005).

In ontogeny as in phylogeny, the prefrontal cortex is a late-developing part of the brain. The development of the brain occurs through the interaction of several processes, some of which are completed before birth such as neurulation, cell proliferation, and migration, and others continue into adulthood. Pathology in its development after birth is linked to diseases like schizophrenia. PFC is one of the last regions of the brain to mature based on many indicators of development, and lateral regions of the PFC compared to the primary motor and ventromedial regions are the latest developing areas (Fuster 2002). In fact, the dorsolateral and medial PFC, regions contributing to the MD network, expand to nearly twice their original size after birth. The dorsolateral PFC reaches adult levels of cortical thickness only in early adolescence (Hill et al. 2010). There is an increase in grey matter until about the age of 12 in frontal and parietal cortices, followed by synaptic pruning – the elimination of synapses. The specificity of synaptic pruning, caused by fruitful interactions with the external environment, is thought to be important for cognition (Craik and Bialystok 2006).

Another widely used definition of the prefrontal cortex is based on its anatomical connectivity. Prefrontal cortex is defined as the projection zone of the mediodorsal nucleus of the thalamus, based on the seminal work by Rose and Woolsley (1948). They showed that mediodorsal nucleus projects to anterior and ventral parts of the brain in non-primates; and in primates, this projection zone was the granular part of the frontal cortex. This allowed the establishment of homologies despite the lack of a granular frontal cortex in non-primates, allowing for the use of models like rats and mice in the study of prefrontal cortex. The projection zone definition of the prefrontal cortex has survived, even in the light of substantial changes in our understanding the prefrontal-mediodorsal thalamus connectivity. Evidence suggests that mediodorsal thalamus projects to cortical fields outside of the prefrontal cortex in many species including rats, mice and monkeys (Markowitsch and Pritzel 1979; Preuss 1995). Thalamic nuclei other than the mediodorsal thalamus also project to the prefrontal cortex. Thus, this criterion does not define the prefrontal cortex unequivocally (Uylings et al. 2003).

While anatomical definition of the prefrontal cortex has been debated, it is its unique functionality that is of primary interest. Some of the first hints of its function were from a famous case study of Phineas Gage. In an accident in 1848, a large iron rod was driven through his head destroying his left frontal lobe. The standard presentation is that, although

Gage retained normal memory, speech and motor skills, his personality changed drastically. He became quick-tempered, impatient, and irritable, characteristics he had not previously displayed. His friends described him as 'no longer Gage'. He had previously been a capable and efficient worker, but after the accident he was unable to complete simple tasks. However, careful analysis of primary evidence shows that most descriptions of Gage's psychological changes were exaggerated (Macmillan 2000). This case study led to the examination of more frontal lobe patients, with lesion studies giving valuable insights into prefrontal function.

Lesion studies suggest that the prefrontal cortex is necessary for many tasks and real world activities, especially complex tasks integrating different cognitive processes or multiple task steps. Seminal case studies in the first part of the 20th century linked the frontal lobe to cognition (Ackerly and Benton 1948). However, there were also some conflicting pieces of evidence. Cognitive deficits were not always observed in all patients with frontal lobe damage (Rylander 1939). Hebb concluded at the end of a series of studies that frontal lobe patients did not perform inferiorly on tests of intelligence when compared with patients with posterior lesions. In fact, one patient improved after extensive bilateral frontal lobectomy (Hebb 1939, 1945; Hebb and Penfield 1940). These claims have now largely been discredited. Frontal lobe patients do not seem to have difficulty with basic operations such as adding and subtracting. However, when faced with more complex problems requiring multiple steps, these patients tend to respond impulsively to an early stimulus and fail to execute the other component steps required for problem solution (Luria and Tsvetkova 1964; Luria 1973). These and other studies make a strong case for the conceptualization of the PFC as an 'executive' or a 'cognitive control centre' controlling representation and processing in other brain regions (Norman and Shallice 1980; Desimone and Duncan 1995; Miller and Cohen 2001).

Norman and Shallice model and the implication for functional differences in the frontal lobe

Psychologists Donald Norman and Tim Shallice proposed a framework of attentional control of executive functioning known as the 'Supervisory Attentional System model (SAS)' (Norman and Shallice 1980). This was one of the early influential models in thinking about executive functions and how any behaviour is achieved through a supervisory or monitoring

system. In their own words, ‘we suggest that supervisory attentional mechanisms of limited capacity oversee the operation of the system, monitoring for the success of the activity, and biasing the selection and suppression of component schemas by altering the activation values of those schemas. We specify that such attentional control does not act directly, but only indirectly through the mediation of activation value’ (Norman and Shallice 1980). This model and its concepts are explained below.

Inspired by the SAS model, the ROBBIA programme is an influential set of structured studies of frontal lesions. This programme sought to develop the SAS model in a series of experiments conducted over a period of 10 years. I use this as an example here to illustrate the ideas of specific functions being assigned to different frontal regions. Stuss et al. (1995) explained a theoretical framework for the prefrontal cortex based on the SAS model. Four components for cognitive processing were postulated in this model: (1) cognitive units or modules, (2) schemata, (3) contention scheduling, and (4) the supervisory (attentional) system. All basic cognitive operations were carried out in modules or units, and controlled by schemata, which were routine programs for already learnt skills. The supervisory system handled non-routine behaviours and functions and was the ‘executive’. The interplay between these components is shown in Figure 1.6.

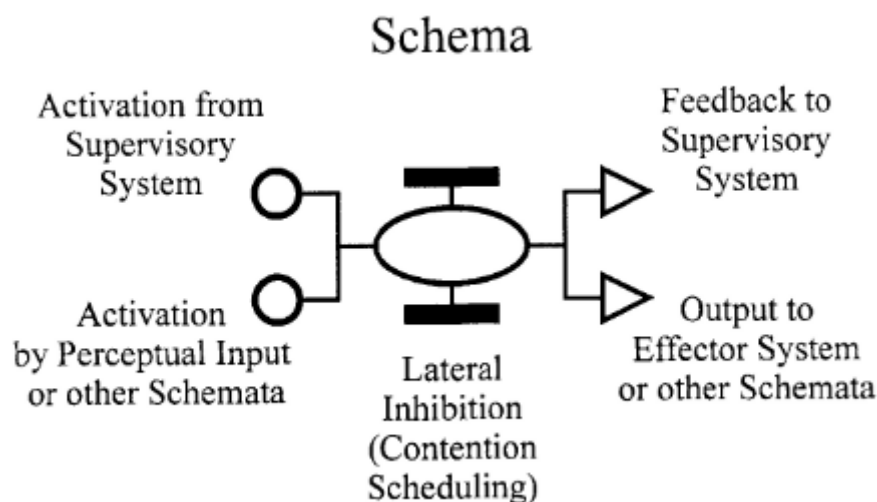


Figure 1.6: The supervisory system. Taken from Stuss et al., 1995. The supervisory system model is shown above with units processing either schemata (routine programs) or non-routine programs under the supervision of an attentional system.

Stuss et al. (1995) proposed 7 different sub-types of this supervisory attentional process – sustaining, concentrating, sharing, suppressing, switching, preparing and setting. They suggest that these attentional processes are present in different regions of the frontal lobe. Integration of information between these regions and from other brain regions can provide a sophisticated control of attention. At an abstract level, this model resembles that of the MD system. While the MD network school of thought proposes these functions be assigned to the network as a whole, the ROBBIA project sought to differentiate the frontal lobe regions based on these different attentional processes.

The ROBBIA project led to several suggestions for functional dissociation in the frontal lobe. Left lateral frontal lobe patients were more likely to show deficits in task setting (Alexander et al. 2005). This was the process of acquiring the correct schema and was tested using task-switching paradigms. Right lateral frontal cortex damage was linked to active monitoring as seen in keeping time in reaction time tasks (Picton et al. 2006). A non-ROBBIA study that looked at category exemplar recall showed poor initial recall, and then excess recall when prompted in these right lateral frontal lobe patients. This was also interpreted to show a monitoring role for this region (Turner et al. 2007). Inferior medial frontal lobe patients also showed impaired performance in switching tasks. This was not limited to the switch but had overall higher error rates, which was linked to motivation and effort (Shallice et al. 2007). While these and other neuropsychological studies have given the impression of functional specialization, they do not negate the potential of interaction between these systems (Stuss and Knight 2013).

Other non-ROBBIA lesion studies have also suggested specific functionality in different frontal lobe regions. Decrease in performance in various tasks has been linked to specific prefrontal lesions: Stroop (Vendrell et al. 1995), Wisconsin card sorting tests where one has to adaptively learn a new stimulus – response mapping when the underlying rule changes in a manner that is invisible to the subject (Milner 1964), attentional shifts (Windmann et al. 2006), sustained attention (Chao and Knight 1995), working memory (Lewinsohn and Libet 1972), and inhibitory control (Aron et al. 2004), to name a few. Many lesion studies have suggested dissociations in the frontal cortex, ROBBIA being a salient example, and they are able to causally link a specific region with a function. However, they cannot directly address the function of the MD network. Lesion studies often suffer in their interpretations due to small sample sizes, lesions not completely overlapping between participants etc. Most notably, lesions often target one of the MD regions rather than the entire network, and they

spill over to neighbouring regions that are not a part of the network. Thus lesion studies can only go so far in addressing the functionality of the MD network.

Alternate theories of frontal lobe function

The meta-analysis of Duncan and Owen (2000; see also Duncan 2006) in particular gave rise to the idea of one common network, encompassing a specific set of frontal and parietal regions, involved in multiple cognitive processes and handling diverse demands (Duncan 2013a). On the other hand, many fMRI studies have also suggested functional dissociations between frontal lobe regions. One proposal is the idea of hierarchical processing and a rostral to caudal gradient in prefrontal regions (Koechlin et al. 2003; Sakai and Passingham 2003; Petrides et al. 2005; Badre and D'Esposito 2007, 2009; Badre 2008). These authors support the cascade model “the LPFC [lateral prefrontal cortex] is organized as a hierarchy of representations originating from premotor cortex and processing distinct signals involved in controlling the selection of appropriate stimulus–response associations” (Koechlin et al. 2003, p. 1181). Accordingly, more posterior regions were hypothesised to process decisions based on sensory input, with more anterior regions processing context based and episodic inputs. A related model was of representational hierarchy, which proposed that the axis represented abstractness of representation rather than the control signal (Badre and D'Esposito 2007). An example of such a study is discussed here in detail (Figure 1.7).

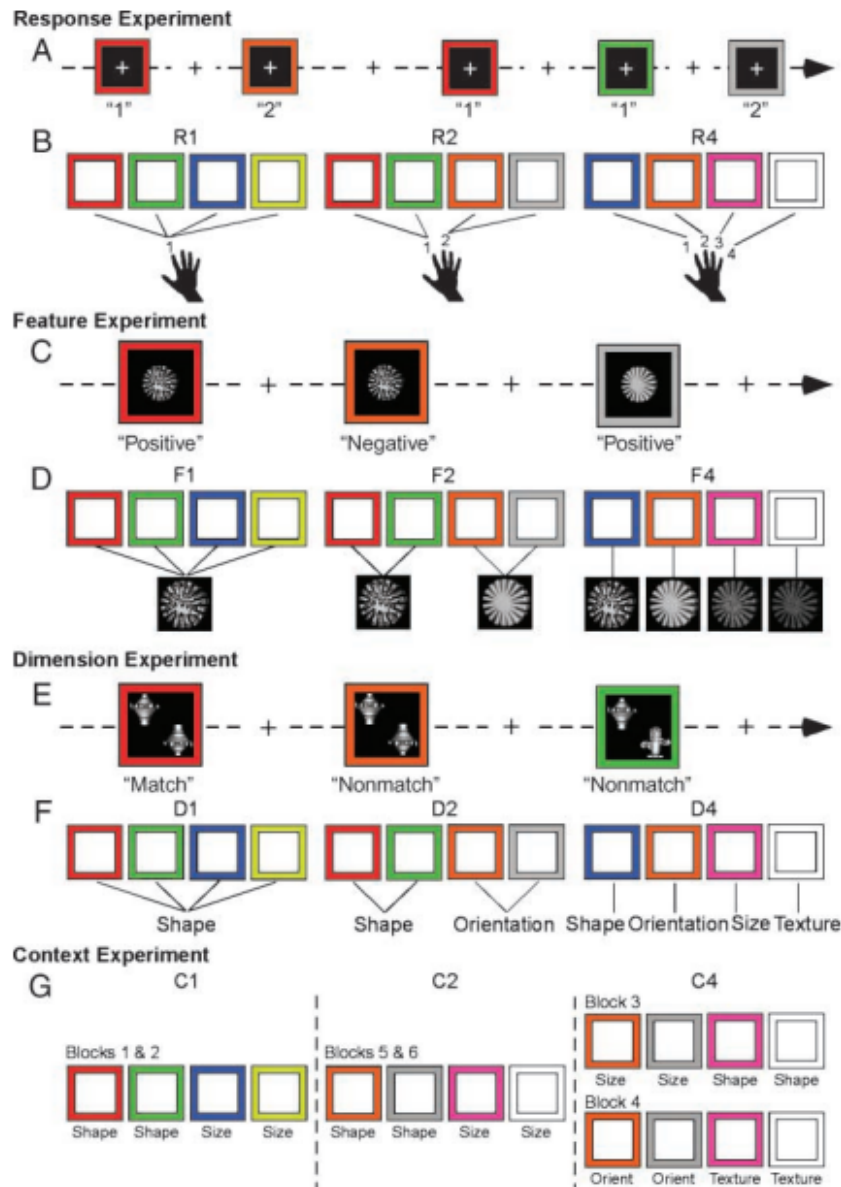


Figure 1.7: Study to investigate representational hierarchy. Reproduced from Badre and D'Esposito (2007). **A-B** depicts the response experiment where the colour of the square determines which finger to use to give the response. In this and the following conditions, the demand is varied parametrically. **C-D**. In the feature experiment, coloured squares were associated with one 'feature' of an enclosed object, e.g., a particular orientation for the orientation dimension or a particular number of spokes for the texture dimension. The subject responded with one button if the square and the target matched, and another if they did not. **(D)**. **E-F** illustrates the dimension experiment, where the colour of the square determines the dimension in which the stimuli need to be matched. **G** shows the context experiment, where the frequency with which each colour is associated with the dimension varies across blocks.

Figure 1.7 shows the study from Badre and D'Esposito (2007). The study included demands with increasing abstractness and each demand was also parametrically modulated. The first was a stimulus-response mapping, where a coloured square determined which finger to use to press a button (response experiment). Second was the feature experiment where the coloured square was associated with a feature of an enclosed object, i.e., one particular orientation (up, down, left or right) for the orientation dimension, or a particular number of spokes on the mottled texture image shown in Figure 1.7C. They had to indicate if the object match the object associated with the coloured square. Third was a dimension experiment where the coloured square indicated the dimension (shape, texture, etc.) to use to determine if the two enclosed stimuli matched along that cued dimension. Last was a context experiment, where depending on the context, a colour could indicate different dimensions. Each experiment also had demand varying parametrically such that all four colours indicated the same rule, two colours indicated one rule and two others indicated a second rule, or all four colours indicated different rules. This was an interesting study that examined both the different types of demand, as well as different magnitudes of each demand.

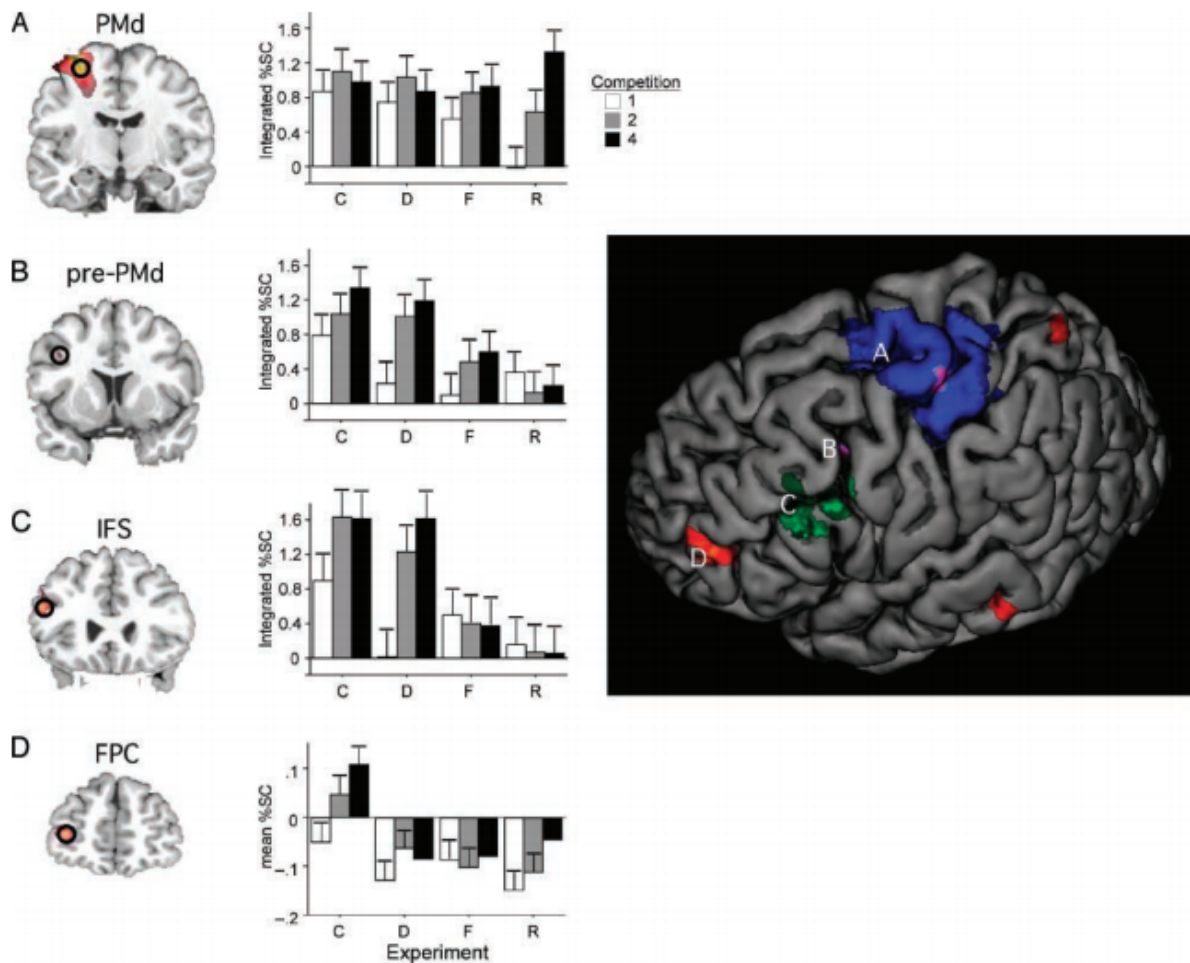


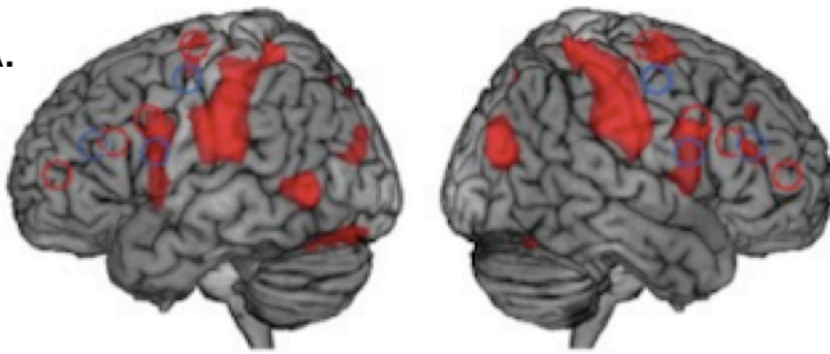
Figure 1.8: Representational hierarchy along the lateral frontal cortex. Reproduced from Badre and D'Esposito (2007). The whole brain results of the parametric variation in four experiments are shown in the right panel. A-D show ROI results for the parametric activation in the four experiments. **A.** The response experiment (R) shows an effect of load ('competition') in dorsal pre-motor area (PMd). **B.** The feature experiment (F) shows a load effect in an anterior dorsal pre-motor area (pre-PMd). **C.** The dimension experiment (D) shows a load effect in the inferior frontal sulcus (IFS). **D.** The context experiment (C) shows a load effect in the frontopolar cortex (FPC).

Figure 1.8 shows the main results for parametric variation in each of the four manipulations. Parametric variation was used to isolate the areas that responded with increased activations on increasing the demand, at each level of abstraction. The response experiment primarily showed a load effect in the dorsal premotor region. The feature experiment showed a load effect in a region anterior to that, the pre-PMd. The dimension experiment showed a load effect in the inferior frontal sulcus (IFS). So far, this provided evidence for the hierarchical

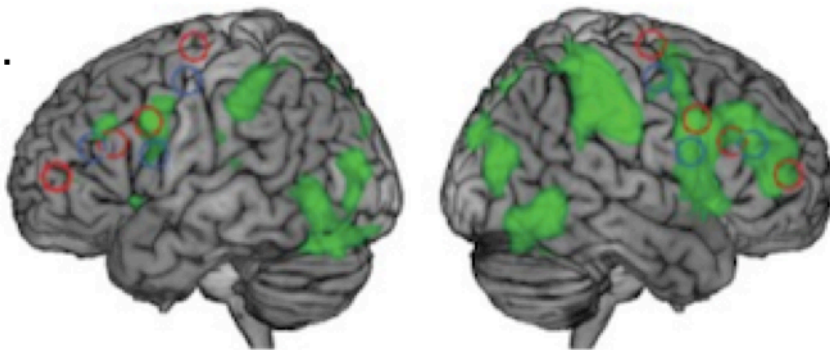
model of increased abstraction activating more anterior parts of the lateral PFC. The context experiment on the other hand showed load effects in the left frontopolar cortex, left posterior IFS/pre-PMd, right DLPFC, bilateral posterior parietal cortex, left inferior temporal cortex, and left fusiform gyrus. In addition, IPS demonstrated sensitivity to increases in competition at all hierarchical levels, with quantitatively more parametric activity for the response and feature experiments. The context experiment as well as the pattern of activation in the IPS would suggest a network of regions, similar to the MD network being activated in order to solve the more demanding conditions in the context experiment.

This rule hierarchy model, with more anterior regions processing more abstract information (Badre and D'Esposito 2007, 2009; Badre 2008), and the information cascade model (Koechlin et al. 2003; Koechlin and Summerfield 2007) with more anterior regions processing context or episodic input (information not currently presented), were tested against the common network of regions hypothesis (Duncan and Owen 2000) in a study by Crittenden and Duncan (2014). The task was to find the odd one out in terms of length among four lines (baseline condition). The different demands were: 1.) Finding the odd one out among 8 lines (8L). 2.) Decrease in the change between the odd one out and the other lines (fine discrimination: FD). 3.) Use of an unnatural stimulus-response mapping (MS). This MS task was earlier shown to activate IPS, FEF, left inferior gyrus and frontal operculum (Jiang and Kanwisher 2003). Here, the FD and MS conditions showed activation in all the multiple-demand regions: IPS, pdLFC, posterior, mid and anterior MFG, AI/FO and preSMA/ACC (Figure 1.9). This was contrary to the rule hierarchy and information cascade models, as they would predict activation in PMd and pre-PMd for FD and MS respectively. The 8L condition did not show activation in the most anterior region of the MFG, largely consistent with the rule hierarchy and information cascade models. Interestingly, single subject data showed that some subjects did show activation in the most anterior regions, even in the 8L condition. These results suggest individual differences, with a dominant pattern of caudal activity, but rostral spread in some participants.

A.



B.



C.

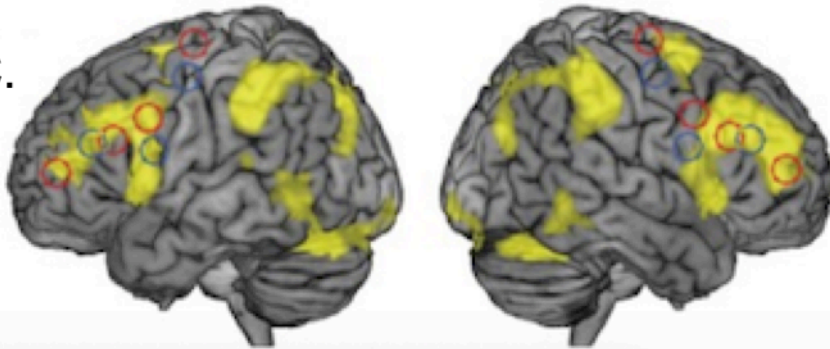


Figure 1.9: Common set of regions across different demands rather than rule hierarchy or information cascade. Whole brain activations for **A.** 8L > 4L contrast, **B.** FD > 4L contrast and **C.** MS > 4L contrast. ROI outlines from Rule Abstraction ROIs (Badre and D'Esposito, 2007) in red outline, Information Cascade ROIs (Koechlin et al., 2003) in blue outline.

The debate still continues between various conflicting theories of prefrontal function, with evidence available for each of them. With studies discussing particular functions of specific MD regions (Postle et al. 1999; Rushworth et al. 2004; Shenhav et al. 2013; Amiez et al.

2016), one school of thought is that of a modular frontal lobe. There are also many studies supporting an MD-like network, albeit with different names and slightly different regions: the cognitive control network (Cole and Schneider 2007), task-activation ensemble (Seeley et al. 2007), task-positive network (Fox et al. 2005), dorsal attention system (Corbetta et al. 1995; Corbetta and Shulman 2002), frontoparietal network (Dosenbach et al. 2007; Yeo et al. 2011), along with studies using graph theoretic approaches (Bullmore and Sporns 2009). There has also been evidence for dissociation of the MD network into frontoparietal and cingulo-opercular networks, with different functions for each (Dosenbach et al. 2006; Nomura et al. 2010; Hampshire et al. 2012; Wallis et al. 2015). The IPS and parts of the MFG are grouped into a frontoparietal network with transient roles in cognitive control, e.g. during task starts or switches; the preSMA/ACC and AI are linked together as the ‘core’ cingulo-opercular network with sustained activity during task blocks (Braver et al. 2003; Dosenbach et al. 2006, 2007, 2008).

MD regions form a network: Evidence from connectivity studies

In resting state studies, participants lie inside the scanner without doing a specific task. The timeseries of voxel activity obtained during this period of rest (or in absence of experimenter assigned task) is correlated to give a measure of functional connectivity between voxels. This method allows us to tap into the underlying connectivity between regions, when an explicit task is not being performed. Thus a brain network isolated by resting state (regions of the network having stronger correlations across time compared to regions outside the network) would imply strong connectivity between the nodes of the network. While these studies lack experimental control, they have consistently isolated networks such as the default mode and frontoparietal networks comprised of regions in the frontal and parietal cortices across many studies (Dosenbach et al. 2007; Power et al. 2011; Yeo et al. 2011; Cole et al. 2014; Assem et al. 2019). The resting state studies provided evidence for MD regions being functionally connected and working as a unit, allowing information exchange and integration (Power et al. 2011; Yeo et al. 2011; Cole et al. 2014; Assem et al. 2019).

Power et al., (2011) used a large cohort ($n = 105$) to study the sub-graphs of existing graph networks, using resting state. They discovered three task-positive networks: the dorsal attention, frontoparietal and cingulo-opercular. They also found a second cingulo-opercular, more posterior and dorsal to the task positive one, resembling the salience network. Yeo and colleagues (2011) used a cohort of 1000 participants to define seven core network

parcellations, which could also be broken into 17 finer networks. One of the seven networks was the frontoparietal network, which comprised of all the MD regions, including AI/FO and preSMA/ACC. Thus, the division of the MD regions into frontoparietal and cingulo-opercular networks (Dosenbach et al. 2006, 2008) may be influenced by cingulo-opercular regions that form their own network, but are also closely adjacent to MD sections of the cingulo-opercular regions.

As functional connectivity is based on intrinsic anatomical connectivity, there is a correlation between task-based networks and resting state networks. However, there are also task-general and task-specific changes in the networks, not explained by resting state networks (Cole et al. 2014). A large meta-study of 10449 experiments showed a set of functionally flexible regions involved in many different tasks (Yeo et al. 2015). These regions were used as seed regions in an independent resting state data set of 1000 subjects to find a functionally flexible network, which resembled the multiple-demand network (Fedorenko et al. 2013) to a high degree. They also proposed heterogeneity within these regions: while the regions support multiple cognitive components, each does so to a *different* degree (Yeo et al. 2015).

Other task based functional connectivity studies have also isolated a frontoparietal network (Cole et al. 2014; Soreq et al. 2019). In a working memory study, a frontoparietal network cluster was active during encoding, maintenance and probe stages. The univariate activity in these frontoparietal regions showed an increase for higher loads during encoding but not during maintenance or probe. The intra-cluster functional connectivity increased with higher load during all three stages (Soreq et al. 2019).

Function of the MD regions: Evidence from multi-voxel patterns

As discussed in the previous sections, large cohort resting studies isolated multiple brain networks, including an MD-like network. Task-based studies showed co-activation but also functional differences between these MD regions. In this section, I will discuss studies investigating MD function by looking at voxel activation patterns in these regions (Stiers et al. 2010; Woolgar, Hampshire, et al. 2011; Woolgar, Thompson, et al. 2011; Nee and Brown 2012; Nelissen et al. 2013; Cole et al. 2016; Wisniewski et al. 2016). Haxby et al. (2001) showed that distributed activity patterns in ventral temporal cortex predicted the category of stimuli being processed. This important study paved the way for further investigations of representation as seen in patterns of voxel activity. Linear classifiers integrate evidence

provided by each voxel separately and thus will show above-chance classification only if some voxels are individually sensitive to the dimension of interest, and so inferences can also be made on sets of voxels (Norman et al. 2006). Inferences can be made about specific voxels, by examining their contribution (weight) on the linear classifier, for each condition. In addition, MVPA can also be used to study functional interdependence of voxels in a region that shows univariate activation for two different conditions. Patterns of voxels can help disambiguate representations (Peelen and Downing 2006). This methodology was extended to representational similarity analysis (RSA), which allows information at a more abstract level to be compared across different modalities (Kriegeskorte et al. 2008). Corresponding activity patterns measured with two different techniques would result in the same information being extracted. This approach attempts to bypass fundamental differences between different techniques: for example, invasive electrophysiology measures the electrical activity of single cells, while fMRI measures the hemodynamic response function of brain activity. Although these signals are related (Bandettini and Ungerleider 2001; Logothetis et al. 2001; Goense and Logothetis 2008), a formal comparison of them can be difficult. In RSA, activity patterns associated with each pair of conditions (Haxby et al. 2001a) are compared to obtain a representational dissimilarity matrix (RDM) characterizing a representation geometry, abstracted away from the spatial layout of the unit used for analysis (voxel, neuron, electrode etc.). This allows comparison of representational geometries across regions, as well as comparison of information representation across species and modality of measurement (Kriegeskorte et al. 2008; Nili et al. 2014). One such comparison that is gaining traction is the comparison of brain representation with that of computational models (Khaligh-Razavi and Kriegeskorte 2014; Holroyd et al. 2018).

Stiers and colleagues (2010) conducted comparisons of activity patterns in each of four tasks: a switching task, an inhibition task, a working memory task and a gambling task. Using the regions that were activated in all four tasks, they isolated MD regions as those with positive signal change modulated by the difficulty of the task condition (i.e., main effect of difficult over easy conditions in switching, working memory, and gambling), and positive signal change without consistent modulations across tasks. Within these MD regions, they found voxels with different task preferences. Are these commonly activated regions involved in processes that are also common to all tasks: response control functions, attention mechanisms, interference handling, error monitoring, memory retrieval, etc.? While that is a

definite possibility, the different task preferences of these voxels could also suggest that these regions participate in the actual task processing.

The functional preferences of voxels allows for the study of underlying representations in MD regions, and this information representation speaks to the function of these regions. Woolgar and colleagues (2011) systematically investigated the nature of representation in each of the MD regions in a simple stimulus-response task (Figure 1.10). The background colour indicated which of two rules was to be applied, and that was to be integrated with the position of the square (the stimulus) in order to get the appropriate response (which of the four buttons to press). In lateral frontal cortex, IPS and AI/FO, rule was the most strongly represented followed by the position of the square (not present in AI). As two colours were used to cue each rule, the actual colour of the cue could also be decoded, and was represented in AI and IPS. Response was least represented, not reaching significance in any region. Strong rule representation fits with univariate studies of the MD region (Duncan and Owen 2000; MacDonald et al. 2000; Brass and Yves von Cramon 2004; Hampshire et al. 2011). While the study did not explicitly test for hierarchy in representation, there was little evidence for only the most anterior parts representing the rule. Different multi-voxel patterns could reflect differing emphasis on multiple executive processes. Various demands are being processed in these regions, and the information is exchanged and integrated within this highly connected ‘central executive’.

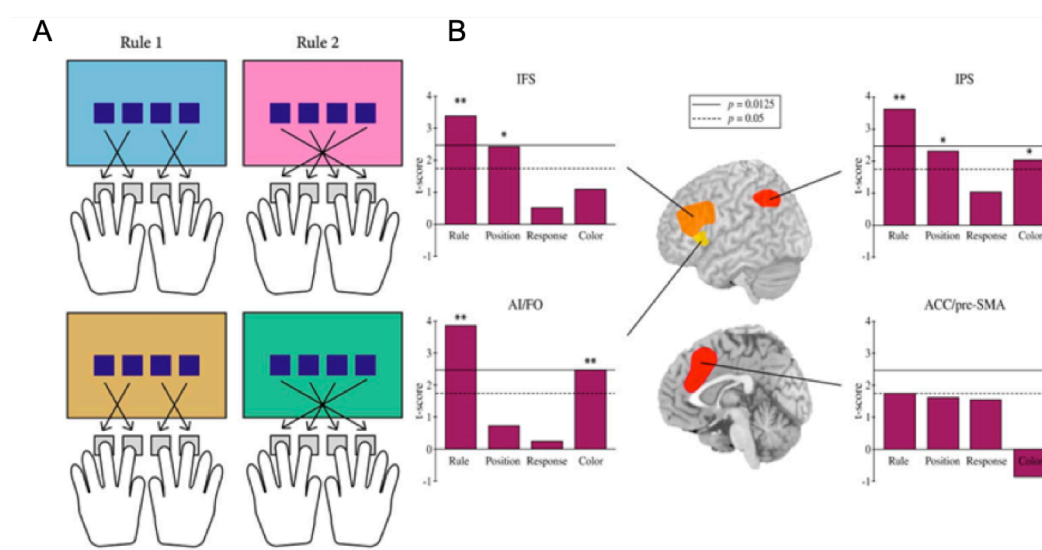


Figure 1.10: Representation of stimuli, rules and responses in the MD network.

Reproduced from Woolgar et al. (2011). **A.** Participants had to press a button according to the

rule cued by the background colour. Two rules were used, and two colours cued each rule. **B.** Classification accuracies were tested against zero in a one-sample t-test and a t-score was calculated with positive values indicating above chance classification. Dotted line corresponds to $p = 0.05$ and solid line to Bonferroni-corrected p-value of 0.0125.

The theory for the function of the MD network deals with dividing problems into subparts, constructing ‘attentional episodes’ (Duncan 2013a). Another aspect of cognition is to maintain these multiple sub-goals with their correct structural relations (Miller et al. 1960). Attentional episodes in a task can be defined at different levels of abstraction, e.g., selecting a response rule and then applying this rule to a specific current stimulus (Badre and D’Esposito 2009). Adaptive coding or flexible coding of information in the prefrontal cortex has often been proposed as a mechanism for constructing these episodes, and fast learning of complex new tasks or applications to novel tasks (Duncan 2001; Cole et al. 2010). Cole et al. (2010) examined the possibility that this adaptability relies upon a compositional scheme of rule representation within the lateral prefrontal cortex, in which new task representations can be constructed from different combinations of familiar rule representations allowing for rapid representation of a wide variety of novel tasks. Each trial was cued with three distinct rules that had to be integrated in order to get the final answer. For example, in ‘SWEET–SAME–LEFT INDEX’ participants decided if each of 2 stimuli presented – for example, grape and apple – were sweet, and pressed their left index finger if they both gave the same answer (left middle finger otherwise). The semantic rule specified the category to be used in order to judge the stimuli – ‘are they sweet?’ The decision rule specified if the semantic rule had to be applied to both or just one of the stimuli. The response rule indicated the decision-response mapping, which finger to use if the decision rule was satisfied. Thus, the three rules had to be appropriately combined to complete the tasks. Decision rules were the focus of analysis as these were the most integrated of rule types – they received information from semantic rules and sent information to response rules. Participants learnt these rules in specific combinations outside the scanner, and were then tested on these practised combinations as well as novel combinations. A searchlight approach showed decision rule representation in parts of left anterior and posterior PFC, anterior insula, right inferior parietal, lateral and medial PFC. Semantic rules showed up in similar regions as well. Importantly, there was successful decoding of decision rules using classifiers trained on the practised blocks and tested on novel blocks. In spite of the practised tasks being non-overlapping (each rule was learnt in a

separate block) and not knowing that novel combinations would be presented, the participants seemed to have used a compositional approach that was also used for the novel task. This compositionality suggests that complex tasks were learnt by breaking the problem into subparts and constructing the component episodes.

While the above study showed similar compositional representations in practised and novel tasks, the question still remained of whether these representations contributed to behaviour. In three of these novelty-sensitive regions located within distinct areas of left anterior LPFC, activity patterns were differentiable for correct and error trials, and decision rule was only decodable in correct trials (Cole et al. 2016). This suggests a contribution of these task representations to successful task performance.

Many studies have demonstrated representation of abstract task-related information in MD regions (Haynes et al. 2007; Bode and Haynes 2009; Greenberg et al. 2010; Nee and Brown 2012; Crittenden et al. 2016). In a free decision making task, task parameters were decodable in both lateral and medial PFC (Haynes et al. 2007). In a task-switching paradigm, the processing of task-set/cue information was investigated across time using a finite impulse response (FIR) model. The cue identity was first decodable in the visual cortex, then in the IPS and last in the ventral LPFC (Bode and Haynes 2009). Waskom and colleagues (2014) used a slow event-related design where the participants had to judge the similarity between two sequentially presented stimuli based on the rule cued at the beginning of the trial. They showed significant rule decoding in regions surrounding the inferior frontal sulcus, and in the IPS at about 3s after stimulus onset, whereas the occipito-temporal cortex showed information at about 5s after stimulus onset. Rule information was also seen in anterior and posterior MFG, frontopolar cortex, inferior frontal gyrus, AI/FO, and posterior superior frontal sulcus, thus in most of the MD regions.

To further understand the nature of processing and representation in a task set, task preparation has been compared with short-term memory storage. Following on from univariate studies that largely find the same areas of LPFC and PPC (posterior parietal cortex) in both (see Ikkai and Curtis 2011 for a review), Muhle-Karbe et al. (2017) used MVPA to compare the same cue of a stimulus-response task when the cue was used for implementation versus when the cue was to be memorised but not implemented. During the instruction phase, the inferior frontal junction area (IFJ), IFS and IPS showed similar levels of cue decoding for both implementation and memorisation blocks, but in the following delay period, this representation was only maintained for the implementation blocks. Thus these

MD regions are likely to be involved transforming semantic task knowledge into a temporary task set for achieving goal-directed behaviour, rather than simply holding in short-term storage.

Selecting task-relevant information from irrelevant information is an important part of goal-directed behaviour. Thus, one would expect task representations in MD regions to be stronger for the task-relevant information. Indeed, MD regions show strong representation of task related information and almost none for irrelevant and distractor information (Erez and Duncan 2015; Woolgar et al. 2015). In a cued categorisation task, where the participants indicated if an object from the cued category was present or absent in the subsequent display, all MD regions discriminated between targets and nontargets (Erez and Duncan 2015). There was also discrimination between two types of nontargets – ones that could be targets on other trials and ones that were never targets. There was however no decoding between the visual categories of two different targets. Thus in the stimulus phase, once the behavioural status of the stimuli has been established – whether it was a target or not – the visual category of the stimuli itself became irrelevant and was not represented strongly enough to be detected by voxel patterns. In MD regions, Woolgar et al. (2015) showed preferential encoding of target identity over that of the distractor, only in the high perceptual difficulty condition (degraded stimuli). The increased representation in MD regions could reflect increased attention to compensate for the degraded stimuli, thus supporting the processing of task-relevant information in downstream areas such as visual cortex (Desimone and Duncan 1995; Miller and Cohen 2001). Alternatively, higher perceptual coding in the MD regions may reflect a more careful and considered decision-making process, or a perceptual process supported by reverberant processing between frontoparietal and visual systems necessary for low quality inputs (Dehaene et al. 2003; Gilbert and Sigman 2007). Preferential encoding of target object versus distractor object also extends to target features versus distractor features within the same object (Jackson et al. 2017).

Flexible processing in MD regions would mean the task representations should be able to change with changing task parameters. As one example, task set representation increases when an additional reward is associated with successful completion of the trial (Etzet et al. 2016). Etzel and colleagues (2016) used two sessions both with the same task-switching task, the first having only non-rewarded trials, and the second with equal number of rewarded and non-rewarded trials. A linear classifier discriminating the two cues was trained on the non-rewarded trials of the first session, and tested on both non-rewarded and rewarded trials in the

second session. The regions that showed increased classification accuracy on rewarded versus non-rewarded trials included regions of the MD network: lateral frontal cortex, medial PFC, and posterior parietal cortex. This increase in representation correlated with an individual subject's behavioural accuracy. Thus MD task representations, that support successful behaviour, are flexible and change to accommodate task parameters.

Neuronal functions and mechanisms in the MD regions

While fMRI studies provide us with a non-invasive methodology to study human brain activations at a whole brain level, the underlying mechanisms of neurons and the dynamics of their activity are lost. Voxelwise representations (discussed in the previous section) have added to our knowledge of the function of MD regions, but each fMRI voxel contains a million or more neurons making it hard to infer mechanism at a neuronal level. Classification accuracies presented in these studies tend to be quite low (Cole et al. 2011; Woolgar, Thompson, et al. 2011; Erez and Duncan 2015; Etzel et al. 2016). This has been linked to functional organization at the neuronal population level (Bhandari et al., 2018; Dubois et al., 2015). Neuronal populations in the same brain region may encode multiple different variables, and decoding some of them may be easier than others (Dubois et al., 2015). Local functional organization and distribution of neural populations in PFC may reduce differences between conditions at the voxel scale measured with fMRI (Guest and Love 2017; Leavitt et al. 2017). In macaques, Tremblay et al. (2015) showed that low-frequency LFPs, which reflect local average activity, yielded lower decoding accuracies compared to single neuron firing rates. Logothetis et al. (2001) showed that LFPs were correlated with the haemodynamic response, thus reduced differences between conditions in LFPs can be extrapolated to low decoding in fMRI. Dubois et al. (2015) examined the coding of face viewpoint and identity information using MVPA of both blood oxygenation level dependent (BOLD) response and single-unit recordings in macaques. MVPA of BOLD patterns only revealed viewpoint information, although both viewpoint and identity were encoded by single neurons. Identity decoding could have suffered because identity-coding neurons were only weakly spatially clustered as compared to viewpoint-coding neurons. Clustering may also enable nearby blood vessels to be strongly driven by neurons selective to one condition, thus creating inhomogeneity in the sampling of the activity of selective neurons within voxels (Kamitani and Tong 2005). Thus MVPA may not give us the whole picture and we need to supplement our understanding with invasive electrophysiology studies. Many of the

mechanistic insights we have about the prefrontal and parietal cortices are through such studies. In this section I will discuss a few of them, primarily using macaques.

The adaptive coding model of prefrontal cortex posits that neurons have the capacity to code many different types of information. In any given task context, neurons adapt to code only information relevant for immediate behaviour (Duncan 2001). This mechanism could facilitate flexible cognition. One possibility of how this adaptive coding could be realised is through mixed-selective neurons. Mixed-selective neurons are high dimensional and encode a particular combination of task variables (Rigotti et al. 2013). Results from many of the prefrontal neuronal studies can be understood and interpreted through these frameworks.

The ability of adaptation was well demonstrated in a study by Rao et al. (1997). In their task (Figure 1.11A), monkeys were first shown a sample stimulus whose identity they had to remember during the subsequent ‘what’ delay. This was followed by the presentation of two stimuli at two different locations, one of them being the same sample stimulus from before. The monkeys had to remember the location of the sample stimulus in the succeeding ‘where’ delay. Lastly they were shown a screen with four locations and had to make a saccade to the place where the sample had appeared. While there were exclusively object identity selective neurons and location selective neurons, around half of the recorded neurons coded for object identity during the ‘what’ delay and location during the ‘where’ delay. This shows the adaptable coding of prefrontal neurons with information encoded changing with task context (Figure 1.11B).

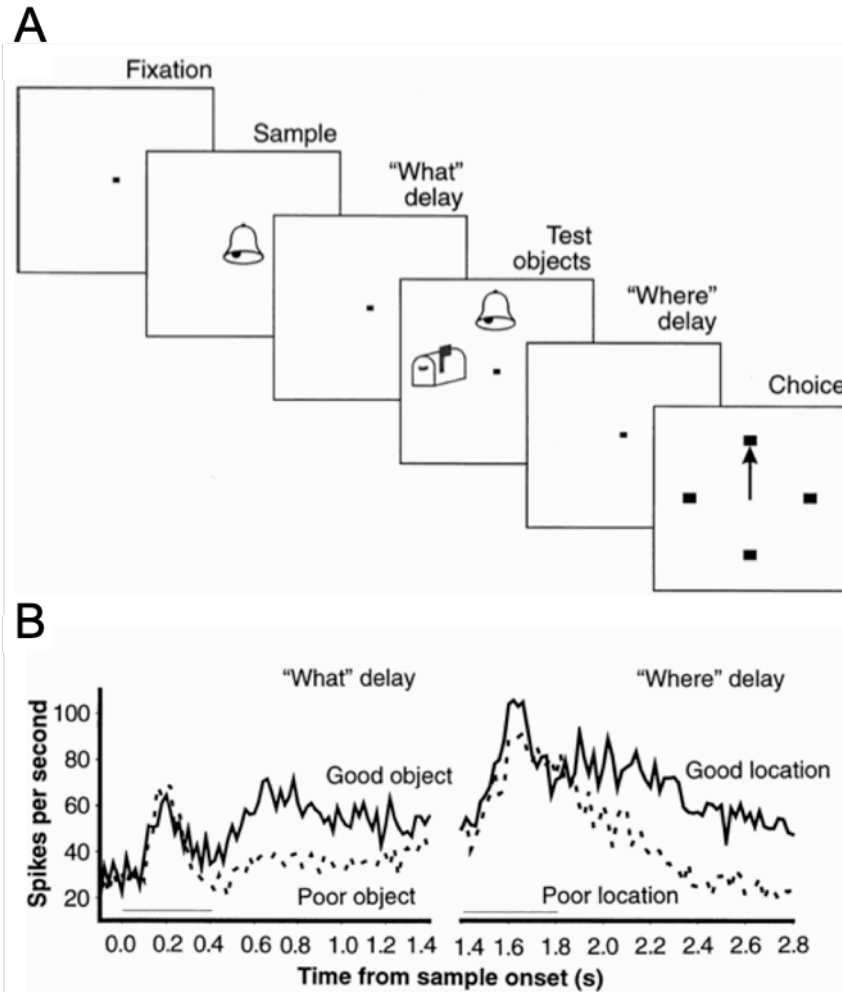


Figure 1.11: The same neurons adaptively coding for object identity and location. Taken from Rao et al., 1997. **A.** An example trial. Trials began with a sample stimulus presented at the centre. The identity of this stimulus had to be maintained across the first – ‘what’ – delay. This was followed by the presentation of two stimuli, one of which was the sample stimulus. Location of the sample stimulus was to be remembered across the second – ‘where’ – delay. Monkeys were then probed for the location of the sample stimulus, to which they made a saccade (arrow). **B.** Activity of sample prefrontal neuron that codes for object identity during the ‘what’ delay and location during the ‘where’ delay. The two grey bars represent the presentation of sample (left) and test objects (right), respectively.

Another theory put forth was that of dynamic organization of prefrontal cells into opponent circuits providing a general mechanism for flexible decision-making in different task contexts (Machens et al. 2005). In a vibrotactile comparison task (Machens et al., 2005), monkeys were presented with two tactile vibration frequencies sequentially and they had to decide if

the first frequency (f_1) was greater than the second (f_2). Two groups of cells were found, one group with activity monotonically increasing with vibration frequency, the other with activity monotonically decreasing. These two opposing populations of neurons interacted flexibly to provide the correct response. As discussed before, one of the cornerstones of prefrontal function is the ability to select task-relevant information. This property could also be explained by opponent coding (Kusunoki et al. 2010). Animals were trained to associate three cues with three stimuli. For a given cue, the following stimulus could be the associated stimulus (target, requiring a saccadic response), a stimulus paired with a different cue (nontarget on this trial, but a target on other trials), or a nontarget that was paired with no cue and thus never served as a target. The largest set of neurons showed target preference, with the strongest responses to the current target, intermediate activity for a nontarget that was a target on other trials, and lowest activity for nontargets that were never targets. Second most frequent was a reverse, anti-target pattern. This pattern was evident in both ventral and dorsal regions of the lateral prefrontal cortex.

