The role of the frontoparietal cortices in

feature-selective attention

Jade B. Jackson, BSc Psychology (Honours)

Department of Cognitive Science

ARC Centre of Excellence in Cognition and its Disorders

Perception in Action Research Centre

Faculty of Human Sciences

Macquarie University, Sydney, Australia

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of

Philosophy (PhD)

June 2016

Table of Contents

v

Summary

In everyday life our brains receive a huge amount of information from our senses. Since we cannot process it all, we must select the information that is most relevant for our current activity and prioritise it over other, irrelevant, information. This 'selective attention' process is critical for making sense of the world around us and for functioning successfully within it.

Certain regions of frontal and parietal cortex have long been implicated as sources of selective attention in the brain. In particular, the 'adaptive coding hypothesis' proposes that certain neural populations adjust to selectively code the information that is required for current behaviour (Duncan 2001). This may serve as a source of bias, supporting related information processing across the rest of the brain (Desimone and Duncan 1995, Miller and Cohen 2001). The primary aims of this thesis were to investigate if adaptive coding provides a mechanism for feature-selective attention in the frontoparietal cortices and whether this in turn modulates responses across the rest of the brain.

In Chapter 1, I present an overview of the relevant literature. In Chapter 2 I used multi-voxel pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data to demonstrate that regions of the frontoparietal cortices prioritise coding of task-relevant features over equivalent irrelevant feature information. In Chapter 3 I replicated this finding and examined the adaptive response of these regions in greater detail, investigating the extent to which single voxels in these regions can be re-used to code information across multiple tasks. In Chapter 4 I developed a paradigm that tracks the extent to which irrelevant information interferes with behavioural performance, and tested the causal contribution of the right dorsolateral prefrontal cortex (dlPFC) to this task using

transcranial magnetic stimulation (TMS). In Chapter 5 I used this paradigm to examine the behavioural and neural consequences of disrupting activity in the frontal lobe, using concurrent TMS-fMRI. Using MVPA, I found that disrupting the right dlPFC with TMS affected the multi-voxel coding of both relevant and irrelevant feature information in frontal, parietal and visual cortices. Finally, in Chapter 6, I discuss the implications of these results in a broader context and suggest some areas for future research. Together, the experiments presented in this thesis advance the understanding of flexible mechanisms employed in the frontoparietal cortices in the context of feature-selective attention.

Author Note

This thesis has been prepared in the form of a thesis by publication. As such, there is a degree of repetition, particularly in the chapter introductions, and references are given at the end of each chapter.

The formatting and referencing style adopted throughout this thesis reflects the requirements of the APA Publication Manual (6th edition). The language reflects English (UK) spelling aside from Chapter 2 which uses English (US) to comply with the requirements of the journal it was submitted to.

Author Statement

I, Jade B. Jackson, certify that the research presented in this thesis entitled "The role of the frontoparietal cortices in feature-selective attention" is an original piece of research and has been written by me. I also certify that this thesis has not been previously submitted for a degree nor has it been submitted to any university or institution other than Macquarie University.

Any sources of information or assistance are cited appropriately and acknowledged.

The research presented in this thesis was approved by the Macquarie University Human Research Ethics Committee (reference numbers: HE27NOV2009-R00180, 5201300541, 5201400585) and the Reading University Human Research Ethics Committee (reference number: UREC 15/45).

Signed,

Jade B. Jackson

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my primary supervisor Alexandra Woolgar. I think that at times a PhD can feel like a lonely endeavour, but thanks to the incredible enthusiasm, support, patience and motivation that I received from Alex, this was not my experience. I consider myself very lucky to have had such an enriching experience and to have learnt so much from her. Thank you to Alex for all of our discussions and debates, as well as her guidance and expertise throughout my candidature. Alex has really inspired me and I have learnt so much from her. I could not have imagined a better mentor for my PhD study.

I would also like to thank my co-supervisor Anina Rich for being so generous with her time and for providing me with endless guidance and encouragement. Also thank you to my associate supervisor Mark Williams for his positivity and support. In addition, thank you to Eva Feredoes for all that she taught me about TMS-fMRI, and for the continuing advise and support. I am really grateful for the opportunity to have worked on the TMS-fMRI project together.