Do population responses of neurons also exhibit features of adaptive coding? To answer this question, Stokes et al. (2013) used a time-resolved population-level correlation analyses exploring how context is encoded and maintained in primate prefrontal cortex and used in flexible decision-making. Monkeys performed a conventional cue-target association task, with distinct cue, delay, and choice phases. The neural tuning profiles in prefrontal cortex changed depending on the context/task phase. As a population, once a choice stimulus was presented, stimulus-specific coding evolved into a different decision-dependent state based on the cue. These results support the prefrontal cortex adaptation theory to accommodate changes in behavioural context (Duncan 2001). Such activity, not cross-generalizing across time and context, also supports the mixed-selectivity theory (Rigotti et al. 2013). Neurons changing their pattern of stimulus selectivity during different task operations were shown by many studies (Sigala et al. 2008; Warden and Miller 2010; Stokes et al. 2013; Naya et al. 2017).

Mante et al. (2013) also investigated context dependence in the prefrontal cortex as a population. Monkeys were instructed by a contextual cue to discriminate the motion or the colour of a random-dot stimulus, and indicate their choice with a saccade to one of two targets. Depending on the context, monkeys were rewarded for choosing the target matching the current direction of motion (motion context) or the current colour (colour context) of the random dots. Activity of example units showed dependency on many task variables, motion

coherence in motion trials, colour coherence in colour trials, and context (motion or colour) in all trials, providing evidence for mixed selectivity (Rigotti et al., 2013) and adaptive coding (Duncan, 2001). Notably, the population activity showed comparable deflections along the motion and colour axes, whether they were relevant or not. A trained recurrent network of nonlinear neurons solved a task analogous to the one solved by the monkeys. The population of these network units also showed dynamic patterns reflecting both task variables. Similarly, in a delayed match-to-sample categorization task (Meyers et al. 2008), the population activity of PFC neurons showed some sample identity information (irrelevant to the task), but higher sample category information (relevant to the task). Once again the PFC code was constantly changing with time and did not generalise between the sample stimulus and the decision phases.

With more evidence for adaptive and flexible coding, the classical idea of persistent coding of stimuli by prefrontal neurons underlying working memory (Fuster 1973; Funahashi et al. 1989; Romo et al. 1999; Constantinidis et al. 2018) is being replaced with the idea that the content of working memory is decodable from a population or ensemble activity (Machens et al. 2010; Barak et al. 2013; Mendoza-Halliday and Martinez-Trujillo 2017; Murray et al. 2017). The same has been modelled using linear attractor models and non-linear recurrent neural networks (Machens et al. 2005; Barak et al. 2013). Barak et al. (2013) systematically studied computational models underlying working memory. A common theoretical framework is that of an attractor neural network that exhibits many intrinsically stable activity states, sustained by mutual excitation between neurons coding for a particular stimulus or its behaviourally relevant attribute. Thus the assume time-invariant tuning of neurons and a fixed connectivity of the network (Hopfield 1982; Machens et al. 2005). Another class of models assumes randomly connected neurons that are not tuned to any particular task, and task performance is purely based on adjustment of network readout (Sussillo and Abbott 2009). Intermediate between these models are networks that start out random, but are trained by allowing changes in the connectivity weights between neurons to optimise task performance. Comparing these three models in a delayed vibrotactile discrimination task indicated that the random network did a good job of both performing the task and matching certain aspects of real data, much better than the linear attractor. The intermediate model provided a better description of the data, although none of the models matched all features of the data. The authors suggest that prefrontal networks may begin in a

random state relative to the task and initially rely on modified readout for task performance, but with further training more tuned neurons with less time-varying responses would emerge.

Like PFC neurons, parietal neurons also encode multiple task variables and participate in selective attention and working memory (Chafee and Goldman-Rakic 1998; Crowe et al. 2004, 2005; Buschman and Miller 2007; Chafee and Crowe 2012). As one example, Crowe et al. (2005) recorded neural activity in area 7a (posterior parietal cortex) of the primate brain during a visual maze task in which monkeys mentally traced a path through the maze without moving their eyes. The direction of the followed path could be decoded from the population activity of the parietal neurons. When the monkeys covertly processed a turn in the path, the population representation of path direction shifted in the direction of the turn. While there has been a particular connection between parietal cortex and spatial tasks, it has also been associated with cognitive signals such as rules (Stoet and Snyder 2004) or abstract categories (Freedman and Assad 2006; Swaminathan and Freedman 2012; Sarma et al. 2015). In a memory-based visual-discrimination task, parietal neurons jointly encode sensory, cognitive and decision-related information (Ibos and Freedman 2017).

The nature of prefrontal and parietal cortical circuits could shed light on their functions. To address this question, Katsuki and colleagues (2014) examined the strength of intrinsic functional connectivity between neurons sampled in PFC and PPC by using cross-correlation analyses of simultaneous recordings from monkeys trained to perform working memory tasks. In both areas, effective connectivity declined with distance between neurons. Pairs of PPC neurons shared a larger percentage of their functional inputs (a larger peak at zero lag in the cross-correlation) when they were located at short (≤ 0.3 mm) distances, compared to pairs of PFC neurons recorded at equivalent distances. This effect was not accounted for by differences in firing rate and was present in the fixation period alone, prior to the start of the trial. No obvious hierarchical pattern of information processing is known to be present between the PPC and PFC, whose connections are more reciprocal (Barbas and Pandya 1989; Cavada and Goldman-Rakic 1989; Felleman and Van Essen 1991). While functional specializations of these areas are not completely understood, anatomical and effective connectivity within and between these regions can help shed light on the possibilities of functional differences in future studies. Timescales of intrinsic fluctuations in spiking activity across areas showed a hierarchical ordering, with sensory areas exhibiting short timescales, parietal areas mid, and prefrontal areas exhibiting longest timescales (Murray et al. 2014). This could reflect the duration over which input is integrated, reflecting increased signal-to-

noise ratio in short-term memory storage or decision-making computations at the higher rungs of the hierarchy.

While there has been an emphasis on spikes, numerous studies have tackled related questions using local field potentials (LFPs) as recorded from single electrodes as well as through non-invasive methods (EEG - electroencephalogram and MEG – magneto encephalogram), analysed either in time or frequency domains. Compared to spiking of single neurons, the LFP is a more integrated signal, combining electrical activity over a region of tissue up to at least several mm (Buzsáki et al. 2012), and relates closely to signals from human brain imaging, including functional magnetic resonance imaging (fMRI) (Logothetis et al. 2001; Goense and Logothetis 2008) and scalp-recorded EEG. LFPs represent activity in a local network of neurons, reflecting synaptic excitatory and inhibitory activity. They are influenced by spiking, hyperpolarisation, changed intrinsic membrane properties of neurons (Marder et al. 1996; Buzsáki et al. 2012), etc. These signals have also been studied in the frequency domain and oscillatory activity in the beta and gamma frequencies has often been linked to working memory load (Howard et al. 2003; Honkanen et al. 2015).

LFP signals in the frequency domain have been associated with many aspects of cognition, including working memory and attention (Buschman and Miller 2007; Gregoriou et al. 2009; Liebe et al. 2012; Salazar et al. 2012; Tremblay et al. 2015; Helfrich et al. 2018; see Helfrich and Knight 2016 for a review). Gregoriou et al. (2009) showed that attention to a stimulus in their joint receptive field leads to enhanced oscillatory coupling between the frontal eye field (FEF) and area V4, particularly at gamma frequencies. This coupling seemed to be initiated by FEF and was time-shifted by about 8 to 13 ms across a range of frequencies. They suggest that the expected conduction and synaptic delays between the areas, and time-shifted coupling at gamma frequencies may optimize the postsynaptic impact of spikes from FEF to V4, improving cross-area communication with attention. Salazar et al. (2012) showed task-dependent and content-specific synchronization of activity across the frontoparietal network during visual working memory, in the range of 10-30 Hz. In a delayed-match-to-sample task, parietal and prefrontal neurons were simultaneously recorded. The time-frequency coherence spectrum on correct trials for frontoparietal LFP pairs was computed. Patterns of synchronization were found only in stimulus-selective parietal neurons. Buschman et al. (2012) found rule-specific increases in synchrony at beta (19-40 Hz) frequencies between electrodes in the PFC. Individual PFC neurons synchronized to the ensemble LFP according to the current rule (colour versus orientation). The ensemble encoding the behaviourally

dominant orientation rule showed increased alpha (6-16 Hz) synchrony when the weaker colour rule was applied. They suggest that beta-frequency synchrony is the mechanism by which the relevant rule ensemble is selected, while alpha-frequency synchrony deselects a stronger but currently irrelevant ensemble.

Lundqvist et al. (2016), in their gamma and beta bursts frequency model, posited that gamma bursts of variable time and frequency in single trials were the signature of encoding and re-activation of memory. They analysed LFPs and spiking from the prefrontal cortex (PFC) of monkeys during a working memory task. They observed brief bursts of gamma oscillations (45–100 Hz) varying in time and frequency, during encoding and re-activation of sensory information. Beta oscillations (20–35 Hz) were more representative of the default state, interrupted by encoding and decoding. These beta bursts were also brief and variable. Only those neurons whose activity reflected encoding/decoding also correlated with changes in gamma burst rate. Thus, they interpret these results as gamma bursts gating access to working memory. This also supports the hypothesis that working memory is characterised by discrete oscillatory dynamics and spiking, rather than persistent activity.

While LFPs in the frequency domain have shed light on possible mechanisms of cognition, this signal in the time domain has been investigated in a few studies yielding interesting insights. Kim et al. (2018) studied PFC in mice performing a task in which different stimuli predicted rewards at different delays. Spikes and cell membrane potentials from pyramidal neurons across layers in PFC were measured. In some cases, changes in predicted delay were reflected in sustained changes of membrane potential without associated changes in firing rate. Persistent change in membrane potential was robust to intracellular perturbations, but could be terminated by an external stimulus (reward). These findings suggest that reward prediction is in part maintained via synaptic mechanisms. In macaques, Cosman et al. (2018) simultaneously recorded neurons in prefrontal cortex and ERPs (event related potentials) from the scalp over extrastriate visual cortex to track the processing of salient distractors during a visual search task. When the salient distractor was successfully ignored, they observed robust suppression of distractor representations, involving the same neurons as in target selection. Critically, the distractor suppression in ERP over the extrastriate visual cortex emerged on average 47 ms after the suppression observed in FEF neurons, showing that FEF is responsible for modulating processing in extrastriate visual areas. The trace of this distractor suppression is found even in extracellular potentials and can even be measured on the scalp.

Précis

This thesis investigates the frontoparietal ‘multiple-demand’ (MD) network that is involved in the processing of diverse demands. Chapter 2 studies the interplay between functional differences as well as common activation/co-recruitment of multiple nodes within the MD network, using a maze task performed by healthy adults in an fMRI scanner. The task included conditions with added demands: complexity (two simple tasks – simple maze and simple arrow – done together), time pressure (planning a simple maze before the display fades away), and reward (time pressure trials with an additional reward on correct completion). Quantitative differences between MD regions were most prominent in the simple tasks. There was widespread co-recruitment in the conditions with additional demands.

Using large ROIs in univariate analysis can give us an idea of the general activity patterns of these regions. However, this approach loses information with respect to task related representation, as measured by voxel-wise patterns that rely on heterogeneity of these regions. Are these regions heterogeneous in terms of task preferences, and are they variable across subjects? To answer these questions, in Chapter 3 we localized the MD network in individual subjects using three functional localiser tasks: spatial working memory, verbal working memory, and Stroop. We tested if voxels selected by any of these localisers could give us a better estimate of task representations in a criterion task. The extent of activation patterns, their specificity, and their consistency across runs revealed a similar picture of variable and distributed activity for the three tasks used to identify these MD voxels. 200 voxels with the highest activations in each of the MD regions were identified using each of the three tasks. These voxels captured the underlying neural representations equally well in a criterion task (a rule-based judgement task), showing the multiple-demand nature of the subject-specific voxels captured using any of the functional localisers.

The idea of increased MD network activation in more demanding tasks, and its function of selecting and representing task-relevant information have largely been studied separately; with the former through univariate contrasts and the latter through MVPA. Chapter 4 investigates task representations, as measured by linear classification of voxelwise patterns, and whether they are modified due to external motivation (here monetary reward) in a cued categorization task. There were higher univariate activations in the MD regions in the rewarded trials. We observed an increase with reward in overall activity across the

frontoparietal control network when the cue was presented, reflecting cognitive effort when the context is set for a task. Multivariate pattern analysis (MVPA) further showed that behavioral status information for the objects presented was conveyed across the network. However, in contrast to our prediction, reward did not increase the discrimination between behavioral status conditions in the stimulus epoch of a trial when object information was processed depending on a current context.

fMRI analysis suffers from low temporal resolution. The dynamics of activity are lost, and the neuronal mechanism can only be inferred because the BOLD signal is an indirect measure of neuronal activity. Thus to study mechanism at the neuronal level, invasive electrophysiological studies become necessary. Chapter 5 examines neural activity in frontal and parietal regions more directly through single unit activity and LFPs, recorded in awake monkeys during a spatial working memory task. The animal had to search for a pre-determined target location in a spatial array through a series of trial and error attempts. Once the target was found, the animal had to select it three more times for additional rewards. Single neurons in both PFC and PPC showed dynamic coding of target information, rather than persistent coding of the discovered target in working memory; in line with adaptive coding and the concept of mixed-selectivity. LFPs on the other hand, held the information constant through time and trial phases. Here we get insight into possible mechanisms for integration in the MD network. The firing rates of the neurons are dependent on the interaction of target location and trial phase, allowing the use of location information to complete different cognitive processes in different trial phases for successful performance. In contrast to dynamic neural firing, the information in the extracellular matrix provides a background of stable information as seen in the time domain signal of the LFP. However, these LFP results were largely impacted by the choice of the reference voltage.

Together, the results show an integrated system, responding as a network to a wide variety of task demands. There were functional differences between the different regions of the network, possibly due to preferential access to different task features among the different MD regions. This was especially more prominent in simple tasks. Due to a high degree of connectivity between these regions, co-recruitment may reflect information exchange reducing observable differences between the regions, even more so in demanding tasks. These multiple-demand voxels can be isolated in single subjects through any task demand and their activity generalises to represent task-relevant information in a completely different task. The increase in their activation can be linked to task processing, as seen by flexible

representation of task information in the context of an increased reward. The neurons in these regions integrate the task information non-linearly and provide hints of how flexible cognition may be achieved. Neuronal activity is dependent on a conjunction of task features, allowing for the use of task information in appropriate behavioural cognitive processes and responses. This conjunction coding suggests a mechanism of information integration and processing in the MD network. LFPs represent the linear summation of the task variables, thus representing stable task information through time and trial phases. However, this stability is lost when using an average channel reference. Whether the stable task signal exists in LFP (representing an average of activity in the local neuronal circuit), and how it relates to the activity seen in univariate and multivariate analysis of fMRI signal is yet to be understood. Future studies will determine if and in what way the stable LFP interacts with the time-varying signal of neurons to achieve flexible goal-directed behaviour.

Chapter 2

Progressive recruitment of the frontoparietal multiple demand system with increased task complexity, time pressure and reward

Introduction

To explain the MD network's involvement in tasks of many different kinds, its role in constructing mental control programs has been proposed (Duncan, 2010, 2013). In general, each step of a complex task requires assembly of multiple cognitive fragments, appropriately related and bound to their roles. Through widespread connections throughout the brain, MD regions may play a central role in constructing such “attentional episodes” (Duncan, 2013).

Though MD activity increases with many kinds of task difficulty (Camilleri et al., 2018; Duncan, 2006; Fedorenko et al., 2013; Nyberg et al., 2003), it is unclear how its component regions function individually and together to address cognitive challenges. Some prior data suggest a pattern of partial differentiation between MD regions at low cognitive load, which progressively disappears as load increases (Duncan, 2001). In verbal tasks, for example, MD activity may be stronger in the left hemisphere at low load, but increasingly bilateral as load increases (Wager and Smith 2003; Rottschy et al. 2012). On the lateral frontal surface, various findings suggest progressive anterior spread of activity as task rules become more hierarchical or complex (Badre and D'Esposito 2007, 2009; Badre 2008; Crittenden and Duncan 2014; Badre and Nee 2018). More broadly, there have been many suggestions of relative functional specializations for different MD regions, though with little consensus emerging across studies (Dosenbach et al. 2007; Nomura et al. 2010; Hampshire et al. 2011; Yeo et al. 2015). A combination of relative specialization but also substantial co-recruitment is expected if MD regions link together multiple cognitive contents into attentional episodes. An attentional episode will be a rich cognitive structure, combining linked stimulus inputs, goals, information from semantic memory, potential actions and rewards and so forth. Given their different connectivity, MD regions will differ in their immediate access to these different domains of information, opening the door to relative specialization in tasks with different content (Duncan, 2001). Such relative specialization may be especially visible at

low load. Linking information into the correct structures, however, will require substantial information exchange and integration, suggesting widespread co-recruitment. Co-recruitment of the entire network may become most visible as load and overall activity increase. Relative specialization accompanied by substantial information exchange is also suggested by studies using multivoxel pattern analysis, with relevant task features widely decodable across MD regions, but with some quantitative differentiations (Cole et al. 2016; Crittenden et al. 2016; Woolgar et al. 2016)

In the present study, we used a maze task to study patterns of specialization and spread across a set of *a priori* MD regions defined by common response to diverse task demands (Fedorenko et al. 2013). In the simplest task versions, participants either planned a 4-move route in a spatial maze, or responded to two arrows that accompanied the maze. In these simple conditions, we expected the strongest evidence for differentiation between MD regions. Two MD regions in particular – the posterior/dorsal LFC (pdLFC) and the IPS – have especially strong links to spatial processing, including eye movements, with the MD region in pdLFC lying close to the location of the human frontal eye field (Pierrot-Deseilligny et al. 1991; Corbetta et al. 1998). Thus we expected strong activity in IPS and pdLFC even for the simple tasks, and recorded saccades to assess their potential contribution to such activity.

Next, we wished to examine the effects of increased task complexity. Though complexity is a broad term, here we considered simply the number of items or cognitive operations involved in the task. On this definition, there is much evidence that increased task complexity drives increased activity across the entire MD network, for example with increasing *n* in *n*-back working memory tasks (Hampshire et al., 2012; Hampshire et al., 2011; Owen, 1997; Postle et al., 1999). To increase complexity, we required maze and arrow tasks to be solved simultaneously. Within the MD network, we expected activity to strengthen and extend with increasing complexity. In particular, following prior results related to rule complexity (Badre and D'Esposito 2007, 2009; Crittenden and Duncan 2014), we expected increasing anterior spread along lateral frontal cortex.

To compare with complexity, we used two further manipulations. The first was time pressure, chosen as a candidate for enhancing the sense of challenge and accompanying autonomic arousal. Activity in several MD regions, e.g. insula and ACC, has been shown to correlate with increasing blood pressure, during both *n*-back working memory tasks and tasks where participants were told to apply physical pressure (Critchley et al. 2003). Such results have

been used to link activity to cognitive effort and autonomic arousal. Supporting a link to autonomic function, activity in similar MD regions is correlated with accuracy in a heartbeat detection task (Critchley et al. 2004). Subjective reports during electrical stimulation of the ACC can include a sense of impending challenge and the need to overcome it (Parvizi et al. 2013). Here we manipulated time pressure by comparing self-paced problem solving with problems visible only for a restricted time.

Our final manipulation was reward. The ACC has been studied extensively in the context of reward processing (Rushworth et al. 2004; Shenhav et al. 2013). In the behaving monkey, ACC neurons code multiple aspects of reward prediction and receipt (Procyk et al. 2000; Hadland et al. 2003; Kennerley and Wallis 2009; Matsumoto et al. 2015). In human functional magnetic resonance imaging (fMRI), ACC activity increases with reward magnitude (Knutson et al. 2005), though strong effects of reward can also be seen in many other MD regions (Padmala and Pessoa 2011; Dixon and Christoff 2012; Botvinick and Braver 2015). To compare reward effects across MD regions, on some trials we added the possibility of a substantial monetary reward.

In line with a link to autonomic function, pupil size has been correlated with activity in insula and other MD regions (Paulus et al. 2015). In general, pupil size increases with cognitive load (Alnaes et al. 2014; Zekveld et al. 2014), and has also been correlated with speed pressure (Murphy et al. 2016) and reward magnitude (Chiew and Braver 2014; Gergelyfi et al. 2015; Muhammed et al. 2016). Here we measured changes in pupil size for each of our task manipulations, and compared profiles of MD activity with profiles of pupil dilation.

In summary, in the context of planning solutions to a spatial maze, we examined manipulations of complexity, time pressure and reward. Across these 3 manipulations, our aim was to examine similarities and differences in the pattern of strengthening and extension of activity across the MD network.

Methods

Participants

25 participants (13 female) between the ages of 18-40 years (mean 27) took part in the study. One was excluded owing to his falling asleep in the scanner. A sample size of 24 was determined before data collection, as typical for functional imaging studies. All participants were right handed with normal or corrected to normal vision and had no history of neurological or psychiatric illness. To avoid interference with pupil measurement, participants with glasses were avoided but participants with contact lenses were allowed. The study was conducted with approval of the Cambridge Psychology Research Ethics Committee. All participants gave written informed consent and were reimbursed for their time.

Task Details

Participants had to solve a simple maze task and variations of it in the scanner (Figure 2.1A). They had to plan a route from a start location (marked by a green dot) to a goal location (marked by a magenta dot), avoiding blocked paths. The goal location was always at one of the four corners of the 4 x 4 maze; the start location was either in one of the central four positions (Figure 2.1B, left) or in an edge position (Figure 2.1B, right), such that the correct route always required four steps, including 1, 2 or 3 turns. In trials with a central start there were five path blocks, three around the start dot, one around the goal and one in between. In the case of an edge start, there were 4 path blocks, 2 around the start, one around the goal, and one in between. The maze spanned 15.2° visual angle horizontally and 10° vertically. Start and goal dots were 0.5° visual angle in diameter. Each trial was split into planning and execution phases (Figure 2.1A). In the planning phase, the maze was presented and participants generated their route then pressed a button to indicate that the plan was complete. There followed a fixed delay of 2 s, and then on 25% of trials only, an execution phase. On these trials the maze reappeared, but without the blocked paths. Participants used their previously planned route to move the green dot from the start position to the goal. Omission of blocked paths ensured that the previously planned route had to be remembered and used. Participants moved the green dot in the maze in any of the 4 directions using a custom-made button box with 4 buttons arranged like the arrow keys on a keyboard. Note that, for fMRI analysis, we focused just on planning phase data; execution was required on a subset of trials simply to ensure that participants did plan as instructed. As the occurrence of the execution

phase was unpredictable, our design ensured that routes were planned on all trials, confirmed by high accuracy when execution was required (see Results).

There were 6 different conditions, each indicated by a 1 s written cue at trial onset, followed by a 1 s interval and then the maze for the planning phase. In the baseline condition (cue “none”), the planning phase was a maze without start and goal dots or blocked paths, which appeared for one second and then disappeared. Participants had to press a button after the maze disappeared. They were instructed to do this as soon as possible. There were no baseline trials with execution, as there was no problem to be solved in this condition. In the simple maze condition (cue “maze”), participants carried out the maze task as described above. In the simple arrow condition (cue “arrow”), the baseline maze appeared with 2 arrows (0.27° visual angle) on either side of the maze. Each arrow pointed in one of four directions – up, down, left, or right. Participants pressed a button once they had memorised the directions of both the arrows. On execution trials, the baseline maze without the arrows reappeared, and the participants pressed buttons to indicate the directions of the two arrows (arrow on the left of the maze first). In the complex condition (cue “both”), the participants had to solve both arrow and maze tasks together. A maze with start and goal dots, blocks, and 4 arrows positioned within the maze was presented (Figure 2.1C). Participants planned a route from the start to the goal, as they did in the simple maze condition, and at the same time remembered the direction of two arrows encountered along the route (but not the two others). Two other arrows were included so as to not bias the planning of the route towards where the arrows were. In the execution phase, participants pressed buttons to move from start to goal in the maze, and when the route reached an arrow, made an additional button press to indicate arrow direction (Figure 2.1C). The green dot only moved for responses that were part of the maze task route. In the time pressure condition (cue “fade”), the task was similar to the simple maze condition, but at the planning phase, the maze faded away into the white background over a period of 0.3 to 1.3 s. The maze immediately started fading linearly from the moment it was presented. The fade time was fixed for each participant, as determined in a pre-scanning session (see Pre-Scanning Session below). Note that, even though the maze faded, participants could still take as long as required to complete their plan, and then pressed the button to proceed. Usually, this happened after the maze had fully disappeared. Finally, the reward condition (cue “fade+£1”) was the same as the time pressure condition, except that, if the trial had an execution phase, a reward was given for correct completion (£1 per

trial). The time pressure was present in the reward condition to make sure the participants did not study the maze longer, in order to be more accurate.

In all conditions, execution trials were followed by a 1 sec feedback display. In the maze task, the goal circle turned yellow to indicate correct completion of the trial. A green tick indicated correct trials in the arrow task. In the maze task, an error was scored for a keypress not moving along an open path. In the arrow task, an error was scored for a keypress in the wrong direction. In the complex condition, these rules were combined, with the rule for the arrow task implemented whenever the sequence of responses reached an arrow position.

Errors resulted in immediate termination of the trial, and presentation of error feedback in the form of a red cross. A 1 s inter-trial interval separated consecutive trials.

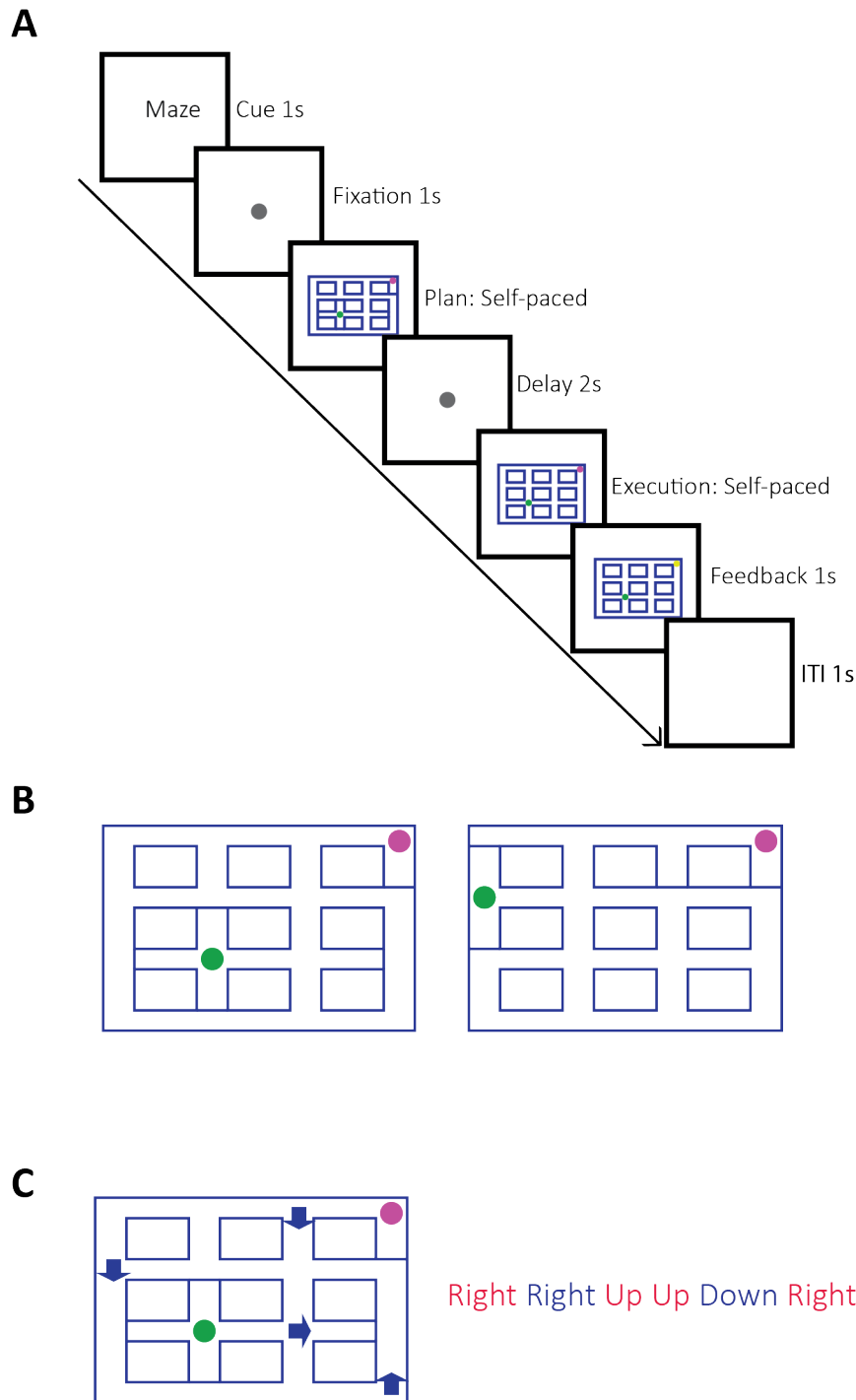


Figure 2.1: A. An example simple maze trial. The trial begins with a cue (1 s), followed by a fixation period (1 s) and then the planning phase. A blue maze is presented with green and magenta dots respectively indicating the starting position and goal. Blue lines extending across some paths indicate blocked paths. Participants press a button once they have planned a route from the start to the goal, avoiding the blocked paths. Following a delay of 2sec, on 25% of the trials, there is an execution phase. The maze reappears with the start and the goal dots, but not the blocked paths. The participants execute their previously planned route and

are given feedback for 1sec. **B. Different start positions.** Example planning phase mazes with start position in centre (left) or at the edge (right). Each maze was designed with only one open path adjacent to the start position, and only one open path adjacent to the goal. All routes were 4 steps. Routes had one, two or three turns. **C. Complex maze.** Example maze from the complex condition (arrow and maze task to be done together). Start, goal, blocked paths and arrows are present. The correct sequence of responses is shown to the right, with red indicating responses for the maze and blue indicating responses to arrows. Trials were only considered correct if all responses were made in the correct order.

Pre-Scanning Session

A practice session took place before the scan. The participants carried out 25 simple maze trials, and 2 of each of the other conditions. 40% of the average planning time of the last 20 simple maze trials (excluding incorrect trials) was set as the fade time to be used in the scanner session. All trials in the pre-scanning session had an execution phase, and the participants were told that this was not always the case in the scanning session.

Scanning Session

The scanning session included a structural scan, eye-tracking calibration and two functional runs. Each run consisted of 192 trials, 32 trials of each of the 6 conditions, presented in a random order. 8/32 were execution trials, except in the baseline condition, which had no execution phase. There were three breaks in each run. During this break, the money earned in the reward trials and the number of execution trials completed correctly were presented on the screen. The task resumed on a button press by the participants. The average \pm SD EPI time for each run of the task was 24.8 ± 1.83 min. The task was written and presented using Psychtoolbox3 (Brainard 1997) and MatLab (The MathWorks). Stimuli were projected on an MRI-compatible screen inside the scanner.

Data Acquisition

fMRI data were acquired using a Siemens 3T Prisma scanner with a 32-channel head coil. We used a multi-band imaging sequence with a multi-band factor of 3, acquiring 2mm isotropic voxels (Feinberg et al. 2010). Other acquisition parameters were: TR = 1.1s, TE = 30ms, 48 slices per volume with a slice thickness of 2 mm and no gap between slices, in plane resolution 2×2 mm, field of view 205 mm, and flip angle 62° . T1-weighted multiecho MPRAGE (van der Kouwe et al. 2008) high-resolution images were also acquired for all participants (voxel size 1 mm isotropic, field of view of $256 \times 256 \times 192$ mm, TR = 2530ms,

TE = 1.64, 3.5, 5.36, and 7.22ms). The voxelwise root mean square across the four MPRAGE images was computed to obtain a single structural image.

Eye Tracking

The diameter and position of the participant's left pupil were continuously measured using an iView X Version 2.8.26 eye tracker (SensoMotoric Instruments SMI, Teltow, Germany). A 9-point spatial calibration was performed before the task began. For each trial, the cue, fixation, planning, delay, execution and ITI phase onset markers were sent to the eye tracker from the MatLab task script. The raw data obtained from the tracker were converted to a text file using SMI IDF Converter 3.0.15. For each participant, the mean pupil size during the planning phase was calculated for each trial, and then averaged across all trials of the same condition. The BeGaze software from SMI was used to automatically detect fixation, blink and saccade events. The number of saccades per second for the planning phase was calculated for each condition and participant.

fMRI analysis

All analysis of fMRI data was performed using SPM12 (Wellcome Department of Imaging Neuroscience, London, England; www.fil.ion.ucl.ac.uk), and the Automatic Analysis (aa) toolbox (Cusack et al. 2014). Initial processing included motion correction and slice time correction. The structural image was coregistered to the Montreal Neurological Institute (MNI) template, and then the mean EPI was coregistered to the structural. The structural was then normalised to the template via a nonlinear deformation, and the resulting transformation was applied on the EPI volumes. Spatial smoothing of FWHM = 5mm was performed for whole brain analyses only.

Separately for each of the 6 conditions, we created planning and execution regressors (planning only for the baseline condition), leaving cue, fixation, delay and ITI as part of the implicit baseline. For each of the four maze conditions (simple maze, complex, time pressure and reward conditions), we separately modelled routes of 1, 2 and 3 turns from the central start position (see Figure 2.1B left), and routes with 1 and 2 turns from the outer start position (see Figure 2.1B right). The mean response to each maze condition was created by averaging these 5 separate route regressors. This was done to remove potential route effects, if any, from the condition contrasts presented in the analysis. Planning and execution phases were taken as lasting from stimulus onset to response (planning phase - button press indicating the end of planning; execution phase – final button press). To create regressors, measured

durations for each phase on each trial were convolved with the canonical hemodynamic response function. Analyses concerned just regressors from the planning phase of each trial, with execution regressors, along with the 6 movement parameters and run means, included as covariates of no interest. Note that, for trials with an execution phase, planning data were included whether or not execution was correct.

Our primary analysis focused on *a priori* MD regions of interest (ROIs). For this analysis, we used a template for the MD network ROIs in MNI space as defined in Fedorenko et al., (2013) (see t-map at <http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem>). Based on the regional divisions in the template, we selected the anterior, middle, and posterior parts of the middle frontal gyrus (aMFG, mMFG, and pMFG, respectively), a posterior dorsal region of the LFC (pdLFC), AI-FO, preSMA/ACC, and IPS, symmetrically defined in the left and right hemispheres (Figure 2.4A). Using the MarsBaR toolbox (<http://marsbar.sourceforge.net>; Brett et al. 2002) for SPM 12, beta estimates for each regressor of interest were averaged across runs and across voxels within each ROI, separately for each participant and condition. Contrasts between conditions were performed using t-tests. To supplement ROI analysis, selected contrasts were also performed voxelwise across the whole brain, corrected for multiple comparisons using the false discovery rate (FDR), $p < .05$ (Benjamini and Hochberg 1995).

In a supplementary analysis we repeated critical contrasts in a set of more focused ROIs. Resting state connectivity analyses often produce a “frontoparietal control” network resembling the MD network (Schaefer et al., 2018; Yeo et al., 2011). To isolate “frontoparietal control” voxels, we conjoined each of our MD ROIs with the corresponding frontoparietal control network from Yeo et al. (2011; 7-network parcellation). This conjunction produced regions of overlap in each of our *a priori* MD ROIs (minimum overlap 482 voxels, maximum 7965 voxels, separated by hemisphere). Analyses were then repeated on these reduced “conjunction” ROIs.

Results

Behaviour

Reaction times (RTs) for the planning phase of each condition, along with accuracies for the execution phase, are shown in Table 1. All RTs were timed from display onset until response.

In the baseline condition, it was display offset that triggered the response.

Planning RT was shortest in the baseline condition, and longest in the complex condition, which required planning of responses for both tasks (simple maze and simple arrow). A repeated measure ANOVA showed a main effect of condition ($F_{5, 23} = 87.59, p < 0.001$). A post hoc Tukey-Kramer multiple comparisons test showed that RTs in the baseline condition were significantly faster than all other conditions (all $p < 0.001$). The complex condition had significantly longer RTs than all other conditions (all $p < 0.001$). There was no RT difference between the simple arrow, simple maze, time pressure, and reward conditions (all $p > 0.05$).

Accuracies of the execution trials were higher in the simple maze and simple arrow compared to the complex and reward (plus time pressure) conditions. A repeated measures ANOVA showed a main effect of condition ($F_{4, 23} = 6.95, p < 0.001$). A post hoc Tukey-Kramer multiple comparisons test showed that accuracies in the complex condition were significantly lower than simple maze and simple arrow ($p < 0.01, p < 0.02$, respectively). Accuracy rate for the reward (plus time pressure) condition was also lower than the simple maze and simple arrow conditions (both $p < 0.02$). Accuracy in the time pressure condition was not significantly different from any other condition ($p > 0.05$). The overall high execution accuracies, and the unpredictability of the execution trials indicated that planning was indeed undertaken.

As noted above, maze routes differed in the number of turns required (1, 2 or 3), and imaging analyses were designed to remove these effects from comparisons between conditions. To examine the effect of the number of turns on planning time, we ran two-way repeated measures ANOVA with condition (simple maze, complex, time pressure and reward) and route complexity as factors (1, 2, or 3 turns). The analysis showed a significant effect ($F_{2, 46} = 15.45, p < 0.001$), with mean planning times for 1-, 2- and 3-turn routes of $1.94 \pm 0.1s$, $2.11 \pm 0.1s$ and $2.27 \pm 0.1s$ respectively. Post hoc tests with Tukey-Kramer multiple comparisons tests showed a shorter RT for the 1-turn routes compared to both 2-turn and 3-turn routes (both $p < 0.01$), and for 2-turn routes compared to 3-turn ($p < 0.01$).

Table 1. Planning phase RTs and execution phase accuracies in each condition. Values are means \pm standard errors.

	Baseline	Simple maze	Simple arrow	Complex	Time pressure	Reward
RT (sec)	1.11 \pm 0.06	1.56 \pm 0.08	1.53 \pm 0.07	3.16 \pm 0.20	1.58 \pm 0.09	1.60 \pm 0.09
Accuracy (% correct)		97.1 \pm 0.75	96.6 \pm 1.12	91.1 \pm 1.68	91.9 \pm 2.00	88.02 \pm 2.41

Imaging

Imaging analyses concerned just the planning phase of each trial. For each ROI, contrasts of the 5 active tasks against baseline are shown in Figure 2.2. Note that results for each ROI are averaged over left and right hemispheres, with ROIs illustrated on the left hemisphere in the centre panel. Separate results for the two hemispheres are shown in Supplementary Figure A2.1. The most striking difference between ROIs concerned the contrast of simple tasks against baseline. In both hemispheres, this contrast was significant in IPS, pdLFC, preSMA/ACC and pMFG, with little hint of activation for remaining ROIs. The same remained true in the complex condition, though the contrast with baseline was now numerically positive in all ROIs of both hemispheres. With time pressure, significant activity against baseline extended to include all ROIs. This activity in all ROIs was further strengthened in the reward condition.

To provide an initial statistical overview, we entered these contrasts against baseline into a three-way repeated measures ANOVA with factors condition (5), ROI (7) and hemisphere (2). The analysis showed a significant main effect of condition ($F_{4, 92} = 13.43, p < 0.001$), reflecting the generally increasing pattern of activity from left to right within each panel of Figure 2.2. There was also a main effect of ROI ($F_{6, 138} = 77.64, p < 0.001$), indicating much stronger overall activation for some ROIs (especially pdLFC, IPS) than others (especially AI/FO, mMFG and aMFG). Though activations were broadly similar on the two sides (Figure S1), there was also a significant main effect of hemisphere ($F_{1, 23} = 25.44, p < 0.001$). Finally there were interactions between condition and ROI ($F_{24, 552} = 10.03, p < 0.001$), condition and hemisphere ($F_{4, 92} = 26.93, p < 0.001$), ROI and hemisphere ($F_{6, 138} = 4.68, p <$

0.001), and between condition, ROI and hemisphere ($F_{24, 552} = 8.05, p < 0.001$), all indicating a degree of functional differentiation between ROIs.

Specific contrasts were then used to isolate the effects of individual variables (Figure 2.3). Again results were broadly similar in left and right hemispheres, though with some quantitative differences (see Supplementary Figure A2.2). In the following sections, we report significant effects averaged across hemispheres, accompanied by results of an ANOVA separating the two.

SM-B = Simple Maze – Baseline
 SA-B = Simple Arrow – Baseline
 C-B = Complex – Baseline
 TP-B = Time Pressure – Baseline
 R-B = Reward – Baseline

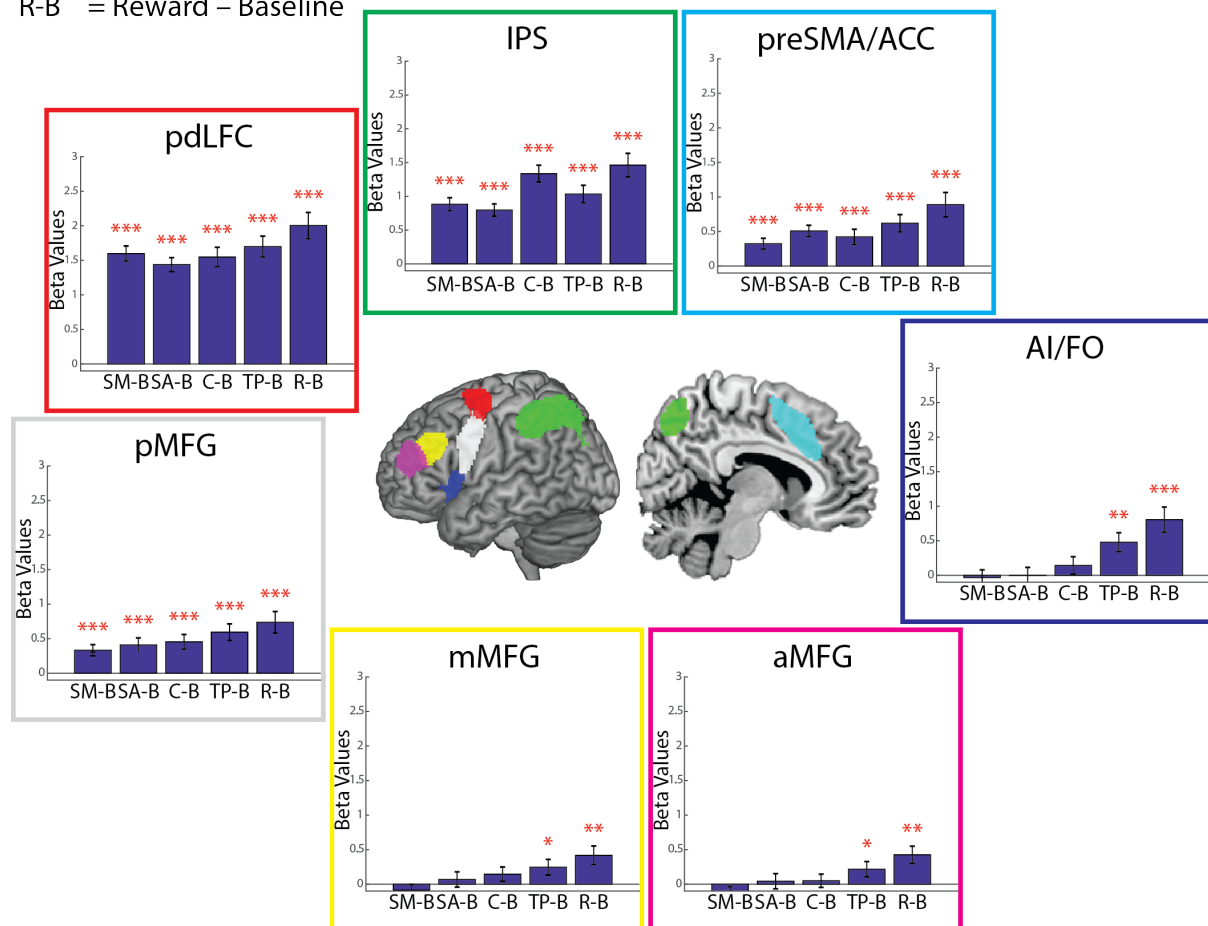


Figure 2.2: Functional differences across MD ROIs more prominent in simple tasks.

Separate panels for each ROI show contrasts (beta values) for each condition minus the baseline condition, with standard error of the mean. Red asterisks show significance for a

one-tailed t-test against zero (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$). Results are averaged across hemispheres, with ROIs (centre panel) illustrated on the left hemisphere.

Simple tasks versus baseline

To examine brain activity associated with the simplest versions of our tasks, we used a contrast of simple tasks (mean of simple maze and simple arrow) against baseline. In Figure 2.3, results for each MD ROI are shown in the leftmost column of each panel. For this contrast, there was significant univariate activity in only a part of the MD network, and in particular, within LFC, only for the most posterior regions. The contrast was significant for pdLFC, pMFG, preSMA/ACC and IPS ($t_{23} = 15.50$, $p < 0.001$, $d = 2.55$, $t_{23} = 4.49$, $p < 0.001$, $d = 0.84$, $t_{23} = 5.60$, $p < 0.001$, $d = 0.81$, $t_{23} = 10.06$, $p < 0.001$, $d = 1.44$ respectively), but with little hint of activity in AI/FO, mMFG or aMFG ($p > 0.1$ for all).

A repeated measures ANOVA for this contrast had factors ROI (7) and hemisphere (2). Confirming the above selectivity, there was a highly significant main effect of ROI ($F_{6, 23} = 84.01$, $p < 0.001$). In addition, the analysis showed main effects of hemisphere ($F_{1, 23} = 48.66$, $p < 0.001$) and an interaction between the two ($F_{6, 23} = 7.02$, $p < 0.001$). Though the main pattern of results was generally similar on the two sides (Supplementary Figure S2), only the left hemisphere showed significant activity in the pMFG.

ST-B = Simple Tasks – Baseline
C-ST = Complex – Simple Tasks
TP-SM = Time Pressure – Simple Maze
R-TP = Reward – Time Pressure

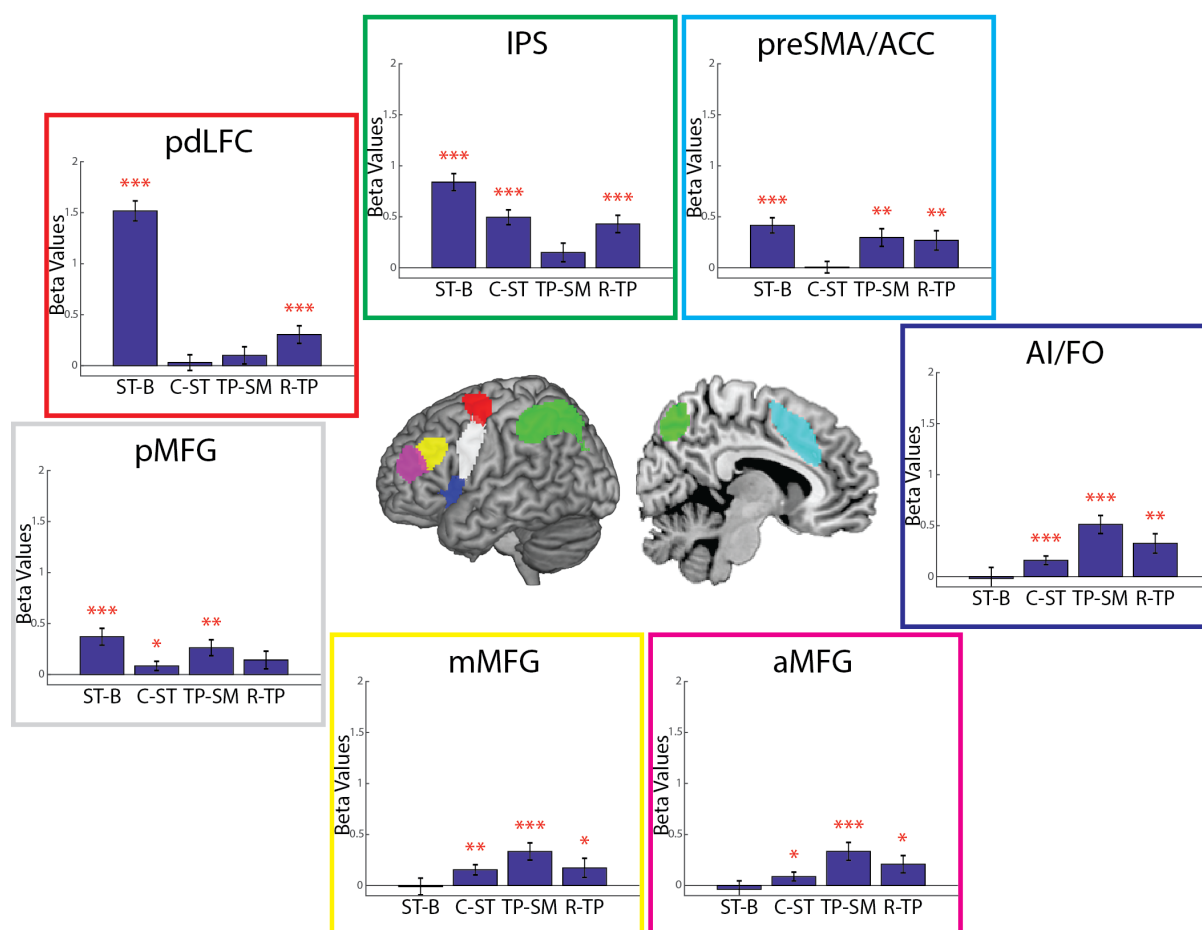
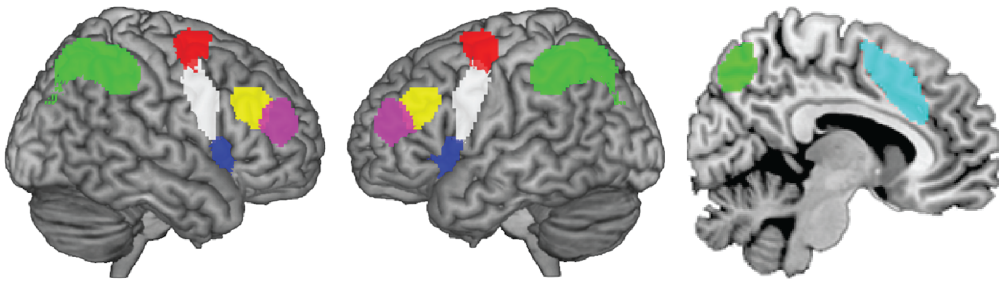


Figure 2.3: Increased co-recruitment with increased demand, challenge and reward.

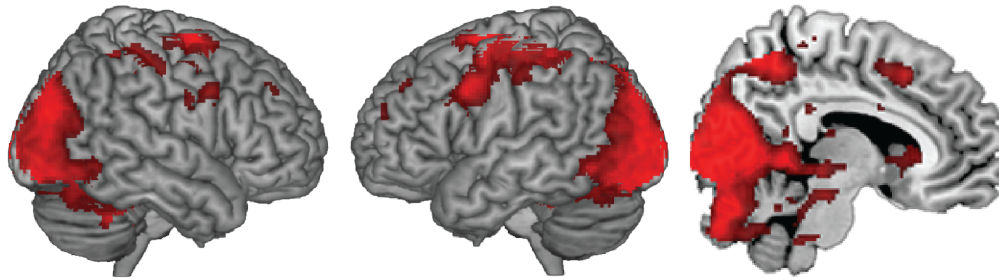
Separate panels for each ROI show effects (beta values) for each contrast: average of simple tasks vs. baseline, complex vs. average of simple tasks, time pressure vs. simple maze, and reward vs. time pressure; with standard error of the mean. Red asterisks show significance for a one-tailed t-test against zero (* = p<0.05, ** = p<0.01, *** = p<0.001). Results are averaged across hemispheres, with ROIs (centre panel) illustrated on the left.

A complementary whole brain analysis for this contrast (Figure 2.4B) confirmed the above picture, with activity significant in posterior dorsal LFC, parietal cortex, and preSMA, accompanied by the expected substantial recruitment of visual areas.

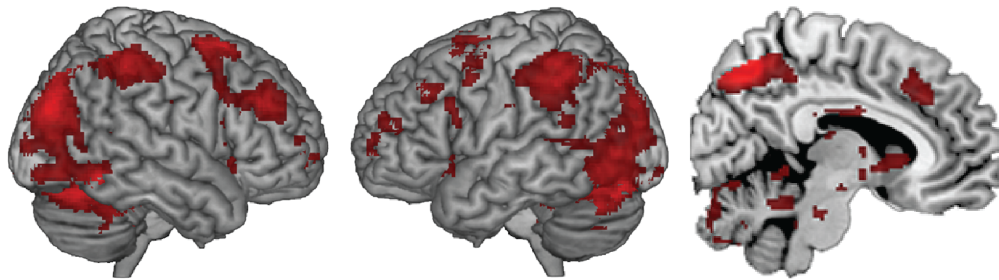
A MD Template



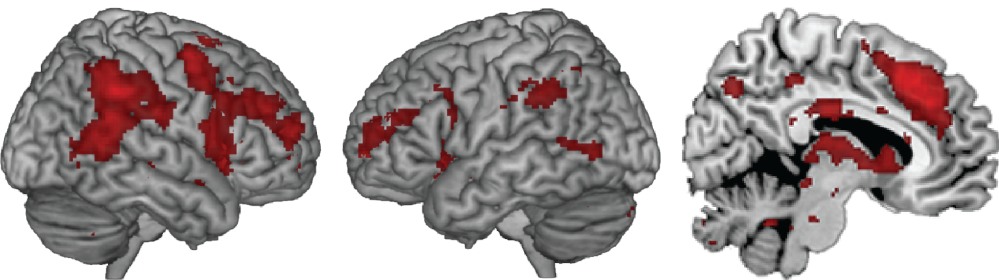
B Simple Tasks versus Baseline



C Complex versus Simple Tasks



D Time Pressure versus Simple Maze



E Reward versus Time Pressure

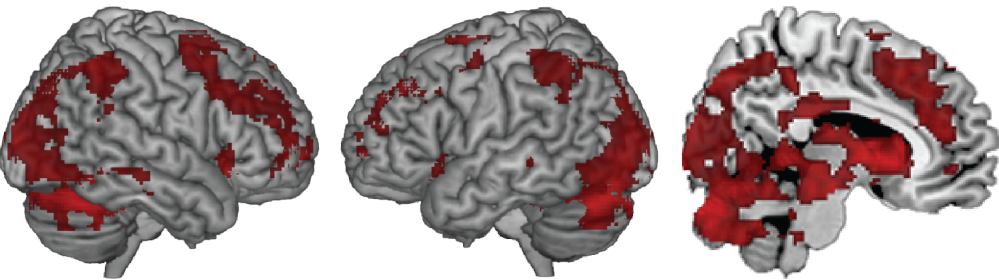


Figure 2.4: Whole brain contrasts. A. Templates of the MD system used for the ROI analysis. B-E. t-statistics associated with significant contrasts of: B. average of simple

tasks vs. baseline; **C.** complex vs. average of simple tasks; **D.** time pressure vs. simple maze; **E.** reward vs. time pressure. Contrasts are corrected for multiple comparisons using FDR ($p < 0.05$). Maps do not include activity in the dorsal part of the parietal lobe, as data for this region were not acquired for 4 participants. Note that ROI results do not change on excluding those 4 participants.

Complexity

A contrast of the complex condition versus the mean of the two simple tasks was used to examine effects of complexity. Results for each MD ROI are shown in the second column of each panel in Figure 2.3. In IPS and pMFG, activity already present in the simple tasks was further increased in the complex condition, especially in IPS ($t_{23} = 6.92$, $p < 0.001$, $d = 1.14$ for IPS and $t_{23} = 1.84$, $p = 0.039$, $d = 0.22$ for pMFG). With increasing complexity, activity also appeared in three new MD regions, AI/FO and mMFG and aMFG ($t_{23} = 3.71$, $p < 0.001$, $d = 0.49$, $t_{23} = 3.09$, $p < 0.01$, $d = 0.38$, $t_{23} = 2.00$, $p < 0.05$, $d = 0.21$ respectively). pMFG showed significant activity only in the left hemisphere, while mMFG and aMFG showed activity only in the right (Supplementary Figure S2). For pdLFC and preSMA/ACC, activity already present in the simple tasks remained at the same level in the complex condition ($p > 0.1$ for both), though this conclusion is tempered by whole brain analysis (see below).

A repeated measures ANOVA for this contrast showed a main effect of ROI ($F_{6, 23} = 20.96$, $p < 0.001$), an interaction of ROI and hemisphere ($F_{6, 23} = 2.77$, $p < 0.05$), but no main effect of hemisphere. Again, the pattern of results was broadly similar on the two sides, with the differences noted above for pMFG, mMFG and aMFG (Supplementary Figure S2).

For this contrast, whole brain analysis (Figure 2.4C) showed increased activity throughout the MD network, including regions that were and were not activated by the simple tasks alone. A small region of significant increase was observed in the preSMA/ACC, though not sufficient (and/or too anterior) to determine results in the entire ROI. Similarly, in the posterior dorsal frontal cortex, whole brain analysis suggested additional activity in the complex condition, but immediately anterior to the *a priori* pdLFC ROI.

Time pressure

Figure 2.3 shows results for the time pressure manipulation in the third column of each panel. Most MD regions showed increased activity with increasing time pressure ($t_{23} = 5.82$, $p < 0.001$, $d = 1.26$, $t_{23} = 3.84$, $p < 0.001$, $d = 0.74$, $t_{23} = 3.94$, $p < 0.001$, $d = 0.67$, $t_{23} = 3.35$, $p <$

0.01, $d = 0.59$, $t_{23} = 3.45$, $p < 0.01$, $d = 0.65$ for AI, aMFG, mMFG, pMFG and preSMA/ACC respectively). The exceptions were pdLFC and IPS that did not show this effect ($p > 0.05$ in each case), though for IPS, the effect was significant in the right hemisphere (Supplementary Figure S1). With time pressure, in particular, aMFG was now strongly added to the active network.

Repeated measures ANOVA for this contrast showed a main effect of ROI ($F_{6, 23} = 9.44$, $p < 0.001$), a main effect of hemisphere ($F_{1, 23} = 5.15$, $p < 0.05$), and an interaction between the two ($F_{6, 23} = 3.09$, $p < 0.05$).

The corresponding whole brain map (Figure 2.4D) confirmed this picture, with activity now extending into anterior regions of MFG and AI. Whole brain maps again showed increased activity in a region immediately anterior to the pdLFC ROI. Additional bilateral activity was seen in posterior temporal cortex.

Reward

To isolate the effect of potential reward, we contrasted the “reward” condition with the “time pressure” condition. These conditions had matched time pressure and differed only in whether correct performance was rewarded. In Figure 2.3, results for this contrast are shown in the rightmost column of each panel. Significant increases with reward appeared in all ROIs except pMFG ($t_{23} = 3.43$, $p < 0.01$, $d = 0.61$; $t_{23} = 2.46$, $p < 0.02$, $d = 0.43$; $t_{23} = 1.84$, $p < 0.05$, $d = 0.35$; $t_{23} = 2.82$, $p < 0.01$, $d = 0.43$; $t_{23} = 5.05$, $p < 0.001$, $d = 0.80$; $t_{23} = 3.50$, $p < 0.001$, $d = 0.42$, for AI, aMFG, mMFG, preSMA/ACC, IPS and pdLFC, respectively; for pMFG, $p = 0.056$, $d = 0.26$).

Repeated measures ANOVA for this contrast showed a main effect of ROI ($F_{6, 23} = 8.18$, $p < 0.001$), but not of hemisphere. There was also a significant interaction ($F_{6, 23} = 4.33$, $p < 0.05$). In all three MFG regions, reward effects were significant on the right but not the left (Supplementary Figure S2).

The corresponding whole brain map (Figure 2.4E) confirmed this picture, with the map showing the entire MD network, especially in the right hemisphere. Again, there was a further activity increase immediately anterior to the *a priori* pdLFC ROI. Increased activity was also seen in more rostral parts of the ACC, along with visual cortex and multiple subcortical structures including ventral and dorsal striatum, in line with well-known roles in reward processing.

Conjunction ROIs

In a supplementary analysis, contrasts were repeated for reduced ROIs, formed by the intersection of our MD ROIs with a frontoparietal control network from Yeo et al. (2011). Results are shown in Figure 2.5. Again, only IPS, pdLFC, pMFG and preSMA/ACC were more active in the simple tasks compared to the baseline. Using these more focused ROIs, all regions showed increased activity in the complex contrast, including pdLFC and preSMA/ACC, which did not show this effect when the entire ROI was used. All regions also showed significant effects of time pressure (previously absent in pdLFC and IPS) and reward (previously absent in pMFG). With these more focused ROIs, there was a clarified picture of selective activity in simple tasks, coupled with increasing activity across the entire network with increase in complexity, time pressure and reward.

ST-B = Simple Tasks – Baseline
 C-ST = Complex – Simple Tasks
 TP-SM = Time Pressure – Simple Maze
 R-TP = Reward – Time Pressure

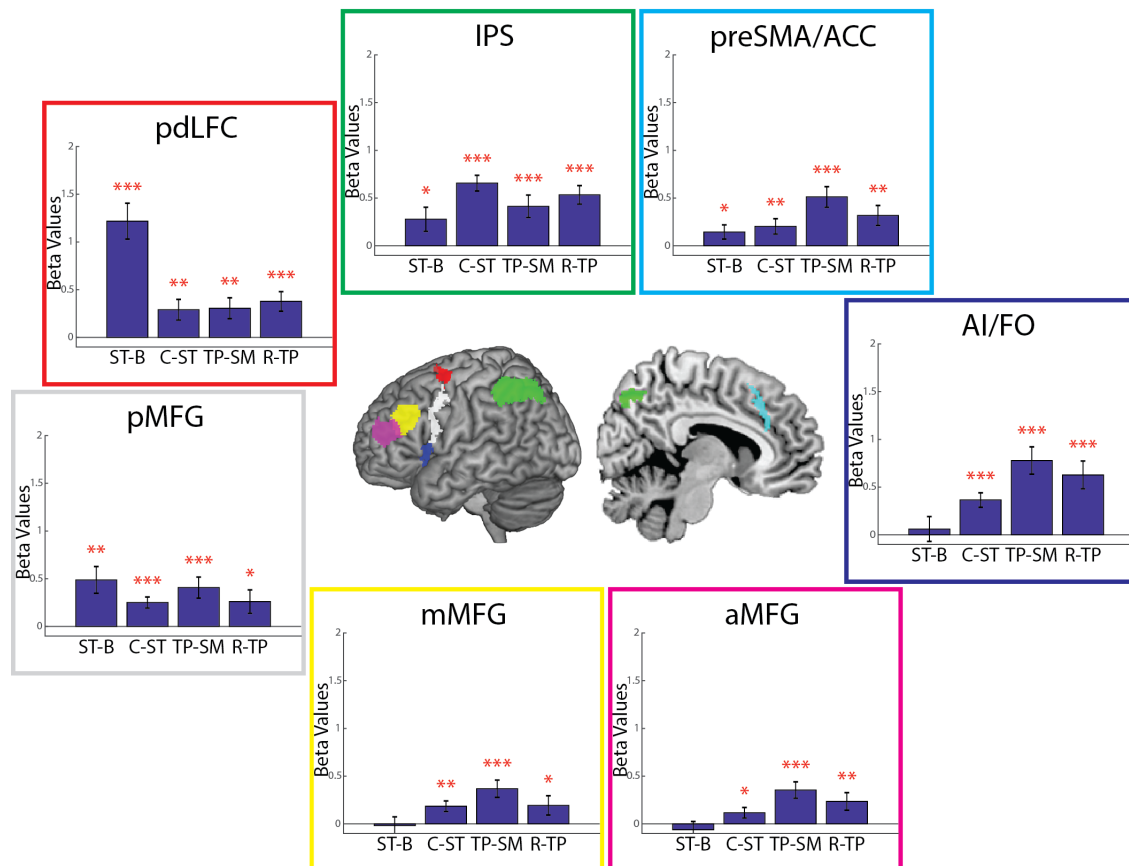


Figure 2.5: Reduced ROIs highlighting co-recruitment with increased demand, challenge and reward. Data as in Figure 2.3 for reduced ROIs (overlap of *a priori* MD ROIs

with frontoparietal control network from Yeo et al., 2011). Reduced ROIs (centre panel) are illustrated on the left hemisphere.

Eye tracking

For each condition, Table 2 shows the mean increase in pupil size between the initial fixation phase of each trial (1 s following the task cue; see Figure 2.1) and the planning phase. A one-way repeated measures ANOVA showed a main effect of condition ($F_{5, 23} = 10.62, p < 0.001$). Post hoc Tukey-Kramer multiple comparisons test showed that pupil dilation in the reward condition was higher than all other conditions ($p < 0.005$ for comparison with all other conditions, $d = 1.12, 1.36, 1.28, 1.46, 0.56$ for baseline, simple maze, simple arrow, complex, and time pressure respectively). No other differences were significant, though a planned comparison between the time pressure and simple maze conditions showed a trend in the predicted direction ($t_{23} = 1.84, p = 0.074$). Pupil sizes during the initial fixation period were not significantly different between conditions.

Saccade rates (number of saccades per second) for each condition are also shown in Table 2. A one-way repeated measures ANOVA showed a main effect of condition ($F_{5, 23} = 15.10, p < 0.001$). Similar post hoc tests as for pupil size analysis showed that the baseline condition had the lowest saccade rate ($p < 0.001, d = 1.34$ and $d = 1.11$ for comparison with simple maze and complex, $p < 0.05, d = 0.91$ and $d = 0.72$ with simple arrow and reward; and $p = 0.13, d = 0.50$ with time pressure). Saccade rates were higher in the simple maze condition compared to time pressure and reward ($p < 0.001, d = 0.67, d = 0.65$), probably because, in the latter, the maze disappeared while planning was still in progress. Saccade rates were similar in simple maze, simple arrow and complex conditions ($p > 0.1$).

Table 2. Planning phase: pupil size (increase from fixation period) and saccade rate. Values are means \pm standard errors.

	Baseline	Simple maze	Simple arrow	Complex	Time Pressure	Reward
Size increase (mm)	0.10 \pm 0.07	0.11 \pm 0.05	0.07 \pm 0.04	0.06 \pm 0.03	0.27 \pm 0.11	0.55 \pm 0.11
Saccade Rate (saccades/sec)	2.57 \pm 0.14	3.20 \pm 0.15	3.43 \pm 0.15	3.31 \pm 0.14	2.93 \pm 0.16	3.01 \pm 0.12

Discussion

MD regions are recruited in many different cognitive tasks. Previously it has been suggested that they construct the successive steps of a mental control program, each step binding together the required stimuli, responses, rules into the correct combinations. Across MD regions, previous data suggest a degree of specialization especially at low load, with progressive recruitment of the entire network as load increases. While differentiation between MD regions may reflect preferential access to different task features, co-recruitment may reflect increasing information integration and exchange – matching the findings of multivoxel pattern analysis, with widespread but quantitatively varying representation of multiple task features across MD regions (Woolgar et al. 2016). To examine differentiation and co-recruitment, we used a spatial maze and manipulations of complexity, time pressure and reward, comparing profiles of activity across an *a priori* set of MD regions.

As predicted, the strongest differentiation between MD regions came from the comparison of simple tasks against baseline. For this contrast, the results suggested recruitment of a largely posterior frontal network, including pdLFC and pMFG, accompanied by IPS and preSMA/ACC (Figure 2.3). The MD regions examined here show increased activity in many tasks, some with no evident spatial element (Fedorenko et al. 2013). At the same time, as noted earlier, it is likely that regions overlapping with or close to pdLFC and IPS have strong involvement in spatial operations, and indeed, much imaging evidence indicates a relative specialization of these regions for spatial tasks (Nee et al., 2013; Owen et al., 2005; Rottschy et al., 2012; Wager and Smith, 2003). With close proximity and functional connections to other spatial regions, pdLFC and IPS may be the first to be strongly recruited in a spatial task, with additional MD regions added as demands increase. Beyond this specific role in spatial operations, it is also possible that posterior frontal and parietal regions are the first to be recruited in many kinds of task, with activity spreading anteriorly as demands increase. This would match previous results suggesting anterior spread with increased rule complexity (e.g. Badre and D'Esposito, 2007, 2009; Crittenden and Duncan, 2014). More work is needed to establish whether the current results are specific to tasks with a strong spatial element, or whether they are more general and apply to other cognitive domains as well.

In the complex condition, as predicted, there was widespread increase of activity across the MD network. This increase was especially clear in the reduced ROIs produced by overlapping our *a priori* MD regions with the Yeo et al. (2011) frontoparietal control network

(Figure 2.5). As compared to simple tasks, activity now spread forward along the lateral frontal surface, and appeared in AI/FO. Of those regions already recruited in the simple tasks, activity further increased in IPS and pMFG, though for pdLFC and preSMA/ACC, this was visible only in the reduced ROIs.

Time pressure was introduced to make the simple maze task more challenging, while keeping computational operations much the same. Based on the literature, we thought it possible that AI/FO and preSMA/ACC would be the most sensitive to time pressure, and indeed, the effect was numerically strongest in AI/FO. More broadly, however, the pattern of results resembled the effects of complexity, with strengthened activity in several of those regions already involved in the simple tasks (pMFG, right IPS and preSMA/ACC), maintained activity in others (pdLFC), and now strong recruitment of all remaining MD regions (AI/FO, mMFG and aMFG). Again, increases were more consistent in reduced ROIs formed from overlap with the Yeo et al. (2011) frontoparietal control network. In line with these findings, other studies have implicated bilateral LFC, ACC and AI in processing speeded stimuli (Peelle et al. 2004; Loose et al. 2006).

Similarly, we had anticipated that selected MD regions - especially preSMA/ACC – might be most sensitive to reward, but instead, reward effects were widespread throughout the MD network. The whole brain contrast in the reward condition also showed activation in subcortical regions - including ventral and dorsal striatum, both involved in reward processing (Liu et al. 2011; Bartra et al. 2013; Clithero and Rangel 2014) - and spreading forward from preSMA/ACC into reward-related rostral ACC (de la Vega et al. 2016). Another notable feature of the whole brain map – shared with the time pressure condition – was activity around the temporoparietal junction. Activity in this region has often been associated with unusual, important events (Downar et al. 2002; Krall et al. 2015), perhaps including the challenge that was introduced with the disappearance of the maze to be solved. Widespread MD recruitment with increasing reward is broadly consistent with a number of previous studies (Padmala and Pessoa 2011; Dixon and Christoff 2012).

Regarding pupil size, we found a significant increase over baseline only in the reward condition, with some additional evidence for the predicted increase in the time pressure condition. Our results are consistent with those of Chiew & Braver (2014), Gergelyfi et al. (2015) and Muhammed et al. (2016), who all found pupil size to be correlated with reward magnitude. Greatest pupil dilation in the reward condition was associated with strongest response throughout the MD system (Figure 2.2).

Though our three manipulations all showed a broad pattern of MD co-recruitment, there were also clear quantitative differences between ROIs. For each manipulation, ANOVA showed significant effects of ROI, with additional differences in exact activity pattern between hemispheres. For complexity, for example, the strongest effect was seen in the IPS. Plausibly, the IPS has most immediate access to the spatial features of the task, and thus most sensitivity to the increases spatial demand of interleaving maze and arrow processing. For time pressure, in contrast, the strongest effect was seen in AI/FO, accompanied by the trend to increased pupil dilation mentioned above. Among MD regions, perhaps AI/FO, as we had predicted, is most closely associated with autonomic arousal, showing the dominant response as arousal increases. At the same time, effective cognition requires multiple task features to be integrated. For example, when time pressure and reward drive autonomic arousal and mental effort, this effort must be specifically linked to the spatial operations of maze and arrow processing. Such integration, we suggest, is reflected in broad MD co-recruitment for all task manipulations.

An important question is the potential contribution of eye movements to our results. Plausibly, in particular, both pdLFC and IPS might contribute to control and production of saccades, with pdLFC lying close to the location of the human frontal eye field (FEF) (Pierrot-Deseilligny et al. 1991; Corbetta et al. 1998) and the IPS containing the parietal eye field (Muri et al. 1996). Saccades of course were more frequent in all active conditions compared to baseline, somewhat resembling the activity profile of pdLFC. It seems unlikely, however, that activity profiles in IPS, and perhaps also pdLFC, can be entirely explained by overt saccadic activity. For example, the rate of saccade production was well matched in complex and simple tasks, while complexity strongly increased IPS activity. Saccades may well have contributed, however, especially to activity in pdLFC, along with possible differences in planning of both executed and non-executed saccades.

In this light, it is intriguing that, in whole brain analysis, we found increased activity linked to complexity and time pressure in a region just anterior to our pdLFC ROI. In recent analysis of data from the Human Connectome Project, we have found evidence for an MD region lying not in the human FEF, but just anterior to it (Assem et al. 2019). Given individual variability between participants, and uncertainties in coregistration, it is possible that, in the current study, the pdLFC ROI derived from previous data was dominated, not by the intended MD region, but by FEF lying just posterior to it. Higher spatial-resolution fMRI data could be used to address this possibility.

Across the MD system, our results illustrate the balance between relative specialization, especially at low load, and progressive co-recruitment as demand or motivation increase. Across MD regions, there were clear quantitative differences in response to our three manipulations of complexity, time pressure and reward. At the same time, the broad pattern of progressive MD recruitment was much the same across manipulations. At each step of a mental control program, different kinds of mental content must be integrated to produce the required cognitive operations. Especially at low load, differentiation between MD regions may reflect preferential access to different task features. As task demands or motivation increase, however, there is increasing information exchange and integration, reflected in increasingly widespread activity across the MD network.

Chapter 3

Individual-subject functional localization increases univariate activation but not multivariate pattern discriminability in the ‘multiple-demand’ frontoparietal network

Introduction

Multiple studies have provided consistent evidence for the involvement of a large distributed network of frontal and parietal regions in cognitive control and flexible goal-directed behavior (Desimone and Duncan 1995; Duncan and Owen 2000; Duncan 2006b, 2010b; Stiers et al. 2010; Duncan 2013b; Fedorenko et al. 2013). This network has been termed the ‘multiple-demand’ (MD) network (Duncan, 2006), and it closely resembles other networks that have been associated with control processes such as the cognitive control network (e.g. Cole and Schneider, 2007), task-activation ensemble (Seeley et al. 2007), and task-positive network (Fox et al., 2005). The MD network includes the intraparietal sulcus (IPS), the anterior-posterior axis of the middle frontal gyrus (MFG), the anterior insula and adjacent frontal operculum (AI/FO), the pre-supplementary motor area (pre-SMA) and the dorsal anterior cingulate (ACC) (Duncan 2010b, 2013b). A primary characteristic of this network is an increase in activity with increased demand, especially seen through functional magnetic resonance imaging (fMRI) blood oxygenation level dependent (BOLD) activity, across a variety of cognitive domains such as working memory, task switching, inhibition, math, and problem solving (Dove et al. 2000; Cole and Schneider 2007; Fedorenko et al. 2013; Shashidhara, Mitchell, et al. 2019).

The development of multivoxel pattern analysis (MVPA) methods (Haxby et al. 2001b; Haynes and Rees 2006; Kriegeskorte et al. 2006) has further led to a variety of findings related to the representation of multiple aspects of cognitive control across the MD network. These include attentional effects, adaptive coding, and coding of target features and task rules (Woolgar, Thompson, et al. 2011; Nee and Brown 2012; Nelissen et al. 2013; Erez and Duncan 2015; Etzel et al. 2016; Wisniewski et al. 2016). MVPA allows for a fine-grained investigation of distributed patterns of activity and the information that is conveyed in these

patterns related to different experimental conditions and their respective cognitive constructs. However, its use in the frontoparietal MD network has been limited by overall low decoding levels, or discriminability between conditions, compared to other brain systems (Bhandari et al. 2018). This could be, at least in part, due to differences between individuals in respect to the spatial organization of the network. Indeed, it has been demonstrated that across the fronto-parietal lobes, several close by regions have been fractionated into different networks in resting state studies, with different boundaries in individuals (Yeo et al. 2011; Glasser et al. 2016; Schaefer et al. 2018). Nevertheless, many MVPA studies of the MD network used a group template to define regions of interest (ROIs) for all subjects to investigate task-related representations. Such a group template is robust and easily comparable across studies. However, it provides high sensitivity at the cost of specificity to individual differences, as it might not accurately identify regions in individual subjects due to both anatomical and functional differences (Brett, Johnsrude, et al. 2002; Nieto-Castañón and Fedorenko 2012; Fedorenko et al. 2013).

An alternative to the group template is using an independent functional localiser task to establish subject-specific ROIs. In this approach, participants perform a short task in addition to the main task in the scanning session, and the data from this task is used to localize regions-of-interest to be tested with data from the main task. This method is commonly used in vision research (Reddy and Kanwisher 2007; Erez and Yovel 2014; Lafer-Sousa et al. 2016; Weiner et al. 2018). For example, specific tasks are used to identify regions in individuals that are recruited for face processing (Kanwisher et al. 1997; Berman et al. 2010) and object processing (Malach et al. 1995). Task contrasts such as faces versus scrambled faces or objects versus scrambled objects are applied, and ROIs can be identified by the experimenter as clusters of activity in individual subjects. However, this alternative of manually defining ROIs by the experimenter using a functional localiser is subjective and therefore may be prone to biases, inaccuracies, and reduced reproducibility (Krishnan et al. 2006; Garrison et al. 2015).

To overcome the limitations of both the group template and the individual manually-defined regions, a hybrid group-constrained subject-specific approach has been proposed for use in the language system (Fedorenko et al. 2010, 2012), and later expanded to the ventral pathway of the visual system (Julian et al. 2012) and theory of mind regions (Paunov et al. 2019). In this approach, independent subject-specific localiser data are collected in addition to the main experimental task, then the thresholded contrast data from this task are masked with a group template of regions and only the voxels that were responsive to the localiser task within the

group template are used for further analysis in the main experimental task. The advantage of this approach is the use of a group template that ensures targeting of similar areas for all participants, as well as refining this localization by subject-specific activations within these areas. It therefore offers an objective experimenter-independent definition of subject-specific regions that does not require manual region definition. Importantly, it supports comparability across different studies because selected voxels in individual subjects are constrained to a group template. Using this group-constrained subject-specific ROIs approach has been shown to increase the detected univariate BOLD response associated with contrasts of interest in language-related areas compared to when a group template was used as ROIs (Fedorenko et al. 2010). More generally, the benefit of using individually defined ROIs for univariate results has been shown for the visual system (Saxe et al. 2006) and has been modelled and demonstrated using simulation data (Nieto-Castañón and Fedorenko 2012).

The hybrid group-constrained subject-specific approach has been subsequently used for the MD network in studies that used both univariate (Blank et al. 2014; Blank and Fedorenko 2017; Mineroff et al. 2018; Paunov et al. 2019) and multivariate measures (Erez and Duncan 2015; Shashidhara and Erez 2019) related to control processes. A spatial working memory task that has been previously demonstrated to robustly recruit the MD network, was used as a localiser. In this localiser task, a highly demanding condition is contrasted with an easier version of the same task to identify the network in individual subjects, then constrained by an anatomical or group-average functional template to define the subject-specific ROIs.

However, it remains an open question whether using these refined ROIs at the single subject level has any benefit for multivariate results in the MD network, similarly to the benefits that were previously reported for univariate results (Saxe et al. 2006; Fedorenko et al. 2010; Nieto-Castañón and Fedorenko 2012). Using the MD group template for MVPA means that many voxels outside the individually-defined functional MD regions are included in the analysis, which may not express the domain-general characteristics of the MD network. In fact, they may be part of other nearby brain systems (Yeo et al. 2011; Glasser et al. 2016; Schaefer et al. 2018), and their inclusion in the multivariate analysis may potentially mask out pattern-based differences between the experimental conditions of interest. If this is indeed the case, then the identification of the MD network in individual subjects has the potential to critically improve our ability to detect the neural signature of control processes as measured by multivariate methods, thus substantially increasing the benefit of using MVPA in cognitive control research. On the other hand, it has been previously demonstrated in the visual system and using simulations that even voxels outside the regions of increased

univariate activity contribute to multivoxel discrimination (Haxby et al. 2001b; Kriegeskorte et al. 2006). This implies that using more finely-defined subject-specific ROIs instead of the large and broad group template may potentially reduce discrimination levels. A related point concerns the size of ROIs. Increased decoding levels have been previously linked with increased size of ROI, at least in the visual system (Eger et al. 2008; Walther et al. 2009; Said et al. 2010), implicating the importance of controlling for the ROI size for MVPA. Using the group-constrained subject-specific ROIs allows for such control by selecting a fixed number of voxels from each ROI with the largest localiser contrast values. Such control for ROI size enables the comparison of decoding levels between the different regions within the network, which vary in size, as well as with regions outside this network. This, however, should not come at the expense of reduced decodability, if indeed the use of smaller ROIs reduces pattern discriminability. Overall, existing data provide only limited evidence regarding the link between the use of subject-specific ROIs and multivoxel pattern measures in the MD network.

In the current study, we build on the previously reported findings and ask whether using functionally-defined subject-specific ROIs affects multivoxel pattern results in the MD network. Because the recruitment of the MD network at the group level is observed across a range of cognitive domains (Fedorenko et al. 2013), different tasks can be potentially used as localisers. We therefore also ask whether different localisers may have different effects on multivariate results. To address these questions, we use three localiser tasks and an independent rule-based criterion task. The localiser tasks are spatial working memory, verbal working memory, and a Stroop-like task, which have all been previously shown to consistently recruit MD regions (Fedorenko et al. 2013). We first assess the reliability and variability of the level of recruitment of the MD network by the localisers. We then assess the benefit of the subject-specific ROIs for univariate results in the independent criterion task, aiming to replicate and generalize previous findings. Finally, we systematically test for the effect of using the subject-specific ROIs defined by each of the three localisers on multivariate results in the independent criterion task. We provide an important empirical evidence for the effect of using subject-specific ROIs on the ability to detect task-related representations in the MD network as reflected in distributed patterns of activity.

Methods

Participants

A total of 25 healthy participants (18 female, mean age 23.8 years) took part in the study. Three participants were excluded because of movements larger than 5 mm during at least one of the scanning runs, and one participant was excluded due to slice by slice variance larger than 300 after slice time correction in more than two runs. In addition, two participants were excluded due to technical problems with the task scripts. Lastly, one participant was excluded to maintain the balance of the order of localisers across participants. This participant was chosen randomly out of the participants who had the same order of localisers and prior to any data analysis beyond pre-processing. Overall, 18 participants were included in the analysis. All participants were right-handed and had normal or corrected-to-normal vision. Participants were either native English speakers or had learnt English at a young age and received their education in English. Participants gave written informed consent prior to participation and received a monetary reimbursement at the end of the experiment. Ethical approval was obtained from the Cambridge Psychology Research Ethics Committee.

Experimental paradigm

The study consisted of three localiser tasks and one rule-based similarity judgement task. The localisers were: a spatial working memory (WM) task, a verbal working memory task and a Stroop task, variations of which have previously been shown to recruit the MD network (Fedorenko et al. 2013). The rule-based task was used as a criterion task to test for univariate effects and rule decoding using MVPA, using both subject-specific ROIs based on activation data from the localisers and the group template. Participants practiced all tasks before the start of the scanning session. During scanning, participants performed two runs of each localiser followed by four runs of the rule-based task. The two runs of the same localiser always followed each other, and the order of the three localisers was balanced across participants. The average total scanning session duration was 105 minutes.

Localiser tasks and criterion task

The spatial WM and verbal WM localiser tasks were adapted from Fedorenko et al. (2013) and the Stroop task was adapted from Hampshire et al. (2012). The localisers were chosen based on their consistent recruitment of the MD network as has been shown by Fedorenko et al. (2013). The localisers all followed a blocked design. Each run contained 10 blocks,

alternating between Easy (5 blocks) and Hard (5 blocks) task conditions. There were no indications for the start or end of each block. The localiser tasks were designed to be used with a contrast of Hard versus Easy conditions, and in order to keep them as short as possible they did not include fixation blocks. The first run always started with an Easy block, and the second with a Hard block. All blocks lasted for 32 seconds, leading to a total run duration of 5 minutes and 20 seconds.

All tasks were coded and presented using Psychtoolbox (Brainard 1997) for MatLab (The MathWorks, Inc.). Stimuli were projected on a 1920 x 1080 screen inside the scanner, and participants used a button box, with one finger from each hand to respond.

Localiser 1: spatial WM task

In the spatial WM task (Figure 3.1A), each trial started with an initial fixation dot (0.5 s), followed by a 3x4 grid with either one (Easy condition) or two (Hard condition) highlighted cells. The highlighted cells were displayed over four seconds, with different cells highlighted every one second, leading to an overall four (Easy condition) or eight (Hard condition) highlighted cells in each grid. In a subsequent two-forced choice display, two grids with highlighted cells were presented on the right and left sides of the screen. Participants pressed a button (left or right) to indicate which grid matched the previously highlighted cells. After a response window of 3.25 s, a feedback was presented for 0.25 s. The correct grid appeared an equal number of times on the right and left. Overall, each trial was 8 s long, and each task block contained four trials.

Localiser 2: verbal WM task

The verbal WM task (Figure 3.1B) followed a similar design to the spatial WM task. Following fixation (0.5 s), participants were presented with four consecutive screens containing one (Easy condition) or two (Hard condition) written digits. In a following two choice display, participants indicated the correct sequence of digits by pressing a button. The two answer options were displayed at the center of the screen, one above the other for ease of reading. The left button was used to choose the sequence on top, while the right button was used for the sequence at the bottom. The correct sequence appeared an equal number of times on the top and bottom. Following a response window of 3.25 s, participants were given feedback at the end of each trial (0.25 s). Each trial was 8 s long, and each task block contained four trials.

Localiser 3: Stroop task

The third localiser was a variation of the Stroop task (Figure 3.1C). On each trial, following a fixation dot (0.5 s), participants were presented with a test word, which was the name of a color, written in color at the top of the screen. In the Easy condition, the ink color was the same as the color name (congruent) (e.g. the word ‘green’ written in green ink), and in the Hard condition the ink color and the color name were different (incongruent) (e.g. the word ‘red’ written in green ink). Participants had to indicate the ink color (rather than the written color name) by choosing one of two answer options at the bottom of the screen, displayed at the same time. The answer options were color name words, and their ink color was different from its name. Participants had to choose the word, i.e. the written color name (regardless of the ink color), that matched the ink color of the test word at the top. Therefore, participants had to switch between attending to the ink color of the test word (ignoring the written color name) to detecting the matching written color name (and ignoring the ink color) out of the two answer options at the bottom. We used a total of six colors throughout the task. In the congruent condition, the ink color of the answer options was chosen randomly, excluding the color name (and ink color) of the test word (e.g. the options for the above congruent test word example could be the word ‘blue’ written in brown ink and ‘green’ written in purple ink, with the latter being the correct answer). In the incongruent trials, the ink color of one of the answer options matched the color name (and not the ink color) of the test word. On half of the incongruent trials, it was the correct answer that had the same ink color as the test word color name (for example, if the test word is ‘red’ written in green, then a correct answer could be ‘green’ written in red). On the other half, it was the incorrect option with ink color the same as the test word color name (e.g. if the test word is ‘red’ written in green, then the incorrect option could be ‘purple’ written in red, see example in Figure 3.1C). This was done to further increase the conflict between stimuli and thus the difficulty level of the hard blocks while ensuring that the ink color of the test word cannot be used when choosing an answer. A total of six colors were used in this task (red, green, blue, orange, purple, and brown). Participants had 1.25 s to view the stimuli and respond, after which they received feedback for 0.25 s. Each trial was 2 s long, and blocks consisted of 16 trials each.

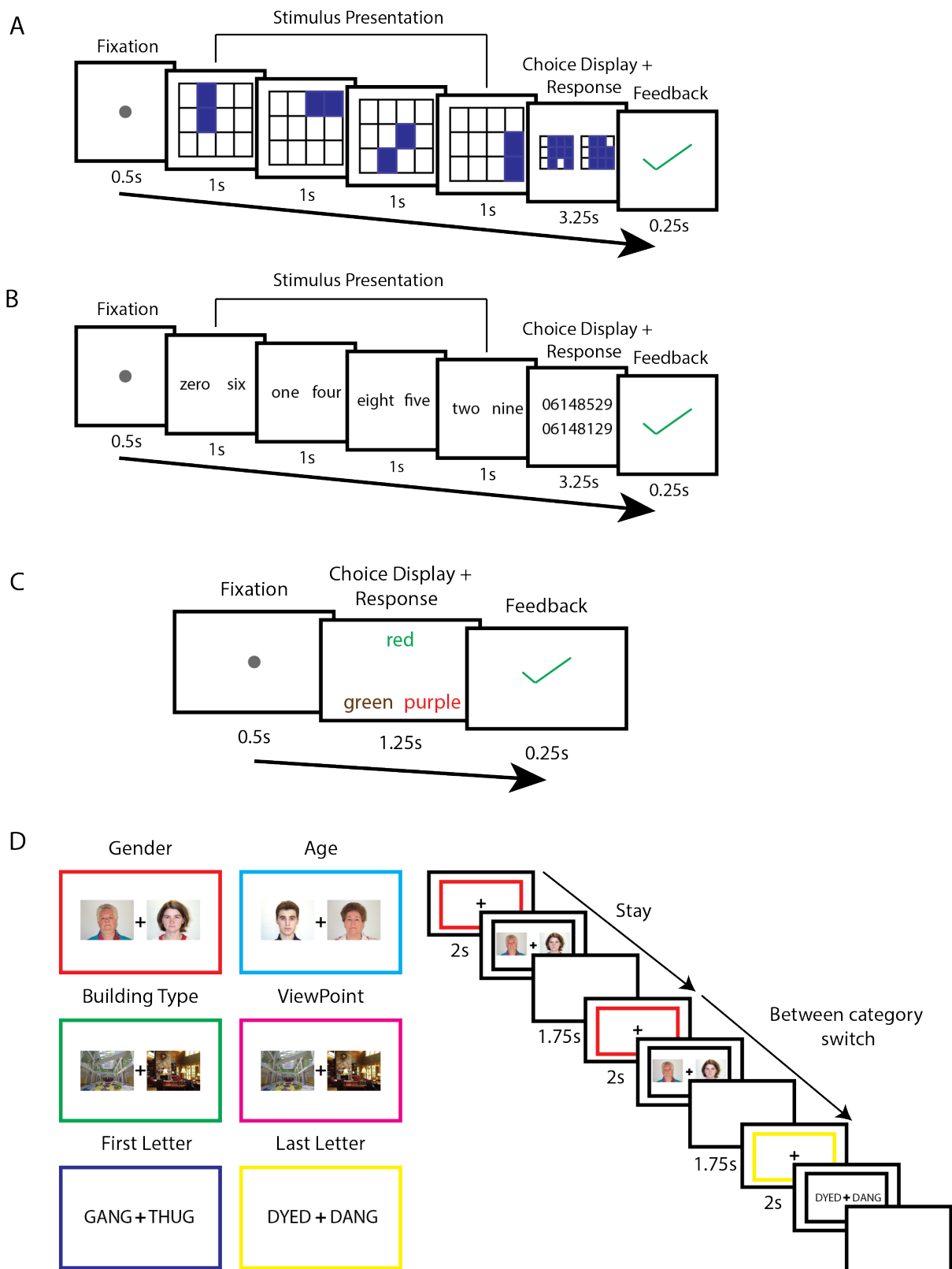


Figure 3.1. Schematic overview of the Hard condition of the three localiser tasks and the rule-based similarity judgement task. **A: Spatial working memory task.** Participants were presented with highlighted cells in a 3x4 grid, on four consecutive screens. In the Easy condition, one cell was highlighted at a time, and in the Hard condition two cells were

highlighted in each screen. They selected the grid with the correctly highlighted cells in a subsequent two-forced choice display. They received feedback after each trial. Positive feedback was indicated by a green tick, and negative by a red cross. **B: Verbal working memory task.** A design similar to the spatial working memory task was used, with written digits instead of a grid. **C: Stroop task.** In each trial, three color names were presented. Participants selected the answer option (out of two options at the bottom) that described the ink color of the test word on top. In the Easy condition, the ink color of the test word matched its color name (congruent), and in the Hard condition they were different (incongruent). In this example of a Hard trial, the correct answer is the word ‘green’ written in brown, and the ink color of the distractor (the word ‘purple’ written in red ink) matches the color name of the test word, thus increasing the difficulty level. See the text (2.3.3) for more details. For all localiser tasks, feedback was presented at the end of each trial. **D: Rule-based similarity judgement task.** The six rules with the corresponding colored frames, paired by the category domain that they should be applied on (left). In each trial, a colored frame indicated the rule, followed by two images. Participants indicated whether the images are the ‘same’ or ‘different’ based on the rule. Transitions between trials could be either the same rule and category domain (‘Stay’), a switch of rule within the same category domain (‘within-category Switch’), or a switch of rule between category domains (‘between-category Switch’). In this example (right), participants indicated whether the two faces have the same gender or not in the first trial, followed by a Stay trial and a between-category Switch trial.

Criterion rule-based similarity judgment task

The rule-based similarity judgement task was a variation of a task previously used by Crittenden and colleagues (2016, 2015) and Smith et al. (2018). This was chosen as the criterion task because it allowed for testing both univariate and multivariate effects. Univariate effects were addressed using a task switching aspect of the task. Multivariate effects were addressed using rule decoding, which has been previously observed across the MD network using this task. Additionally, this task enabled a more detailed investigation of the potential effect of using individual ROIs on the difference between two types of rule discriminations.

Prior to the start of the scanning session, participants learned to associate colored frames with six rules (Figure 3.1D). In each trial, participants indicated whether two displayed images were the same or different based on the given rule. The six rules were applied on stimuli from

three different category domains (faces, buildings, words), with two rules per category. The rules and category domains were: (1) Gender (male/female, red frame) and (2) Age (old/young, light blue frame) applied on faces; (3) Building type (cottage/skyscraper, green frame) and (4) Viewpoint (seen from the outside/inside, magenta frame) applied on buildings; (5) First letter (dark blue frame) and (6) Last letter (yellow frame) applied on words and pseudo-words.

Each trial began with a colored frame (2 s) that indicated the rule to be applied. This was followed by two stimuli presented to the left and right of a fixation cross (Figure 3.1D) and the cue replaced by a black frame. Participants had to respond ‘same’ or ‘different’ based on the rule by pressing the left or right button, and response mapping was counterbalanced across subjects. The task was self-paced and the stimuli were on the screen until a response was made. Each trial was followed by an inter-trial interval (ITI) of 1.75 s. The colored frame was 14.96° visual angle along the width and 11.60° along the height. The choice stimuli were displayed at 3.68° eccentricity from the center and were 6.0° and 4.5° along the width and height.

We used an event-related design, with 12 trials per rule in each run. Out of these, half (6) of the trials had ‘same’ as correct response and half (6 trials) had ‘different’ as correct response. Out of the 6 trials with ‘same’ as correct response, half (3 trials) also had ‘same’ as correct response if the other rule for the category was applied, therefore identical responses using either rule for this category. The other half (3 trials) had ‘different’ as correct response using the other rule in this category, therefore different responses for the two rules. A similar split was used for the 6 trials with ‘different’ as the correct response. To decorrelate the cue and stimulus presentation phases, 4 out of 12 trials of each rule were chosen randomly to be catch trials, in which the colored frame indicating the rule was shown, but was not followed by the stimuli.

The task included switches between rules that were used for the univariate analysis. Trial switches were defined based on the rules in two consecutive trials. In ‘Stay’ trials, the previous trial had the same rule and category domain. For ‘within-category Switch’ trials, the previous trial had the same category domain (faces, buildings, or letters) but a different rule (e.g. age vs. gender for faces). For ‘between-category Switch’ trials, the previous trial had a different category domain and therefore necessarily a different rule. There was an equal number of Stay, within-category Switch and between-category Switch trials (24). The order of the trials was determined pseudo-randomly while balancing the types of switches. Each run included a total of 73 trials, with an extra trial at the beginning of the run that was

required for the balancing of the switch types. This trial was assigned a random rule and was excluded from the analysis.

To address multivariate effects, we used decoding of rule pairs in the task. Each pair of rules out of the six could then be referred to as either ‘within-category’, i.e., applied on the same category domain, or ‘between-category’, i.e., applied on different category domain. The idea being, while all rules may be decoded across the frontoparietal network, the ‘between-category’ rules might be more distinct (i.e. higher decoding levels) than ‘within-category’ rules (Crittenden et al. 2016). To avoid confounding the rule representation with visual information in the task used by Crittenden et al. 2016, we used a variant of this task, with separate cue and stimulus presentation phases in each trial (Smith et al. 2018).

The participants practiced the task prior to the scanning session until they learned the rules. The practice consisted of two parts. During the first part, trials included feedback, while during the second part feedback was omitted. There was no feedback during the scanning session runs of this task. During the scanning session, after completing the localiser tasks, the participants were asked to state the six rules to make sure they remembered the rules before starting the rule-based task. They were also shown the rules again if they requested it. The task was self-paced, with an average duration of each run of 5 minutes and 52 seconds (SD = 29.1 seconds).

fMRI data acquisition

Participants were scanned in a Siemens 3T Prisma MRI scanner with a 32-channel head coil. A T2*-weighted 2D multiband Echo-Planar Imaging, with a multi-band factor of 3, was used to acquire 2 mm isotropic voxels (Feinberg et al., 2010). Other acquisition parameters were: 48 slices, no slice gap, TR = 1.1 s, TE = 30 ms, flip angle = 62° , field of view (FOV) = 205 mm, in plane resolution: 2 x 2 mm. In addition to functional images, T1-weighted 3D multi-echo MPRAGE (van der Kouwe et al., 2008) structural images were obtained (voxel size 1 mm isotropic, TR = 2530 ms, TE = 1.64, 3.5, 5.36, and 7.22 ms, FOV = 256 mm x 256 mm x 192 mm). A single structural image was computed per subject by taking the voxelwise root mean square across the four MPRAGE images that are generated in this sequence.