I am very fortunate to have been given the opportunity to complete my PhD at the Department of Cognitive Science, Macquarie University. I have had such a positive experience thanks to the friendly and kind people in this department. Thank you to the members of the Perception in Action Research Centre for interesting discussions and invaluable feedback. Thank you to Lesley McKnight, Marcus Ockenden, Teri Roberts and Lisa Yen for helping me with the many forms and technical challenges. Thank you to all of my colleagues and friends for making the office a fun and stimulating environment.

A special mention to Lina Teichmann, Jon Horlyck, Tijl Grootswagers and Denise Moerel who make me laugh every day, and for whom I am incredibly grateful for to have as friends. I thank them for all of the good times, for which there are too many to list, and

Chapter 1

Introduction

At every moment our brains receive far more input than they can process at once. Because of this, we need to prioritise a subset of the available information and ignore the rest, a process known as *selective attention*. Additionally, the selective attention system needs to be flexible, to cope with the changing input and goals of everyday life. Consider, for example, a situation where you are in an airport trying to reach your flight. You must attend to the relevant inputs (flight details on the board) to find your gate number (current goal) in order to reach your flight on time (end goal). In order to do this effectively, you need to ignore irrelevant details (e.g. other flight information, announcements, background noise, etc) that are bombarding your senses. Having established the gate number you now need to attend to other aspects of the environment (e.g. direction signs) while ignoring information that was previously relevant (flight board). This need to prioritise processing of some aspects of input over others, depending on our current goal, is a constant requirement that our brains are usually able to cope with efficiently.

In some cases, we may wish to prioritise processing of a single feature of an object. For example, if you were to use a welding torch to weld two pieces of metal together, you would need to first focus your attention on the colour of the flame produced to ensure that it is at the correct temperature. Next you would need to switch your focus of attention to the width of the beam, currently ignoring the colour of the flame, in order to adjust the width depending on the size of the metals you are welding together. In everyday

life, we encounter a variety of situations requiring this ability to flexibly select the relevant inputs for the current task, whether this is at the level of listening for a flight announcement amongst distractors or at the level of focusing on the relevant feature of a single object.

The neural underpinnings of flexible selective attention are thought to involve the frontal and parietal cortices. Evidence for this comes from several converging methods. Neuropsychological research has pointed to the frontal and parietal lobes, and in particular the prefrontal cortex (PFC), as critical to adaptive goal-directed behaviour. Damage to this region often results in inflexible behaviour both in humans (e.g. Luria, 1966; Manes et al., 2002) and in non-human primates (Rossi, Bichot, Desimone, & Ungerleider, 2007). Evidence from neurostimulation studies has also demonstrated the importance of frontal cortex, alongside the parietal cortices, in enabling goal-directive behaviour via top-down modulation of earlier processing areas (e.g. Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Higo, Mars, Boorman, Buch, & Rushworth, 2011; Lee & D'Esposito, 2012; Miller, Vytlacil, Fegen, Pradhan, & D'Esposito, 2011; Morishima et al., 2009; Ruff et al., 2006; Taylor, Nobre, & Rushworth, 2007; Zanto, Rubens, Thangavel, & Gazzaley, 2011). In human neuroimaging, studies similarly converge that cognitive control seems to be implemented by a network of regions in frontal and parietal cortices (Cole & Schneider, 2007; Corbetta & Shulman, 2002; Duncan, 2010; Duncan & Owen, 2000; Fedorenko, Duncan, & Kanwisher, 2013; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008).

A large body of evidence implicates both the frontal and parietal cortex in cognitive control. However, the mechanisms by which flexible selective attention, a key component of cognitive control, is realised, requires further investigation. One influential proposal, the adaptive coding hypothesis (Duncan, 2001) theorises that neurons in certain

higher brain regions dynamically adjust their responses to code for information that is currently relevant. In doing so, they may provide a source of bias to other, more specialised, brain regions, to support preferential processing of the relevant information in those regions as well (Desimone & Duncan, 1995). Thus, adaptive coding provides a possible neural mechanism for the implementation of selective attention.