Data Analysis

Preprocessing

Pre-processing was performed using the automatic analysis (aa) pipelines (Cusack et al. 2014) and SPM12 (Wellcome Trust Centre for Neuroimaging, University College London,

London) for MatLab. The data were first motion corrected by spatially realigning the EPI images. The images were then unwarped using the fieldmaps, slice time corrected and co-registered to the structural T1-weighted image. The structural data were normalized to the Montreal Neurological Institute (MNI) template using nonlinear deformation, after which the transformation matrix was used to normalize the EPI images. For the univariate ROI analyses, the localiser data were pre-processed without smoothing. For the whole-brain random-effects analysis and voxel selection for subject-specific ROIs, the localiser data were smoothed with a Gaussian kernel of 5 mm full-width at half-maximum (FWHM). No smoothing was applied to the data of the rule-based criterion task.

General linear model (GLM)

A general linear model (GLM) was estimated per participant for each localiser task. Regressors were created for the Easy and Hard task blocks and were convolved with a canonical hemodynamic response function (HRF). Run means and movement parameters were used as covariates of no interest. The resulting β -estimates were used to construct the contrast of interest between the Hard and Easy conditions, and the difference in β -estimates ($\Delta\beta$) was used to estimate the activation evoked by each localiser.

For the rule-based criterion task, two different GLMs were used to address univariate and multivariate effects separately. The GLM for the univariate analysis was based on types of switches between trials. A GLM was estimated for each participant using cue phase regressors for the three switch types: Stay, within-category Switches and between-category Switches. Additional stimulus phase regressors were used separately for each switch type, from stimulus onset to response, but these were not analyzed further. A similar GLM was used for multivariate analysis, based on the six rule types instead of switch types. The duration of the regressors was 2 s for all conditions in both models. The regressors were convolved with the canonical HRF and the six movement parameters and run means were included as covariates of no interest.

Localisers activity patterns

We used both random-effects whole-brain and ROI analyses to assess and compare the activation levels and recruitment of the MD network by the localisers. All measures used voxel data of the Hard versus Easy contrast and the resulting difference in beta estimates ($\Delta\beta$), computed across the two runs of each localiser as well as separately for each run

when required and as detailed below for the specific analyses. Additional Hard versus Easy contrasts were computed using combinations of different localisers.

For ROI analysis, we used a template for the MD network, derived from an independent task-based fMRI dataset (Fedorenko et al., 2013, <http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem>). The template is bilateral with an equal number of voxels in each hemisphere for each ROI. ROIs and their respective number of voxels per hemisphere include the anterior insula (AI, 992 voxels), posterior/dorsal lateral frontal cortex (pdLFC, 1132 voxels), intraparietal sulcus (IPS, 4260 voxels), the anterior, middle and posterior middle frontal gyrus (MFG) (621, 712 and 1269 voxels, respectively) and the pre-supplementary motor area (preSMA, 1247 voxels). The group-level ROI analyses were computed using the MarsBaR SPM toolbox (Brett, Anton, et al. 2002a).

Measures to compare activation patterns of the localisers

We conducted several analyses to examine and compare the spread and similarity of activation patterns of the localisers.

Whole-brain spatial spread of activity patterns of localisers across subjects

We conducted a whole-brain analysis to examine the spread of activation patterns across individual participant data. For each voxel we computed the number of participants with significant activations by applying FDR ($p < 0.05$) across all voxels and all participants. This yielded a whole-brain map in which voxel data represents the number of participants with significant activation.

Correlation measures to compare activation patterns of the localisers

To quantify and compare activity patterns of the three localisers, we used Fisher-transformed Pearson correlation of $\Delta\beta$ estimates (Hard – Easy) across voxels. Contrast data were estimated for each run separately and correlations were computed across all voxels in each MD ROI. First, we computed the reliability of activation patterns. For each subject and localiser we correlated the $\Delta\beta$ estimates of the two runs, separately for MD ROIs and then averaged. Second, to compare activity patterns between subjects, we then estimated the similarity in activity patterns between subjects. For each subject, each localiser and each MD ROI we computed the correlation of $\Delta\beta$ estimates between the first run of that subject and the second run of another subject of the same localiser. For each subject, this was computed 17 times, using all other subjects and averaged across them to get between-subject correlation of activity, separately for each of the three localisers. Lastly, to estimate the similarity in

activity patterns between the different localisers, we computed for each subject the correlation between the first runs of each pair of localisers.

Individual MD localization using the group-constrained subject-specific approach

Individual subject ROIs were defined using the group-constrained subject-specific approach. For each localiser, we used the Hard versus Easy contrast data across the two runs to obtain $\Delta\beta$ estimates. Then, for each ROI, the 200 voxels with the largest $\Delta\beta$ estimates were selected. The number of voxels that were selected was defined prior to any data analysis. The selected voxels were then used to test for the effects of subject-specific ROIs and choice of localiser on both the univariate and multivariate activity as measured in the rule-based criterion task. We compared measures of activity using both subject-specific ROIs based on the different localisers and the group template (i.e., using all voxels within each ROI), as well as between localisers.

Univariate analysis of the criterion task

For the univariate activity in the criterion task, we used two contrasts with varying cognitive demand to test for the effect of subject-specific ROIs and choice of localiser across the MD network. For each subject, the GLM cue regressors across all four runs were used to compute two univariate contrasts: within-category Switch versus Stay trials, and between-category Switch versus Stay trials. The results were then averaged across hemispheres and subjects. Similarly to the Hard vs. Easy contrast in the localiser tasks, activity across the MD network is expected to increase with increased demand.

Multivoxel pattern analysis (MVPA)

We used rule decoding in the criterion task to test for the effect of subject-specific ROIs and choice of localiser on multivariate activity. The decoding analysis focused on the task rules during the cue phase, when only colored frames appear on the screen, therefore avoiding any confounds of the subsequent stimuli. Classification accuracy was computed using a support vector machine classifier (LIBSVM library for MATLAB, $c=1$) implemented in the Decoding Toolbox (Hebart et al. 2015). A leave-one-run-out cross-validation was employed to compute pairwise classifications for all task rule combinations (15 in total), and classification accuracy was averaged across all folds for each pair of rules. The average accuracy of all rule pairs, as well as separately for within- and between-category rule pairs, were computed for each subject and ROI. For each subject, classification accuracies were computed using subject-specific ROIs based on the three independent localisers' data

separately, as well as using all voxels within each ROI (i.e. group template). Decoding accuracies above chance (50%) were then averaged across hemispheres and subjects and were tested against zero using one-tailed t-tests.

Since all localisers were chosen because their robust and consistent recruitment of the MD network, combinations of localisers could also be used to define subject-specific ROIs. To test for multivariate results in subject-specific ROIs defined using data combined from two different localisers, we repeated the decoding analysis using contrast data comprised from two runs, one of each localiser. We used the first run of each localiser and created three combination contrasts from pairs of localiser tasks: spatial WM + verbal WM, spatial WM + Stroop, and verbal WM + Stroop. We also used all 6 localiser runs to create a spatial WM + verbal WM + Stroop contrast. The combination contrasts were used to define subject-specific ROIs for the decoding analysis in a similar way to the individual localisers' data.

One advantage of using group-constrained subject-specific ROIs for multivariate analysis in particular is that it allows to keep the number of voxels in each ROI fixed and controlled.

This may be important as it has been previously demonstrated in the visual system that the size of ROI may affect decoding accuracy levels (Eger et al. 2008; Walther et al. 2009; Said et al. 2010). To test for potential effect of ROI size on the decoding results, the MVPA was repeated using a range of ROI sizes (50, 100, 150, 200, 250, 300, 350 and 400 voxels).

To ensure that our decoding results did not depend on the choice of classifier, we repeated the MVPA using a representational similarity analysis (RSA) approach, with linear discriminant contrast (LDC) as a measure of dissimilarity between rule patterns (Nili et al. 2014; Carlin and Kriegeskorte 2017). Cross-validated Mahalanobis distances were calculated for all 15 pairwise rule combinations and averaged to get within- and between-category rule pairs, for each ROI and participant, using all the voxels in the ROI and subject-specific ROIs defined using the different localiser tasks. For each pair of conditions, we used one run as the training set and another run as the testing set. This was done for all pairwise combinations of the 4 runs and LDC values were then averaged across them. Larger LDC values indicate more distinct patterns of the tested conditions, while the LDC value itself is non-indicative for level of discrimination. The choice of using LDC rather than LD-t (associated t-value) meant that we could meaningfully look at differences between distances, and particularly the distinction between within- and between-category rule discriminations in the criterion task. We therefore used the difference of between- and within-category rule pairs compared to 0 as indication for representation of rule information.

Statistical testing and code

To compare performance of the localisers to the group levels as well as to each other, we first used a repeated measures ANOVA as a statistical model with factors as appropriate for each question and as detailed in the Results. The main factor of interest in all these ANOVAs is the localiser factor (with levels for the three localisers and the group template when required), and we report all the results related to this factor and its interactions, but not other interactions that are not of interest for our questions. To directly test for our research questions which may not have been captured by the ANOVA model, we also used separate *t*-tests for each localiser compared to the group template when appropriate, as well as for all possible pairs of localisers, both corrected for multiple comparisons.

We used an alpha level of .05 for all statistical tests. Bonferroni correction for multiple comparisons was used when required, and the corrected *p*-values and uncorrected *t*-values are reported. To quantify the evidence for difference in pattern discriminability when using the group template and subject-specific ROIs defined by functional localisers, we conducted a complementary Bayes factor analysis (Rouder et al. 2009). We used JZS Bayes factor for one-sample *t*-test and $\sqrt{2}/2$ as the Cauchy scale parameter, therefore using medium scaling. The Bayes factor is used to quantify the odds of the alternative hypothesis being more likely than the null hypothesis, thus enables the interpretation of null results. A Bayes factor greater than 3 is considered as some evidence in favor of the alternative hypothesis. All analyses were conducted using custom-made MATLAB (The Mathworks, Inc) scripts, unless otherwise stated.

Results

Behavioural results

The mean accuracies and reaction times (RT) for the Easy and Hard conditions of the spatial WM, verbal WM, and Stroop localiser tasks are listed in Table 1. As expected, there was a significant increase in RT during the Hard compared to the Easy condition for all localisers (two-tailed paired t-test: spatial WM: $t_{17} = 10.03$, $p < 0.001$; verbal WM: $t_{17} = 25.01$, $p < 0.001$; Stroop: $t_{17} = 9.89$, $p < 0.001$), as well as a significant decrease in accuracy (spatial WM: $t_{17} = 8.65$, $p < 0.001$; verbal WM: $t_{17} = 6.95$, $p < 0.001$; Stroop: $t_{17} = 7.47$, $p < 0.001$). These results confirmed that the task manipulation of Easy and Hard conditions worked as intended.

Accuracy levels for the rule-based criterion task were high (mean \pm SD: 95.3% \pm 2.3), indicating that the participants were able to learn the different rules and apply them correctly. The mean accuracies (mean \pm SD): 93.5 \pm 3.4, 95.7 \pm 3.2, 94.1 \pm 4.2, 95.4 \pm 3.3, 95.4 \pm 4.6, 97.7 \pm 3.0, and the mean RTs (mean \pm SD): 1.33 \pm 0.4 s, 1.46 \pm 0.5 s, 1.55 \pm 0.5 s, 1.55 \pm 0.6 s, 1.27 \pm 0.4 s, 1.30 \pm 0.3 s, for age, gender, building type, viewpoint, first letter, and last letter rules, respectively. Since the focus of this study was to compare imaging results across different methods rather than linking them to behavior or making inference about the underlying cognitive construct, we did not analyze the accuracy levels and the RTs for the criterion task any further.

Table 1. RTs and accuracies in each condition. Values are means \pm standard errors.

	Spatial WM		Verbal WM		Stroop	
	Easy	Hard	Easy	Hard	Easy	Hard
RT (sec)	1.15 \pm 0.05	1.63 \pm 0.06	1.02 \pm 0.04	1.91 \pm 0.04	0.78 \pm 0.02	0.90 \pm 0.02
Accuracy (% correct)	94.2 \pm 1.3	71.4 \pm 2.5	96.8 \pm 0.9	79.2 \pm 2.6	94.4 \pm 0.9	68.5 \pm 4.1

Whole brain and ROI univariate analysis

To test for the recruitment of the MD network, a whole-brain random effects analysis was conducted for the Hard versus Easy contrasts of each localiser task (Figure 3.2A). The whole-brain patterns of activity clearly showed recruitment of the MD network by all localisers.

Areas of increased activity included the anterior-posterior axis along the middle frontal gyrus, anterior insula, and the area anterior to the FEF on the lateral surface; preSMA on the medial surface; and IPS on both the lateral and medial surfaces. An additional visual component was observed, as expected from the nature of the tasks. The pattern of activity for the Stroop localiser was sparser, in particular on the right hemisphere.

An ROI analysis further confirmed the recruitment of the MD network with increased task difficulty (Figure 3.2B). All localisers showed a significant increase in activation for the Hard compared to the Easy condition in each of the ROIs with Bonferroni correction (3) for multiple comparisons (one-sample two-tailed t-test against 0: $t_{17} \geq 2.57$, $p \leq 0.02$ for all). A three-way repeated measures ANOVA with factors task (spatial WM, verbal WM, Stroop), ROI (7) and hemisphere (left, right) revealed a main effect of task ($F_{2, 34} = 3.96$, $p = 0.028$). There was a significant main effect of ROIs ($F_{6, 102} = 7.6$, $p < 0.001$), indicating some differences in activity levels between the ROIs. Although the recruitment of the two hemispheres was broadly similar, there was also an effect of hemisphere ($F_{1, 17} = 15.54$, $p = 0.001$), with larger activity on the left than the right hemisphere. There were significant interactions between tasks and ROIs, tasks and hemispheres, ROIs and hemispheres and a three-way interaction ($F > 4.1$ $p < 0.009$). Post-hoc pairwise comparisons with Bonferroni correction (3 comparisons) across ROIs demonstrated a marginal difference in activity between the verbal WM and Stroop task ($t_{17} = 2.64$, $p = 0.051$), but the overall activity did not differ between the spatial WM and verbal WM tasks or the spatial WM and Stroop tasks ($t_{17} \leq 1.25$, $p \geq 0.3$ for both). Overall, all the tasks recruited the MD network, with some differences in the activation patterns across ROIs for the different localisers. Importantly, the univariate results confirmed a significant increase in activity in the MD network with increased task difficulty for all three localisers, as expected and designed.

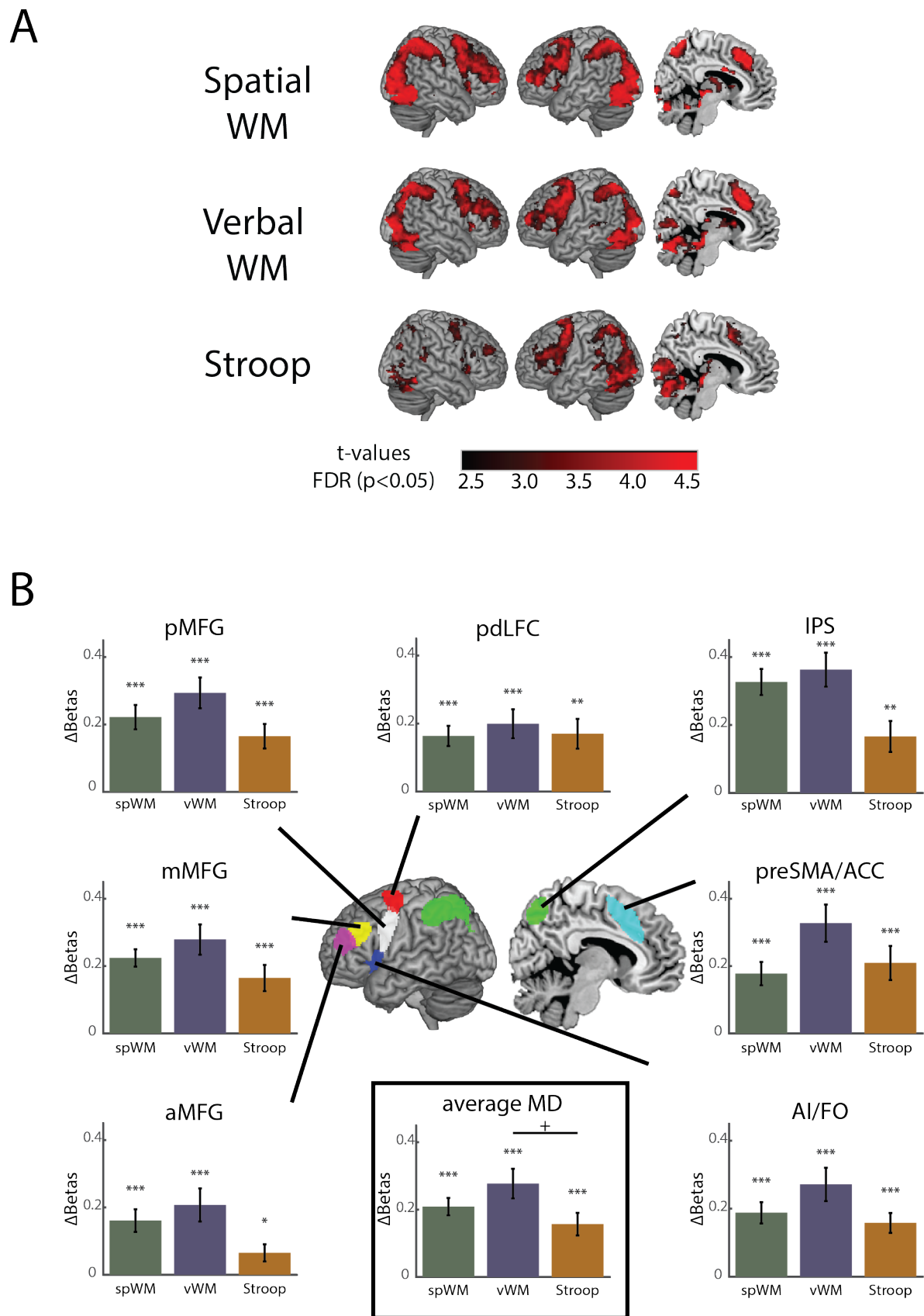


Figure 3.2. Increased activity across the MD network with increased difficulty level for all three localisers. A: Whole-brain t-maps of the contrast between Hard and Easy

conditions for the localiser tasks. WM = working memory. t-maps are FDR corrected, $p < 0.05$. **B:** Univariate results for the contrast between the Hard and Easy conditions for the three localiser tasks, per MD ROI and averaged across ROIs. Significance levels above zero in each ROI and for each localiser (with Bonferroni correction for 3 comparisons) are shown to demonstrate recruitment of the MD network. For the average across MD ROIs, pairwise statistical testing between localisers were also done using a paired two-tailed t-test with Bonferroni correction (3 comparisons). Error bars indicate SEM. The MD template that was used for ROI analysis is shown in the middle for reference (adapted from Fedorenko et al. 2013). spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, + $p < 0.06$.

Comparisons of activity patterns between localisers

We further quantified and compared the variability of activity patterns across subjects and localisers. To test for variability across subjects, two methods were used: A whole-brain overlay of subjects' activation per voxel and correlations of activity across voxels between runs of the same subject versus different subjects. Variability across localisers was tested by correlating the activity across voxels between runs of the same localiser versus different localisers.

Variability of activity patterns across subjects

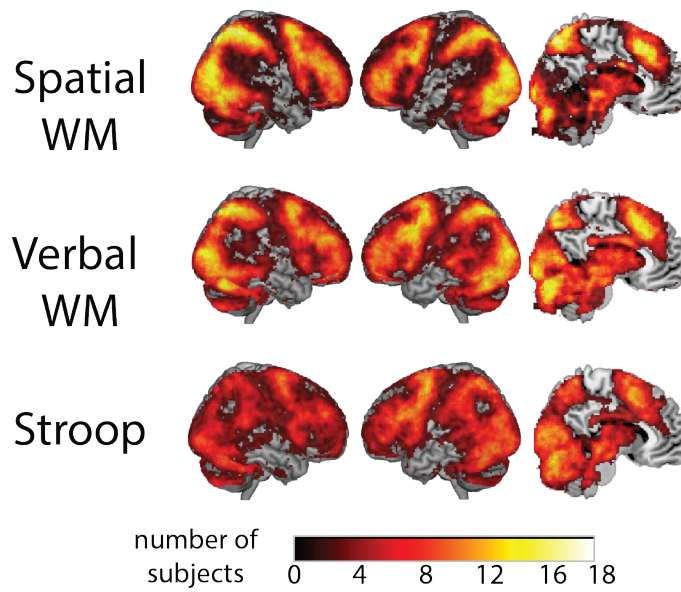
While the group averages of the Hard versus Easy contrasts of the three localisers closely resembled the MD template, there was substantial variability between activation patterns of individual participants. To visualize this spread of activations across individuals, we computed a whole-brain overlay map for each localiser (Figure 3.3A). Each voxel shows the number of participants with significant activations after applying FDR ($p < 0.05$) correction across voxels and subjects. The overlay maps show peaks of activation for many subjects in areas similar to those observed in the group averages. However, it is also clear from these maps that there is substantial variability across subjects with activations that extend to large parts of the frontal and parietal cortices, as well as the visual cortex as expected from the visual nature of the tasks.

Correlations of activity of the different localisers for each subject and between subjects further demonstrated that activity patterns are subject-specific. For each localiser, we computed the reliability of activation as the correlation between runs of the same localiser and same subject using all voxels within each ROI of the MD template. These reliabilities were compared to correlation between two runs from different subjects and same localiser. The correlations were averaged across ROIs to get the correlations of activity across the MD network (Figure 3.3B). The reliability of activation patterns for each subject was high for all localisers (Mean \pm SD: spatial WM: 0.72 ± 0.13 ; verbal WM: 0.75 ± 0.14 ; Stroop: 0.53 ± 0.34), and their similarity was substantially lower for different subjects (Mean \pm SD: spatial WM: 0.19 ± 0.05 ; verbal WM: 0.15 ± 0.04 ; Stroop: 0.08 ± 0.06). A two-way repeated measures ANOVA with task (3: spatial WM, verbal WM and Stroop), and correlation type (2: within- or between-subjects) as factors showed that the within-subject reliabilities were significantly larger than the between-subject correlations ($F_{1, 17} = 322.4, p < 0.001$), providing another support for the variability of activity patterns across subjects. There was also a significant main effect of task ($F_{2, 34} = 8.5, p < 0.001$), with no interaction between the factors ($F_{2, 34} = 2.3, p = 0.11$). Post-hoc pairwise comparisons with Bonferroni correction (3 comparisons) showed that the Stroop task had lower correlations than the spatial WM ($t_{17} = 2.8, p = 0.037$) and verbal WM ($t_{17} = 3.34, p = 0.012$) tasks.

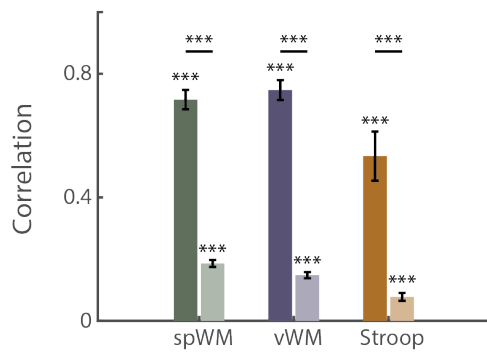
Variability of activity patterns across localisers

To assess the similarity of activity patterns between localisers, we computed the within-subject correlation of activity pattern across voxels between the first runs of pairs of localisers (Figure 3.3C). As expected, there was a substantial positive correlation between the localisers (Mean \pm SD: spatial WM with verbal WM: 0.67 ± 0.36 ; spatial WM with Stroop: 0.27 ± 0.43 ; verbal WM with Stroop: 0.46 ± 0.33), with correlations between the working memory tasks and between verbal WM and Stroop being well above 0 (two-tailed t-test against zero corrected for 3 comparisons: $t > 6, p < 0.001$), and a marginally significant correlation between the spatial WM and Stroop correlation following Bonferroni correction ($t_{17} = 2.65, p = 0.051$). These correlations were significantly smaller than the within-subject within-localiser reliabilities (two-tailed t-test of the average reliabilities across localisers vs. the average correlations of pairs of localisers: $t_{17} = 9.37, p < 0.001$). These reduced correlations demonstrated that although all localisers recruited the MD network, there was some spatial variability in their activation patterns.

A



B



C

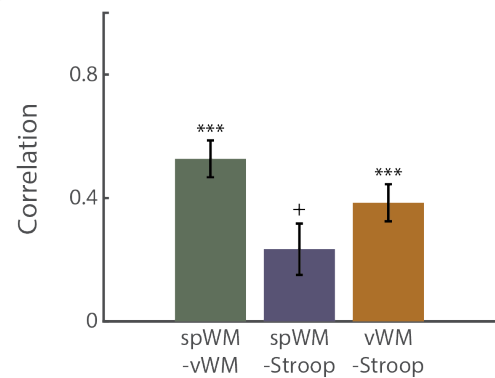


Figure 3.3. Activity pattern are variable across subjects and localisers. A: Activation patterns of single subjects only partially overlap, demonstrating variability across participants. The color of each voxel shows the number of subjects that had significant activation in that voxel in the Hard versus Easy contrast, with the color bar indicating number of subjects up to 18 (sample size), thresholded at 1 subject. **B.** Correlation of activity within runs of the same subject (reliabilities, darker bars) and runs of different subjects (lighter bars) for each localiser, averaged across MD ROIs. **C.** Within-subject correlation of activity patterns between pairs of localisers, averaged across MD ROIs. Pearson correlations are presented, while Fisher transformed correlations were used for statistical inference. Asterisks above bars show significance levels (two-tailed t-test against zero, corrected for 3 comparisons). Asterisks above horizontal lines between bars show significance levels of

differences (paired two-tailed t-test against zero, corrected for 3 comparisons). Error bars indicate SEM. + $p < 0.06$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Subject-specific ROIs and univariate activity in the rule-based criterion task

The main aim of our study was to test for the effect of subject-specific ROIs on univariate and particularly multivariate activity measures in the MD network. We used the rule-based criterion task to extract both univariate and multivariate measures and tested whether using subject-specific ROIs using the independent localisers' data affects the ability to identify activity as expected in the MD network, and whether such changes depend on the choice of localiser.

For the univariate activity, we computed two task switch contrasts in the criterion task: the more demanding between-category Switch versus Stay trials, and the less demanding within-category Switch versus Stay trials. The analysis was done using subject-specific ROIs based on the localisers' data as well as using all voxels within each ROI and averaged across all MD ROIs (Figure 3.4). A three-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels), contrast type (within-category Switch versus Stay, between-category Switch versus Stay) and ROI (7) was set to test for the effect of using subject-specific ROIs and localiser choice on activity levels. There was a main effect of task ($F_{3, 51} = 14.21$, $p < 0.001$) and post-hoc tests with Bonferroni correction for 6 comparisons showed that activity when using the group template was lower than when using subject-specific ROIs, for all localisers ($t_{17} > 3.5$, $p < 0.02$). The activation using subject-specific ROIs based on the spatial WM was lower than that of verbal WM ($t_{17} = 3.40$, $p = 0.02$) and other comparisons between localisers were not significant ($t_{17} > 1.4$, $p > 0.9$). As expected for activity in the MD network, there was a main effect of contrast type ($F_{1, 17} = 12.12$, $p = 0.003$), with activity for the more demanding between-category Switch being larger than the within-category Switch. There was also an interaction between task and contrast type ($F_{3, 51} = 7.92$, $p < 0.001$), but post-hoc tests with Bonferroni correction for 4 comparisons showed that activity for the between-category Switch was larger than for the within-category Switch for all localisers and when the group template was used ($t_{17} > 2.9$, $p < 0.036$). Overall, these results demonstrate that subject-specific ROIs leads to an increase in the observed univariate results, with similar effects for the different localisers.

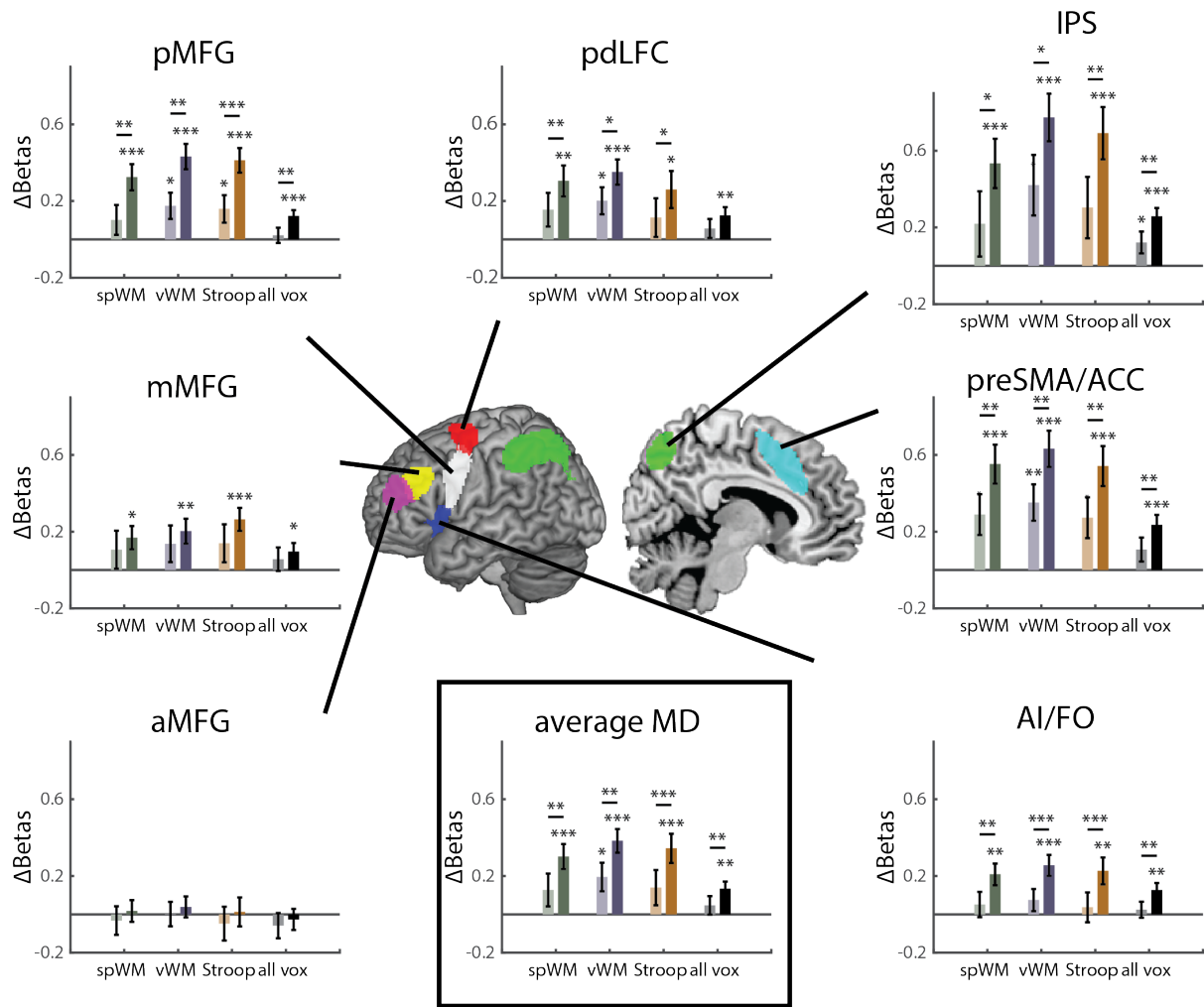


Figure 3.4. Univariate activity in the criterion task increases when using subject-specific ROIs compared to the group template. Within- (lighter bars) and between- (darker bars) category versus Stay trials contrast values using subject-specific ROI defined by the different localiser tasks and using all voxels (group template, no localiser), per ROI and averaged across MD ROIs. Δbeta values were larger when using subject-specific ROIs compared to when using the group template, and similar for all three localisers. Between-category Switch activity was larger than the within-category Switch for all localisers and when using all voxels, as expected for the MD network. spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. Significant activation levels are shown above bars (two-tailed t-test against 0). For visualization, asterisks above horizontal lines between bars show significance levels of differences (paired two-tailed t-test against zero). Error bars indicate SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Subject-specific ROIs and rule decoding in the rule-based criterion task

We next tested for the effect of subject-specific ROIs on decoding levels in the rule-based criterion task across the MD network. We computed the decoding accuracy of all pairwise discriminations between rules when using each localisers' data to select the 200 most responsive voxels within each ROI for each subject and when using all the voxels in each ROI. Overall decoding accuracies above chance (50%) across all MD ROIs and rule-pairs were 2.85 ± 14.53 , 2.71 ± 14.75 , 3.17 ± 15.04 , 5.08 ± 15.34 for subject-specific ROIs defined using the spatial WM, verbal WM and Stroop localisers, and when using the group template, respectively (mean \pm SD). These decoding levels were similar to previous studies that used a similar experimental paradigm (Crittenden et al. 2015; Smith et al. 2018). We first tested whether overall decoding accuracy across all pairs of rules differed between the localisers and compared to when all voxels within each ROI were used. A two-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels) and ROI (7) showed a main effect of task ($F_{3, 51} = 2.9$, $p = 0.042$). However, none of the post-hoc tests to compare pairs of tasks survived correction (Bonferroni correction for 6 comparisons, $t_{17} < 2.7$, $p > 0.09$). There was a main effect of ROI ($F_{6, 102} = 2.5$, $p = 0.025$) but no interaction with task ($F_{18, 306} = 0.8$, $p = 0.7$). Overall, decoding across all pairs of conditions was similar for all localisers and when the group template was used.

Our choice of criterion task has enabled us to not only test for the effect of localiser type on the overall discrimination between rules, which might be too coarse to depict, but also for the relative decoding levels of the two types of discriminations, thus potentially picking up more subtle effects. The criterion task included discriminations between rules applied on the same category (within-category discriminations, e.g., between the gender and age rules, both applied on the faces category) and discriminations between rules applied on different categories (between-category discriminations, e.g., between the gender rule applied on the faces category and the viewpoint rule applied on the building category). To get this more fine-grained picture of the effect of subject-specific ROIs using localiser data on rule decoding, as well as when the groups template is used, we split the decoding accuracy to within- and between-category rule discriminations (Figure 3.5). A three-way repeated measures ANOVA with task (spatial WM, verbal WM, Stroop, all-voxels), distinction type (within-category, between-category) and ROI (7) as within-subject factors was set to test for differences between localisers related to decoding of the two distinction types. There was no main effect of task ($F_{3, 51} = 1.0$, $p = 0.4$), and no interaction between distinction type and task

($F_{3, 51} = 0.9, p = 0.4$). To test for differences between each of the localisers and the group template as well as pairs of localisers based on our pre-defined questions, additional t-tests between all pairs of tasks were conducted, but no significant differences were found ($t_{17} < 1.8, p > 0.5$, corrected for 6 comparisons). There was no main effect of distinction type ($F_{1, 17} = 3.2, p = 0.089$), though a numerical trend was consistent with the previously reported results (Crittenden et al. 2016). There was no main effect of ROI or interaction of ROI with distinction type or task ($F < 1.8, p = 0.09$). Taken together, this indicates that decoding results in the criterion task were similar for all localisers as well as when the group template was used, similar to the results obtained across all pairs of conditions.

Based on these ANOVA results and to further establish that using localiser data for subject-specific ROIs did not change decoding levels, we conducted a Bayes factor analysis (Rouder et al. 2009), separately for each functional localiser compared to decoding with all voxels. First, the difference in classification accuracy between the between- and within-category distinctions averaged across ROIs for each functional localiser was compared to the classification accuracies with all voxels using a paired two-tailed t-test. The t-value was then entered into a one-sample Bayes factor analysis with a Cauchy scale parameter of 0.7. The Bayes factors for the spatial WM, verbal WM and Stroop localiser when compared to the decoding levels with all voxels were 0.40, 0.56 and 0.79, respectively. These results demonstrate little evidence for difference in decoding levels when subject-specific ROIs and the group template are used.

Individual-subject localization of the MD network for MVPA can be done not only using one localiser task, but also using two runs of two different localiser, thus benefitting from the activation patterns evoked by both localiser to more robustly identify MD-like voxels. To test for the effect of such an approach on decoding results, we repeated the analysis using subject-specific ROIs based on localiser data combined from two different localisers (spatial WM + verbal WM, spatial WM + Stroop, verbal WM + Stroop). A three-way repeated measures ANOVA with task (4: three combination contrasts and all-voxels), distinction type (within/between category) and ROI (7) as within-subject factors did not show a main effect of task ($F_{3, 51} = 2.5, p = 0.07$), similarly to the results when individual localisers were used to define subject-specific ROI. There was no effect of distinction type ($F_{1, 17} = 3.3, p = 0.08$), despite a numerical trend, and there was no interaction of task and distinction type ($F_{3, 51} = 1.4, p = 0.25$). Similarly, subject-specific ROI were also defined using all localisers with one contrast using all 6 localiser runs (spatial WM + verbal WM + Stroop). A three-way repeated

measures ANOVA with task (2: one combination contrast and all-voxels), distinction type (within/between category) and ROI (7) as within-subject factors showed no main effect of task or distinction type or an interaction of the two ($F < 3.2, p > 0.05$). Overall, these results indicate that using combinations of localiser tasks to define subject-specific ROIs yielded decoding results similar to the ones obtained when using the group template, at least with the range of localisers that we used here.

To ensure that our results are robust and not limited to the choice of classifier, we repeated the analysis using a representational similarity analysis (RSA) approach (Kriegeskorte et al. 2008; Nili et al. 2014) and linear discriminant contrasts (LDC). An LDC value between pair of conditions indicates the level of their discriminability, with larger values meaning better discrimination. The difference between the average LDC of all the between-category rule pairs and all the within-category rule pairs (Δ LDC) was calculated per participant and per ROI, for both subject-specific ROIs and group template, with Δ LDC larger than 0 as an indication for rule information. For all localisers as well as when all voxels within the group template were used, the Δ LDC was greater than zero (two tailed: $t_{17} > 3.3, p < 0.016$, with Bonferroni correction for multiple (4) comparisons), indicating that the distributed patterns of activity conveyed rule information. This was comparable to the trend of distinction type effect seen in the SVM analysis. Importantly, a two-way repeated measures ANOVA with Δ LDC as dependent variable and task (spatial WM, verbal WM, Stroop, all-voxels) and ROI (7) as within-subject factors showed no main effect of task ($F_{3, 51} = 0.3, p = 0.7$) and no interaction ($F_{18, 306} = 1.1, p = 0.35$), further supported by individual t-tests of each localiser compared to the group template as well as all possible pairs of localisers ($t_{17} < 0.8, p > 0.8$, Bonferroni corrected for 6 comparisons). These results indicate that discriminability was similar when using the three localisers to define ROIs in individual subjects and when using all voxels within the group template, in line with the SVM decoding results.

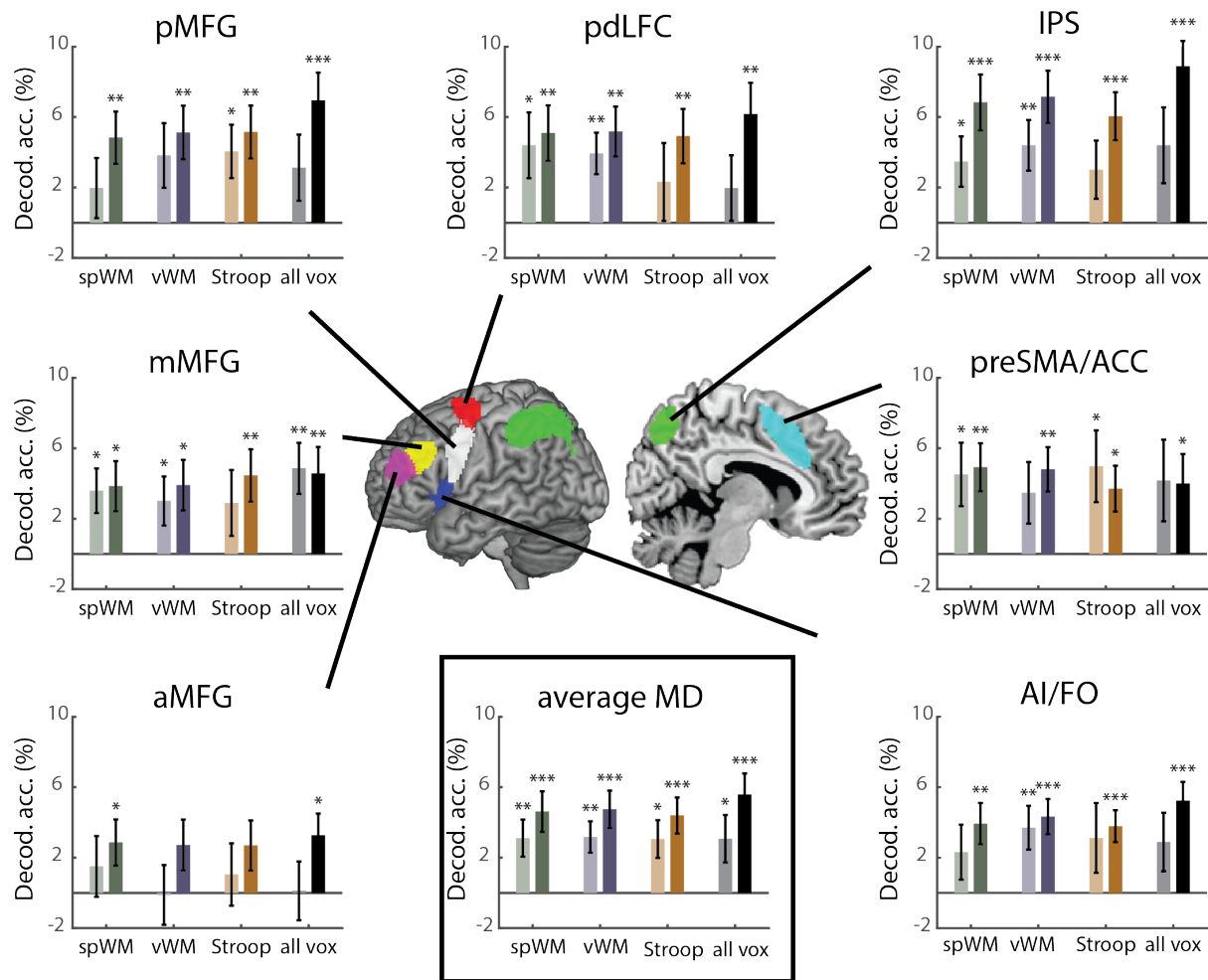


Figure 3.5. Decoding accuracy for within- and between-category rule pairs. Within- (lighter bars) and between- (darker bars) category rule decoding accuracy values above chance (50%) for subject-specific ROIs defined using the different localiser tasks and using all voxels (group template), per ROI and averaged across MD ROIs. Decoding accuracies were similar for all three localisers and when using all voxels within each template ROI, with similar differences between within- and between-category rule decoding (see Text (3.5) for statistical details). spWM = spatial working memory, vWM = verbal working memory. pdLFC = posterior/dorsal lateral prefrontal cortex, IPS = intraparietal sulcus, preSMA = pre-supplementary motor area, AI = anterior insula, aMFG = anterior middle frontal gyrus, mMFG = middle frontal gyrus, pMFG = posterior middle frontal gyrus. To demonstrate rule decoding, significant decoding accuracy above chance (50%) is shown (one-tailed t-test against 0). Error bars indicate SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Effect of ROI size on decoding results

In order to examine whether decoding results depend on the number of selected voxels in subject-specific ROIs, as has been observed in the visual system, we performed MVPA for the decoding accuracies across all MD ROIs using different ROI sizes. ROI sizes ranged from 50 to 400, in steps of 50. A three-way repeated measures ANOVA was done with factors: task (spatial WM, verbal WM, Stroop), ROI size (8) and distinction type (2: within- and between-category). There was a main effect of ROI size ($F_{7, 119} = 11.6, p < 0.001$), no main effect of task ($F_{2, 34} = 0.1, p = 0.9$) and no interaction between ROI size and task ($F_{14, 238} = 0.7, p = 0.7$), with the latter indicating that the difference between ROI sizes was the same for the different localisers. We then pooled the data across the three localisers and rule distinctions in order to visualize the main effect of ROI size (Figure 3.6). Classification accuracies tended to be lower for the smaller ROI sizes, and particularly for 50 and 100 voxels, but overall decoding levels were stable with similar decoding accuracies for ROI size of 150 voxels and above. For all ROI sizes, classification accuracy was above chance ($t > 3.4, p < 0.003$).

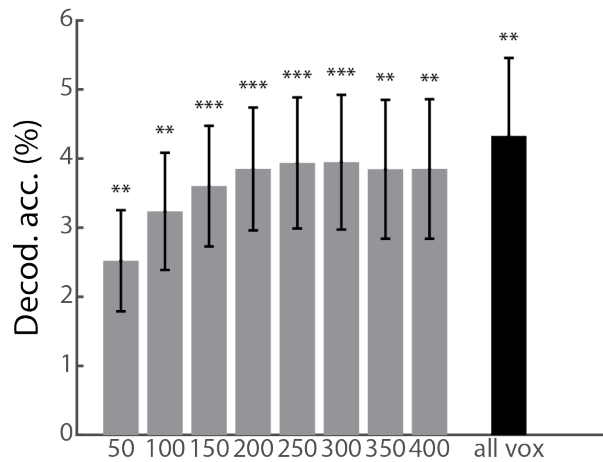


Figure 6. Decoding accuracy for rule distinction for different ROI sizes. Decoding accuracy above chance (50%) is presented for average of all the localisers, MD ROIs and distinction types. The decoding accuracy level using all the voxels in the MD ROIs (group template) is shown for reference at the rightmost bar. Error bars indicate SEM. ** = $p < 0.01$, *** = $p < 0.001$.

Discussion

In this study we tested for the effect of using individually-defined ROIs within the cognitive control frontoparietal MD network compared to a group template on both univariate and multivariate results in a rule-based criterion task. We systematically tested three localiser tasks (spatial WM, verbal WM, and Stroop) and used a group-constrained subject-specific approach to define ROIs at the single subject level using a conjunction of the independent localiser data and a group mask (Fedorenko et al. 2010). The primary benefit of the proposed approach is the use of both a group template that ensures spatial consistency across participants, as well as individual-subject activation patterns within the group template that provide more focused targeting of MD voxels. We showed a clear benefit for using the individual ROIs compared to when using the group template for univariate contrasts of activity. For multivariate discriminability measures, however, the results were similar for the individually-defined ROIs and for the group template, with no clear benefit, or cost, for the subject-specific ROIs approach. Despite differences between the localisers in their spatial activation patterns of the MD network at the individual subject level, we observed similar performance for both univariate and multivariate measures for all three localisers, demonstrating that the choice of localiser did not make a difference to the obtained results. Overall, our results demonstrate that using individually-defined ROIs is a useful way to maintain, or even increase in the case of univariate activity, the sensitivity of broad group templates with an added specificity of activations at the individual subject level.

Our univariate results show that the observed activity is larger when ROIs are defined at the individual level compared to using a group template. These findings replicate and generalize previous findings in other systems such as language and vision, as well as in simulated data (Saxe et al. 2006; Fedorenko et al. 2010; Nieto-Castañón and Fedorenko 2012). When using the group template, many voxels outside the subject-specific functional signature of the studied region at the individual level are included in the analysis. These voxels average-out the overall activity level, reducing its sensitivity to detect actual changes (Nieto-Castañón and Fedorenko 2012). Previous studies showed that boundaries between parcellated networks varied across individuals (Yeo et al. 2011; Glasser et al. 2016; Schaefer et al. 2018), further supporting the idea that overall average of activity within a group template may not capture changes that can be observed when ROIs are localized in individual subjects.

Our main focus in this study was the multivariate results in the MD network and whether individually-localized ROIs will result in a benefit, as seen in pattern discriminability

measures, similar to the one observed in the univariate results. Most studies that used MVPA for fMRI data across the frontoparietal network used a group template as ROIs, implemented in a variety of ways. These include, among others, using all voxels within the functionally-defined group-average MD ROIs (Woolgar, Hampshire, et al. 2011; Muhle-Karbe et al. 2017), defining areas of interest based on univariate or multivariate effects of part of the data of the main task and testing on another (Ester et al. 2015; Etzel et al. 2016), centering spheres on peak activation loci (Fox, Snyder, Barch, et al. 2005), resting-state networks (Cole et al. 2013), searchlight algorithm (Cole et al. 2016), and anatomical landmarks in conjunction with group-level univariate contrast (Curtis et al. 2005). Because boundaries between specialized areas and networks vary between individuals (Yeo et al. 2011; Glasser et al. 2016; Schaefer et al. 2018), and specifically for the MD network (Fedorenko et al. 2012), the individual ROIs approach has the potential to improve our ability to reliably discriminate patterns of activity within this network. Our results, however, showed that this is not the case. In contrast to the univariate results, pattern discriminability was largely similar when using the individual ROIs and the group template. Our criterion task was chosen to allow for more subtle distinctions between between-category and within-category rule pairs that have been previously observed across the MD network (Crittenden et al. 2016). In the task used by Crittenden et al. (2016), the stimuli were presented at the same time as the colored frames, therefore the between-category and within-category rule pairs were confounded with the category of the stimuli. To avoid this confound, a later study used a variation of the task presenting only colored frames first, and the stimuli later without the cue (Smith et al. 2018). Using this cue and stimulus separation resulted in no difference between the between- and within-category rule pairs, despite strong decoding overall across the MD network. We investigated whether the subject-specific ROIs might show more subtle distinctions between rule pairs. However, the differences between the between-category and within-category rule pairs were similar when using the individual ROIs and the group template, with no benefit for the subject-specific ROIs. Importantly, using the individual ROIs for MVPA did not lead to reduced discriminability. The preserved accuracy levels demonstrated that the increased localization at the individual subject level did not come at a cost of reduced decodability and maintained the sensitivity to detect task-related neural representations. We note that these results do not depend on the choice of classifier. To ensure the robustness of our results, we used linear discriminant contrast (LDC) in addition to SVM to measure discriminability and differences between localisers and observed similar results. Therefore, the similar

multivariate performance when using individual ROIs and the group template was independent of the choice of the MVPA method.