The focus of this thesis is to investigate whether adaptive coding in human frontoparietal cortex could operate as a mechanism for selective attention. I focus particularly on feature-selective attention: the selection of relevant over irrelevant features of single objects (Chen, Hoffmann, Albright, & Thiele, 2012). In this chapter I start by reviewing the literature that implicates frontal and parietal cortex in goal-directed behaviour. Next, I review some of the prominent models of attention, which sit within a broad framework of executive control. I then focus particularly on the adaptive coding hypothesis (Duncan, 2001) and consider the evidence for restricted regions of the frontoparietal cortices as candidates for adaptive coding in the human brain. Finally, I explore how adaptive coding in this network could provide a mechanism for feature-selective attention. I conclude with an outline of the thesis, outlining my research questions about the implementation of feature-selective attention in the human brain.

Frontal and parietal lobes as critical for goal-directive behaviour

Selective attention is necessary to prioritise the contents of capacity-limited networks in favour of task-relevant representations. This supports successful performance. The process is thought to rely on "top-down" modulation, in which signals from higher cortical regions bias processing in earlier cortical regions. Top-down modulation may

support selection through both enhancement of task-relevant, and suppression of task-irrelevant, neural activity in specialised processing regions such as the visual cortex (e.g. Desimone & Duncan, 1995; Gazzaley, Cooney, McEvoy, Knight, & D'esposito, 2005; Kanwisher & Wojciulik, 2000; Knight, Staines, Swick, & Chao, 1999; Shimamura, 2000). The frontal cortex was one of the first areas to emerge as a potential source of top-down signals due to the widespread and reciprocal anatomical connectivity between subregions of the frontal lobes and multiple levels of the visual system (Miller & Cohen, 2001; Webster, Bachevalier, & Ungerleider, 1994). In addition, a wealth of neuropsychological studies has connected frontal damage to deficits in executive functions, which involve impairment of selective processing (e.g. Barceló & Knight, 2002; Bornstein, 1986; Drewe, 1974; Eslinger & Grattan, 1993; Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Milner, 1963; Nelson, 1976; Perret, 1974; Robinson, Heaton, Lehman, & Stilson, 1980; Stuss, Floden, Alexander, Levine, & Katz, 2001; Vendrell et al., 1995). In this section I will review a selection of the neuropsychological data that points towards a role of the frontal, and to a lesser extent the parietal, cortices in executive functions. I will then discuss neuroimaging studies that provide further evidence that the frontal and parietal lobes are crucial for controlling attention. Finally, I will review neurostimulation data that link both frontal and parietal cortices with top-down modulation of earlier processing regions.

Although several cognitive deficits have been associated with damage to the frontal cortex, a common denominator is inflexibility. For example, patients with frontal lobe damage are typically unable to resist interference from stimuli, or aspects of stimuli, that would normally be suppressed or ignored (Chao & Knight, 1995). The Wisconsin Card Sorting Test (WCST, Grant & Berg, 1948) provides a measure of this inflexibility in frontal patients, in which the subject has to remember, and shift when needed, the

categorising principle of a series of figures (Milner, 1963). Individuals must deduce the rule by which the cards should be sorted (rather than being told the rule explicitly). After the initial rule is learned successfully, the examiner changes the rule so that the old rule must be rejected, the new rule discovered, and a switch made from using the old rule to the new. The ability to exhibit such flexible readjustment is proposed to be a central characteristic of executive function (Duncan, 2001). Individuals with frontal lobe damage tend to persist in sorting items according to the previous and now inappropriate rule, exhibiting difficulty with shifting to what is currently relevant. This has been observed in humans with frontal damage (e.g. Barceló & Knight, 2002; Bornstein, 1986; Drewe, 1974; Eslinger & Grattan, 1993; Janowsky et al., 1989; Milner, 1963; Nelson, 1976; Robinson et al., 1980) as well as non-human primates with lesions in the PFC (Dias, Robbins, & Roberts, 1996).

Another popular assessment for attentional control is the Stroop task (Stroop, 1935). In this task participants are asked to name the colour of the ink of (a) colour words that are printed in a congruous coloured ink (e.g. the word "green" printed in green ink, the congruent condition) and (b) colour words that are printed in incongruous coloured ink (e.g., the word "green" printed in blue ink, the incongruent condition) (Egner, 2007). Thus participants must attend to the ink colour and suppress information derived from the word. The robust effect, however, is that participants are slower and less accurate on the incongruent condition relative to the congruent one, suggesting that the information in the word interferes with the naming of the colour of the ink. This "congruency effect" provides a measure of interference by the irrelevant dimension. Frontal patients have been shown to perform poorly in this task, for example by making more errors in the incongruent condition, in comparison to healthy controls (Stuss et al., 2001; Vendrell et al., 1995) and patients with non-frontal lesions (Perret, 1974). Data from frontal lobe

patients with both the WCST and the Stroop, then, indicate a potentially causal role for the frontal cortex in selecting behaviourally relevant information over that which is irrelevant and distracting.