One possible explanation for the similarity of multivariate results obtained for the individual ROIs and the group template is that voxelwise distributed pattern across the entire MD template captured the information related to rule decoding well, in the criterion task that we used, with no need for further refinement of the voxels selected for this decoding. Previous studies indeed showed that multivoxel discrimination may also be driven by voxels outside the focused regions of increased univariate activity (Haxby et al. 2001b; Kriegeskorte et al. 2006). Another related point is the relatively high decoding accuracies that we observed when using the group template. Decoding levels were at a level similar to what has been shown to be the base rate for the frontoparietal network (Bhandari et al. 2018), therefore potentially limiting our ability to identify increases in decoding levels. Notably, an important limitation of our data, is the use of only one criterion task due to limited time in the scanner. It is possible that multivariate benefits or costs of the individual ROIs will be observed for other tasks, and future studies will be required to generalize our results in that respect. We note, that since our analysis focused on the fixed-duration cue phase of the trials, our results are well controlled and not driven by behavioral responses such as reaction times. More generally, the underlying factors that contribute to pattern discriminability in the frontoparietal network are not yet well understood (Bhandari et al. 2018), with some previous data showing clear limitations of fMRI decodability compared to what is observed in single-unit data in other brain systems (Dubois et al. 2015). Our data provide another tier of evidence to better understand the relationship between the spatial organization and activity of the MD network at the micro and macro levels and pattern decodability using fMRI.

In this study, we systematically tested three localisers: a spatial WM, verbal WM and a Stroop task. We used tasks that capture a core cognitive aspect associated with the MD network and have the potential to be used as general functional localisers. As expected, and in line with previous results (Fedorenko et al. 2013), all three localisers showed increased activity in the MD network, thus confirming their suitability to serve as localiser tasks. Activation patterns of the localisers in individual subjects were highly reliable, as reflected in high correlations across voxels between the two runs of each localiser. In contrast, these correlations were substantially reduced across subjects, as is also shown in the whole-brain overlay maps, demonstrating the need for a subject-specific ROI definition approach. In line with previous data (Fedorenko et al. 2013), there were substantial correlations between activation patterns of the different localisers (computed within each subject), providing

further support these correlations were lower than those between runs of the same localiser. The Stroop task evoked weaker, and less MD-focused pattern of activity, and had lower reliability between runs. These differences compared to the other two localisers were small and in some cases only marginally significant, but may imply that the Stroop task captured the MD network less well. The spatial differences in recruitment between the localisers may reflect differential functional preferences for cognitive demands and constructs across MD regions (Dosenbach et al. 2007; Nomura et al. 2010; Crittenden et al. 2016; Assem et al. 2019; Shashidhara, Mitchell, et al. 2019). Specifically, the two working memory tasks might reflect a more similar cognitive construct compared to the Stroop task that involves conflict monitoring and inhibition. Another possible explanation for the difference could be related to the difficulty manipulation in the tasks. While increase in difficulty level in the working memory tasks was simply controlled by increasing the number of highlighted cells in the grid or numbers, this manipulation in the Stroop task was operationally less well defined.

An important aspect of the individual ROI approach that we used is the ability to control for the ROI size. It has been previously shown that increased ROI size leads to increased classification levels in the visual system, highlighting the need to control for ROI size when comparing results across ROIs (Eger et al. 2008; Walther et al. 2009; Said et al. 2010; Erez and Yovel 2014). It was not clear whether this is the case for MD regions, which are different from visual regions in multiple respects, and whether more generally the choice of this parameter affects decoding levels. In our data (Figure 6), we observed slightly lower levels of classification for the smaller ROI sizes (50 and 100 voxels), but these stabilized for ROI sizes of 150 voxels or more, in line with previous reports (Erez and Duncan 2015; Shashidhara and Erez 2019). This does not necessarily mean that such high dimensionality is required to reach maximal decodability without over-fitting, an issue that other studies have looked into more formally (Ahlheim and Love 2018). Importantly, controlling for ROI size may be essential when comparing MD regions to each other as well as to other brain systems, such as visual areas, further emphasizing the importance of using a method that enables such control.

Different variations of the individual ROI approach that we used here can be designed for future studies, offering a balance between the need of a consistent definition of regions across participants, and perhaps studies, and the localization at the individual participant level. Such variations can be designed depending on the research question, and can be used with both univariate and multivariate analyses while avoiding double-dipping (Kriegeskorte et al. 2009). For example, data from two or more localiser tasks can be combined, as we demonstrated here. Combining data across localisers could lead to capturing core parts of the

MD system, thus reflecting the multiple-demand nature of the selected voxels (Duncan 2010b, 2013b; Fedorenko et al. 2013; Assem et al. 2019). An even more cognitively diverse variation can be a localiser that consists of multiple tasks within the same run with a similar manipulation of difficulty level. On the other end of this scale, a localiser task can be designed to target a specific cognitive aspect of interest, and constraining activation patterns by a group template will ensure that the areas of interest are within the boundaries of the MD network.

In summary, we used three independent localiser tasks to define subject-specific ROIs and test the effect of using the individual ROIs compared to a group template on univariate and multivariate effects in a rule-based criterion task. The univariate results in the criterion task greatly benefitted from using individual ROIs compared to using the group template. In contrast, multivoxel task-related representations did not vary with localisers were similar for the subject-specific ROIs and the group template, and for all localisers, with no benefit of increased pattern discriminability, as well as no cost of reduced discriminability. The group-constrained individually-defined ROIs offer a refined and targeted localization of each participant's MD regions based on the individual's unique functional pattern of activity, while ensuring that similar brain regions are studied in all participants. Pushing forward towards standardization in the field, our study provides important empirical evidence for researchers using both univariate and multivariate analysis of fMRI data to study the functional organization of the MD network.

Chapter 4

Reward motivation does not modulate coding of behaviourally relevant category distinctions across the frontoparietal cortex

Introduction

A fundamental aspect of flexible goal-directed behavior is the selection and integration of information depending on a current goal to determine its relevance to behavior and lead to a decision. In non-human primates, single-cell data from the lateral prefrontal cortex, as well as parietal cortex, provide detailed evidence for the coding of task-relevant information. It has been shown that neural activity contains information about the context, also referred to as cue or task-set, as well as the integrated information of cue and a subsequent input stimulus, such as task-related categorical and behavioral decision (Freedman et al. 2001; Wallis et al. 2001; Kusunoki et al. 2010; Kadohisa et al. 2013; Mante et al. 2013; Stokes et al. 2013). In the human brain, a network of frontal and parietal cortical regions, the ‘multiple-demand’ (MD) network (Fedorenko et al. 2013; Mitchell et al. 2016), has been shown to be involved in information selection and integration, and more generally in control processes. This network is associated with multiple aspects of cognitive control, such as spatial and verbal working memory, math, conflict monitoring, rule-guided categorization and task switching, (Fedorenko et al. 2013; Vergauwe and Cowan 2015; Cole et al. 2016). The MD network spans the anterior-posterior axis of the middle frontal gyrus (MFG); posterior dorso-lateral frontal cortex (pdLFC); the anterior insula and frontal operculum (AI/FO); the pre-supplementary motor area and the adjacent dorsal anterior cingulate cortex (preSMA/ACC); and intraparietal sulcus (IPS) (Duncan 2010). Multiple neuroimaging studies demonstrated that distributed patterns of activity across the MD network measured by functional magnetic resonance imaging (fMRI) reflected a variety of task-related information. These include task sets, behavioral relevance and task-dependent categorical decisions (Li et al. 2007; Woolgar, Hampshire, et al. 2011; Woolgar, Thompson, et al. 2011; Erez and Duncan 2015; Woolgar et al. 2015; Wisniewski et al. 2016; Muhle-Karbe et al. 2017). In contrast, sensory areas such as the high-level general object visual region, the lateral occipital complex (LOC), as well as the primary visual cortex, contain information about the visual properties and categorization of

stimuli, with weaker, or non-existing, task effects (Harel et al. 2014; Bugatus et al. 2017; Hebart et al. 2018).

With growing interest in recent years in the link between cognitive control and motivation, it has been proposed that motivation enhances control processes by sharpening representation of task goals and prioritizing task-relevant information across the frontoparietal network and other regions associated with cognitive control (Simon 1967; Kruglanski et al. 2002; Botvinick and Braver 2015; Etzel et al. 2016). In line with this idea, it has been shown that motivation, usually manipulated as monetary reward, increases task performance (Padmala and Pessoa 2010, 2011). Neuroimaging studies linked increased activity with reward in frontoparietal regions across a range of tasks, including working memory (Pochon et al. 2002; Taylor et al. 2004), selective attention (Mohanty et al. 2008; Krebs et al. 2012), response inhibition (Padmala and Pessoa 2011), and problem solving (Shashidhara, Mitchell, et al. 2019).

Although the accumulating evidence at the behavioral and neural level in humans are consistent with this sharpening and prioritizing account (Wallace 1960; Simon 1967; Kruglanski et al. 2002; Pessoa 2009; Braver 2012; Chiew and Braver 2014), they do not directly address the effect of motivation on the coding of task-related information and selection and integration processes. Some support for this idea comes from single-neuron data recorded from the prefrontal cortex of non-human primates: reward was associated with greater spatial selectivity, enhanced activity related to working memory and modulated task-related activity based on the type of reward (Watanabe 1996; Leon and Shadlen 1999; Kennerley and Wallis 2009). A more direct evidence in humans was recently demonstrated by Etzel et al. (2016). They showed that reward enhances coding of task cues across the frontoparietal cortex, and suggested that task-set efficacy increases with reward. Subsequently, Hall-McMaster et al. (2019) used electroencephalogram (EEG) and demonstrated that this effect of reward on the coding of task cues was particularly evident when a switch in context was required (Hall-McMaster et al. 2019). They also provided some evidence that the representation of features relevant for a given task is enhanced when the reward level is high. However, since their results were obtained using EEG, the spatial specificity of such reward effects are limited. Therefore, it remains unclear whether the previously reported facilitative effect of reward on representation across the frontoparietal cortex is limited to preparatory cues, or whether reward also enhances the coding of

behaviorally relevant information, when the cue and a subsequent stimulus are integrated, leading to the behavioral decision thus supporting goal-directed flexible behavior.

Following the sharpening hypothesis, in this study we asked whether reward motivation enhances the representation of behaviorally relevant information, as determined by the integration of cue and stimulus input. Furthermore, previous studies have associated reward with decreased conflict in interference tasks (Padmala and Pessoa 2011; Stürmer et al. 2011; Krebs et al. 2013), suggesting that any effect of reward may be particularly important for high-conflict items, in other words, a conflict-contingent effect. We therefore also asked whether such facilitative effect of reward is selective for highly conflicting items. We recently showed that behaviorally relevant, but not irrelevant, category distinctions of objects were coded across the MD network (Erez and Duncan 2015). In contrast, such differences were not observed in the LOC. Here, we used a similar cued detection categorization task while participants' brain activity was measured using fMRI. Participants detected whether an object from a cued visual category (target category) was present or absent. On each trial, one of two categories was cued, and objects from those two categories could be either Targets, or nontargets with high behavioral conflict, as they could be targets on other trials (High-conflict nontarget). An additional category was never cued, serving as nontarget with low behavioral conflict (Low-conflict nontarget). This design created three levels of behavioral status (Targets, High-conflict nontargets, Low-conflict nontargets). Critically, following this integration process, the relevant information that is expected to be represented across the MD network is the behavioral status of a given category, rather than the visual category itself (Erez and Duncan 2015). Therefore, the behaviorally relevant category distinctions were pairs of categories with different behavioral status. We used multivariate pattern analysis (MVPA) to measure representation of the behaviorally relevant category distinctions as reflected in distributed patterns of response in the *a priori* defined MD network. To manipulate motivation, on half of all trials a substantial monetary reward was offered. We tested whether the neural pattern discriminability between the behaviorally relevant category distinctions increased with reward, and whether this effect was selective for the distinction between Targets and High-conflict nontargets. A common view posits that top-down signals from the frontoparietal MD network to the visual cortex play an important role in the processing of task-related information. Therefore, to test for the effect of reward motivation on the representation of the behavioral status distinctions in the visual cortex, we conducted

similar analyses in the high-level general object visual region, the lateral occipital complex (LOC).

Materials and Methods

Participants

24 participants (13 females), between the ages of 18-40 years (mean age: 25) took part in the study. Four additional participants were excluded due to large head movements during the scan (greater than 5 mm). The sample size was determined prior to data collection, as typical for neuroimaging studies and in accordance with counter-balancing requirements of the experimental design across participants. A similar sample size showed sufficient power to detect representation of behavioral status in a previous study (Erez and Duncan 2015). All participants were right handed with normal or corrected-to-normal vision and had no history of neurological or psychiatric illness. The study was conducted with approval by the Cambridge Psychology Research Ethics Committee. All participants gave written informed consent and were monetarily reimbursed for their time.

Task Design

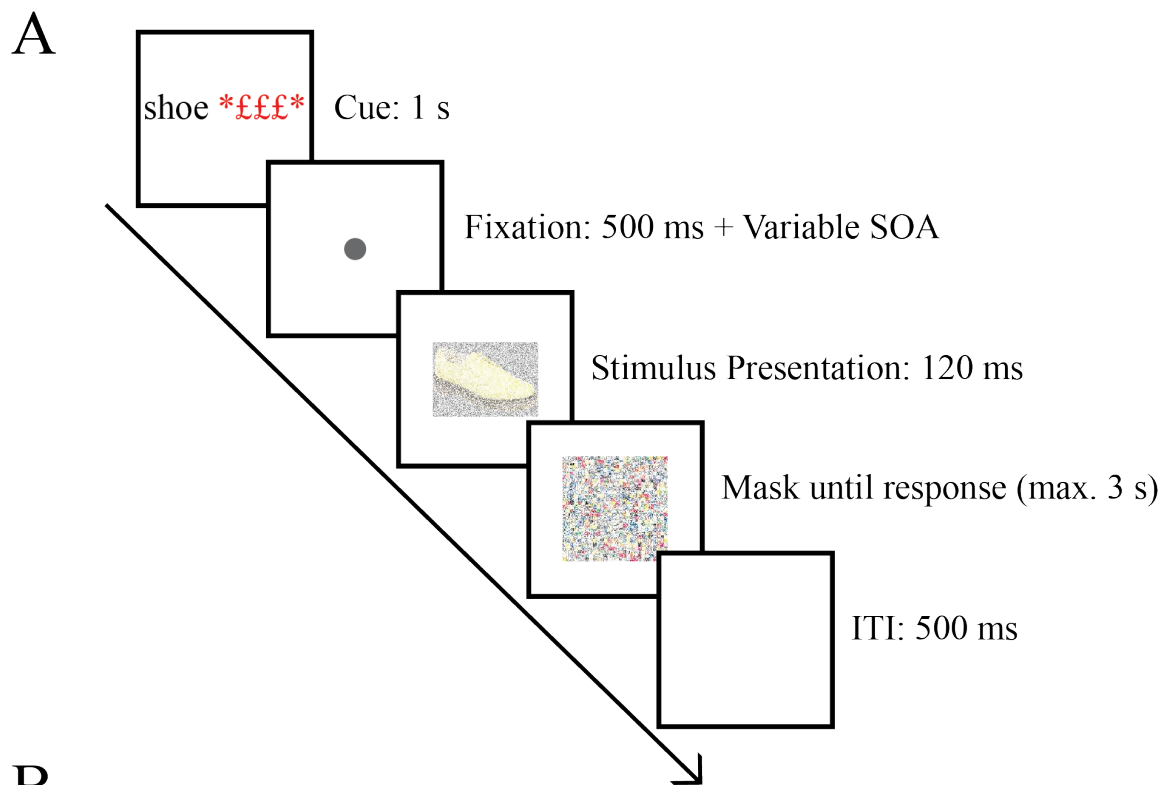
Participants performed a cued categorization task in the MRI scanner (Figure 4.1A). Our primary question concerned the representation during the stimulus epoch of a trial where cue and stimulus are integrated, and we therefore designed the task accordingly. At the beginning of each trial, one of three visual categories (sofas, shoes, cars) was cued, determining the target category for that trial. Participants had to indicate whether the subsequent object matched this category or not by pressing a button. For each participant, only two of the categories were cued as targets throughout the experiment. Depending on the cue on a given trial, objects from these categories could be either Targets (T), or nontargets with high conflict (as they could serve as targets on other trials). The third category was never cued, therefore objects from this category served as Low-conflict nontargets. This design yielded three behavioral status conditions: Targets, High-conflict nontargets and Low-conflict nontargets (Figure 4.1B). The assignment of the categories to be cued (and therefore serve as either Targets or High-conflict nontargets) or not (and serve as Low-conflict nontargets) was counter-balanced across participants.

To manipulate motivation, half of the trials were cued as reward trials, in which participants had the chance of earning £1 if they completed the trial correctly and within a time limit. To assure the incentive on each reward trial, four random reward trials out of 40 in each run were assigned the £1 reward. To avoid longer reaction times when participants try to maximize their reward, a response time threshold was used for reward trials, set separately

for each participant as the average of 32 trials in a pre-scan session. The participants were told that the maximum reward they could earn is £24 in the entire session (£4 per run), and were not told what the time threshold was. Therefore, to maximize their gain, participants had to treat every reward trial as a £1 trial and respond as quickly and as accurately as possible, just as in no-reward trials.

Each trial started with a 1 s cue, which was the name of a visual category that served as the target category for this trial. On reward trials, the cue included three red pound signs presented next to the category name. The cue was followed by a fixation dot in the center of the screen presented for 0.5 s and an additional variable time of either 0.1, 0.4, 0.7 or 1 s, selected randomly, in order to make the stimulus onset time less predictable. The stimulus was then presented for 120 ms and was followed by a mask. Participants indicated by a button press whether this object belonged to the cued target category (present) or not (absent). Following response, a 1 s blank inter-trial interval separated two trials. For both reward and no-reward trials, response time was limited to a maximum of 3 s, after which the 1 s blank inter-trial interval started even when no response was made. For reward trials, an additional subject-specific response time threshold was used as mentioned above to determine whether the participants earned the reward or not, but this time threshold did not affect the task structure and was invisible to the participants.

We used catch trials to decorrelate the BOLD signals of the cue and stimulus phases. 33% of all trials included cue followed by fixation dot for 500 ms, which then turned red for another 500 ms indicating the absence of the stimulus, followed by the inter-trial interval.









	Target (T)	Non-target Inconsistent (NI)	Non-target Consistent (NC)
Cue: shoe	shoe 	sofa 	car 
Cue: sofa	sofa 	shoe 	car 

Figure 4.1: Experimental paradigm. A. An example of a trial. A trial began with a cue (1 s) indicating the target category, followed by 500 ms fixation period. Reward trials were cued with three red £ symbols next to the target category. After an additional variable time (0.4,

0.7, 1.0 or 1.3 s), an object was presented for 120 ms. The object was then masked (a scramble of the all the stimuli used), until response or for a maximum of 3 s. The participants pressed a button to indicate whether the object was from the cued category (Target trials) or not (Nontarget trials). **B. Experimental conditions.** For each participant, two categories served as potential targets depending on the cue, and a third category never served as target. Here as an example, shoes and sofas are the cued categories and cars as the uncued category. In the Target trials, the presented object matched the cued category. In the High-conflict nontarget trials, the object did not match the cued category, but was from the other cued category, therefore could serve as a target on other trials. In the Low-conflict nontarget trials, the presented object was from the category that was never cued. Overall, this design yielded three levels of behavioral status: Targets, High-conflict nontargets, and Low-conflict nontargets. The design was used for both no-reward and reward conditions.

Stimuli

Objects were presented at the center of the screen on a grey background. The objects were 2.95° visual angle along the width and 2.98° visual angle along the height. Four exemplars from each visual category were used. Exemplars were chosen with similar colors, dimensions, and orientation across the categories. All exemplars were used an equal number of times in each condition and in each run to ensure that any differences between the experimental conditions will not be driven by the variability of exemplars. To increase the task demand, based on pilot data, we added Gaussian white noise to the stimuli. The post-stimulus mask was generated by randomly combining pieces of the stimuli that were used in the experiment. The mask was the same size as the stimuli and was presented until a response was made or the response time expired.

Structure and Design

Each participant completed 6 functional runs of the task in the scanner (mean duration \pm SD: 6.2 ± 0.13 min). Each run started with a response-mapping instructions screen (e.g. left = target present, right = target absent), displayed until the participants pressed a button to continue. Halfway through the run, the instructions screen was presented again with the reversed response mapping. All trials required a button response to indicate whether the target was present or absent, and the change of response mapping ensured that conditions were not confounded by the side of the button press. Each run included 104 trials. Out of these, 8 were dummy trials following the response mapping instructions (4 after each

instructions screen), and were excluded from the analysis. Of the remaining 96 trials, one-third (32 trials) were cue-only trials (catch trials). Of the remaining 64 trials, 32 were no-reward trials and 32 were reward trials. Of the 32 no-reward trials, half (16) were cued with one visual category, and half (16) with the other. For each cued category, half of the trials (8) were Target trials, and half of the trials (8) were nontarget trials, to assure an equal number of target (present) and nontarget (absent) trials. Of the nontarget trials, half (4) were High-conflict nontargets, and half (4) were Low-conflict nontargets. There were 4 trials per cue and reward level for the High- and Low-conflict nontarget conditions, and 8 for the Target condition, with the latter split into two regressors (see General Linear Model (GLM) for the Main Task section below). A similar split was used for reward trials. An event-related design was used and the order of the trials was randomized in each run. At the end of each run, the money earned in the reward trials and the number of correct trials (across both reward and no-reward trials) were presented on the screen.

Functional Localisers

In addition to the main task, we used two other tasks in order to functionally localize MD regions and LOC in individual participants using independent data. These were used in conjunction with ROI templates and a double-masking procedure to extract voxel data for MVPA (See ROI definition for more details).

To localize MD regions, we used a spatial working memory task (Fedorenko et al. 2013). On each trial, participants remembered 4 locations (Easy condition) or 8 locations (Hard condition) in a 3X4 grid. Each trial started with fixation for 500 ms. Locations on the grid were then highlighted consecutively for 1 s (1 or 2 locations at a time, for the Easy and Hard conditions, respectively). In a subsequent two-alternative forced-choice display (3 s), participants had to choose the grid with the correct highlighted locations by pressing the left or the right button. Feedback was given after every trial for 250 ms. Each trial was 8 s long, and each block included 4 trials (32 s). There was an equal number of correct grids on the right and left in the choice display. Participants completed 2 functional runs of 5 min 20 sec each, with 5 Easy blocks alternated with 5 Hard blocks in each run. We used the contrast of Hard vs. Easy blocks to localize MD regions.

As a localizer for LOC we used a one-back task with blocks of objects interleaved with blocks of scrambled objects. The objects were in grey scale and taken from a set of 61 everyday objects (e.g. camera, coffee cup, etc.). Participants had to press a button when the

same image was presented twice in a row. Images were presented for 300 ms followed by a 500 ms fixation. Each block included 15 images with two image repetitions and was 12 s long. Participants completed two runs of this task, with 8 object blocks, 8 scrambled object blocks, and 5 fixation blocks. The objects vs. scrambled objects contrast was used to localize LOC.

Scanning Session

The scanning session included a structural scan, 6 functional runs of the main task, and 4 functional localizer runs – 2 for MD regions and 2 for LOC. The scanning session lasted up to 100 minutes, with an average 65 minutes of EPI time. The tasks were introduced to the participants in a pre-scan training session. The average reaction time of 32 no-reward trials of the main task completed in this practice session was set as the time threshold for the reward trials to be used in the scanner session. All tasks were written and presented using Psychtoolbox3 (Brainard 1997) and MatLab (The MathWorks, Inc).

Data Acquisition

fMRI data were acquired using a Siemens 3T Prisma scanner with a 32-channel head coil. We used a multi-band imaging sequence (CMRR, release 016a) with a multi-band factor of 3, acquiring 2 mm isotropic voxels (Feinberg et al. 2010). Other acquisition parameters were: TR = 1.1 s, TE = 30 ms, 48 slices per volume with a slice thickness of 2 mm and no gap between slices, in plane resolution 2×2 mm, field of view 205 mm, flip angle 62° , and interleaved slice acquisition order. No iPAT or in-plane acceleration were used. T1-weighted multiecho MPAGE (van der Kouwe et al. 2008) high-resolution images were also acquired for all participants, in which four different TEs were used to generate four images (voxel size 1 mm isotropic, field of view of $256 \times 256 \times 192$ mm, TR = 2530 ms, TE = 1.64, 3.5, 5.36, and 7.22 ms). The voxelwise root mean square across the four MPAGE images was computed to obtain a single structural image.

Data and Statistical Analysis

The primary analysis approach was multi-voxel pattern analysis (MVPA), to assess representation of behaviorally relevant category distinctions with and without reward. An additional ROI-based univariate analysis was conducted to confirm the recruitment of the MD network. Preprocessing, GLM and univariate analysis of the fMRI data were performed using SPM12 (Wellcome Department of Imaging Neuroscience, London, England; www.fil.ion.ucl.ac.uk), and the Automatic Analysis (aa) toolbox (Cusack et al. 2014).

We used an alpha level of .05 for all statistical tests. Bonferroni correction for multiple comparisons was used when required, and the corrected p-values and uncorrected t-values are reported. All t tests that were used to compare two conditions were paired due to the within-subject design. A one-tailed t test was used when the prediction was directional, including testing for classification accuracy above chance level. All other t tests in which the *a priori* hypothesis was not directional were two-tailed. Additionally, effect size (Cohen's d_z) was computed. All analyses were conducted using custom-made MATLAB (The Mathworks, Inc) scripts, unless otherwise stated.

Preprocessing

Initial processing included motion correction and slice time correction. The structural image was coregistered to the Montreal Neurological Institute (MNI) template, and then the mean EPI was coregistered to the structural. The structural image was then normalized to the MNI template via a nonlinear deformation, and the resulting transformation was applied on the EPI volumes. Spatial smoothing of FWHM = 5 mm was performed for the functional localizers data only.

General Linear Model (GLM) for the Main Task

We used GLM to model the main task and localizers' data. Regressors for the main task included 12 conditions during the stimulus epoch and 4 conditions during the cue epoch. Regressors during the stimulus epoch were split according to reward level (no-reward, reward), cued visual category (category 1, category 2), and behavioral status (Target, High-conflict nontarget, Low-conflict nontarget). To assure an equal number of target present and target absent trials, the number of Target trials in our design was twice the number of High-conflict and Low-conflict nontarget trials. The Target trials included two repetitions of each combination of cue, visual category and exemplar, with a similar split for reward trials. These two Target repetitions were modelled as separate Target1 and Target2 regressors in the GLM to make sure that all the regressors were based on an equal number of trials, but were invisible to the participants. All the univariate and multivariate analyses were carried out while keeping the two Target regressors separate to avoid any bias of the results, and they were averaged at the final stage of the results. Overall, the GLM included 16 regressors of interest for the 12 stimulus conditions. Each regressor was based on data from all correct trials in the respective condition in each run (up to 4 trials). To account for possible effects of reaction time (RT) on the beta estimates because of the varying duration of the stimulus

epoch, and as a consequence their potential effect on decoding results, these regressors were modelled with durations from stimulus onset to response (Woolgar et al. 2014). This model scales the regressors based on the reaction time, thus the beta estimates reflect activation per unit time and are comparable across conditions with different durations. Regressors during the cue epoch included both task and cue-only (catch) trials and were split by reward level and cued category, modelled with duration of 1 s. Cue regressors were based on 16 trials per regressor per run. As one-third of all trials were catch trials, the cue and stimulus epoch regressors were decorrelated and separable in the GLM. Regressors were convolved with the canonical hemodynamic response function (HRF). The 6 movement parameters and run means were included as covariates of no interest.

GLM for the Functional Localisers

For the MD localizer, regressors included Easy and Hard blocks. For LOC, regressors included objects and scrambled objects blocks. Each block was modelled with its duration. The regressors were convolved with the canonical hemodynamic response function (HRF). The 6 movement parameters and run means were included as covariates of no interest.

Univariate Analysis

We conducted an ROI analysis to test for the effect of reward on overall activity for the different behavioral status conditions and cues. We used templates for the MD network and for LOC as defined below (see ROI definition). Using the MarsBaR toolbox (<http://marsbar.sourceforge.net>; Brett et al. 2002) for SPM 12, beta estimates for each regressor of interest were extracted and averaged across runs, and across voxels within each ROI, separately for each participant and condition. For the MD network, beta estimates were also averaged across hemispheres (see ROI definition below). Second-level analysis was done on beta estimates across participants using repeated measures ANOVA. The data for the Target condition was averaged across the two Target1 and Target2 regressors, separately for the no-reward and reward conditions.

ROI Definition

MD network template. ROIs of the MD network were defined *a priori* using an independent data set (Fedorenko et al. 2013; see t-map at <http://imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem>). These included the anterior, middle, and posterior parts of the middle frontal gyrus (aMFG, mMFG, and pMFG, respectively), a posterior dorsal region of the lateral frontal cortex (pdLFC), AI-FO, pre-SMA/ACC, and IPS, defined in the

left and right hemispheres. The visual component in this template is widely accepted as a by-product of using largely visual tasks, and is not normally considered as part of the MD network. Therefore, it was not included in the analysis. The MD network is highly bilateral, with similar responses in both hemispheres (Fedorenko et al. 2013; Erez and Duncan 2015). We therefore averaged the results across hemispheres in all the analyses.

LOC template. LOC was defined using data from a functional localizer in an independent study with 15 participants (Lorina Naci, PhD dissertation, University of Cambridge). In this localizer, forward- and backward-masked objects were presented, as well as masks alone. Masked objects were contrasted with masks alone to identify object-selective cortex (Malach et al. 1995). Division to the anterior part of LOC, the posterior fusiform region (pFs) of the inferior temporal cortex, and its posterior part, the lateral occipital region (LO) was done using a cut-off MNI coordinate of $Y=-62$, as previous studies have shown differences in processing for these two regions (MacEvoy and Epstein 2011; Erez and Yovel 2014).

Voxels selection for MVPA

To compare between regions within the MD network and between sub-regions in LOC, we controlled for the ROI size and used the same number of voxels for all regions. We used a dual-masking approach that allowed the use of both a template, consistent across participants, as well as subject-specific data as derived from the functional localizers (Fedorenko et al. 2010; Shashidhara, Spronkers, et al. 2019). For each participant, beta estimates of each condition and run were extracted for each ROI based on the MD network and LOC templates. For each MD ROI, we then selected the 200 voxels with the largest t-value for the Hard vs. Easy contrast as derived from the independent subject-specific functional localizer data. This number of voxels was chosen prior to any data analysis, similar to our previous work (Erez and Duncan 2015). For each LOC sub-region, we selected 180 voxels with the largest t-values of the object vs. scrambled contrast from the independent subject-specific functional localizer data. The selected voxels were used for the voxelwise patterns in the MVPA for the main task. The number of voxels that was used for LOC was smaller than for MD regions because of the size of the pFs and LO masks. For the analysis that compared MD regions with the visual regions, we used 180 voxels from all regions to keep the ROI size the same.

Multivoxel pattern analysis (MVPA)

We used MVPA to test for the effect of reward motivation on the discrimination between the task-related behavioral status pairs. Voxelwise patterns using the selected voxels within each

template were computed for all the task conditions in the main task. We applied our classification procedure on all possible pairs of conditions as defined by the GLM regressors of interest during the stimulus presentation epoch, for the no-reward and reward conditions separately (Figure 4.1B). For each pair of conditions, MVPA was performed using a support vector machine classifier (LIBSVM library for MATLAB, $c=1$) implemented in the Decoding Toolbox (Hebart et al. 2015). We used leave-one-run-out cross-validation in which the classifier was trained on the data of five runs (training set) and tested on the sixth run (test set). This was repeated 6 times, leaving a different run to test each time, and classification accuracies were averaged across these 6 folds. Classification accuracies were then averaged across pairs of different cued categories, yielding discrimination measures for three pairs of behavioral status (Targets vs. High-conflict nontargets, Targets vs. Low-conflict nontargets, and High-conflict vs. Low-conflict nontargets) within each reward level (no-reward, reward). Because the number of Target trials in our design was twice the number of High-conflict and Low-conflict nontarget trials, each discrimination that involved a Target condition was computed separately for the two Target regressors (Target1 and Target2) and classification accuracies were averaged across them.

The Target and High-conflict nontarget pairs of conditions included cases when both conditions had an item from the same visual category as the stimulus (following different cues), as well as cases in which items from two different visual categories were displayed as stimuli (following the same cue). To test for the contribution of the visual category to the discrimination, we split the Target vs. High-conflict nontarget pairs of conditions into these two cases and the applied statistical tests accordingly.

Whole brain searchlight pattern analysis

To test whether additional regions outside the MD network show change in discriminability between voxelwise patterns of activity of behavioral status conditions when reward is introduced, we conducted a whole-brain searchlight pattern analysis (Kriegeskorte et al. 2006). This analysis enables the identification of focal regions that carry relevant information, unlike the decoding based on larger ROIs, which tests for a more widely distributed representation of information. For each participant, data was extracted from spherical ROIs with an 8 mm radius, centered on each voxel in the brain. These voxels were used to perform the same MVPA analysis as described above. Thus, for each voxel, we computed the classification accuracies for the relevant distinctions, separately for the no-reward and reward conditions. These whole-brain maps were smoothed using a 5 mm

FWHM Gaussian kernel. The t-statistic from a second level random-effects analysis on the smoothed maps was thresholded at the voxel level using FDR correction ($p < 0.05$).

Results

Behaviour

RT for three behavioral status conditions, Target, High-conflict nontarget and Low-conflict nontarget, in the no-reward trials were 589 ± 98 ms, 662 ± 103 ms, and 626 ± 107 ms, respectively (mean \pm SD); RTs for these conditions in the reward trials were 541 ± 99 ms, 614 ± 99 ms, 585 ± 97 ms, respectively (mean \pm SD). A two-way repeated measures ANOVA with motivation (no-reward, reward) and behavioral status as within-subject factors showed a main effect of motivation ($F_{1, 23} = 40.07, p < 0.001$), with reward trials being shorter than no-reward trials, as expected from the experimental design in which response was required within a time limit to receive the reward. An additional main effect of behavioral status ($F_{2, 23} = 50.97, p < 0.001$) was observed, with no interaction between reward and behavioral status ($F_{2, 23} = 0.63, p = 0.54$). Subsequent post-hoc tests with Bonferroni correction for multiple comparisons showed that RTs for Target trials were faster than High-conflict and Low-conflict nontarget trials ($t_{23} = 10.03, p < 0.001, d_z = 2.05$; $t_{23} = 5.17, p < 0.001, d_z = 1.06$ respectively), and Low-conflict nontarget trials were faster than the High-conflict ones ($t_{23} = 4.96, p < 0.001, d_z = 1.01$), as expected from a cued target detection task. Overall accuracy levels were high (mean \pm SD: $92.51\% \pm 0.08\%$). Mean and SD accuracy rates for the Target, High-conflict nontarget and Low-conflict nontarget conditions in the no-reward trials were $91.2\% \pm 5.8\%$, $89.1\% \pm 8.8\%$, and $96.6\% \pm 3.8\%$, respectively; and for the reward trials they were $94.2\% \pm 5.0\%$, $87.8\% \pm 8.7\%$, $96.1\% \pm 4.4\%$, respectively. A two-way repeated measures ANOVA with motivation and behavioral status as within-subject factors showed no main effect of motivation ($F_{1, 23} = 0.49, p = 0.49$), confirming that the added time constraint for reward trials did not lead to drop in performance. There was a main effect of behavioral status ($F_{2, 23} = 29.64, p < 0.001$) and an interaction between motivation and behavioral status ($F_{2, 23} = 5.81, p < 0.01$). Post-hoc tests with Bonferroni correction for multiple comparisons showed larger accuracies for Low-conflict nontargets compared to Targets and High-conflict nontargets (Two-tailed t-test: $t_{23} = 5.64, p < 0.001, d_z = 1.15$; $t_{23} = 5.50, p < 0.001, d_z = 1.12$ respectively) in the no-reward trials, as expected given that the Low-conflict nontarget category was fixed throughout the experiment. In the reward trials, performance accuracies were larger for Target compared to High-conflict nontarget ($t_{23} = 4.45, p < 0.001, d_z = 0.91$) and Low-conflict nontarget compared to High-conflict ones ($t_{23} = 5.92, p < 0.001, d_z = 1.2$) and similar between Targets and Low-conflict nontargets ($t_{23} =$

2.49, $p = 0.06$). Accuracies for Target trials were larger for the reward trials compared to no-reward ($t_{23} = 2.92$, $p = 0.008$, $d_z = 0.61$), indicating a possible behavioral benefit of reward. There was no difference between reward and no-reward trials for High-conflict and Low-conflict nontargets ($t_{23} < 1.1$, $p > 0.1$, for both).

Activity across the MD network during the cue epoch

To address our primary research question, the analysis focused on the stimulus epoch. However, to get a full picture of the data and for comparability with previous studies that showed increase in cue information, we also report the results for the cue epoch here. The analysis focuses on the MD network only and not the LOC since no object stimuli were presented at this epoch of the trial.

We first tested for a univariate effect of reward during the cue phase (averaged across the β estimates of the two cues) across all MD regions. A two-way repeated measures ANOVA with reward (2: no-reward, reward) and ROI (7) as factors showed a main effect of reward ($F_{1, 23} = 13.75$, $p = 0.001$) with increased activity during the reward trials compared to the no-reward trials. There was also a main effect of ROI ($F_{6, 138} = 6.44$, $p < 0.001$) and an interaction of reward and ROI ($F_{6, 138} = 6.67$, $p < 0.001$). Post-hoc tests showed that all regions except aMFG showed increased activation for reward trials compared to no-reward trials (Two tailed, Bonferroni corrected for 7 comparisons: $t_{23} > 3.06$, $p < 0.04$, $d_z > 0.62$ for all ROIs except aMFG; $t_{23} = 2.38$, $p = 0.18$, $d_z = 0.49$ for aMFG). Overall, the MD network showed a strong univariate reward effect during the cue epoch.

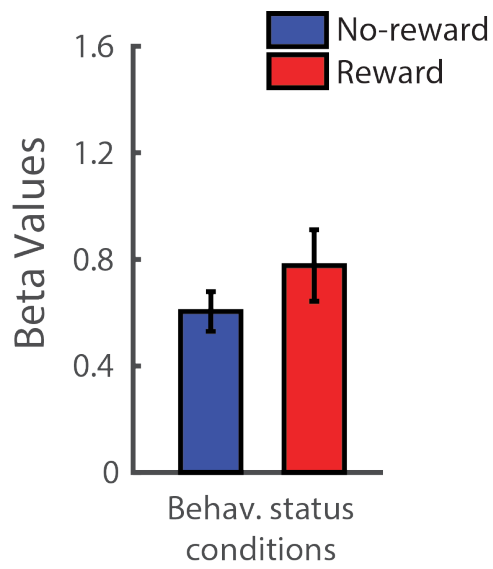
We next asked whether the cues were decodable as measured using MVPA, and whether decoding levels increased with reward as has been previously reported (Etzel et al. 2016; Hall-McMaster et al. 2019). Decoding between the two cues separately for the two reward levels were computed in each of the MD ROIs. A two-way repeated measures ANOVA with reward (2) and ROI (7) as factors showed no main effects or interactions ($F < 1.9$, $p > 0.08$). Decoding levels averaged across all MD ROIs were (mean \pm SD) 51.71% \pm 4.39% and 50.55% \pm 6.69% for the reward and no-reward conditions, respectively. There was no significant cue decoding above chance for both no-reward and reward conditions (One-tailed t-test, reward: $t_{23} = 1.9$, $p = 0.07$, $d_z = 0.39$; no-reward: $t_{23} = 0.4$, $p = 0.7$, $d_z = 0.07$). Overall, we found that cue information could not be decoded in any of the MD ROIs and in both no-reward and reward conditions.

Lastly, we conducted a complementary whole-brain searchlight analysis to test whether cue decoding was observed in other regions beyond the MD network. A second level random-effects analysis of cue decoding, separately for the no-reward and reward conditions, did not reveal any additional regions that showed cue decoding (FDR correction $p < 0.05$). An additional searchlight analysis was set to test for increase in cue decoding with reward, but similar results were obtained, with no voxels surviving FDR correction ($p < 0.05$). Altogether, our results show that despite substantial increases in overall univariate activity with reward during the cue epoch across the MD network, the cues in our study were not decodable in both no-reward and reward conditions.

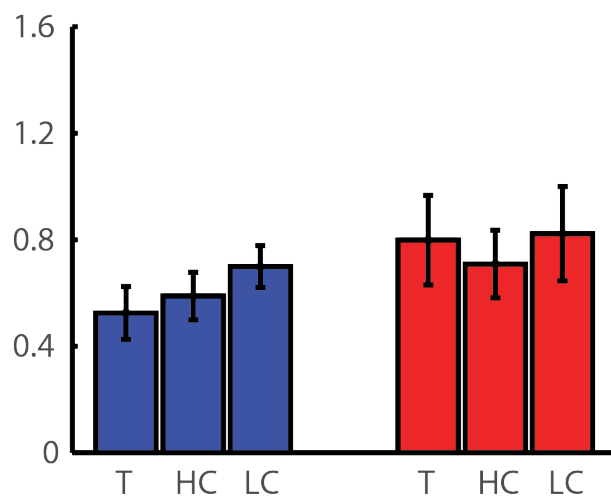
Univariate activity in the MD network during the stimulus epoch

We started our analysis for the stimulus epoch by testing for the effect of reward motivation on the overall activity in MD regions, and whether such effect is different for the three behavioral status conditions. We used averaged β estimates for each behavioral status (Target, High-conflict nontarget, Low-conflict nontarget) and reward level (no-reward, reward) in each of the MD ROIs (Figure 4.2). A three-way repeated measures ANOVA with reward (2), behavioral status (3) and ROI (7) as within-subject factors showed no significant main effect of reward ($F_{1, 23} = 3.37, p = 0.079$). There was an interaction of reward level and ROI ($F_{6, 138} = 5.02, p = 0.001$), with only the AI/FO showing reward effect following post-hoc tests and Bonferroni correction for multiple (7) comparisons ($t_{23} = 3.88, p = 0.005, d_z = 0.79$). The IPS showed a reward effect that did not survive multiple comparisons ($t_{23} = 2.58$, uncorrected $p = 0.016$, corrected $p = 0.11, d_z = 0.53$). Importantly, there was no main effect of behavioral status ($F_{2, 46} = 0.97, p = 0.57$) and no interaction of reward and behavioral status ($F_{2, 46} = 0.51, p = 0.61$). Overall, the univariate results indicated similar levels of activity for the three behavioral status conditions. While we expected an increase in univariate activity in many MD regions with reward (Padmala and Pessoa 2011; Dixon and Christoff 2012; Shashidhara, Mitchell, et al. 2019), we observed such an increase only in the AI/FO.

A



B



C

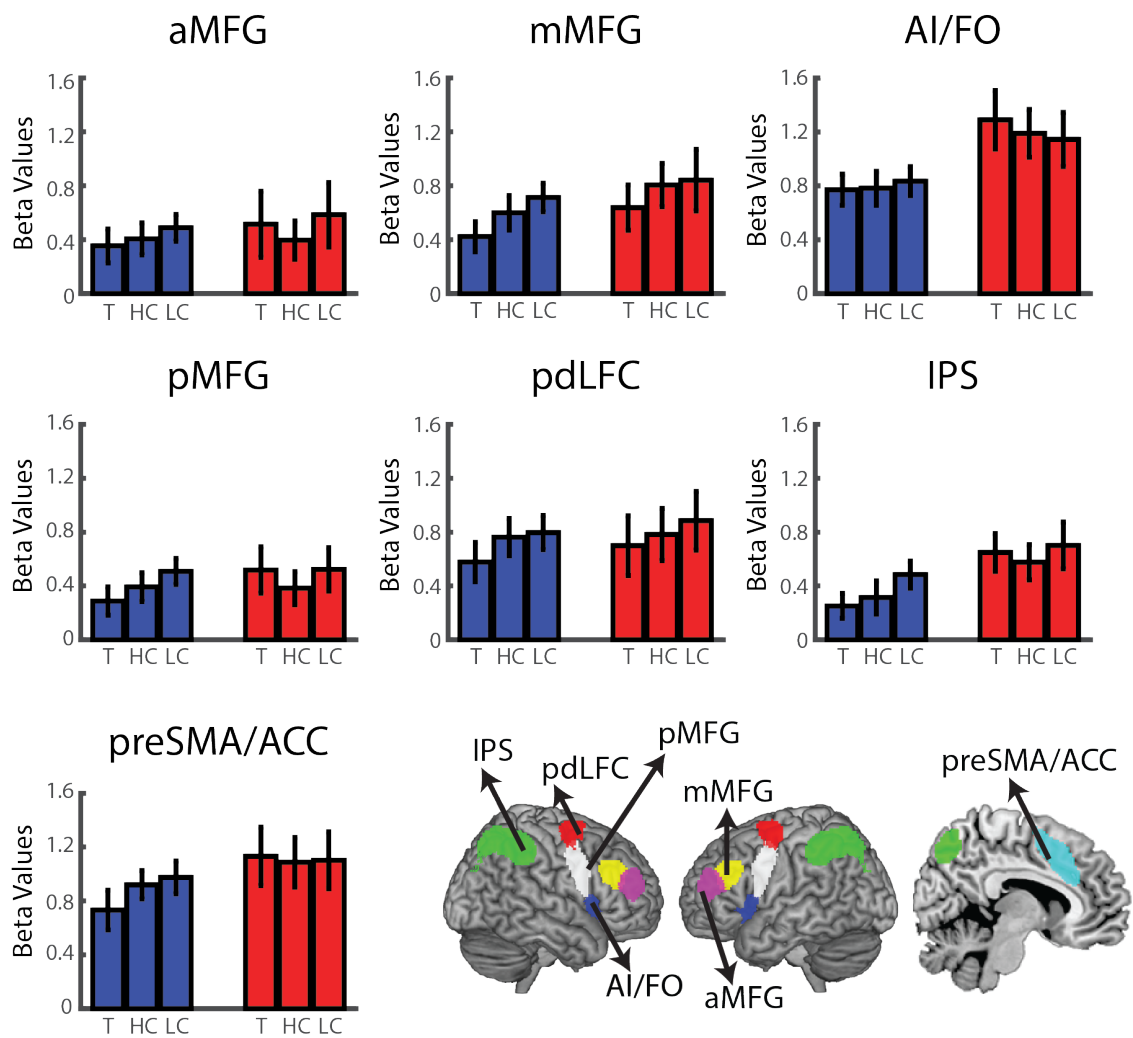


Figure 4.2: Univariate activity across the MD network during the stimulus epoch. A. Univariate results averaged across all MD regions. Results are averaged across the behavioral status conditions for no-reward (blue bar) and reward (red bar) conditions, showing strong recruitment of the network and no increase with reward. **B. Average univariate activity across the MD network** is shown separately for each behavioral status condition for no-reward (blue bars) and reward (red bars) conditions. Activity is similar for the three behavioral status conditions and does not increase with reward. T: Target, HC: High-conflict nontarget, LC: Low-conflict nontarget). **C.** Univariate results for the individual MD regions showing similar results for all regions. Post-hoc tests showed that only activity in AI increased with reward. The MD network template is shown for reference. pdLFC: posterior/dorsal lateral prefrontal cortex, IPS: intraparietal sulcus, preSMA: pre-supplementary motor area, ACC: anterior cingulate cortex, AI: anterior insula, FO: frontal operculum, aMFG, mMFG, pMFG: anterior, middle and posterior middle frontal gyrus, respectively. Errors bars indicate S.E.M.

Effect of reward motivation on discrimination of behaviourally relevant category distinctions in the MD network

Our main question concerned the representation of task-related behavioral status information across the MD network and its modulation by reward, and we used MVPA to address that. For each participant and ROI we computed the classification accuracy above chance (50%) for the distinctions between Target vs. High-conflict nontarget, Target vs. Low-conflict nontarget and High-conflict vs. Low-conflict nontargets, separately for no-reward and reward conditions (Figure 4.3). The analysis was set to test for discrimination between behavioral status conditions within each reward level, and whether these discriminations are larger when reward is introduced compared to the no-reward condition. A three-way repeated-measures ANOVA with reward (2), behavioral distinction (3) and ROI (7) as within-subject factors showed no main effect of ROI ($F_{6, 138} = 0.97, p = 0.45$) or any interaction of ROI with reward and behavioral distinction ($F < 1.16, p > 0.31$). Therefore, the classification accuracies were averaged across ROIs for further analysis (Figure 4.3A). First, we looked at the overall discrimination of behavioral status pairs. Averaged across the three pairs of behavioral status, decoding accuracies were (mean \pm SD) $51.4\% \pm 2.8\%$ and $51.8\% \pm 3.5\%$ for the no-reward and reward conditions, respectively. Decoding levels were above chance (50%) for both the no-reward and reward trials (one-tailed t-test against chance, corrected for 2 comparisons,

reward: $t_{23} = 2.5$, corrected $p = 0.02$, $d_z = 0.5$; no-reward: $t_{23} = 2.34$, corrected $p = 0.03$, $d_z = 0.48$). The decoding levels above chance for the individual pairs of behavioral status for the no-reward and reward conditions are summarized in Table 1. Overall, our results show that on average behaviorally relevant categorical distinctions are represented across the MD network in both no-reward and reward conditions, with some differences between individual pairs of behavioral status.

In our critical analysis we tested for the modulatory effect of reward on the discriminability between pairs of behavioral status. In contrast to our prediction, a two-way repeated measures ANOVA with reward (2) and behavioral distinction (3) as within-subject factors showed no main effects of reward or behavioral distinction ($F_{1, 23} = 0.26$, $p = 0.6$; $F_{2, 46} = 1.37$, $p = 0.26$, respectively), and no interaction of the two ($F_{2, 46} = 0.74$, $p = 0.48$). To test for the specific prediction that reward might increase discrimination for the high conflict pair of conditions that may not have been picked up by the ANOVA, we compared decoding levels for the Target vs. High conflict nontarget for the no-reward and reward conditions. Further in contrast to our prediction, classification accuracy was not larger in the reward trials compared to the no-reward trials for the Target vs. High-conflict nontarget distinction (One-tailed paired t-test: $t_{23} = 1.07$, $p = 0.15$, $d_z = 0.22$). In summary, although the average decoding levels of behavioral status were above chance across the MD network, we did not find increases in decodability with reward.

To test for other brain regions that may have shown increased pattern discriminability with reward beyond the MD network, we conducted a complementary whole-brain searchlight analysis. In a second-level random-effects analysis of behavioral status classification maps (average across the three pairs of behavioral status) of reward vs. no-reward conditions, none of the voxels survived an FDR threshold of $p < 0.05$. A separate searchlight analysis for classification of Targets vs. High-conflict nontargets showed similar results, with no voxels surviving FDR correction ($p < 0.05$). Therefore, our data did not reveal any brain regions that showed the predicted increase in discriminability with reward.

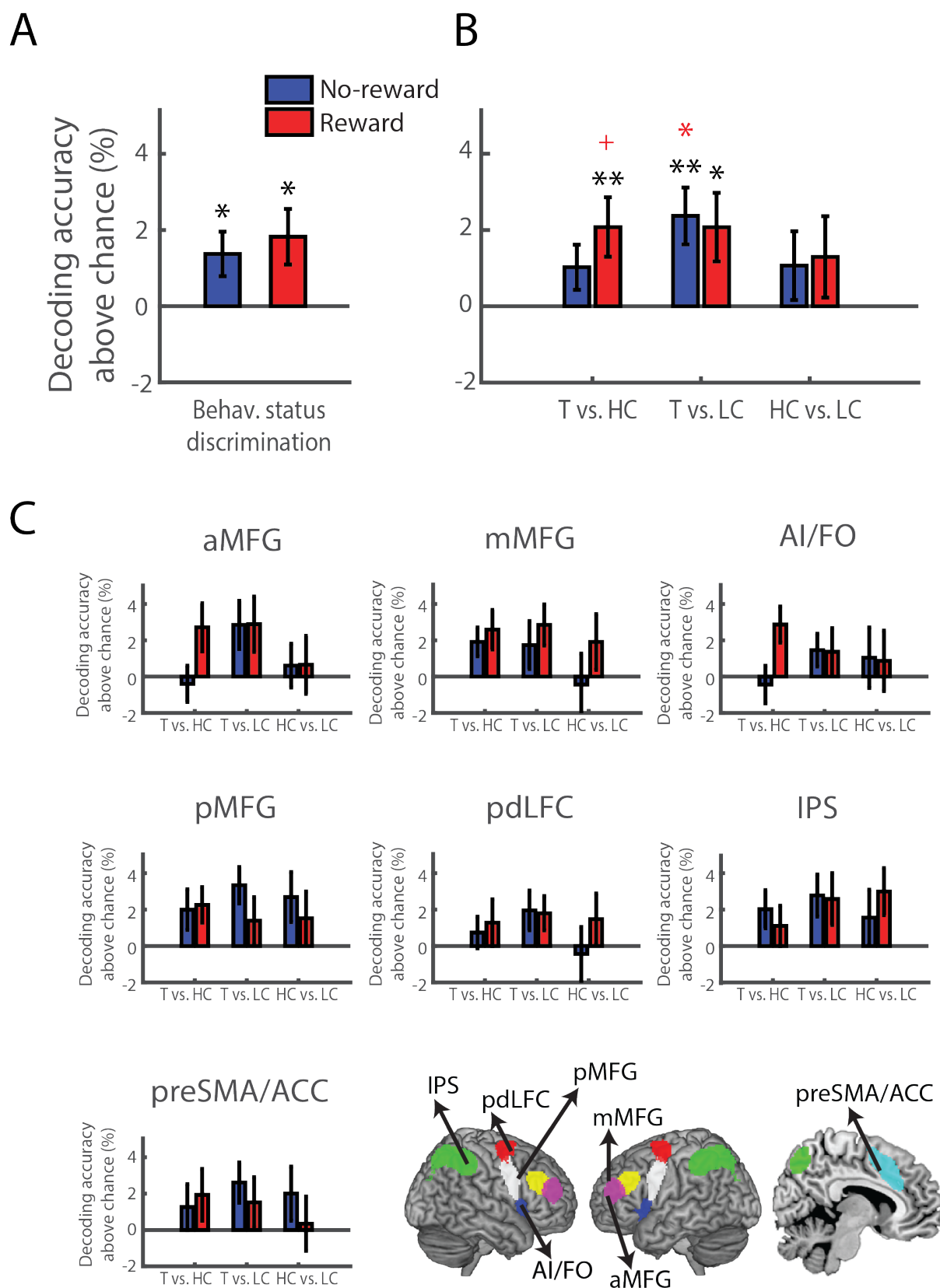


Figure 4.3: Reward does not modulate distinctions of behavioral status across the MD network. **A.** Classification accuracy is presented as percentage above chance (50%),

averaged across all MD regions and behavioral status pairs, for no-reward (blue bars) and reward (red bars) trials. Behavioral status was decodable but not modulated by reward. Asterisks above bars show significant decoding above chance (One-tailed, Bonferroni corrected for 2 comparisons). **B.** The data in A is shown separately for the three distinctions of Target vs. High-conflict nontarget, Target vs. Low-conflict non-target, and High- vs. Low-conflict nontargets. T: Target, HC: High-conflict nontarget, LC: Low-conflict nontarget. Asterisks above bars show one-tailed significant discrimination between behavioral categories above chance without correction (black), and Bonferroni corrected for multiple (6) comparisons (red). See Table 1 for details. **C.** Decoding results are shown for the individual MD regions. The MD network template is shown for reference. pdLFC: posterior/dorsal lateral prefrontal cortex, IPS: intraparietal sulcus, preSMA: pre-supplementary motor area, ACC: anterior cingulate cortex, AI: anterior insula, FO: frontal operculum, aMFG, mMFG, pMFG: anterior, middle and posterior middle frontal gyrus, respectively. Errors bars indicate S.E.M. + $p < 0.06$, * $p < 0.05$, ** $p < 0.01$.

Table 1: Decoding accuracies for pairs of behavioral status across the MD network.

t values are for a one-tailed t-test against change of 50%. Corrected p values were obtained using Bonferroni correction for 6 comparisons. + $p < 0.06$, * $p < 0.05$, ** $p < 0.01$

Distinction	Reward level	Mean \pm S.E.M (%)	t_{23} value	Uncorrected p	corrected p	Effect size
Target vs.	No-reward	50.9 \pm 0.6	1.43	0.083	0.5	0.29
High-conflict nontarget	Reward	52 \pm 0.8	2.56	** 0.009	+ 0.053	0.52
Target vs.	No-reward	52.4 \pm 0.7	3.24	** 0.002	* 0.011	0.66
Low-conflict nontarget	Reward	52.1 \pm 0.9	2.23	* 0.02	0.11	0.45
High-conflict vs. Low-conflict nontargets	No-reward	50.9 \pm 0.9	1.32	0.2	1	0.20
	Reward	51.4 \pm 1.1	0.96	0.1	0.6	0.27

Effects of reward motivation on behaviourally relevant category distinctions in LOC

It is widely accepted that the frontoparietal MD network exerts top-down control on visual areas, contributing to task-dependent processing of information. To test for reward effects on decoding of categorical information based on behavioral status in the visual cortex, we performed similar univariate and multivariate analyses during the stimulus epoch in the high-level general object visual region, the lateral occipital complex (LOC), separately for its two sub-regions, LO and pFs. We first conducted univariate analysis to test for an effect of reward and behavioral status on overall activity in LOC, which did not show a change in BOLD response with reward. A four-way repeated-measures ANOVA with reward (2), behavioral status (3), ROI (2), and hemisphere (2) as within-subject factors showed no main effect of reward ($F_{1,23} = 0.3$, $p = 0.6$), no main effect of ROI ($F_{1,23} = 3.7$, $p = 0.07$), or hemisphere ($F_{1,23} = 3.95$, $p = 0.06$). There was an interaction of reward and ROI ($F_{1,23} = 7.3$, $p = 0.01$), but post-hoc tests with correction for multiple (2) comparisons showed that activity was not larger for reward compared to no-reward trials in both LO and pFs (Two-tailed t-test: $t_{23} = 1.45$, $p = 0.16$, $d_z = 0.3$; $t_{23} = 0.94$, $p = 0.36$, $d_z = 0.2$; for LO and pFs, respectively). There was a main effect of behavioral status ($F_{2,46} = 8.73$, $p < 0.001$), but no interaction of behavioral status and reward ($F_{2,46} = 0.56$, $p = 0.57$). Altogether, the univariate results show that reward did not lead to increased activity in LOC for any of the behavioral status conditions.

We then tested for the representation of the task-related behavioral status conditions in LOC (Figure 4.4). Decoding levels averaged across all pairs of behavioral status and the two LOC ROIs were above chance for both no-reward and reward conditions (mean \pm SD: $3.23\% \pm 3.64\%$ and $3.76\% \pm 4.14\%$ for the no-reward and reward conditions, respectively. One-tailed t-test against chance, corrected for 2 comparisons, reward: $t_{23} = 4.45$, corrected $p < 0.001$, $d_z = 0.9$; no-reward: $t_{23} = 4.34$, corrected $p < 0.001$, $d_z = 0.9$). Importantly, decoding levels were not larger for the reward conditions compared to the no-reward conditions for any of the behavioral status distinctions, with similar results for both LO and pFs. A four-way repeated-measures ANOVA with reward (2), behavioral distinction (3), ROIs (2) and hemispheres (2) as within-subject factors showed no main effect of reward ($F_{1,23} = 0.34$, $p = 0.56$) or interaction of reward and ROI ($F_{1,23} = 1.14$, $p = 0.29$). No other main effects or interactions were significant ($F < 3.15$, $p > 0.05$). Overall, these results demonstrate that reward did not modulate the coding of the task-related behavioral status distinctions in LOC.

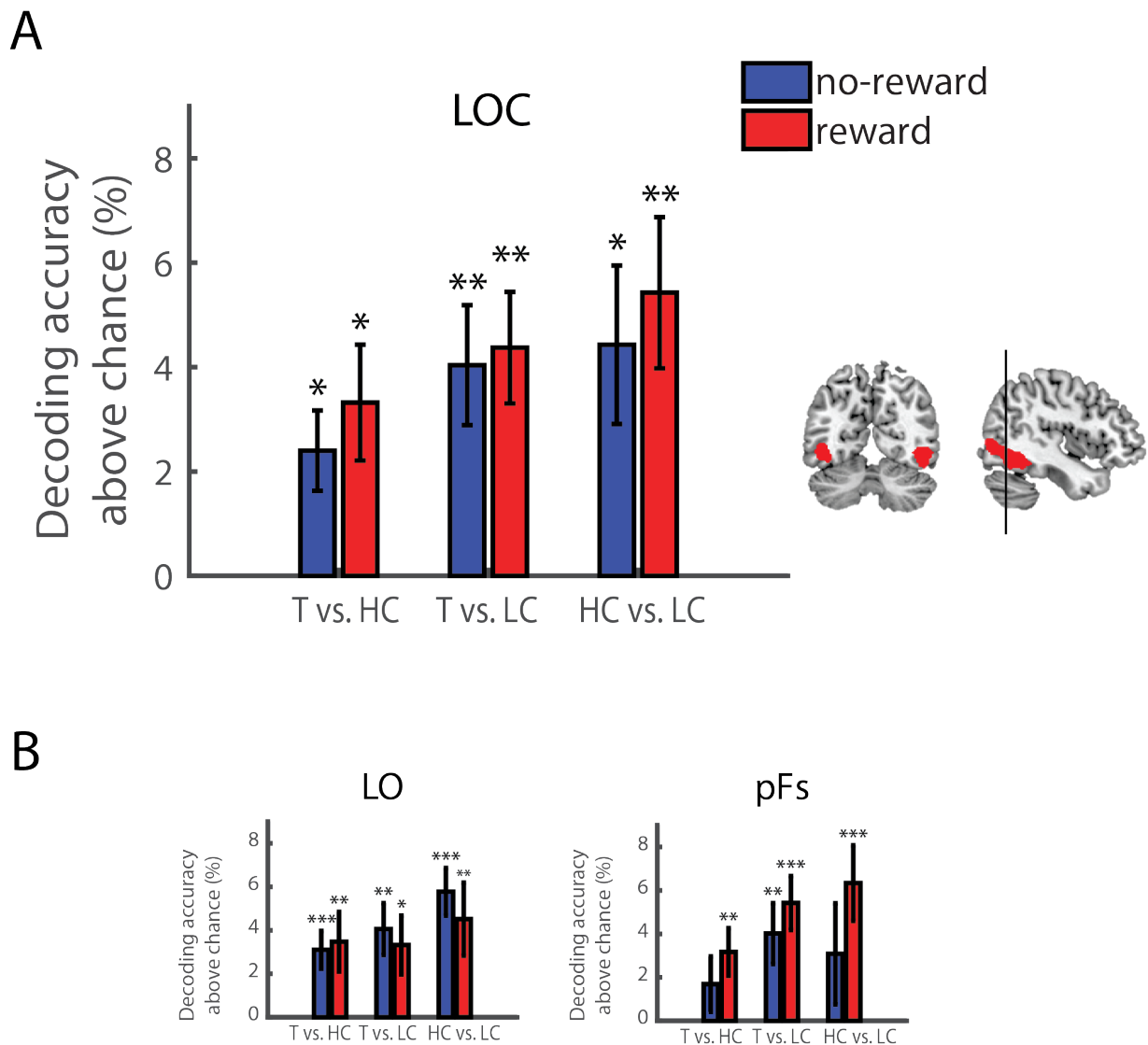


Figure 4.4: Reward motivation does not increase coding of behavioral status in LOC. A. Classification accuracy averaged across all behavioral status pairs is presented as percentage above chance (50%), averaged across LO and pFs and both hemispheres. Classification accuracies for no-reward (blue bars) and reward (red bars) conditions are similar and above chance. Asterisks above bars show significant decoding level above chance, (one-tailed Bonferroni corrected for 2 comparisons). B. Classification accuracies are similar for all three behavioral status distinctions. T: Target, HC: High-conflict nontarget, LC: Low-conflict nontarget. Asterisks above bars show one-tailed significant discrimination between behavioral categories above chance without correction (black), and corrected for multiple (6) comparisons (red). C. Classification accuracies for LO and pFs are presented separately, averaged across hemispheres. The LOC template is shown on sagittal and coronal planes, with a vertical line dividing it into posterior (LO) and anterior (pFs) regions. Errors bars indicate S.E.M. + $p < 0.06$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Conflict-contingent vs. visual category effects

An important aspect of the Target and High-conflict nontarget conditions in this experiment was that they both contained the same visual categories, which could be either a target or a nontarget (Figure 4.1B). Therefore, the Target vs. High-conflict nontarget pairs of conditions in our decoding analysis included cases where the stimuli in the two conditions were items from different visual categories (e.g. shoe and sofa following a ‘shoe’ cue), as well as cases where the two stimuli were items from the same visual category (e.g. shoe following a ‘shoe’ cue and a ‘sofa’ cue). We further investigated whether the representation in the MD network and in the LOC was driven by the task-related high conflict nature of the two conditions or by the different visual categories of the stimuli, and whether there was a facilitative effect of reward which is limited to the representation of the visual categories. For each participant, the decoding accuracy for this behavioral status distinction was computed separately for pairs of conditions in which the stimuli belonged to the same visual category (different cue trials), and for pairs in which the stimuli belonged to different visual categories (same cue trials). This analysis was conducted by selecting 180 voxels for both MD and LOC ROIs, to keep the ROI size the same. For both MD and LOC regions, there was no interaction with ROI or hemisphere, therefore accuracy levels were averaged across hemispheres and ROIs for the MD network and LOC (repeated measures ANOVA with reward (2), distinction type (2, same or different visual category), ROIs (7 for MD, 2 for LOC) and hemispheres (2, just for LOC) as within-subject factors: $F < 3.15$ $p > 0.05$ for all interactions with ROI and hemisphere). Figure 4.5 shows Target vs. High-conflict nontarget distinctions separately for same and different visual categories for no-reward and reward conditions, for both the MD and LOC.

We next tested for the effect of reward and distinction type (same or different visual category) on decoding levels in each the two systems. In the MD network, a two-way repeated measures ANOVA with reward (2) and distinction type (2) as factors showed no main effect of reward ($F_{1,23} = 0.92$, $p = 0.35$) and no effect of category distinction or their interaction ($F_{1,23} = 2.9$, $p = 0.1$; $F_{1,23} = 0.1$, $p = 0.8$, respectively). These results show that there was no effect of reward on high conflict items that may be specific for the distinction between visual categories. In contrast, a similar ANOVA for LOC showed a main effect of distinction type ($F_{1,23} = 25.9$, $p < 0.001$) and no effect of reward or their interaction ($F_{1,23} = 0.05$, $p = 0.8$; $F_{1,23} = 0.46$, $p = 0.50$, respectively). Together, these results demonstrate that

representation was driven by visual categories in LOC, but not in the MD network. To further establish this dissociation between the two systems, we used a three-way repeated measures ANOVA with distinction type (2, same or different visual category), reward (2), and brain system (2, MD or LOC) as within-subject factors. There was no main effect of brain system ($F_{1,23} = 1.2, p = 0.3$), allowing us to compare between the two systems. An interaction between distinction type and system ($F_{1,23} = 16.7, p < 0.001$) confirmed that decoding levels in the two systems were affected differently by visual category. Critically to our research question, reward did not lead to increased decoding levels in either of the systems (no main effect of reward or interactions with reward: $F < 0.77, p > 0.39$).

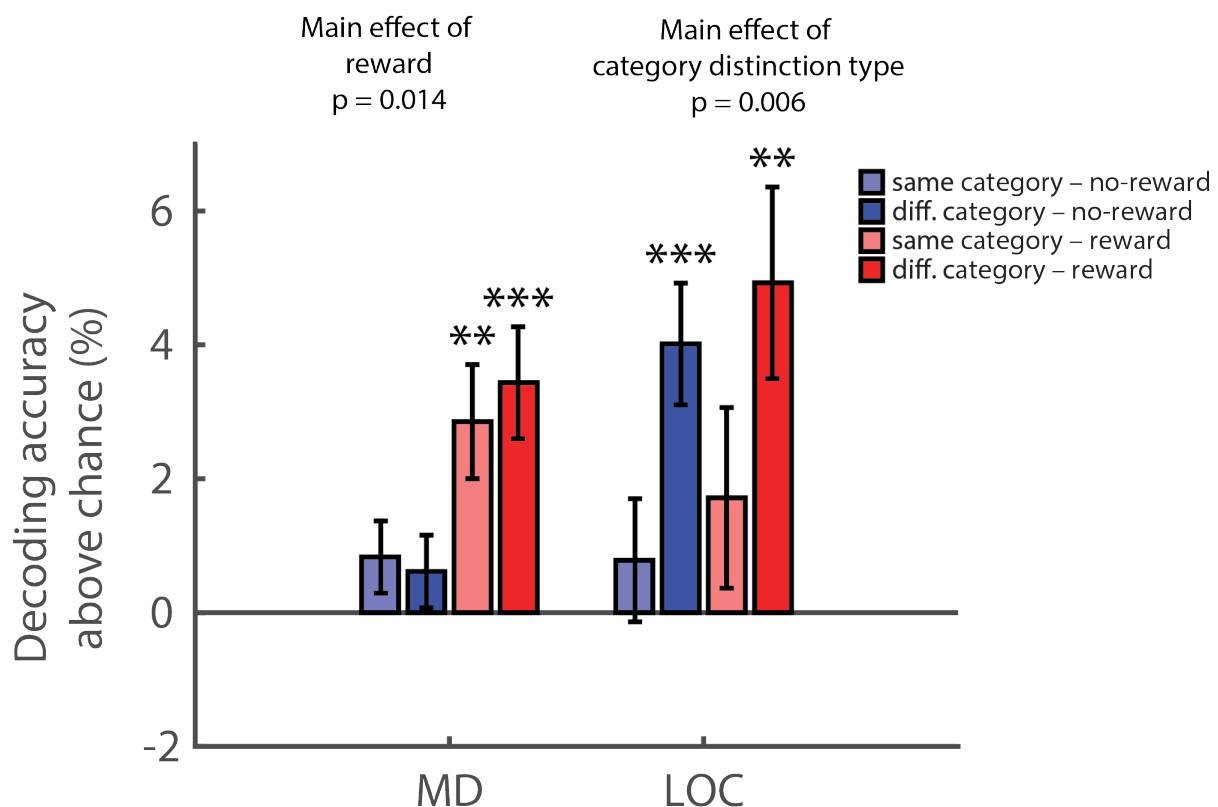


Figure 4.5: Decoding of highly conflicting behavioral status distinctions in the MD network and LOC. Classification accuracies above chance (50%) are presented for no-reward and same-visual-category distinctions (light blue), no-reward and different-visual-category distinctions (dark blue), reward and same-visual-category distinctions (light red), and reward and different-visual-category distinctions (dark red), separately for the MD network and the LOC, averaged across regions and hemispheres in each system. In the MD network, neither reward nor visual category modulated the discrimination of Target vs. High-conflict nontarget. In contrast, classification accuracies in the LOC are larger when the

displayed objects are from two different visual categories compared to when they belong to the same visual category, irrespective of the reward level. Asterisks above bars show one-tailed significant discrimination above chance without correction (black), and corrected for multiple (6) comparisons (red). Significant main effects of visual category are shown above the bars of each system. Errors bars indicate S.E.M. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Discussion

In this study we used a cued target detection task to test for the effect of reward motivation on the coding of task-related behaviorally-relevant category distinctions in the frontoparietal MD network as reflected in distributed patterns of fMRI data. Participants detected whether an item belonged to a cued visual category. Two visual categories served as either targets or nontargets, depending on the cue. A third category was never cued and therefore was never a target, creating three levels of behavioral status: Targets, High-conflict nontargets, and Low-conflict nontargets. During the cue epoch of reward trials, activity across the MD network increased, possibly reflecting increased cognitive effort when a context is set for a trial. Using MVPA, we showed that information about the behavioral status during the subsequent stimulus epoch of the trial was represented across the MD network. However, in contrast to our prediction, motivation, in the form of monetary reward, did not enhance the distinctions between the three behavioral status conditions across the MD network. Additionally, we did not find evidence for a selective facilitative effect of reward on discriminability of highly conflicting items (competition-contingent effect). In the LOC, information about the behavioral status of the presented stimuli was primarily driven by visual categories and was not modulated by motivation.

Previous reports showed an enhancement effect of motivation on overall activity in the frontoparietal control network (Padmala and Pessoa 2011; Dixon and Christoff 2012; Botvinick and Braver 2015). Recently, it was demonstrated that cue decoding increased with motivation (Etzel et al. 2016), and in particular when task rules change from one trial to another (Hall-McMaster et al. 2019). Whether reward also modulates the representation of task-related information that is processed while cue and stimulus information is integrated remained unclear. These two effects of reward are complementary to one another, and are both key aspects of cognitive control and essential when reaching a decision. If reward enhances cue coding, then it would be reasonable to hypothesize that it may also facilitate the integration process of the cue and the subsequent stimulus that leads to successful completion of the task. Furthermore, previous studies have suggested that motivation particularly affects conditions of high conflict. Padmala and Pessoa (2011) reported a decrease in interference with reward in response inhibition tasks. Reward also reduced incongruency effect in the Stroop task compared to non-rewarded trials (Krebs et al. 2013) and enhanced error monitoring (Stürmer et al. 2011). Thus, we predicted that reward motivation will enhance the representation of task information, and that this effect may be specific for highly conflicting

items. Based on our previous work that showed representation of behavioral status information across the frontoparietal cortex (Erez and Duncan 2015), we used three behavioral status levels and their distinctions to test these predictions. While the overall representation of behavioral status across the MD network replicated our previous findings (Erez and Duncan 2015), our results did not show an increase in representation with reward, in contrast to our predictions. Additionally, we did not observe a selective increase in representation for the highly conflicting items, namely Targets vs. High-conflict nontargets. Recently, Hall-McMaster et al., (2019) showed some increases in task-relevant stimulus features information when reward levels were high, using distributed patterns in EEG data. We did not observe such changes in our data, and the difference in results may possibly be due to differences in the design, as well as the limited time window where such differences were observed in EEG that cannot be detected with the low temporal resolution fMRI data. More generally, several reasons can provide potential explanations for the results obtained in our study, showing no facilitative effect of reward on pattern discriminability of behavioral status. Indeed, it is possible that the effect of reward is limited to cue decoding, as has been previously demonstrated (Etzel et al. 2016), and does not extend to the stimulus phase when information is processed based on the cue. However, we cannot rule out other possible explanations, including that our reward manipulation was not sufficiently strong to make a difference to pattern discriminability, insufficient power, and multiple factors that result in overall low decoding levels across the frontoparietal cortex (Bhandari et al. 2018) making small effects hard to detect with current MVPA methods.

Our predictions were based on the sharpening and prioritization account, which postulates that motivation leads to a sharpened neural representation of relevant information depending on the current task and needs. Previous neurophysiological evidence provide support for this aspect: reward has been associated with firing of dopaminergic neurons (Schultz et al. 1997; Bayer and Glimcher 2005), and dopamine has been shown to modulate tuning of prefrontal neurons and to sharpen their representations (Vijayraghavan et al. 2007; Thurley et al. 2008; Ott and Nieder 2016). The prioritization aspect can be related to the expected value of control (EVC) theory (Shenhav et al. 2013) and reward-based models for the interaction of reward and cognitive control, essentially a cost-benefit trade-off (Botvinick and Braver 2015). Cognitive control is effortful and hence an ideal system would allocate it efficiently, with a general aim of maximizing expected utility. Despite the appeal of this account, our results did not show experimental support for this view. At the behavioral level, we observed some

evidence for such a benefit of reward. Accuracy levels of performance in the task for Target trials were higher in the reward compared to the no-reward condition. Additionally, while in the no-reward condition Target trials were less accurate than Low-conflict nontargets, in the reward condition there were no differences between them. We did not observe a similar benefit in reaction times. This could be due to the time threshold that we used for reward trials, which reduced the reaction time on all the reward trials, and may have masked an interaction with reward.

Relatedly, we did not observe univariate differences in activity between targets and the two nontargets across the MD network, and in contrast to previously reported data (Hampshire et al. 2007, 2009). In our design, the targets were twice as frequent as High and Low-conflict nontargets, which could have resulted in lower activity of the targets compared to the other two due to a frequency effect (Braver et al. 2001; Hampshire et al. 2009). Another possible univariate effect could have been a higher activation for more demanding condition, the High-conflict nontargets. The target, frequency and difficulty effects together could have led to no univariate differences between the behavioral status conditions, as also seen in our previous work (Erez and Duncan 2015).

The visual categorization aspect of our task allowed us to investigate effects of reward on representation in LOC compared to the MD network, and in particular whether there is a specific effect of reward that is driven by visual differences. In the MD network, decoding levels were similar between conditions with the same visual category and different visual category, and there was no modulation by reward in any of them. This does not mean that there is no visual information in the MD network (Stokes et al. 2013), but rather that we did not observe it in our fMRI study. In contrast, the discrimination in LOC was driven by the visual categories, as expected in the visual cortex, with Targets and High-conflict nontargets being discriminable only when items belonged to two different visual categories. While it is widely agreed that the frontoparietal cortex exerts top-down effects on visual areas, there is no clear prediction as to whether any effects of reward should be observed in the visual cortex. Our results provide evidence that the effects of reward were not present in LOC. Although previous studies have shown differences in representations between pFs and LO (Jiang et al. 2007; Li et al. 2007; Harel et al. 2014), our results were similar for both regions. Although our primary question addressed the representation of task-related information during the integration of stimulus and cue, we also tested for an effect of reward in the cue epoch. The use of catch trials ensured that the cue and stimulus GLM regressors were

appropriately decorrelated. The overall univariate activity across the MD network increased with reward during the cue epoch, possibly reflecting an increase in cognitive effort due to the reward. However, we did not observe cue decoding above chance, in contrast to previous reports (Etzel et al. 2016; Hall-McMaster et al. 2019). One reason for this difference in decoding results may be related to the design of our task. We used words of the category names as cues, which appeared together at the same time with the no-reward/reward indication – the reward trial cues had additional red pound signs. Our cues allowed for high task performance (compared to using abstract symbols, which is more difficult) rather than maximizing the decodability between cues. The visually salient reward signal that was presented simultaneously with the cues may have masked the decoding of the cue. Other reasons for the different results may be differences in the areas that were used for the analysis (Etzel et al. 2016), and differences in patterns that can be identified using fMRI and EEG data (Hall-McMaster et al. 2019).

In summary, we asked whether reward motivation leads to increased representation of task-related information across the frontoparietal network and the LOC. We found that information about behavioral status was present across the MD network. However, in contrast to the prediction based on the sharpening and prioritization account of reward effects, we did not find an increase in representation levels with reward. In the LOC, we observed representation of behavioral status, but this was driven by visual category information and was not modulated by reward. With growing interest in the interaction between control processes and motivation, our study provides an important experimental evidence for the underlying neural mechanisms and the potential limitations of the effects of reward, thus contributes another tier to the accumulating knowledge in the field which is critical for the development of computational models.

Chapter 5

A new working memory signal in frontoparietal local field potentials

Introduction

The MD network is a set of frontoparietal regions involved in multiple cognitive tasks, as seen in functional imaging studies in humans. Mitchell and colleagues (2016) sought to identify a macaque counterpart to the human MD system using fMRI connectivity in 35 rhesus macaques under anaesthesia. Using whole brain connectivity analyses they identified seven clusters across frontoparietal and insular cortex comparable to human MD regions and one additional cluster in the lateral fissure. These regions include areas around the arcuate sulcus, principal sulcus and intraparietal sulcus, the regions usually chosen to study prefrontal and parietal neurons. Electrophysiology studies, unlike fMRI, use one task in which the animals are extensively trained. Using several tasks to isolate the MD regions, as in the fMRI definition of the network, would be very difficult. Thus, in order to study the function of prefrontal and parietal neurons underlying the MD network, a spatial working memory task was chosen.

Working memory has been studied extensively in behaving animals. In non-human primates, information maintenance has classically been linked to persistent, stable firing in neurons of the frontal and parietal cortex (Fuster and Alexander, 1971; Funahashi et al., 1989).

Although, in recent years much evidence has emerged of dynamic activity, with single neurons carrying information only for brief periods of a working memory delay (Barak et al., 2010; Schmitt et al., 2017), or more radically, changing their pattern of stimulus selectivity during different task operations (Warden and Miller, 2010; Rigotti et al., 2013; Stokes et al., 2013; Naya et al., 2017; Sigala et al., 2008). This is in line with the mixed-selectivity hypothesis – neurons encode conjunctions of task features (Rigotti et al., 2013).

Computational approaches have explained these dynamic firing rates using linear stable subspaces, emphasising the stable ensemble or population information of dynamic single neuron activities (Machens et al. 2010; Mendoza-Halliday and Martinez-Trujillo 2017; Murray et al. 2017) and modelled the same using linear attractor models and non-linear recurrent neural networks (Machens et al. 2005; Barak et al. 2013). Other accounts have also

proposed synaptic mechanisms for the so called ‘silent’ working memory, to explain information storage (Stokes, 2015; Mongillo et al., 2008; Lundqvist et al., 2016).

While the primary emphasis of this work has been spiking activity, much task-related information can also be decoded from features of the local field potential, extracted either in time or frequency domains (Buschman et al., 2012; Tremblay et al., 2015; Kim et al., 2018). In the frequency domain, oscillatory activity in the beta and gamma frequencies has often been linked to working memory load (Howard et al. 2003; Honkanen et al. 2015). An integration of frequency information with single trial time information was used by Lundqvist et al. (2016). In their gamma and beta bursts frequency model, they posit gamma bursts of variable time and frequency in single trials as a signature of encoding and re-activation of memory.

Compared to spiking of single neurons, the LFP is a more integrated signal, combining electrical activity over a region of tissue up to at least several mm (Buzsaki et al., 2012). It relates closely to signals from human brain imaging, including fMRI (Logothetis et al., 2001) and scalp-recorded EEG. The LFP is affected by many sources, conspicuously including excitatory and inhibitory synaptic inputs, but also by spikes, after-hyperpolarizations, changes in intrinsic membrane properties and more (Buzsaki et al., 2012; Marder et al., 1996). Given its integration of many signals, including hyperpolarizing synaptic inputs, the LFP can carry information invisible in spiking output (Kim et al., 2018). Existing interpretations and models of information storage in the context of dynamic neuronal firing, find a stable working memory trace in a network signal – often using dimensionality reduction methods to get an integrated signal (Machens et al. 2010; Mendoza-Halliday and Martinez-Trujillo 2017; Murray et al. 2017). LFP is one such signal that integrates multiple different types of electrical activity in a local circuit. Thus, LFPs could contain a more stable information trace.

Here we compared the strength and stability of information coding in spikes and LFPs, recorded from awake macaques carrying out a spatial working memory task. The animals were presented with a 5 choice array. Through a process of trial and error, the animals found one of the locations that led to positive feedback and reward. This was the explore phase of the problem. In three subsequent trials, they returned to that target location for further rewards – the exploit phase. Neural activity was recorded in prefrontal and inferior parietal cortex. We examined spikes and LFP signals for target location information in the exploit phase, where the animal knew the target. Crucially, we analysed LFP data not in the

frequency domain, as in much current work, but simply as a time-varying voltage recorded at each electrode. To study the dynamic nature of the signal, we correlated target location preference across time. Spiking activity patterns changed rapidly over time and phase of the trial, as expected. In contrast, in LFPs a strong signal of target location remained largely constant across time. This target information arose post positive feedback in the explore phase and remained stable throughout the exploit phase. These LFP recordings used either the recording chamber or the headpost as reference. Both these reference locations were further away from the electrodes, than is ideal. The more distant the reference, there is a greater chance of volume conduction to be a part of the signal. The choice of the reference voltage, as well as its possible impacts on the results are further investigated and discussed.

Methods

Subjects

Two male rhesus monkeys (*Macaca mulatta*), each weighing 13 kg performed the task. The experiments were performed in accordance with the Animals (Scientific Procedures) Act 1986 of the UK; all procedures were licensed by a Home Office Project License obtained after review by Oxford University's Animal Care and Ethical Review committee, and were in compliance with the guidelines of the European Community for the care and use of laboratory animals (EUVD, European Union directive 86/609/EEC).

Task

In each session, the animal completed a series of problems, each continuing until 4 rewards had been obtained (Figure 5.1A). In each trial the animal was shown a visual array of 5 squares or circles. Circles and squares were used in alternate problems throughout the session, and change in shape (along with additional cues, see below) served to indicate the onset of a new problem. In each problem, one location was randomly defined as the target. On each trial, the monkey selected one location by touching it. If the location touched on the current trial was the predefined target, it was followed by positive feedback and reward; and if it was a nontarget, the touch led to a negative feedback signal and no reward. Thus, to begin each problem, the animal had to work through a series of trials, sampling locations until the target was discovered. After target discovery, the animal was given further trials until the target location had been touched 3 more times, after which the problem ended and a new target location was randomly selected.

Details of events on each trial are illustrated in Figure 5.1B. Before the trial began, the screen showed the display of 5 black squares or circles (each 5.7 x 5.7 deg visual angle, centred 11.4 deg from fixation), accompanied by a central white fixation point (FP). Three conditions had to be met for onset of the trial - completion of an inter-trial interval, fixation at the centre of the screen (window 7.6 x 7.6 deg), and holding down a start key. When these conditions were met, the FP turned red, indicating the start of the trial. Following a delay of 1.2 to 2.0 s, the FP switched to cyan (go signal) to indicate that a response could be made. The animal then released the start key and touched one of the locations (touch required within 1.8 s of go signal). After the touch had been held for 0.35 to 0.45 s, the selected black square/circle turned either green (target) or red (nontarget). This feedback signal remained for 0.3 s, and if the touched location was the target, a drop of soft food (reward) was delivered 0.05 to 0.15 s

after feedback offset. The trial was aborted without reward if the monkey released the key or broke fixation prior to the go signal, or if screen touch was not held until feedback.

Different inter-trial intervals and displays indicated different stages of the problem. For successive trials before each reward, there was a minimum period of 0.7-0.9 s before the next trial could begin, with the screen showing the central white FP and display of location markers. After a rewarded trial, there was a period of only the white FP, lasting 3.2-3.5 s, followed by appearance of the location markers when the trial was initiated. To indicate the end of a problem, the screen remained blank for 3.3-3.6 s.

The period from 500 ms before to the trial onset (FP turning red) was the pre-fixation phase (Prefix). Only the white FP was displayed during this period in the exploit trials. The period from trial onset to 1 s after was the fixation phase (Fix). The location markers were displayed in this period and the animal was fixating at the centre. From the onset of the go signal to 1 s after was the movement phase (Mov), which included the animal releasing the start key and reaching to touch one of the 5 displayed locations. The period from the onset of the feedback signal to 1 s after was the feedback phase (Fb). This included a reward in all the trials, as only successful trials were analysed.

Task events were controlled by REX real-time data acquisition and laboratory control software (developed by the National Institutes of Health), with displays presented on a 17.5 inch LED touch screen placed in front of the animal's chair. The start key was attached to the front of the chair.

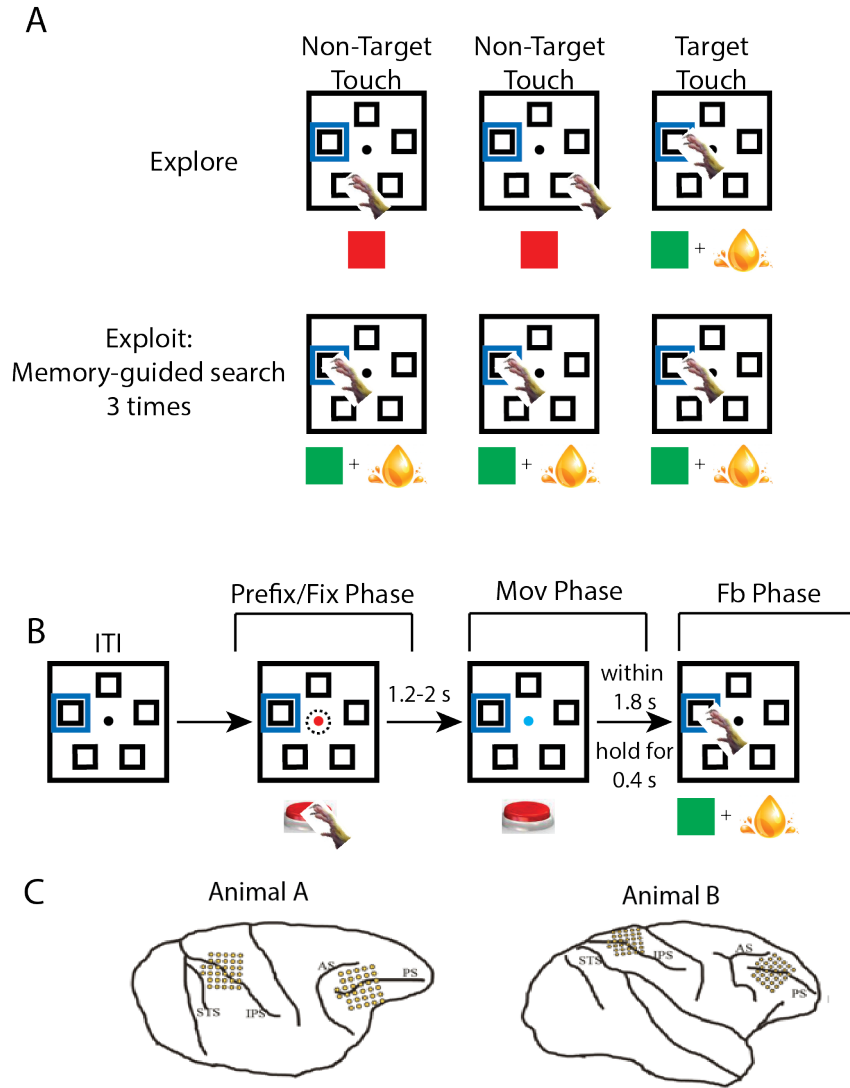


Figure 5.1: A. Spatial working memory task. The rewarded location is indicated by a blue square, not present on actual display. The animal initially finds this location through trial and error, i.e., the explore phase. It then selects this target location three more times in the exploit phase. **B.** The timeline of a single trial. The trial began when the animal achieved fixation (indicated by the dotted circle), and depressed the start key. The fixation point changed colour to red indicating this onset. After a variable delay of 1.2 – 2 s a go signal was presented, which was the fixation point changing colour from red to cyan. At this signal, the animal released the start key and touched one location. The trial was aborted if the touch was not made within 1.8 s of the go signal, and if the touch was not held for 0.4 s. Following this, feedback was presented by the touched location changing colour. Green indicated positive feedback, and red negative. Positive feedback was followed by reward. **C.** Recording locations for the two animals AS: arcuate sulcus; IPS: intraparietal sulcus; PS: principal sulcus; STS: superior temporal sulcus.

Recordings

Each animal was implanted with a titanium head holder and recording chambers (Gray Matter Research), fixed on the skull with stainless steel screws. Frontal chambers were placed over the lateral prefrontal cortex of the right hemisphere for both monkey A (AP = 33.9, ML = 20.3; AP, anterior-posterior; ML, medio-lateral) and monkey B (AP = 36.2, ML = 58.1). Posterior chambers were placed over the parietal cortex of the right hemisphere for both monkey A (AP = -4.6, ML = 50.6) and monkey B (AP = -3.2, ML = 47.4). Recording locations for each animal are shown in Fig. 5.1C. A craniotomy made under each chamber enabled physiological recording. All surgical procedures were aseptic and carried out under general anaesthesia.

Data were recorded over a total of 79 daily sessions, across both animals. For each chamber, a 32-channel semi-chronic microdrive system (SC-32, Gray Matter Research) was used, interfaced to a multi-channel data acquisition system (Cerebus System, Blackrock Microsystems). Within the microdrive there was an inter-electrode spacing of 1.5 mm. Between sessions, to ensure recording of new cells, electrodes were advanced by a minimum of 62.5 μ m. We advanced microelectrodes until we could isolate neuronal activity before starting the task and did not pre-select the neurons based on their task responses. LFP recordings from separate sessions were treated as separate data sets.

Neural activity was amplified, filtered (300 Hz to 10kHz), and stored for offline sorting and analysis (Offline Sorter, Plexon) to obtain spikes. Continuous data were also recorded separately from 0.3 – 250 Hz sampled at 10, 1 and 0.5 KHz. Different sampling rates were used on different sessions. The voltage on the chamber was used as reference. The broadband data were filtered online using a Butterworth filter. Eye position was sampled at 120 Hz using an infrared eye tracking system (Applied Science Laboratories).

At the end of the experiments, animals were deeply anaesthetized with barbiturate and then perfused through the heart with heparinized saline followed by 10% formaldehyde in saline. The brains were removed for histology and recording locations confirmed (Figure 5.1C).

Data and Analysis

All statistical analyses were carried out using MatLab (MathWorks), with custom-made scripts. The raw LFP voltages were down-sampled to 1 KHz. Only channels with isolated single units were included in the analysis. Second-order notch filters at 50, 100 and 150 Hz were used to remove line noise.

Major analyses used data just from successful exploit trials (in each problem, correct trials following first reward). For each trial, data were extracted for three periods: Prefix/Fix (-500 ms to +1000 ms from trial onset), Mov (0 to +1000 ms from go signal), and Fb (0 to +1000 ms from feedback signal). Mov and Fb periods were completely non-overlapping as the average time between the go signal and feedback onset was 1.221 s for animal A, 1.148 s for animal B. For LFPs, we analysed raw voltage, without subtraction of any baseline, as we were interested in sustained signals extending across trials.

For ANOVAs (Figure 5.3), four phases were used: the Prefix phase (from 0.3 to 0.1 s before the onset of the array), the Fix phase (from 0.1 to 0.3 s after the onset of the array), the Mov phase (from 0.1 to 0.3 s after the onset of the go signal), and the Fb phase (from 0.1 to 0.3 s after the onset of feedback). A 2-way ANOVA with target location (5 levels), and trial phase (4 levels: Prefix, Fix, Mov and Fb) was performed, and proportion of explained variance was measured by partial ω^2 , calculated by the formula:

$$\omega^2 = df_{\text{effect}} \times (MS_{\text{effect}} - MSE) / (SS_{\text{effect}} + (N_{\text{total}} - df_{\text{effect}}) \times MSE)$$

df_{effect} is degrees of freedom for the factor of interest (location, trial phase, interaction), MS_{effect} is the mean square for the factor, SS_{effect} is the sum of squares for the factor, MSE is the mean square error, and N_{total} is the total number of observations (trials). The partial ω^2 were tested against zero, and compared between location, trial phase and interaction (two-tailed t-test).

For the time-resolved correlation analysis, firing rates and raw voltages were estimated in 50 ms non-overlapping windows for Prefix/Fix (-500 ms to +1000 ms from trial onset), Mov (0 to +1000 ms from go signal), and Fb (0 to +1000 ms from feedback signal). They were first averaged across the exploit trials within each problem set and then an average per target location was computed. To correlate spatial preference at each time point with all other time points, we separated the data using an interleaved approach, i.e., dividing the data in session into odd and even problems. For a given pair of locations, we calculated the difference in firing rate/voltage for each neuron/channel, once in odd and once in even problems. We used only non-adjacent locations (1-3, 1-4, 2-4, 2-5, and 3-5) in order to maximise the location discrimination. Across all recorded neurons/channels at a given time window, this produced

two vectors of difference scores, one for odd and one for even problems. Correlation between these vectors for the same time point in the trial shows the reliability of location discrimination across the population, as ascertained by correlating independent sets of data. Correlation for different time points (t_1 , t_2) shows the ability to predict target preference at a time window t_2 , given the preference at t_1 . In other words, correlating different time points, measures stability of target location information across time and trial phases. The similarity of this pattern across time was calculated as a Pearson coefficient that was Fisher-transformed to average across all different non-adjacent location pairs. This average correlation matrix was converted back to Pearson coefficients for ease in interpretation.

A null distribution was estimated from 1000 permutations of randomly shuffling location labels and creating correlation matrices with the same procedure as detailed above. The 95% confidence interval was the threshold used to detect correlation above chance. This was corrected for multiple comparisons using FDR ($p < 0.05$). The LFP and spike correlation were compared using a Fisher r-to-z transformation. The z value was converted to a p-value and corrected for multiple comparisons using FDR ($p < 0.05$).

Results

Behaviour

We obtained data from two adult male rhesus monkeys. In the working memory task (Figure 5.1A), the animal had to discover which of 5 screen locations was the rewarded target location, and then on subsequent trials return to this target for repeated rewards. Animals were extensively trained, with close to perfect performance (Figure 5.2). Key release time (RT) was in the range 220-240 ms for animal A and 240-260 ms for animal B, without much change between explore and exploit phases or between the multiple rewards in the exploit phase. The time from key release to touch (movement time) was around 520 ms for the explore phase and 440 ms for exploit phase in both animals. Thus, movement time decreased in the exploit phases.

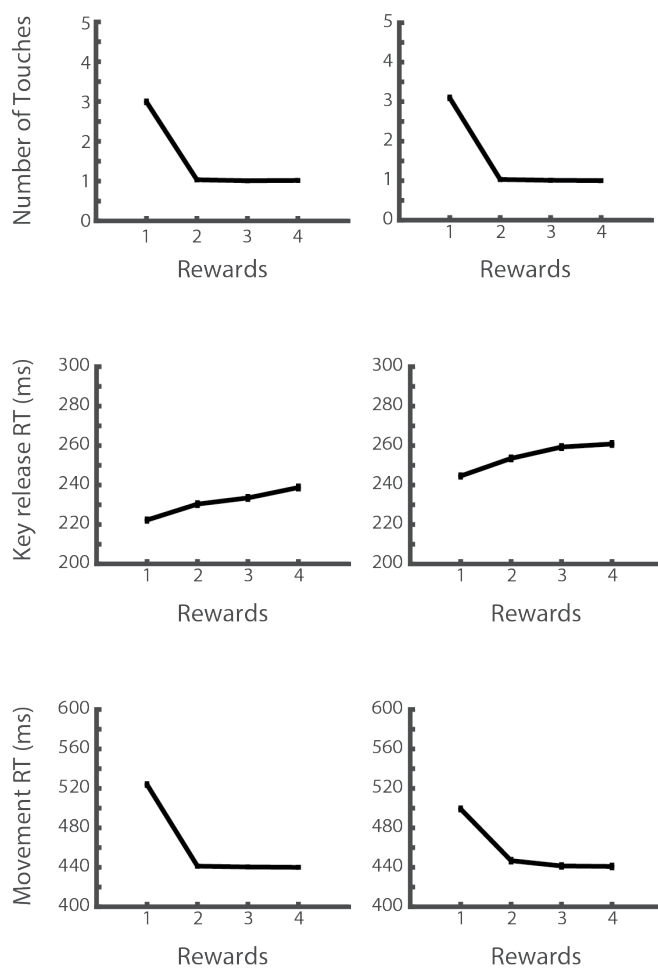


Figure 5.2: Behavioural Data. Top panel: mean number of trials (location touches) per reward. Middle panel: Mean reaction time (RT) in ms to release the key following go signal.

Bottom panel: Mean movement time in ms from key release to object touch. Left column shows data for animal A, and right column for animal B.

Pattern of coding in spikes and LFPs

In each animal, neuronal activity was recorded from the right pre-frontal cortex (PFC), and the right inferior parietal cortex (IPC) (Figure 5.1C). Additional recording sites in superior parietal cortex were aimed to record visuomotor activity, and not considered further in the analyses present in this chapter.

Major data analyses concerned just exploit trials, in which the animal already knew the target location and held it in working memory. Figures 5.3A and B show location tuning curves of six neurons, three each from PFC and IPC. The analysis was done on averaged trials of the memory-guided exploit phase where the animals accessed the previously found target location. Average firing rate in four time windows were calculated for each target location. Tuning curves were plotted by averaging the firing rates within four phases of the trial separately – the Prefix phase (from 0.3 to 0.1 s before the onset of the array), the Fix phase (from 0.1 to 0.3 s after the onset of the array), the Mov phase (from 0.1 to 0.3 s after the onset of the go signal), and the Fb phase (from 0.1 to 0.3 s after the onset of feedback). The highest average firing rate of the five locations was used as an index for target preference. The location with the highest firing rate in the Prefix phase was plotted in the centre of all the tuning curves. This allowed for visual comparison of tuning across trial phases. The locations in array were ordered from 1 to 5 in a clockwise fashion beginning with 1 at the top.

Most neurons showed a change in their tuning curve between phases of the trial, as seen in the first two example neurons in both recorded areas (Figures 5.3A, B). The location information encoded in the firing rates of the neurons was dependent on the trial phase. Very few neurons, such as the last neuron in PFC, showed phase-independent location information. For each of the neurons, an ANOVA was done with location and trial phase as factors. The p-value for location and the interaction of location and trial phase is reported above each plot. The first neurons in both areas show an overall location effect, as well as an interaction effect. The pattern suggests that these neurons are particularly active in one phase of the trial only (here, feedback). The second neurons in PFC and IPC show no overall effect of location and a significant interaction. These show pure conjunction coding, with dissimilar tuning across trial phases. They are active in multiple phases, but prefer different locations in each.

The last neuron in Figure 5.3A shows a location effect, but no interaction effect. This neuron had similar tuning in all trial phases.

A similar analysis was done with LFPs (Figures 5.3C, D). To examine spatial tuning in LFPs, we measured mean voltage at each channel (electrode) across the same analysis windows. The tuning curves for 3 example single channels in PFC (Figure 5.3C) and IPC (Figure 5.3D) were plotted using the most negative voltage during Prefix as preferred location. Many channels showed a similar tuning curve across the phases indicating a stable information code throughout the trial, as seen in the first two channels from both areas. In the first channel from both areas, there is sharp location tuning in all trial phases, without much change in target preference across them. In the second channels, again there is a location effect, with the same location being the most negative one in all trial phases. However, there is an overall pattern of trial phase coding; the tuning curves are transposed, without much change in shape. There were a few channels that did have a more dynamic location preference, resembling the neurons, as seen in the last channel in PFC.

This trend in example units and channels was quantitatively examined by performing an ANOVA on 200ms time windows with locations (5 levels) and trial phases (4 levels: Prefix, Fix, Mov and Fb) for each neuron and channel. Neurons in PFC commonly showed significant effects of trial phase (69% of cells), location (33%), and crucially their interaction (28%). Data for IPC were similar (78, 29, and 31% respectively for phase, location and interaction). Corresponding analyses for LFP data showed a very different pattern, with main effects of location much stronger than interactions. In PFC, 83, 74 and 21% of channels showed significant effects of phase, location and interaction. In IPC, 81, 72 and 19% of channels showed significant effects of phase, location and interaction. Thus the trial phase was the most strongly encoded information in both neurons and single channels. A similar proportion of neurons encoded location and the interaction. However, many more LFP channels encoded location and comparatively few channels encoded the interaction.

The same pattern was also seen in the population average partial omega squared for each factor: location, trial phase and their interaction. Spikes in both PFC and IPC (Figures 5.3E, F) showed significant trial phase information, as measured by partial omega squared (PFC: $\omega^2 = 0.07$, $p < 0.001$; IPC: $\omega^2 = 0.11$, $p < 0.001$), location information (PFC: $\omega^2 = 0.03$, $p < 0.001$; IPC: $\omega^2 = 0.02$, $p < 0.001$), as well as an interaction of the two (PFC: $\omega^2 = 0.03$, $p < 0.001$; IPC: $\omega^2 = 0.03$, $p < 0.001$). Trial phase was higher than location and interaction effects ($p < 0.001$; for both) in PFC and in IPC ($p < 0.001$; for both). Importantly, there was no

difference between the location and interaction effects in PFC ($p > 0.05$) and slightly higher interaction effect compared to location effect in IPC ($p < 0.05$). This interaction of location and trial phase information in single units suggests a non-linear integration of the two signals. LFPs in both recording regions (Figures 5.3G, H) also showed both trial phase (PFC: $\omega^2 = 0.18$, $p < 0.001$; IPC: $\omega^2 = 0.25$, $p < 0.001$) and location information (PFC: $\omega^2 = 0.07$, $p < 0.001$; IPC: $\omega^2 = 0.07$, $p < 0.001$), but no interaction of the two (PFC: $\omega^2 = 0.01$, $p = 0.15$; IPC: $\omega^2 = 0.001$, $p = 0.63$). Similar to spikes, trial phase was higher than location and interaction effects ($p < 0.001$; for both) in PFC and in IPC ($p < 0.001$; for both). However, unlike spikes, there was a much higher location effect compared to its interaction with trial phase in PFC ($p < 0.001$) and in IPC ($p < 0.001$). The lack of this interaction in channels indicates an additive signal that changed due to both location and phase, with a linear combination of the two.

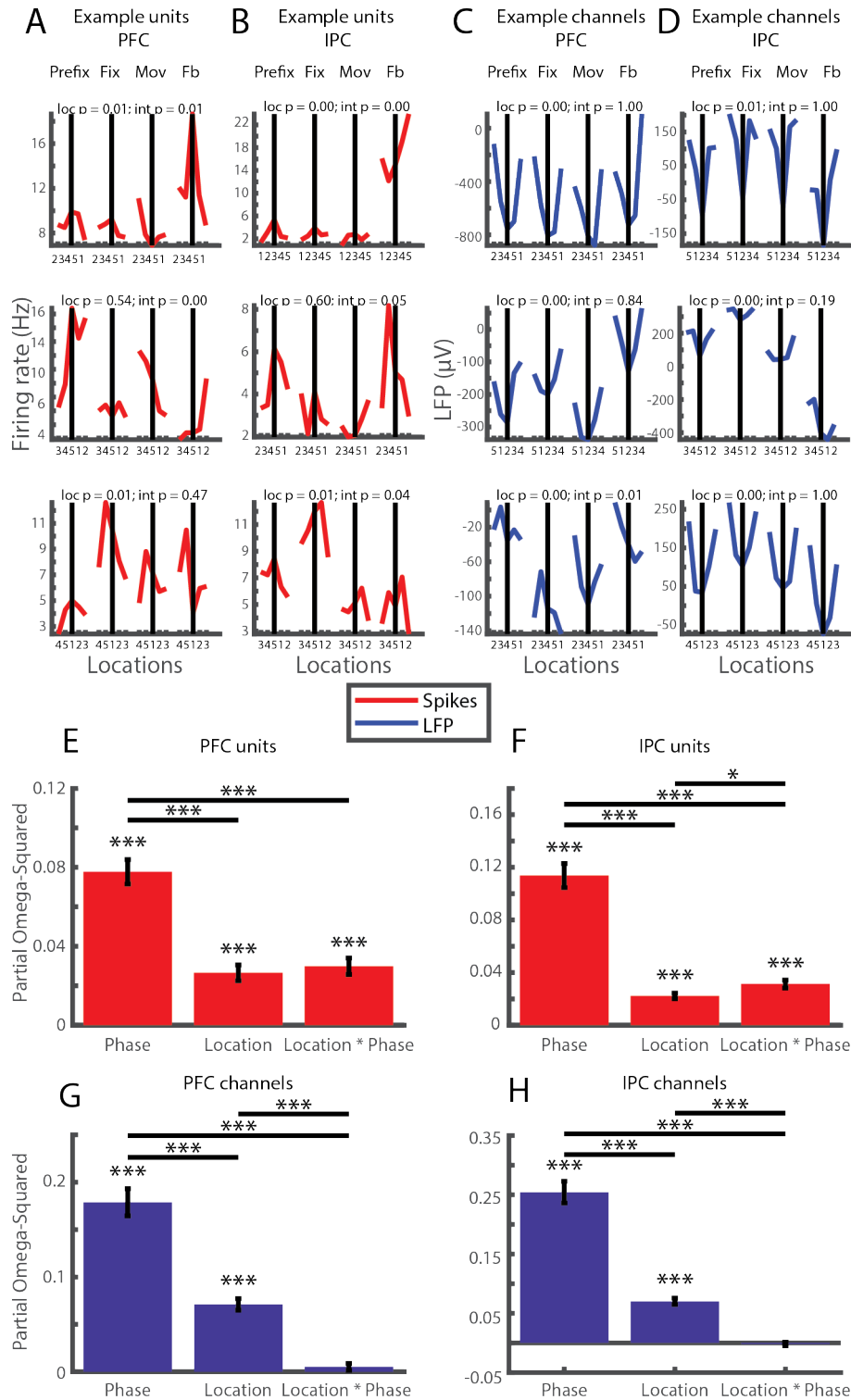


Figure 5.3: **A.** Tuning curves of example units from PFC. Firing rates for four task phases are ordered by target location, with locations in the array labelled 1 to 5 clockwise beginning at top. For each unit, the location with highest firing during Prefix is plotted in the centre for all task phases. **B.** Same as A, for IPC. **C, D.** LFP tuning curves of example single channels in PFC and IPC. Location with most negative voltage during Prefix is plotted in the centre. **E, F.** Average partial ω^2 values for trial phase, location and interaction averaged across units.

Asterisks indicate significant partial ω^2 as well as differences between them. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. **G, H.** Same as E-F for LFP channels.

Stability of information across time

For more detailed examination of temporal stability at the population level, we used a cross-temporal correlation analysis, performed on 50 ms non-overlapping windows covering -500 to +1000 ms from fixation onset, 0 to +1000 ms from go signal, and 0 to +1000 ms from feedback signal. For each neuron and time point, spike data were split into two halves (odd and even problems in the session). For a given pair of locations X and Y, we calculated the difference in firing rate (X-Y) for each neuron, once in odd and once in even problems.

Across all recorded neurons at a given time window, this produced two vectors of difference scores, one for odd and one for even problems. Correlation between these vectors for the same time point in the trial shows the reliability of location discrimination across the whole cell population, while correlation for different time points shows the stability of location coding. Results averaged across location pairs (see Methods) are shown in Figures 5.4A (PFC) and B (IPC), with results for an equivalent analysis of LFPs in Figures 5.4C, D.

For spike data, correlation matrices show a strong diagonal, indicating reliable location discrimination at each time point but little generalisation across time points (Figures 5.4A, B). Temporally specific location coding appeared in a burst following fixation onset, and again during Mov and Fb periods following go signal onset and feedback onset respectively. Though LFP correlation matrices also show a strong diagonal (Figures 5.4C, D), this was accompanied by a broad pattern of positive correlations throughout the matrix, especially spanning the whole Prefix and Fix period.

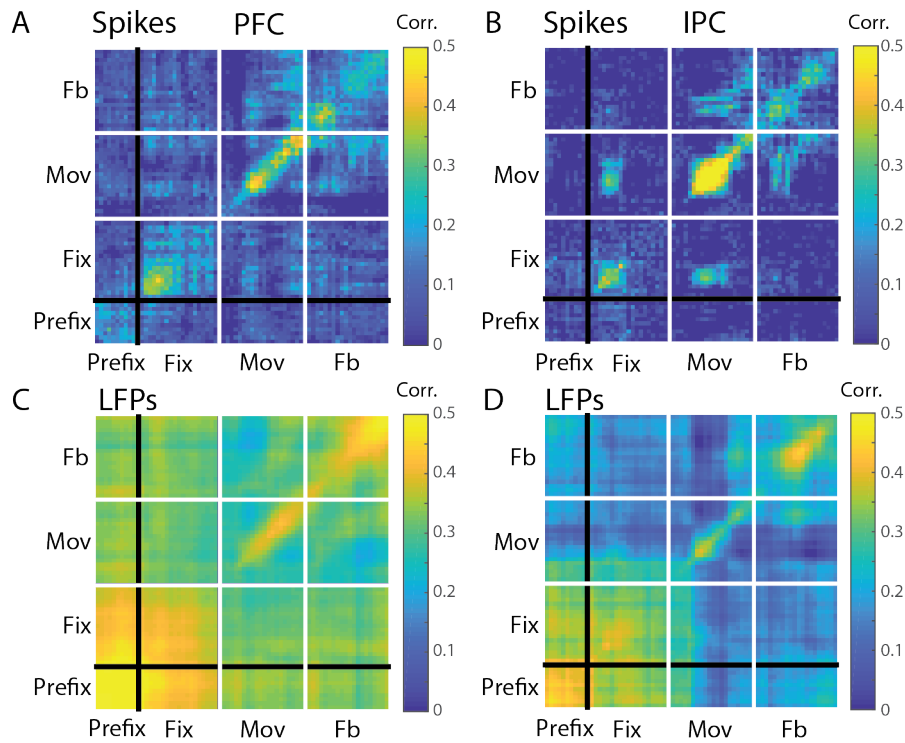


Figure 5.4: **A.** Cross-temporal correlation matrix of target preference across time for PFC spikes. Data have been split into two halves (odd and even trials), and for each half of the data, vectors of location preference for one time window are correlated (Pearson's r) with vectors from the other half of the data, both for the same time point (major diagonal) and all other time points (off diagonal). To make the matrix symmetrical, data have been averaged for the two possible directions of analysis (odd trials time X correlated with even time Y, and the reverse). **B.** As for A, IPC spikes. **C, D.** Same as A, B for LFP data.

To compare these patterns quantitatively we chose four 50 ms reference time windows, Prefix (250 to 200 ms before fixation onset), Fix (200 to 250 ms after fixation onset), Mov (200 to 250 ms after go signal), and Fb (200 to 250 ms after feedback onset). For each reference window, data from Figures 5.4A-D are replotted in Figures 5.5A, B, now showing odd-even vector correlations between the reference time window and all windows across the trial. A Prefix reference plot would be one horizontal or vertical line in the big correlation matrix, corresponding to a Prefix 50 ms time window (250 to 200 ms before fixation onset). As seen in Figures 5.4A-B, for the Prefix period there was little location information in spikes (odd-even correlation close to zero even for the reference period itself). LFPs in contrast showed strong location coding even in the Prefix window, and location preferences even in this early period of the trial remained predictive of preferences through the whole

trial. For other reference time windows, the reliability of information (odd-even correlation for the reference period itself) was comparable for spikes and LFPs, but again, LFP patterns were significantly more stable (Figures 5.5A, B). The red bar below each plot shows significant spike correlation, corrected for multiple comparisons using FDR. This red bar shows significant correlation during Fix, Mov and Fb reference periods, and cross-generalisation of target location information to some other time windows as well. In contrast, LFPs showed very stable target information across the entire trial in both areas, as seen by the blue bars below each plot. Similarly the green bars show that this information is significantly higher in LFPs at many time points, especially in frontal cortex.

Origin of the memory trace

For LFPs, high levels of location selectivity in the pre-trial period as well as its maintenance throughout the trial pose the question of when this information arises in the network. We investigated this by examining location preference on the last explore trial, when the animal was first given positive feedback on touching the target location. In Figures 5.5C and D, a reference vector from the Prefix period in exploit trials, as used in Figures 5.5A, B, was correlated with vectors across the time course of exploit trials (same data as Figures 5.5A, B), but also backwards in time into the final explore trial. For spikes, of course, correlations remained low across explore and exploit trials, since spikes contained little spatial information in the reference period. For LFPs, correlation with the reference period rose rapidly following feedback on the final explore trial, when the animal first learned which location to enter into working memory.

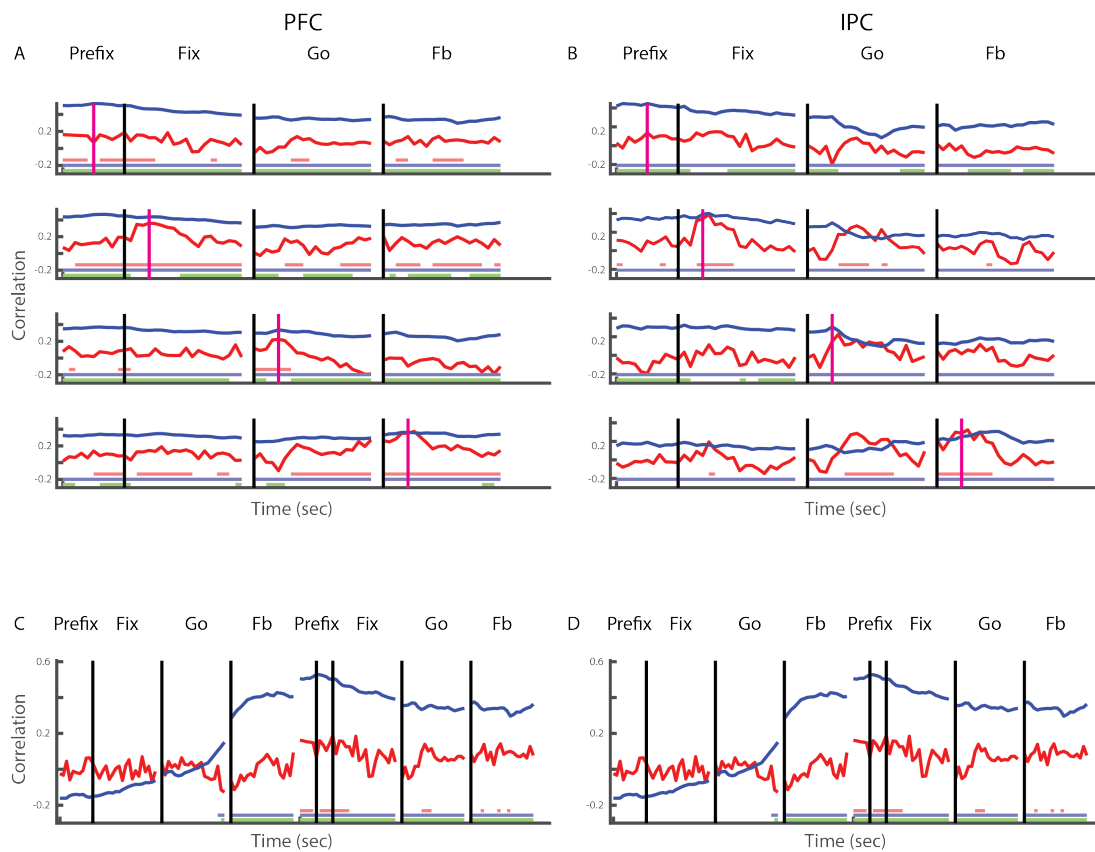


Figure 5.5: Comparison of correlation between spikes and LFPs. **A.** Correlation of target preference across time in PFC for four reference time windows from each trial phase, Prefix (250 to 200 ms before fixation onset), Fix (200 to 250 ms after fixation onset), Mov (200 to 250 ms after go signal), Fb (200 to 250 ms after feedback onset). The spikes are the red plots and LFPs are blue. Black vertical lines indicate trial onset, go signal and feedback onset. Purple line indicates reference window. Red bar below each plot indicates significant periods of above-chance cross-temporal classification, corrected for multiple comparisons using FDR ($p < 0.05$). Blue bar indicates the same for LFPs. Green bar shows periods of significantly larger LFP correlations compared to spikes. Overall the correlation across time during the entire trial was higher in LFPs than spikes. **B.** Same as A for IPC. **C.** Analysis similar to A, extended back into final explore trial. Reference window is in the Prefix (250 to 200 ms before fixation onset) window of the exploit phase as in the top panel of A. The correlation in LFPs in the exploit phase can be traced back to the feedback phase of the last explore trial, when the target location is first discovered.

Discussion

In this study we have examined two different signals in awake macaques during a spatial working memory task: firing rates of single units and LFPs. The single neurons showed dynamic target location coding, where the firing rates were representative of an interaction of target location and trial phase. This suggests that the working memory is encoded such that its contents can be readily used for different behaviours, as appropriate in the different phases of the trial. The LFPs on the other hand, showed stable target information across time and trial phase. This is suggestive of an additive signal, encoding both the target location and trial phase, and not their interaction. The integrated activity of a local circuit of neurons reflected in the LFPs provided a background of stable information, in the context of dynamic single neurons.

The neuronal results here were not very consistent with the classical idea of persistent firing underlying frontoparietal working memory signals (Fuster 1973; Funahashi et al. 1989; Romo et al. 1999; Constantinidis et al. 2018). Rather, the firing patterns fit the idea of adaptive coding (Duncan, 2001) and ‘mixed selectivity’ (Sigala et al. 2008; Warden and Miller 2010; Rigotti et al. 2013; Stokes et al. 2013; Naya et al. 2017). The mixed selectivity hypothesis suggests that prefrontal neurons code for a conjunction of task features in a high dimensional space (Rigotti, 2013), here a conjunction of target location and trial phase. The neurons have to perform different spatial computations, accompanied by different visual inputs and eye and hand behaviour, at different stages of the trial. These functions are possible due to nonlinear or conjunctive coding of location and trial phase.

Dynamic coding in spikes, however, contrasts strongly with stable coding in LFPs, where despite changes in cognitive operations, visual inputs, and behaviour, target location signals remained largely constant. To the best of our knowledge, this is the first study of its kind to investigate working memory in time-varying voltages recorded from macaques. Kim et al. (2018) recorded spikes and cell membrane potentials from pyramidal neurons in mouse PFC. In their task, different stimuli predicted rewards at different delays. Predicted delay was reflected in sustained changes of membrane potential, without associated changes in firing rate. This suggests that certain cognitive operations like reward predictions can be realised via synaptic mechanisms. Though LFPs integrate electrical signals of many kinds, one prominent influence is synaptic input. Critically, LFPs showed high levels of target information in the pre-trial period, which had very little spiking activity. It is conceivable that

in LFPs, we record stable location as well trial phase information, reflecting the synaptic input to a local circuit of neurons that may or may not lead to spiking. In the extracellular matrix, location and phase could be largely additive, meaning that location coding is stable over the different visual inputs, cognitive operations and motor activity of the trial.

The presence of target location information before the trial starts does not in itself confirm a working memory signal. In order to ascertain the nature of this signal, its origin was traced. This signal began at the feedback phase of the last exploit trial, where the animal learned the target location (Figures 5C, D). The origin of this signal at the discovery of the target suggests that it represents a working memory signal, and it remained stable throughout the exploit phase. Linking the strength of this signal to behaviour would have been a valuable addition in examining the function of such a stable signal. However, the task design did not allow for that. Future studies are needed to investigate if LFPs are an epiphenomenon or are indeed needed for behaviour.

In frontal and parietal cortex, our data show very different behaviour for spikes and LFPs. In Figure 5.6 we sketch a possible model accounting for spike and LFP data in these terms. Four example neurons represent an area of tissue contributing to LFP recorded at a single electrode. In each neuron, inputs are received on multiple dendrites. The cell fires only when conjoined location and phase inputs are received on the same dendrite, reflecting the ability of single dendrites to carry out nonlinear AND operations (Hausser et al. 2000; Polsky et al. 2004; Stuart and Spruston 2015). The model assumes that location inputs are sustained throughout the trial, while different phases of the trial corresponding to different segments of behaviour are captured by separate phase inputs.

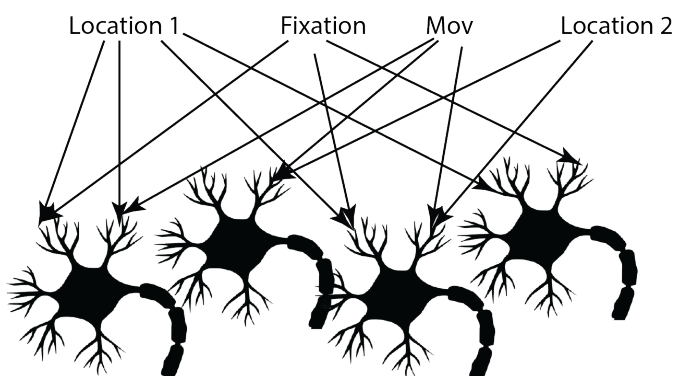


Figure 5.6: Model of location and phase coding in spikes and LFPs. Four neurons illustrate an area of tissue contributing to LFP at one electrode. Location inputs (here just locations 1

and 2) are assumed stable over the trial, while phase inputs (here just Fix and Mov) arrive at corresponding trial periods. In each neuron, inputs are nonlinearly combined (conjunctive input to a single dendrite) to generate spiking output.

A sustained location input reflects a stable working memory signal, arising outside the recorded region and thus not seen in spike data. Though the sources of location and phase inputs are unknown, there are many candidates. Given that we did not find sustained location selectivity in either PFC or IPC, two of the cortical regions most closely linked to working memory, a subcortical source for a stable location signal is one strong possibility. Multiple studies particularly studying the interplay between medio-dorsal thalamus and prefrontal cortex in the context of working memory (Fuster and Alexander 1971; Funahashi et al. 2004; Watanabe and Funahashi 2012), would suggest this region as a candidate region for tonic location input. Fronto-thalamic substrates of cognitive flexibility were investigated in a task-switching task in mice (Rikhye et al. 2018). Thalamic responses showed sustained context-relevant representations, which could be the sustained location input here. The phase input allows this stable signal to be conjoined with a signal indicating what should currently be done with this location information – corresponding to trial segments of fixate, reach and receive feedback. PFC and IPC cells control behaviour by conjoining such multiple sources of input to determine required cognitive and behavioural output.

The model sketched in Figure 5.6 fits several aspects of our data. First, different net synaptic input from two locations across the area of integration for LFPs recorded at one electrode, leading to different net responses to these locations (cf. Figures 5.3C, D). A particular channel may receive more of an e.g. ‘location 1’ input, resulting in a preference for ‘location 1’. The same is true for net input for different task phases. Assuming a stable location input across task phases, from a region such as medio-dorsal thalamus, the location preference would remain constant throughout the trial. Second, there can be different types of integration at single neurons as seen in Figure 5.6. For any given neuron, location preference can be stable across task phases (leftmost neuron). Here the same location is being integrated with different trial phase information on two different dendrites; the neuron would fire for the same location at different trial phases. The second neuron from the left shows firing restricted to a single phase. The third neuron from the left shows different locations being integrated with different task phases, changing location preference in different trial phases. The model is

also consistent with previous data showing how patterns of PFC spiking radically change across successive phases of a task (Sigala et al., 2008).

Recent debates have focused on two complementary mechanisms for working memory – sustained neural firing and short-term synaptic change (Constantinides et al., 2018; Lundquist et al., 2018). Our data add strong evidence for a third – a stable cortical LFP signal in the context of highly dynamic cortical spiking. A stable LFP signal is significant because it relates closely to signals from human neuroimaging, including fMRI and EEG, linking our findings to information decoding from these modalities (King and Daehene, 2014; Kriegeskorte et al., 2008). By combining stable with transient inputs, cortex may provide the critical conjunctive operations needed to change sustained working memory contents into time-varying behaviour.

Reference Voltage for LFP

The above LFP analysis was conducted with headpost as the reference voltage in most sessions and the chamber as the reference voltage in the rest. The reference was subtracted from each channel by the recording system during the session, in an online manner. The chamber and headpost are further away than for example, the guide tube, which is a more robust choice as the reference voltage. We redid the analysis with an average reference (Tremblay et al., 2015). It involved calculating the reference as the average of all electrodes recorded in that session, done offline. This was not straightforward as there were sessions with fewer than 4 channels, which were excluded from this analysis, owing to a large contribution of each channel to the average. Due to low number of channels per session overall, the session average was also not an ideal reference to use for this data. We also used one channel as a reference per session. This reference channel was not included in the analysis, to keep the reference independent. These were done to test for the efficacy of the previously used reference. Both these forms of references led to deviations from the results detailed above. The magnitude and the stability of the correlations decreased, making them similar to the spikes, contrary to the main analysis (Figure 5.5). The population ANOVA was also similar to the spikes suggesting no increased location information compared to the location and trial phase interaction information as seen before (Figure 5.3).

These follow-up analyses question the extent to which the LFP interpretations with headpost and chamber reference hold. More broadly, they raise the question of which is the best reference voltage to use. Ideally the reference voltage should be close enough to capture all

the noise, artifacts, and common signal reaching all the channels in any given session; but far enough away as to not capture the signal local to any channel. Given the low number of active channels in each session, the post hoc references such as average channel and a single channel are not ideal. Latest recording systems use either the guide tube as the reference (Mendoza-Halliday & Martinez-Trujillo, 2017) or an additional wire inserted durally or subdurally. This reference is subtracted from each of the channels in the guide tube during recording. Had the results from all the references converged, we would be more confident in the LFP results. With differences between the different referencing schemes and no ideal reference, our conclusions remain uncertain. While the current dataset may not be ideal to address the question of the stability of LFPs, it remains an interesting and important question that has not been examined in much detail. Future studies are needed to evaluate if LFPs are indeed physiologically relevant in working memory.

Discussion

Together, the results from the above experimental chapters show an integrated multiple - demand system, responding as a network to a wide variety of task demands. Chapter 2 addresses the question of functional differences between the different regions of the network. Functional differences were more prominent in simple tasks and these observable differences were reduced in the more challenging and demanding conditions (Figures 2.3 and 2.5). The more posterior regions being active in the simple tasks would be in line with previous studies suggesting increasingly anterior activations with increased rule complexity (Badre and D'Esposito 2007, 2009; Badre 2008). However, with increased complexity, time pressure and reward, the activity did not localise to more anterior regions, but instead showed stronger recruitment of the entire MD network. One possibility is that the connectivity between these MD regions could be heightened with increased task demands, allowing faster and better information exchange in the more demanding tasks. In a recent study by Soreq et al. (2019), there was increased functional connectivity in the frontoparietal cortex with increasing working memory load, particularly in the maintenance period. More such studies measuring functional connectivity between particular MD regions during task performance, while varying the demand can better explain the interplay between the various MD regions and how that changes with increasing task demand.

Functional differences between the nodes of the network are evident in all different types of task demands. In Chapter 2, all of the contrasts examined: simple tasks versus baseline, complex tasks versus simple tasks, time pressure versus simple maze and reward versus time pressure, showed an ROI effect. There were also ROI differences when comparing all the conditions together. Similarly in Chapter 3, across the three localiser tasks – spatial working memory, verbal working memory and Stroop – there was a main effect of ROI. In Chapter 4 there were again differences between ROIs in the way they responded to reward. Thus quantitative differences in activity do exist between the different regions of the MD network. However, this does not go against the network theory. Different ROIs could in part be having separate functions, but the interconnections between them allow for information exchange and a unified network function.

Different task demands elicit different levels of activity of the MD regions. Comparing the activation for the three contrasts in Chapter 2 (Figures 2.3 and 2.4) showed that reward manipulation most strongly recruits the MD network, followed by time pressure and then

complexity. Similarly in Chapter 3, the two working memory tasks led to a stronger overall MD pattern compared to the Stroop task (Figures 3.2 and 3.3). This could be due to differential increases in task demand in the different contrasts. It could also reflect differential recruitment of the network by different task demands. Future studies are needed that vary demand parametrically and allow direct comparison of different demands to further address this question.

MD regions showed a strong increase in univariate activity with reward in a spatial navigation and the cue phase of a cued categorisation task. In the stimulus phase of the cued categorisation task, AI was the only region that showed a reward effect in post hoc tests using the rather strict Bonferroni correction. AI has previously been implicated in reward function as well (Amiez et al. 2016). There were ROI differences in the amount of activity elicited by reward in both tasks. This is line with some previous studies showing widespread frontoparietal involvement in the interaction of motivation and cognitive control (Padmala and Pessoa 2011; Dixon and Christoff 2012; Botvinick and Braver 2015). There have been suggestions of particular MD regions being involved in reward like the preSMA/ACC (Hadland et al. 2003; Rushworth et al. 2004; Shenhav et al. 2013). One possibility could be a specialised reward-related preSMA/ACC function of determining the amount of control required for any task, based on the reward expected and the cost in terms of cognitive effort (Shenhav et al. 2013). This information is shared with the rest of the network, all regions contributing to increased control in order to secure the increased reward.

Chapter 3 examined these multiple-demand regions at the voxel level. Continuing from the previous chapter that investigated functional differences between large regions, this study looked at task preferences within regions. In single subjects, the voxels responding to multiple different task demands were largely contained within an independent MD group template. Sets of voxels with the highest activity in different task demands all showed similar performance in an independent MVPA analysis. These chosen voxels however, did not show greater information than using all the voxels in the group mask. Previously using subject-specific localisation has shown improved univariate results (Brett, et al. 2002; Saxe et al. 2006; Fedorenko et al. 2010; Nieto-Castañón and Fedorenko 2012). This was also shown in our data, univariate contrasts in the independent criterion task were improved by voxel selection using any of the three functional localisers (Figure 3.4). On the other hand, voxels that did not show a particular functional signature relevant to the discriminated conditions contributed to classification accuracy (Haxby et al. 2001a; Kriegeskorte et al. 2006). For

example, Haxby et al. (2001a) showed that removing voxels from the ventral visual regions that respond most strongly to some visual category did not strongly affect the ability to discriminate that category from other categories. Consequently, selecting the MD voxels in single subjects might not offer the same advantage to pattern analysis as it did to univariate analysis. We do however show evidence for multiple-demand voxels isolated through one task demand generalising to represent task-relevant information in a completely different task.

Overlapping voxels processing task demand and being involved in task processing as shown by their representation of the task-relevant features provides further evidence for the MD function of a general problem solver. This was used in Chapter 4 by investigating how the task representational space changes with increase in reward. In particular the integrated cue and stimulus information was measured, i.e., the discrimination between behavioural status – target, high-conflict nontarget, and low-conflict nontarget. Voxels were selected for each subject based on voxels most active in an independent functional localiser. Here the direct effect of motivation on representation was examined on voxels that were involved in an independent task demand. None of the discriminations between conditions increased with increasing reward, suggesting that the reward effect could be limited to the sharpening in the cue phase.

Previously the increase in target and distractor discrimination with increased perceptual complexity was shown (Woolgar et al. 2015). What underlies the difference in mechanism that leads to such an increase in discrimination in task-relevant information with perceptual complexity, but not so in the case of reward, is not yet known. It is possible that reward is more of a preparatory signal, while other types of task demand lead to changes during task solving, measurable as changes in the representational space.

MD regions show an increase of cue or task set representation with reward (Etzel et al. 2016). They represent the behavioural status of the stimuli in the cued categorisation task, and not the task irrelevant visual category of the stimuli (Erez and Duncan 2015). Chapter 4 showed that MD regions discriminate between target and a high-conflicting nontarget but this discrimination does not increase with reward (Figure 4.3). How the increased cue representation and reward-related univariate activity translates to increased behavioural accuracy or performance is an open question. Kusunoki and colleagues (2010) showed opponent coding of sets of neurons either preferring targets or nontargets that are never cued; with the high-conflicting nontargets' firing rate being in between the two. Reward could

sharpen the tuning curves of individual neurons, thus pushing the target and high-conflicting nontargets further apart, a mechanism similar to the one mediating feature-based attention (Maunsell and Treue 2006). This would affect the discrimination between targets and low-conflicting nontargets as well. However, we did not see any significant increases in discrimination with reward (Figure 4.3). An identical mechanism for reward value and selective mechanism for attention is seen in V1 neurons (Stănişor et al. 2013). Thus, a sharpening effect of the tuning curves could still underlie the increased discrimination of targets and nontargets in the MD regions with reward and increased attention; but the overall low decoding accuracies (Bhandari et al., 2018) and possibly low power in study did not allow us to see this pattern in the representational space as measured by fMRI.

The intricate mechanisms mediating the complex functionality of the MD regions can be studied through responses of single neurons. Chapter 5 shows how single neurons code an interaction of target location and trial phase, allowing the target location information to be used differently in different phases of the trial. The neurons in these frontal and parietal regions integrate the task information non-linearly and provide hints of how flexible cognition may be achieved. Neuronal activity being dependent on a conjunction of task features has been shown before in a variety of tasks (Sigala et al. 2008; Warden and Miller 2010; Rigotti et al. 2013; Stokes et al. 2013; Naya et al. 2017). In the model shown in figure 5.6, we propose one possible mechanism for this integration. Tonic location and trial phase information comes in as input into the prefrontal and parietal regions, possibly from medio-dorsal thalamus (Fuster and Alexander 1971; Funahashi et al. 2004; Watanabe and Funahashi 2012; Rikhye et al. 2018). Inputs falling on the same dendrite are integrated, leading to spiking. If the different trial phases are integrated with the same location, then the target preference remains stable throughout the trial (Figure 5.3A, bottom neuron). Instead, if different target locations are integrated with different phases of the trial, it leads to more dynamic coding of the target (Figures 5.3A and B). One can assume a similar mechanism for the integration of cue and stimulus information in the cued categorisation task in Chapter 4. Dopamine regulating motivation, reward and learning has shown to be concentrated in dendritic spines (Yao et al. 2008; Schultz 2015). Reward could thus modulate the spatial and temporal summation of information at the dendrites through dopamine, thus affecting task processing and behaviour.

This integration mechanism speaks to the broader function of the MD network. Flexible cognition requires integration of different pieces of task-relevant information. The same

stimuli can lead to opposite responses depending on the context or rule. For example, in the cued categorization task, depending on the cue the same stimuli can result in opposite responses. Thus, mixed selectivity or adaptive coding in the prefrontal and parietal neurons (Duncan 2001; Rigotti et al. 2013) as seen in the results of Chapter 5 (Figures 5.3A, B, E, F), could allow such flexibility in behaviour.

In a working memory task, where is the target information maintained if neurons encode the target for short periods of time only? Assuming the tonic inputs are being integrated at the dendrites, the electrical signal in the extracellular matrix – the LFPs – should reflect a linear summation of the inputs. LFPs represent an additive operation on the task variables, thus showing stable task information through time and trial phases (Figures 5.4C, D). A major contributor to these signals is the synaptic transmembrane current, though other sources include spikes, ionic fluxes, and intrinsic membrane oscillations etc. Thus these potentials could be reflecting the synaptic input to a local neuronal network in the form of excitatory and inhibitory potentials that may or may not lead to spiking. The information in the LFP is seen after the target location is discovered, in the absence of spikes, and continues to remain stable as long as it is accessed (Figures 5.5C, D). The question remains as to whether this signal is an epiphenomenon or if it contributes to behaviour. Rat cortical pyramidal neurons in slices were stimulated and recorded from to investigate the relationship between LFPs and spikes (Anastassiou et al. 2011). They found that LFPs could strongly entrain action potentials, particularly for slow (< 8 Hz) fluctuations of the extracellular field. These results suggest that LFPs could impact spiking, through ephaptic coupling. A coordination of spike timing in hippocampal neurons to the local theta oscillation underlies successful memory formation in humans (Rutishauser et al. 2010). The electric fields surrounding the neurons, that are influenced by spiking, can in turn influence neuronal function (O'Keefe and Recce 1993; Womelsdorf et al. 2007; Kayser et al. 2009; Anastassiou et al. 2010, 2011), suggesting a possible mechanism for emergent properties in the brain.

However these LFP results were tempered by the fact they were not replicable with the average reference and a single channel reference scheme. In the semi-chronic recording system, there are few electrodes in any given session making these post-hoc referencing schemes far from ideal. On the other hand, the lower stability of correlations with these references could suggest that the stability comes from a common signal that affects all channels in a given session. Instead of a local linear summation of signals in the PFC, the stable signal may originate elsewhere. One possibility is the medio-dorsal thalamus that is

shown to have contextual information thalamus (Fuster and Alexander 1971; Funahashi et al. 2004; Watanabe and Funahashi 2012; Rikhye et al. 2018). The nature of the LFP signal provides an interesting insight into information coding and could be involved in behaviour as well. Future studies are needed to determine if LFPs in PFC carry a more stable signal than spikes.

Another reason to study LFPs is its close relationship to the BOLD signal (Logothetis et al. 2001; Goense and Logothetis 2008). LFPs have a bearing on the univariate and multivariate results from fMRI studies. Investigating the coding of face viewpoint and identity information using MVPA of both the BOLD response and single-unit recordings in macaques showed only viewpoint information in fMRI, and both viewpoint and identity were encoded by single neurons (Dubois et al. 2015). The lack of identity information in the BOLD could be due to the identity neurons being less clustered, producing a more distributed signal in the extracellular matrix. Future studies are needed to find the transfer function that can infer the electrophysiological mechanisms from information representation in fMRI voxels. This will help integrate the knowledge we have from fMRI studies of the MD network, with the underlying characteristics of the neurons and neural networks.

This thesis has examined the MD network and its function as the general problem solver at various spatial scales. The network handles various different task demands including working memory, inhibition, spatial navigation, time pressure, reward etc. Being a distributed frontoparietal network, its nodes get input from various sources allowing for partial functional specialisation. Its strong interconnections allow for information exchange and functioning of these regions as a network. The mechanism underlying this flexible cognition is the ability to select and process information relevant to the task, and modify task representation / processing based on task demands such as complexity, cognitive effort, reward etc. This flexibility is made possible by integration of both stable and varying inputs from cortical and subcortical regions in order to maintain and utilise the information appropriately for goal-directed behaviour. Integration at the neuronal level and communication between the nodes at the cortical region level can generate cognitive control needed for diverse task performance.

Bibliography

- Ackerly SS, Benton AL. 1948. Report of case of bilateral frontal lobe defect. *Res Publ Assoc Res Nerv Ment Dis.* 27:479–504.
- Ahlheim C, Love BC. 2018. Estimating the functional dimensionality of neural representations. *Neuroimage.* 179:51–62.
- Alexander MP, Stuss DT, Shallice T, Picton TW, Gillingham S. 2005. Impaired concentration due to frontal lobe damage from two distinct lesion sites. *Neurology.* 65:572–579.
- Alnaes D, Sneve MH, Espeseth T, Endestad T, van de Pavert SHP, Laeng B. 2014. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. *J Vis.* 14:1–20.
- Amiez C, Wutte MG, Faillenot I, Petrides M, Burle B, Procyk E. 2016. Single subject analyses reveal consistent recruitment of frontal operculum in performance monitoring. *Neuroimage.* 133:266–278.
- Anastassiou CA, Montgomery SM, Barahona M, Buzsáki G, Koch C. 2010. The effect of spatially inhomogeneous extracellular electric fields on neurons. *J Neurosci.* 30:1925–1936.
- Anastassiou CA, Perin R, Markram H, Koch C. 2011. Ephaptic coupling of cortical neurons. *Nat Neurosci.* 14:217–223.
- Aron AR, Robbins TW, Poldrack RA. 2004. Inhibition and the right inferior frontal cortex. *Trends Cogn Sci.* 8:170–177.
- Assem M, Glasser MF, Essen DC Van, Duncan J. 2019. A Domain-general Cognitive Core defined in Multimodally Parcellated Human Cortex. *bioRxiv.* 517599.
- Badre D. 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci.* 12:193–200.
- Badre D, D’Esposito M. 2007. Functional Magnetic Resonance Imaging Evidence for a Hierarchical Organization of the Prefrontal Cortex. *J Cogn Neurosci.* 19:2082–2099.
- Badre D, D’Esposito M. 2009. Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci.* 10:659–669.

- Badre D, Nee DE. 2018. Frontal Cortex and the Hierarchical Control of Behavior. *Trends Cogn Sci.* 22:170–188.
- Bandettini P, Ungerleider LG. 2001. From neuron to BOLD: New connections. *Nat Neurosci.* 4:864–866.
- Barak O, Sussillo D, Romo R, Tsodyks M, Abbott LF. 2013. From fixed points to chaos: Three models of delayed discrimination. *Prog Neurobiol.* 103:214–222.
- Barak O, Tsodyks M, Romo R. 2010. Neuronal population coding of parametric working memory. *J Neurosci.* 30:9424–9430.
- Barbas H, Pandya DN. 1989. Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J Comp Neurol.* 286:353–375.
- Bartra O, McGuire JT, Kable JW. 2013. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage.* 76:412–427.
- Bayer HM, Glimcher PW. 2005. Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron.* 47:129–141.
- Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B.* 57:289–300.
- Berman MG, Park J, Gonzalez R, Polk TA, Gehrke A, Knaffla S, Jonides J. 2010. Evaluating functional localisers: The case of the FFA. *Neuroimage.* 50:56–71.
- Bhandari A, Gagne C, Badre D. 2018. Just above Chance: Is It Harder to Decode Information from Prefrontal Cortex Hemodynamic Activity Patterns? *J Cogn Neurosci.* 30:1473–1498.
- Blank I, Fedorenko E. 2017. Domain-general brain regions do not track linguistic input as closely as language-selective regions. *J Neurosci.* 37:9999–10011.
- Blank I, Kanwisher N, Fedorenko E. 2014. A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *J Neurophysiol.* 112:1105–1118.
- Bode S, Haynes J-D. 2009. Decoding sequential stages of task preparation in the human brain. *Neuroimage.* 45:606–613.
- Botvinick M, Braver TS. 2015. Motivation and Cognitive Control: From Behavior to Neural

- Mechanism. *Annu Rev Psychol.* 66:80–113.
- Brainard DH. 1997. The Psychophysics Toolbox. *Spat Vis.* 10:433–436.
- Brass M, Yves von Cramon D. 2004. Decomposing Components of Task Preparation with Functional Magnetic Resonance Imaging. *J Cogn Neurosci.* 4:609–620.
- Braver TS. 2012. The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn Sci.* 16:106–113.
- Braver TS, Reynolds JR, Donaldson DI. 2003. Neural Mechanisms of Transient and Sustained Cognitive Control during Task Switching. *Neuron.* 39:713–726.
- Brett M, Anton J-L, Valabregue R, Poline J-B. 2002a. Region of interest analysis using an SPM toolbox. *Neuroimage.* 16.
- Brett M, Anton J-L, Valabregue R, Poline J-B. 2002b. Region of interest analysis using an SPM toolbox - Abstract Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. *Neuroimage.*
- Brett M, Johnsrude IS, Owen AM. 2002. The problem of functional localization in the human brain. *Nat Rev Neurosci.* 3:243–249.
- Bugatus L, Weiner KS, Grill-Spector K. 2017. Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage.* 155:437–449.
- Bullmore ET, Sporns O. 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci.* 10:186–198.
- Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK. 2012. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron.* 76:838–846.
- Buschman TJ, Miller EK. 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science.* 315:1860–1862.
- Buzsáki G, Anastassiou CA, Koch C. 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat Neurosci.* 13:407–420.
- Camilleri JA, Müller VI, Fox PT, Laird AR, Hoffstaedter F, Kalenscher T, Eickhoff SB. 2018. Definition and characterization of an extended multiple-demand network. *Neuroimage.* 165:138–147.
- Campbell AW. 1904. Histological Studies on the Localisation of Cerebral Function. *J Ment*

- Sci. 50:651–662.
- Carlén M. 2017. What constitutes the prefrontal cortex? *Science*. 358:478–482.
- Carlin JD, Kriegeskorte N. 2017. Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Comput Biol*. 13:e1005604.
- Cavada C, Goldman-Rakic PS. 1989. Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *J Comp Neurol*. 287:422–445.
- Chafee M V., Crowe DA. 2012. Thinking in spatial terms: decoupling spatial representation from sensorimotor control in monkey posterior parietal areas 7a and LIP. *Front Integr Neurosci*. 6:112.
- Chafee M V., Goldman-Rakic PS. 1998. Matching Patterns of Activity in Primate Prefrontal Area 8a and Parietal Area 7ip Neurons During a Spatial Working Memory Task. *J Neurophysiol*. 79:2919–2940.
- Chao LL, Knight RT. 1995. Human prefrontal lesions increase distractibility to irrelevant sensory inputs. *Neuroreport*. 6:1605–1610.
- Chiew KS, Braver T. 2014. Dissociable influences of reward motivation and positive emotion on cognitive control. *Cogn Affect Behav Neurosci*. 14:509–529.
- Clithero JA, Rangel A. 2014. Informatic parcellation of the network involved in the computation of subjective value. *Soc Cogn Affect Neurosci*. 9:1289–1302.
- Cole MW, Bagic A, Kass R, Schneider W. 2010. Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. *J Neurosci*. 30:14245–14254.
- Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. 2014. Intrinsic and Task-Evoked Network Architectures of the Human Brain. *Neuron*. 83:238–251.
- Cole MW, Etzel JA, Zacks JM, Schneider W, Braver TS. 2011. Rapid Transfer of Abstract Rules to Novel Contexts in Human Lateral Prefrontal Cortex. *Front Hum Neurosci*. 5:142.
- Cole MW, Ito T, Braver TS. 2016. The Behavioral Relevance of Task Information in Human Prefrontal Cortex. *Cereb Cortex*. 26:2497–2505.
- Cole MW, Reynolds JR, Power JD, Repovs G, Anticevic A, Braver TS. 2013. Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat Neurosci*. 16:1348–

1355.

- Cole MW, Schneider W. 2007. The cognitive control network: Integrated cortical regions with dissociable functions. *Neuroimage*. 37:343–360.
- Constantinidis C, Funahashi S, Lee D, Murray JD, Qi X-L, Wang M, Arnsten AFT. 2018. Persistent Spiking Activity Underlies Working Memory. *J Neurosci*. 38:7020–7028.
- Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL. 1998. A common network of functional areas for attention and eye movements. *Neuron*. 21:761–773.
- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci*. 3:215–229.
- Corbetta M, Shulman GL, Miezin FM, Petersen SE. 1995. Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*. 270:802–805.
- Cosman JD, Lowe KA, Zinke W, Woodman GF, Schall JD. 2018. Prefrontal Control of Visual Distraction. *Curr Biol*. 28:414–420.e3.
- Craik FIM, Bialystok E. 2006. Cognition through the lifespan: mechanisms of change. *Trends Cogn Sci*. 10:131–138.
- Critchley HD, Mathias CJ, Josephs O, O’Doherty J, Zanini S, Dewar B, Cipolotti L, Shallice T, Dolan RJ. 2003. Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence. *Brain*. 126:2139–2152.
- Critchley HD, Wiens S, Rotshtein P, Öhman A, Dolan RJ. 2004. Neural systems supporting interoceptive awareness. *Nat Neurosci*. 7:189–195.
- Crittenden BM, Duncan J. 2014. Task Difficulty Manipulation Reveals Multiple Demand Activity but no Frontal Lobe Hierarchy. *Cereb Cortex*. 24:532–540.
- Crittenden BM, Mitchell DJ, Duncan J. 2015. Recruitment of the default mode network during a demanding act of executive control. *Elife*. 4:e06481.
- Crittenden BM, Mitchell DJ, Duncan J. 2016. Task Encoding across the Multiple Demand Cortex Is Consistent with a Frontoparietal and Cingulo-Opercular Dual Networks Distinction. *J Neurosci*. 36:6147–6155.
- Crowe DA, Averbeck BB, Chafee M V., Georgopoulos AP. 2005. Dynamics of Parietal Neural Activity during Spatial Cognitive Processing. *Neuron*. 47:885–891.

- Crowe DA, Chafee M V., Averbeck BB, Georgopoulos AP. 2004. Neural Activity in Primate Parietal Area 7a Related to Spatial Analysis of Visual Mazes. *Cereb Cortex*. 14:23–34.
- Curtis CE, Cole MW, Rao VY, D’Esposito M. 2005. Canceling Planned Action: An fMRI Study of Countermanding Saccades. *Cereb Cortex*. 15:1281–1289.
- Cusack R, Vicente-Grabovetsky A, Mitchell DJ, Wild CJ, Auer T, Linke AC, Peelle JE. 2014. Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. *Front Neuroinform*. 8:90.
- de la Vega A, Chang LJ, Banich MT, Wager TD, Yarkoni T. 2016. Large-Scale Meta-Analysis of Human Medial Frontal Cortex Reveals Tripartite Functional Organization. *J Neurosci*. 36:6553–6562.
- Deacon T. 1997. *The Symbolic Species*. W. W. Norton & Company, Inc.
- Dehaene S, Sergent C, Changeux J-P. 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc Natl Acad Sci USA*. 100:8520–8525.
- Desimone R, Duncan J. 1995. Neural Mechanisms of Selective Visual Attention. *Annu Rev Neurosci*. 18:193–222.
- Dixon ML, Christoff K. 2012. The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence. *PLoS One*. 7:e51637.
- Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. 2008. A dual-networks architecture of top-down control. *Trends Cogn Sci*. 12:99–105.
- Dosenbach NUF, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RAT, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE. 2007. Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci USA*. 104:11073–11078.
- Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. 2006. A Core System for the Implementation of Task Sets. *Neuron*. 50:799–812.
- Dove A, Pollmann S, Schubert T, Wiggins CJ, Yves von Cramon D. 2000. Prefrontal cortex activation in task switching: an event-related fMRI study. *Cogn Brain Res*. 9:103–109.
- Downar J, Crawley AP, Mikulis DJ, Davis KD. 2002. A Cortical Network Sensitive to

- Stimulus Salience in a Neutral Behavioral Context Across Multiple Sensory Modalities. *J Neurophysiol.* 87:615–620.
- Dubois J, de Berker AO, Tsao DY. 2015. Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *J Neurosci.* 35:2791–2802.
- Duncan J. 2001. An adaptive coding model of neural function in prefrontal cortex. *Nat Rev Neurosci.* 2:820–829.
- Duncan J. 2006a. EPS Mid-Career Award 2004: Brain mechanisms of attention. *Q J Exp Psychol.* 59:2–27.
- Duncan J. 2006b. EPS mid-career award 2004: Brain mechanisms of attention. *Q J Exp Psychol.* 59:2–27.
- Duncan J. 2010a. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci.* 14:172–179.
- Duncan J. 2010b. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci.* 14:172–179.
- Duncan J. 2013a. The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron.* 80:35–50.
- Duncan J. 2013b. The Structure of Cognition: Attentional Episodes in Mind and Brain. *Neuron.* 80:35–50.
- Duncan J, Owen AM. 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23:475–483.
- Eger E, Ashburner J, Haynes J, Dolan RJ, Rees G. 2008. Functional Magnetic Resonance Imaging Activity Patterns in Human Lateral Occipital Complex Carry Information about Object Exemplars within Category. *J Cogn Neurosci.* 20:356–370.
- Erez Y, Duncan J. 2015. Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *J Neurosci.* 35:12383–12393.
- Erez Y, Yovel G. 2014. Clutter Modulates the Representation of Target Objects in the Human Occipitotemporal Cortex. *J Cogn Neurosci.* 26:490–500.
- Ester EF, Sprague TC, Serences JT. 2015. Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron.* 87:893–905.

- Etzel JA, Cole MW, Zacks JM, Kay KN, Braver TS. 2016. Reward Motivation Enhances Task Coding in Frontoparietal Cortex. *Cereb Cortex*. 26:1647–1659.
- Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci USA*. 110:16616–16621.
- Fedorenko E, Hsieh P-J, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. 2010. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J Neurophysiol*. 104:1177–1194.
- Fedorenko E, Nieto-Castañón A, Kanwisher N. 2012. Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia*. 50:499–513.
- Feinberg DA, Moeller S, Smith SM, Auerbach E, Ramanna S, Gunther M, Glasser MF, Miller KL, Ugurbil K, Yacoub E. 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS One*. 5:e15710.
- Felleman DJ, Van Essen DC. 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*. 1:1–47.
- Fox MD, Snyder AZ, Barch DM, Gusnard DA, Raichle ME. 2005. Transient BOLD responses at block transitions. *Neuroimage*. 28:956–966.
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. 2005a. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA*. 5:9673–9678.
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. 2005b. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA*. 5:9673–9678.
- Freedman DJ, Assad JA. 2006. Experience-dependent representation of visual categories in parietal cortex. *Nature*. 443:85–88.
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK. 2001. Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science*. 291:312–316.
- Funahashi S, Bruce CJ, Goldman-Rakic PS. 1989. Mnemonic Coding of Visual Space in the Monkey's Dorsolateral Prefrontal Cortex. *J Neurophysiol*. 61:331–349.
- Funahashi S, Takeda K, Watanabe Y. 2004. Neural mechanisms of spatial working memory:

- Contributions of the dorsolateral prefrontal cortex and the thalamic mediodorsal nucleus. *Cogn Affect Behav Neurosci.* 4:409–420.
- Fuster JM. 1973. Unit Activity in Prefrontal Cortex During Delayed-Response Performance: Neuronal Correlates of Transient Memory. *J Neurophysiol.* 36:61–78.
- Fuster JM. 2002. Frontal lobe and cognitive development. *J Neurocytol.* 31:373–385.
- Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. *Science.* 173:652–654.
- Garrison KA, Rogalsky C, Sheng T, Liu B, Damasio H, Winstein CJ, Aziz-Zadeh LS. 2015. Functional MRI preprocessing in lesioned brains: Manual versus automated region of interest analysis. *Front Neurol.* 6:196.
- Gergelyfi M, Jacob B, Olivier E, Zénon A. 2015. Dissociation between mental fatigue and motivational state during prolonged mental activity. *Front Behav Neurosci.* 9:176.
- Gilbert CD, Sigman M. 2007. Brain States: Top-Down Influences in Sensory Processing. *Neuron.* 54:677–696.
- Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. 2016. A multi-modal parcellation of human cerebral cortex. *Nature.* 536:171–178.
- Goense JBM, Logothetis NK. 2008. Neurophysiology of the BOLD fMRI Signal in Awake Monkeys. *Curr Biol.* 18:631–640.
- Greenberg AS, Esterman M, Wilson D, Serences JT, Yantis S. 2010. Control of Spatial and Feature-Based Attention in Frontoparietal Cortex. *J Neurosci.* 30:14330–14339.
- Gregoriou GG, Gotts SJ, Zhou H, Desimone R. 2009. High-Frequency, Long-Range Coupling Between Prefrontal and Visual Cortex During Attention. *Science.* 324:1207–1209.
- Guest O, Love BC. 2017. What the success of brain imaging implies about the neural code. *Elife.* 6:e21397.
- Hadland KA, Rushworth MFS, Gaffan D, Passingham RE. 2003. The Anterior Cingulate and Reward-Guided Selection of Actions. *J Neurophysiol.* 89:1161–1164.
- Hampshire A, Highfield RR, Parkin BL, Owen AM. 2012. Fractionating Human Intelligence. *Neuron.* 76:1225–1237.

- Hampshire A, Thompson R, Duncan J, Owen AM. 2011. Lateral prefrontal cortex subregions make dissociable contributions during fluid reasoning. *Cereb Cortex*. 21:1–10.
- Hampton AN, O'Doherty J. 2007. Decoding the neural substrates of reward-related decision making with functional MRI. *Proc Natl Acad Sci USA*. 104:1377–1382.
- Harel A, Kravitz DJ, Baker CI. 2014. Task context impacts visual object processing differentially across the cortex. *Proc Natl Acad Sci USA*. 111:E962–E971.
- Haxby J V, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001a. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293:2425–2430.
- Haxby J V, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001b. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex
Downloaded from. *Science*. 293:2425–2430.
- Haynes J-D, Rees G. 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci*. 7:523–534.
- Haynes J-D, Sakai K, Rees G, Gilbert S, Frith CD, Passingham RE. 2007. Reading Hidden Intentions in the Human Brain. *Curr Biol*. 17:323–328.
- Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM. 2018. The representational dynamics of task and object processing in humans. *Elife*. 7:e32816.
- Hebart MN, Gorgen K, Haynes J-D. 2015. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front Neuroinform*. 8:88.
- Hebb DO. 1939. Intelligence in Man after Large Removals of Cerebral Tissue: Report of Four Left Frontal Lobe Cases. *J Gen Psychol*. 21:73–87.
- Hebb DO. 1945. Man's frontal lobes. *Arch Neurol Psychiatry*. 54:10.
- Hebb DO, Penfield W. 1940. Human behavior after extensive bilateral removal from the frontal lobes. *Arch Neurol Psychiatry*. 44:421.
- Helfrich RF, Fiebelkorn IC, Szczepanski SM, Lin JJ, Parvizi J, Knight RT, Kastner S. 2018. Neural Mechanisms of Sustained Attention Are Rhythmic. *Neuron*. 99:854–865.e5.
- Helfrich RF, Knight RT. 2016. Oscillatory Dynamics of Prefrontal Cognitive Control. *Trends Cogn Sci*. 20:916–930.

- Hill J, Inder T, Neil J, Dierker D, Harwell J, Van Essen DC. 2010. Similar patterns of cortical expansion during human development and evolution. *Proc Natl Acad Sci USA*. 107:13135–13140.
- Höller-Wallscheid MS, Thier P, Pomper JK, Lindner A. 2017. Bilateral recruitment of prefrontal cortex in working memory is associated with task demand but not with age. *Proc Natl Acad Sci USA*. 114:E830–E839.
- Holroyd CB, Ribas-Fernandes JJF, Shahnazian D, Silvetti M, Verguts T. 2018. Human midcingulate cortex encodes distributed representations of task progress. *Proc Natl Acad Sci USA*. 115:6398–6403.
- Honkanen R, Rouhinen S, Wang SH, Palva JM, Palva S. 2015. Gamma Oscillations Underlie the Maintenance of Feature-Specific Information and the Contents of Visual Working Memory. *Cereb Cortex*. 25:3788–3801.
- Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities (associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices). *Proc Natl Acad Sci USA*. 79:2554–2558.
- Howard MW, Rizzuto DS, Caplan JB, Madsen JR, Lisman J, Aschenbrenner-Scheibe R, Schulze-Bonhage A, Kahana MJ. 2003. Gamma Oscillations Correlate with Working Memory Load in Humans. *Cereb Cortex*. 13:1369–1374.
- Ibos G, Freedman DJ. 2017. Sequential sensory and decision processing in posterior parietal cortex. *Elife*. 6:e23743.
- Ikkai A, Curtis CE. 2011. Common neural mechanisms supporting spatial working memory, attention and motor intention. *Neuropsychologia*. 49:1428–1434.
- Jackson J, Rich AN, Williams MA, Woolgar A. 2017. Feature-selective Attention in Frontoparietal Cortex: Multivoxel Codes Adjust to Prioritize Task-relevant Information. *J Cogn Neurosci*. 29:310–321.
- Jiang X, Bradley E, Rini RA, Zeffiro T, Vanmeter J, Riesenhuber M. 2007. Categorization training results in shape- and category-selective human neural plasticity. *Neuron*. 53:891–903.
- Jiang Y, Kanwisher N. 2003. Common Neural Mechanisms for Response Selection and Perceptual Processing. *J Cogn Neurosci*. 15:1095–1110.

- Julian JB, Fedorenko E, Webster J, Kanwisher N. 2012. An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage*. 60:2357–2364.
- Kadohisa M, Petrov P, Stokes MG, Sigala N, Buckley MJ, Gaffan D, Kusunoki M, Duncan J. 2013. Dynamic Construction of a Coherent Attentional State in a Prefrontal Cell Population. *Neuron*. 80:235–246.
- Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci*. 8:679–685.
- Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 17:4302–4311.
- Katsuki F, Qi X-L, Meyer T, Kostelic PM, Salinas E, Constantinidis C. 2014. Differences in Intrinsic Functional Organization Between Dorsolateral Prefrontal and Posterior Parietal Cortex. *Cereb Cortex*. 24:2334–2349.
- Kayser C, Montemurro MA, Logothetis NK, Panzeri S. 2009. Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. *Neuron*. 61:597–608.
- Kennerley SW, Wallis JD. 2009. Reward-dependent modulation of working memory in lateral prefrontal cortex. *J Neurosci*. 29:3259–3270.
- Khaligh-Razavi SM, Kriegeskorte N. 2014. Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Comput Biol*. 10.
- Kim E, Bari BA, Cohen JY. 2018. Reward-predictive persistent membrane potential dynamics in prefrontal cortex. *Progr No 50729/CCC14 2018 Neurosci Meet Planner San Diego, CA Soc Neurosci 2018 Online*.
- King JR, Dehaene S. 2014. Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn Sci*. 18:203–210.
- Knutson B, Taylor J, Kaufman M, Peterson R, Glover GH. 2005. Distributed Neural Representation of Expected Value. *J Neurosci*. 25:4806–4812.
- Koechlin E, Ody C, Kouneiher F. 2003. The Architecture of Cognitive Control in the Human Prefrontal Cortex. *Science*. 302:1181–1185.
- Koechlin E, Summerfield C. 2007. An information theoretical approach to prefrontal

- executive function. *Trends Cogn Sci.* 11:229–235.
- Krall SC, Rottschy C, Oberwelland E, Bzdok D, Fox PT, Eickhoff SB, Fink GR, Konrad K. 2015. The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Struct Funct.* 220:587–604.
- Krebs RM, Boehler CN, Appelbaum LG, Woldorff M. 2013. Reward Associations Reduce Behavioral Interference by Changing the Temporal Dynamics of Conflict Processing. *PLoS One.* 8:e53894.
- Krebs RM, Boehler CN, Roberts KC, Song AW, Woldorff M. 2012. The involvement of the dopaminergic midbrain and cortico-striatal-thalamic circuits in the integration of reward prospect and attentional task demands. *Cereb Cortex.* 22:607–615.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA.* 103:3863–3868.
- Kriegeskorte N, Kyle Simmons W, F Bellgowan PS, Baker CI. 2009. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci.* 12:535–540.
- Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci.* 2:4.
- Krishnan S, Slavin MJ, Tran T-TT, Doraiswamy PM, Petrella JR. 2006. Accuracy of spatial normalization of the hippocampus: Implications for fMRI research in memory disorders. *Neuroimage.* 31:560–571.
- Kruglanski AW, Shah JY, Fishbach A, Friedman R, Woo Young Chun, Sleeth-Keppler D. 2002. A theory of goal systems. *Adv Exp Soc Psychol.* 34:331–378.
- Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. 2010. Target Detection by Opponent Coding in Monkey Prefrontal Cortex. *J Cogn Neurosci.* 22:751–760.
- Lafer-Sousa R, Conway BR, Kanwisher N. 2016. Color-Biased Regions of the Ventral Visual Pathway Lie between Face-and Place-Selective Regions in Humans, as in Macaques. *J Neurosci.* 36:1682–1697.
- Leavitt ML, Mendoza-Halliday D, Martinez-Trujillo J. 2017. Sustained Activity Encoding Working Memories: Not Fully Distributed. *Trends Neurosci.* 40:328–346.
- Leon MI, Shadlen MN. 1999. Effect of Expected Reward Magnitude on the Response of Neurons in the Dorsolateral Prefrontal Cortex of the Macaque. *Neuron.* 24:415–425.

- Lewinsohn PM, Libet J. 1972. Pleasant events, activity schedules, and depressions. *J Abnorm Psychol.* 79:291–295.
- Li S, Ostwald D, Giese M, Kourtzi Z. 2007. Flexible Coding for Categorical Decisions in the Human Brain. *J Neurosci.* 27:12321–12330.
- Liebe S, Hoerzer GM, Logothetis NK, Rainer G. 2012. Theta coupling between V4 and prefrontal cortex predicts visual short-term memory performance. *Nat Neurosci.* 15:256–262.
- Liu T, Hospadaruk L, Zhu DC, Gardner JL. 2011. Feature-Specific Attentional Priority Signals in Human Cortex. *J Neurosci.* 31:4484–4495.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature.* 412:150–157.
- Loose R, Kaufmann C, Tucha O, Auer DP, Lange KW. 2006. Neural networks of response shifting: Influence of task speed and stimulus material. *Brain Res.* 1090:146–155.
- Lundqvist M, Herman P, Miller EK. 2018. Dual Perspectives Dual Perspectives Companion Paper: Persistent Spiking Activity Underlies Working Memory, by Christos Constantinidis, Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not. *J Neurosci.* 38:7013–7019.
- Lundqvist M, Rose J, Herman P, Brincat SL, Buschman TJ, Miller EK. 2016. Gamma and Beta Bursts Underlie Working Memory. *Neuron.* 90:152–164.
- Luria AR. 1973. Neuropsychological Studies in the USSR. A Review (Part I). *Proc Natl Acad Sci USA.* 70:959–964.
- Luria AR, Tsvetkova LS. 1964. The programming of constructive activity in local brain injuries. *Neuropsychologia.* 2:95–107.
- MacDonald AW, Cohen JD, Andrew Stenger V, Carter CS. 2000. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science.* 288:1835–1838.
- MacEvoy SP, Epstein RA. 2011. Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci.* 14:1323–1329.
- Machens CK, Romo R, Brincat SL. 2010. Functional, But Not Anatomical, Separation of “What” and “When” in Prefrontal Cortex. *J Neurosci.* 30:350–360.

- Machens CK, Romo R, Brody CD. 2005. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science*. 307:1121–1124.
- Macmillan M. 2000. *An odd kind of fame : stories of Phineas Gage*. MIT Press.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB. 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc NatL Acad Sci USA*. 92:8135–8139.
- Mante V, Sussillo D, Shenoy K V., Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*. 503:78–84.
- Marder E, Abbott LF, Turrigiano GG, Liu Z, Golowasch J. 1996. Memory: Recording Experience in Cells and Circuits. *Proc NatL Acad Sci USA*. 93:13481–13486.
- Markowitsch HJ, Pritzel M. 1979. The prefrontal cortex: Projection area of the thalamic mediodorsal nucleus? *Physiol Psychol*. 7:1–6.
- Matsumoto K, Suzuki W, Tanaka K. 2015. Neuronal Correlates of Goal-Based Motor Selection in the Prefrontal Cortex. *Science*. 301:229–232.
- Maunsell JHR, Treue S. 2006. Feature-based attention in visual cortex. *Trends Neurosci*. 29:317–322.
- Mendoza-Halliday D, Martinez-Trujillo J. 2017. Neuronal population coding of perceived and memorized visual features in the lateral prefrontal cortex. *Nat Commun*. 8:15471.
- Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. 2008. Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex. *J Neurophysiol*. 100:1407–1419.
- Miller EK, Cohen JD. 2001. An Integrative Theory of Prefrontal Cortex Function. *Annu Rev Neurosci*. 24:167–202.
- Miller GA, Galanter E, Pribram K. H. 1960. *Plans and the structure of behavior*. Holt, Rinehart and Winston, Inc.
- Milner B. 1964. Some effects of frontal lobectomy in man, *The Frontal Granular Cortex and Behavior*. McGraw-Hill.
- Mineroff Z, Blank I, Mahowald K, Fedorenko E. 2018. A robust dissociation among the language, multiple demand, and default mode networks: Evidence from inter-region

- correlations in effect size. *Neuropsychologia*. 119:501–511.
- Mitchell DJ, Bell AH, Buckley MJ, Mitchell AS, Sallet J, Duncan J. 2016. A Putative Multiple-Demand System in the Macaque Brain. *J Neurosci*. 36:8574–8585.
- Mitchell DJ, Cusack R. 2018. Visual short-term memory through the lifespan: Preserved benefits of context and metacognition. *Psychol Aging*. 33:841–854.
- Mohanty A, Gitelman DR, Small DM, Mesulam MM. 2008. The Spatial Attention Network Interacts with Limbic and Monoaminergic Systems to Modulate Motivation-Induced Attention Shifts. *Cereb Cortex*. 18:2604–2613.
- Mongillo G, Barak O, Tsodyks M. 2008. Synaptic Theory of Working Memory. *Science*. 319:1543–1546.
- Muhammed K, Manohar S, Ben Yehuda M, Chong TT-J, Tofaris G, Lennox G, Bogdanovic M, Hu M, Husain M. 2016. Reward sensitivity deficits modulated by dopamine are associated with apathy in Parkinson’s disease. *Brain*. 139:2706–2721.
- Muhle-Karbe PS, Duncan J, De Baene W, Mitchell DJ, Brass M. 2017. Neural Coding for Instruction-Based Task Sets in Human Frontoparietal and Visual Cortex. *Cereb Cortex*. 27:1891–1905.
- Muri RM, Vermersch AI, Rivaud S, Gaymard B, Pierrot-Deseilligny C. 1996. Effects of single-pulse transcranial magnetic stimulation over the prefrontal and posterior parietal cortices during memory-guided saccades in humans. *J Neurophysiol*. 76:2102–2106.
- Murphy PR, Boonstra E, Nieuwenhuis S. 2016. Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nat Commun*. 7:13526.
- Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo H, Lee D, Wang XJ. 2014. A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci*. 17:1661–1663.
- Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang XJ. 2017. Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proc Natl Acad Sci USA*. 114:394–399.
- Naya Y, Chen H, Yang C, Suzuki WA. 2017. Contributions of primate prefrontal cortex and medial temporal lobe to temporal-order memory. *Proc Natl Acad Sci USA*. 114:13555–13560.

- Nee DE, Brown JW. 2012. Rostral-caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *Neuroimage*. 63:1285–1294.
- Nee DE, Brown JW, Askren MK, Berman MG, Demiralp E, Krawitz A, Jonides J. 2013. A Meta-analysis of Executive Components of Working Memory. *Cereb Cortex*. 23:264–282.
- Nelissen N, Stokes MG, Nobre AC, Rushworth MFS. 2013. Frontal and parietal cortical interactions with distributed visual representations during selective attention and action selection. *J Neurosci*. 33:16443–16458.
- Nieto-Castañón A, Fedorenko E. 2012. Subject-specific functional localisers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage*. 63:1646–1669.
- Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A Toolbox for Representational Similarity Analysis. *PLoS Comput Biol*. 10:e1003553.
- Nomura EM, Gratton C, Visser RM, Kayser A, Perez F, D’Esposito M. 2010. Double dissociation of two cognitive control networks in patients with focal brain lesions. *Proc Natl Acad Sci USA*. 107:12017–12022.
- Norman DA, Shallice T. 1980. Attention to action: willed and automatic control of behavior. San Diego: University of California Center for Human Information Processing.
- Norman KA, Polyn SM, Detre GJ, Haxby J V. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci*. 10:424–430.
- Nyberg L, Marklund P, Persson J, Cabeza R, Forkstam C, Petersson KM, Ingvar M. 2003. Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia*. 41:371–377.
- O’Keefe J, Recce ML. 1993. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*. 3:317–330.
- Ott T, Nieder A. 2016. Dopamine D2 Receptors Enhance Population Dynamics in Primate Prefrontal Working Memory Circuits. *Cereb Cortex*. 27:4423–4435.
- Owen AM. 1997. The functional organization of working memory processes within human lateral frontal cortex: the contribution of functional neuroimaging. *Eur J Neurosci*. 9:1329–1339.

- Owen AM, McMillan KM, Laird AR, Bullmore ET. 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp.* 25:46–59.
- Padmala S, Pessoa L. 2010. Interactions between cognition and motivation during response inhibition. *Neuropsychologia.* 48:558–565.
- Padmala S, Pessoa L. 2011. Reward Reduces Conflict by Enhancing Attentional Control and Biasing Visual Cortical Processing. *J Cogn Neurosci.* 23:3419–3432.
- Parvizi J, Rangarajan V, Shirer WR, Desai N, Greicius MD. 2013. The Will to Persevere Induced by Electrical Stimulation of the Human Cingulate Gyrus. *Neuron.* 80:1359–1367.
- Passingham RE. 1973. Anatomical Differences between the Neocortex of Man and Other Primates. *Brain Behav Evol.* 7:337–359.
- Paulus FM, Krach S, Blanke M, Roth C, Belke M, Sommer J, Müller-Pinzler L, Menzler K, Jansen A, Rosenow F, Bremmer F, Einhäuser W, Knake S. 2015. Fronto-insula network activity explains emotional dysfunctions in juvenile myoclonic epilepsy: Combined evidence from pupillometry and fMRI. *Cortex.* 65:219–231.
- Paunov AM, Blank IA, Fedorenko E. 2019. Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *J Neurophysiol.* 121:1244–1265.
- Peelen M V., Downing PE. 2006. Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn Sci.* 11:4–5.
- Peelle JE, McMillan C, Moore P, Grossman M, Wingfield A. 2004. Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain Lang.* 91:315–325.
- Pessoa L. 2009. How do emotion and motivation direct executive control? *Trends Cogn Sci.* 13:160–166.
- Petrides M. 2005. Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc B Biol Sci.* 360:781–795.
- Petrides M, Cadoret G, Mackey S. 2005. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature.* 435:1235–1238.

- Picton TW, Stuss DT, Shallice T, Alexander MP, Gillingham S. 2006. Keeping time: Effects of focal frontal lesions. *Neuropsychologia*. 44:1195–1209.
- Pierrot-Deseilligny C, Rivaud S, Gaymard B, Agid Y. 1991. Cortical control of reflexive visually-guided saccades. *Brain*. 114:1473–1485.
- Pochon JB, Levy R, Fossati P, Lehericy S, Poline JB, Pillon B, Le Bihan D, Dubois B. 2002. The neural system that bridges reward and cognition in humans: An fMRI study. *Proc Natl Acad Sci*. 99:5669–5674.
- Postle BR, Berger JS, D'Esposito M. 1999. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc Natl Acad Sci USA*. 96:12959–12964.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. 2011. Functional network organization of the human brain. *Neuron*. 72:665–678.
- Preuss TM. 1995. Do Rats Have Prefrontal Cortex? The Rose-Woolsey-Akert Program Reconsidered. *J Cogn Neurosci*. 7:1–24.
- Procyk E, Ford Dominey P, Amiez C, Joseph J-P. 2000. The effects of sequence structure and reward schedule on serial reaction time learning in the monkey. *Cogn Brain Res*. 9:239–248.
- Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. *Science*. 276:821–824.
- Reddy L, Kanwisher N. 2007. Category Selectivity in the Ventral Visual Pathway Confers Robustness to Clutter and Diverted Attention. *Curr Biol*. 17:2067–2072.
- Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, Fusi S. 2013. The importance of mixed selectivity in complex cognitive tasks. *Nature*. 497:585–590.
- Rikhye R V., Gilra A, Halassa MM. 2018. Thalamic regulation of switching between cortical representations enables cognitive flexibility. *Nat Neurosci*. 21:1753–1763.
- Romo R, Brincat SL, Hernández A, Lemus L. 1999. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*. 399:470–473.
- Rose JE, Woolsley CN. 1948. The orbitofrontal cortex and its connections with the mediodorsal nucleus in rabbit, sheep and cat. *Res Publ -Association Res Nerv Ment Dis*.

27:210–232.

- Rottschy C, Langner R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB. 2012. Modelling neural correlates of working memory: A coordinate-based meta-analysis. *Neuroimage*. 60:830–846.
- Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev*. 16:225–237.
- Rushworth MFS, Walton ME, Kennerley SW, Bannerman DM. 2004. Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci*. 8:410–417.
- Rutishauser U, Ross IB, Mamelak AN, Schuman EM. 2010. Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*. 464:903–907.
- Rylander G. 1939. Personality changes after operations on the frontal lobes: A clinical study of 32 cases. *Acta Psychiatr Neurol Scand*. 20:1–327.
- Said CP, Moore CD, Engell AD, Todorov A, Haxby J V. 2010. Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis*. 10:1–12.
- Sakai K, Passingham RE. 2003. Prefrontal interactions reflect future task operations. *Nat Neurosci*. 6:75–81.
- Salazar RF, Dotson NM, Bressler SL, Gray CM. 2012. Content-specific fronto-parietal synchronization during visual working memory. *Science*. 338:1097–1100.
- Sarma A, Masse NY, Wang XJ, Freedman DJ. 2015. Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat Neurosci*. 19:143–149.
- Saxe R, Brett M, Kanwisher N. 2006. Divide and conquer: A defense of functional localisers. *Neuroimage*. 30:1088–1096.
- Schaefer A, Kong R, Gordon EM, Laumann TO, Zuo X-N, Holmes AJ, Eickhoff SB, Yeo TBT. 2018. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex*. 28:3095–3114.
- Schmitt LI, Wimmer RD, Nakajima M, Happ M, Mofakham S, Halassa MM. 2017. Thalamic amplification of cortical connectivity sustains attentional control. *Nature*. 545:219–223.
- Schultz W. 2015. Neuronal reward and decision signals: from theories to data. *Physiol Rev*. 95:853–951.

- Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science*. 275:1593–1599.
- Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control. *J Neurosci*. 27:2349–2356.
- Semendeferi K, Armstrong E, Schleicher A, Zilles K, Van Hoesen GW. 2001. Prefrontal Cortex in Humans and Apes: A Comparative Study of Area 10. *Am J Phys Anthropol*. 114:224–241.
- Shallice T, Stuss DT, Picton TW, Alexander MP, Gillingham S. 2007. Multiple effects of prefrontal lesions on task-switching. *Front Hum Neurosci*. 1:2.
- Shashidhara S, Erez Y. 2019. Reward motivation modulates representation of behaviorally-relevant information across the frontoparietal cortex. *bioRxiv*. 609537.
- Shashidhara S, Mitchell DJ, Erez Y, Duncan J. 2019. Progressive Recruitment of the Frontoparietal Multiple-demand System with Increased Task Complexity, Time Pressure, and Reward. *J Cogn Neurosci*. 31:1617–1630.
- Shashidhara S, Spronkers FS, Erez Y. 2019. Localizing the “multiple-demand” frontoparietal network in individual subjects. *bioRxiv*. 661934.
- Shenhav A, Botvinick M, Cohen JD. 2013. The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*. 79:217–240.
- Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, Duncan J. 2008. Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proc Natl Acad Sci USA*. 105:11969–11974.
- Simon HA. 1967. Motivational and emotional controls of cognition. *Psychol Rev*. 74:29–39.
- Simon O, Mangin J-F, Cohen L, Le Bihan D, Dehaene S. 2002. Topographical Layout of Hand, Eye, Calculation, and Language-Related Areas in the Human Parietal Lobe. *Neuron*. 33:475–487.
- Smith V, Mitchell DJ, Duncan J. 2018. Role of the Default Mode Network in Cognitive Transitions. *Cereb Cortex*. 28:3685–3696.
- Soon CS, Brass M, Heinze H-J, Haynes J-D. 2008. Unconscious determinants of free decisions in the human brain. *Nat Neurosci*. 11:543–545.

- Soreq E, Leech R, Hampshire A. 2019. Dynamic network coding of working-memory domains and working-memory processes. *Nat Commun.* 10:936.
- Stănişor L, van der Togt C, Pennartz CMA, Roelfsema PR. 2013. A unified selection signal for attention and reward in primary visual cortex. *Proc Natl Acad Sci USA.* 110:9136–9141.
- Stiers P, Mennes M, Sunaert S. 2010. Distributed task coding throughout the multiple demand network of the human frontal-insular cortex. *Neuroimage.* 52:252–262.
- Stoet G, Snyder LH. 2004. Single Neurons in Posterior Parietal Cortex of Monkeys Encode Cognitive Set. *Neuron.* 42:1003–1012.
- Stokes MG. 2015. ‘Activity-silent’ working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn Sci.* 19:394–405.
- Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. 2013. Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron.* 78:364–375.
- Stuart GJ, Spruston N. 2015. Dendritic integration: 60 years of progress. *Nat Neurosci.* 18:1713–1721.
- Stürmer B, Nigbur R, Schacht A, Sommer W. 2011. Reward and punishment effects on error processing and conflict control. *Front Psychol.* 2:335.
- Stuss DT, Knight RT. 2013. *Principles of frontal lobe function.* Oxford University Press.
- Stuss DT, Shallice T, Alexander MP, Picton TW. 1995. A Multidisciplinary Approach to Anterior Attentional Functions. *Ann N Y Acad Sci.* 769:191–211.
- Sussillo D, Abbott LF. 2009. Generating coherent patterns of activity from chaotic neural networks. *Neuron.* 63:544–557.
- Swaminathan SK, Freedman DJ. 2012. Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nat Neurosci.* 15:315–320.
- Taylor SF, Welsh RC, Wager TD, Luan Phan K, Fitzgerald KD, Gehring WJ. 2004. A functional neuroimaging study of motivation and executive function. *Neuroimage.* 21:1045–1054.
- Thurley K, Senn W, Lüscher H-R. 2008. Dopamine Increases the Gain of the Input-Output Response of Rat Prefrontal Pyramidal Neurons. *J Neurophysiol.* 99:2985–2997.

- Tibon R, Greve A, Henson RN. 2018. The missing link? Testing a schema account of unitization. *Mem Cognit.* 46:1023–1040.
- Tremblay S, Doucet G, Pieper F, Sachs A, Martinez-Trujillo J. 2015. Single-Trial Decoding of Visual Attention from Local Field Potentials in the Primate Lateral Prefrontal Cortex Is Frequency-Dependent. *J Neurosci.* 35:9038–9049.
- Turner MS, Cipolotti L, Yousry T, Shallice T. 2007. Qualitatively different memory impairments across frontal lobe subgroups. *Neuropsychologia.* 45:1540–1552.
- Uylings HBM, Groenewegen HJ, Kolb B. 2003. Do rats have a prefrontal cortex? *Behav Brain Res.* 146:3–17.
- van der Kouwe AJW, Benner T, Salat DH, Fischl B. 2008. Brain morphometry with multiecho MPRAGE. *Neuroimage.* 40:559–569.
- Vendrell P, Junqué C, Pujol J, Jurado MA, Molet J, Grafman J. 1995. The role of prefrontal regions in the Stroop task. *Neuropsychologia.* 33:341–352.
- Vergauwe E, Cowan N. 2015. Attending to items in working memory: evidence that refreshing and memory search are closely related. *Psychon Bull Rev.* 22:1001–1006.
- Vijayraghavan S, Wang M, Birnbaum SG, Williams G V, Arnsten AFT. 2007. Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nat Neurosci.* 10:376–384.
- Wager TD, Smith EE. 2003. Neuroimaging studies of working memory: a meta-analysis. *Cogn Affect Behav Neurosci.* 3:255–274.
- Wallace AFC. 1960. : Plans and the Structure of Behavior . George A. Miller, Eugene Galanter, Karl H. Pribram. *Am Anthropol.* 62:1065–1067.
- Wallis G, Stokes MG, Cousijn H, Woolrich M, Nobre AC. 2015. Frontoparietal and Cingulo-opercular Networks Play Dissociable Roles in Control of Working Memory. *J Cogn Neurosci.* 27:2019–2034.
- Wallis JD, Anderson KC, Miller EK. 2001. Single neurons in prefrontal cortex encode abstract rules. *Nature.* 411:953–956.
- Walther DB, Caddigan E, Fei-Fei L, Beck DM. 2009. Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *J Neurosci.* 29:10573–10581.
- Warden MR, Miller EK. 2010. Task-Dependent Changes in Short-Term Memory in the

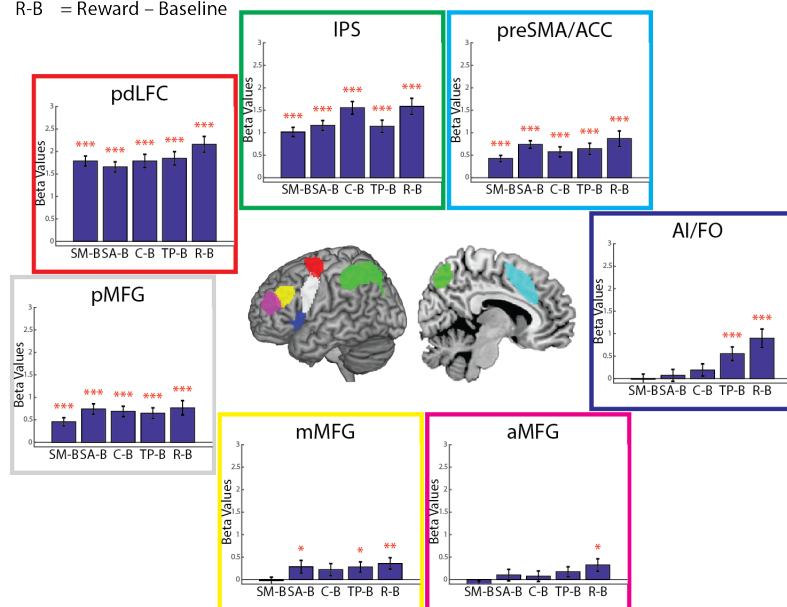
- Prefrontal Cortex. *J Neurosci.* 30:15801–15810.
- Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD. 2014. Frontoparietal representations of task context support the flexible control of goal-directed cognition. *J Neurosci.* 34:10743–10755.
- Watanabe M. 1996. Reward expectancy in primate prefrontal neurons. *Nature.* 382:629–632.
- Watanabe Y, Funahashi S. 2012. Thalamic mediodorsal nucleus and working memory. *Neurosci Biobehav Rev.* 36:134–142.
- Weiner KS, Barnett MA, Witthoft N, Golarai G, Stigliani A, Kay KN, Gomez J, Natu VS, Amunts K, Zilles K, Grill-Spector K. 2018. Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *Neuroimage.* 170:373–384.
- Windmann S, Schneider T, Reczio J, Grobosch M, Voelzke V, Blasius V, Bramer A, Ischebeck W, Janikowski G, Mandrella W, Unger C, Wischnjak L. 2006. Intact emotion-induced recognition bias in neuropsychological patients with executive control deficits. *Cogn Affect Behav Neurosci.* 6:270–276.
- Wisniewski D, Goschke T, Haynes J-D. 2016. Similar coding of freely chosen and externally cued intentions in a fronto-parietal network. *Neuroimage.* 134:450–458.
- Womelsdorf T, Schoffelen J-M, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P. 2007. Modulation of neuronal interactions through neuronal synchronization. *Science.* 316:1609–1612.
- 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.
- Woolgar A, Hampshire A, Thompson R, Duncan J. 2011. Adaptive coding of task-relevant information in human frontoparietal cortex. *J Neurosci.* 31:14592–14599.
- Woolgar A, Jackson J, Duncan J. 2016. Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. *J Cogn Neurosci.* 28:1433–1454.
- Woolgar A, Thompson R, Bor D, Duncan J. 2011. Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage.* 56:744–752.

- Woolgar A, Williams MA, Rich AN. 2015. Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *Neuroimage*. 109:429–437.
- Yao W-D, Spealman RD, Zhang J. 2008. Dopaminergic signaling in dendritic spines. *Biochem Pharmacol*. 75:2055–2069.
- Yeo BTT, Krienen FM, Eickhoff SB, Yaakub SN, Fox PT, Buckner RL, Asplund CL, Chee MWL. 2015. Functional Specialization and Flexibility in Human Association Cortex. *Cereb Cortex*. 25:3654–3672.
- Yeo TBT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, Fischl B, Liu H, Buckner RL. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol*. 106:1125–1165.
- Zekveld AA, Rudner M, Kramer SE, Lyzenga J, Ronnberg J. 2014. Cognitive processing load during listening is reduced more by decreasing voice similarity than by increasing spatial separation between target and masker speech. *Front Neurosci*. 8:88.

Appendix

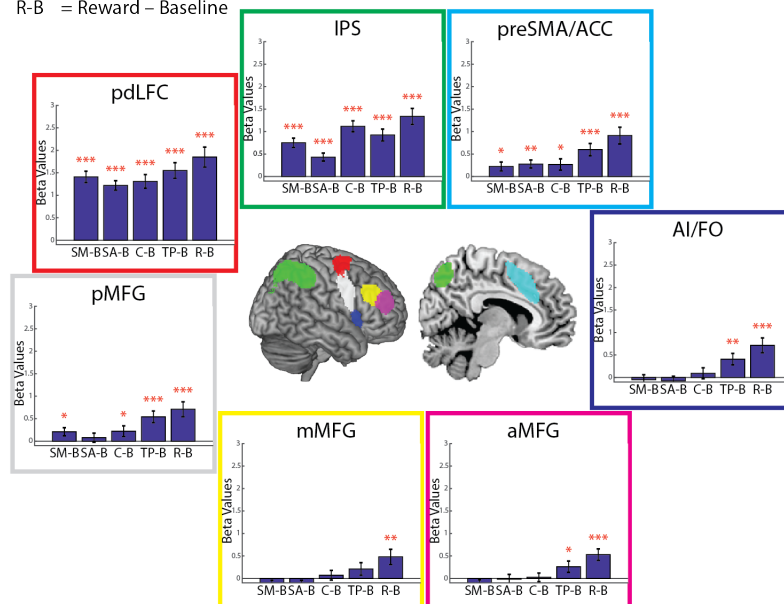
A

SM-B = Simple Maze – Baseline
SA-B = Simple Arrow – Baseline
C-B = Complex – Baseline
TP-B = Time Pressure – Baseline
R-B = Reward – Baseline



B

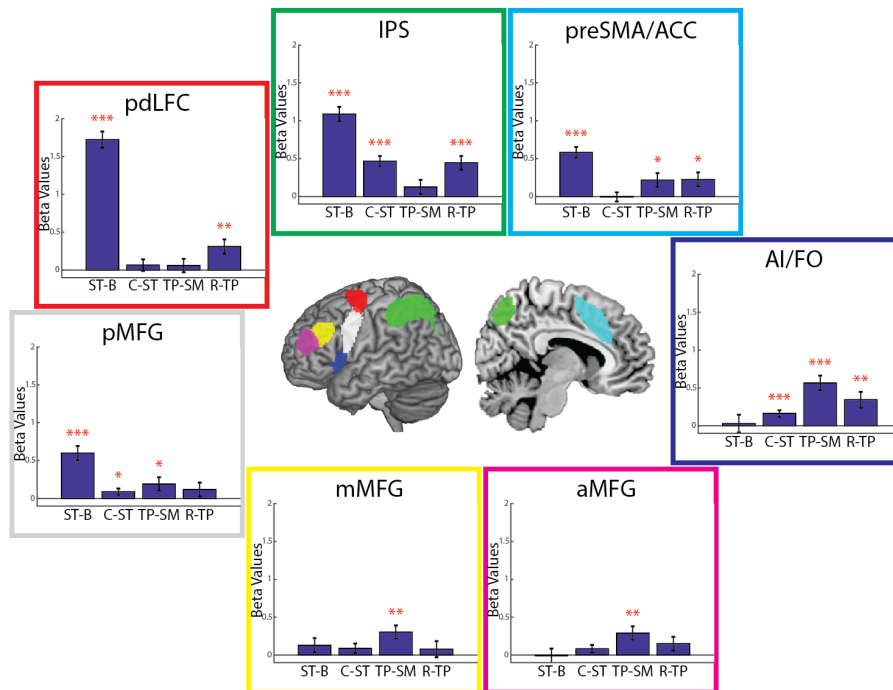
SM-B = Simple Maze – Baseline
SA-B = Simple Arrow – Baseline
C-B = Complex – Baseline
TP-B = Time Pressure – Baseline
R-B = Reward – Baseline



Supplementary Figure A2.1: Data as Figure 2.2, shown separately for A, left and B, right hemisphere.

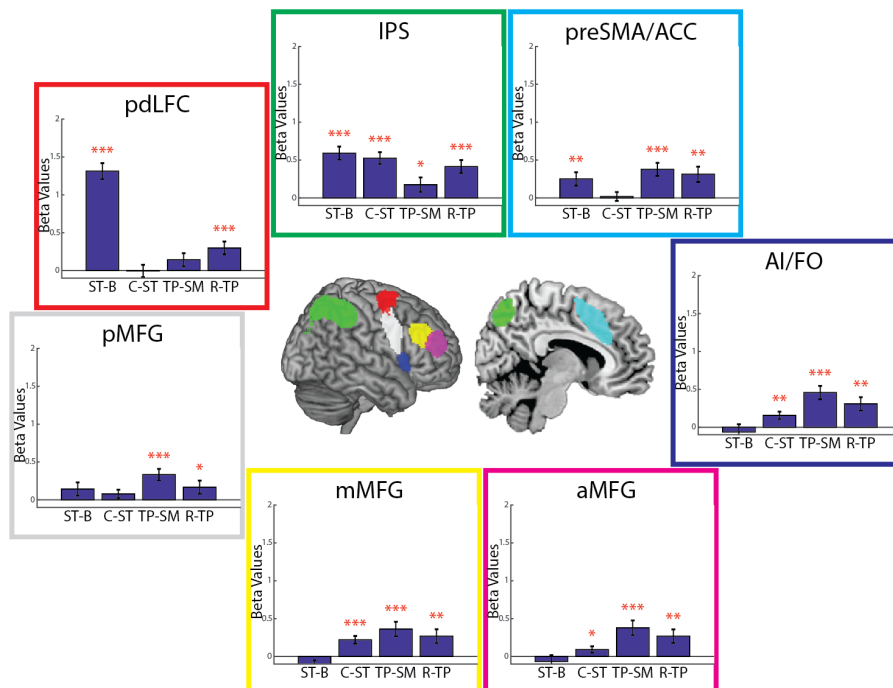
A

ST-B = Simple Tasks – Baseline
C-ST = Complex – Simple Tasks
TP-SM = Time Pressure – Simple Maze
R-TP = Reward – Time Pressure



B

ST-B = Simple Tasks – Baseline
C-ST = Complex – Simple Tasks
TP-SM = Time Pressure – Simple Maze
R-TP = Reward – Time Pressure



Supplementary Figure A2.2: Data as Figure 2.3, shown separately for A, left and B, right hemisphere.