Damage to regions of the parietal cortex has also been shown to affect performance on tasks that require flexibility. This includes (but is not limited to) tasks that involve response conflict (Coulthard, Nachev, & Husain, 2008), visuo-temporal attention (Shapiro, Hillstrom, & Husain, 2002), and aspects of working memory (WM) (e.g. Heide, Blankenburg, Zimmermann, & Kömpf, 1995; Koenigs, Barbey, Postle, & Grafman, 2009). A study by Roca et al., (2009) found that frontal patient deficits on many executive tasks requiring flexible selective attention, including WCST, can be entirely accounted for by deficits in fluid intelligence scores (Roca et al., 2009), which, in turn, are linearly predicted by the extent to which certain regions of frontal and parietal cortex are damaged (Woolgar et al., 2010). This suggests that certain regions of both frontal and parietal cortex may be important for attentional control.

The role of frontal and parietal cortices in a variety of executive control processes has also been informed by neuroimaging data. For example, the PFC is activated in tasks where filtering of irrelevant information is required (Banich, Milham, Atchley, Cohen, Webb, Wszalek, Kramer, Liang, Wright, et al., 2000; Banich, Milham, Atchley, Cohen, Webb, Wszalek, Kramer, Liang, Barad, et al., 2000; Liu, Banich, Jacobson, & Tanabe, 2004; MacDonald, Cohen, Stenger, & Carter, 2000). Activity in the anterior cingulate cortex (ACC) has also been associated with tasks requiring filtering of irrelevant information such as in the Stroop task with greater ACC activation in the incongruent condition compared with the neutral condition (e.g. Bench et al., 1993). Several adaptation studies also have shown that changes to an attended object (Downar, Crawley, Mikulis, & Davis, 2001; Hon, Epstein, Owen, & Duncan, 2006) or attended feature (Thompson &

Duncan, 2009) leads to extensive activation throughout frontal and parietal regions. In addition, frontal and parietal regions have been linked with top-down modulation of responses in posterior regions. For example, Kastner et al. (1999) cued participants to shift attention to a spatial location in expectation of a target stimulus. During the expectation period, several frontal and parietal regions showed increased activity compared to a resting baseline condition. Greater activity was also observed in visual cortices during the expectation period, perhaps implicating frontal and parietal regions in top-down signalling to the visual cortex, although the slow timecourse of fMRI does not allow inference about the order of events. These various studies provide converging evidence that frontal and parietal regions are important in selective attention.

In healthy adults, neurostimulation data have provided causal evidence that top-down influences from frontal and parietal cortices support selection for behaviourally relevant information. The critical role of the frontal and parietal cortices in attentional control has been demonstrated in humans using neurostimulation alone (e.g. transcranial magnetic stimulation, TMS) and also in combination with neuroimaging techniques (e.g. in combination with fMRI).

TMS is a neurostimulation technique in which magnetic stimulation is used to induce an electric field in the brain. TMS delivers short magnetic pulses that temporarily disrupt neural processing at the site of stimulation, creating a temporary, reversible, 'virtual lesion' (Walsh & Cowey, 2000). We can investigate the behavioural consequences of the stimulation, which allows us to infer whether activity at the site of stimulation contributes to a particular cognitive operation. For example, early studies on the human visual system (e.g. Amassian et al., 1989; Amassian et al., 1993) showed that stimulation of the visual cortex can interfere with perception. These early studies demonstrated that this technique has the potential to be used to infer causal structure-function relationships.

If disrupting a region using TMS affects performance of a specific aspect of an attentionally-demanding task, this can provide evidence that the region is involved in attention. For example, disruption of the dorsolateral PFC (dlPFC) with TMS impairs performance in a high load WM task, but only when irrelevant information is present in the paradigm (Sandrini, Rossini, & Miniussi, 2008). This suggests that the dlPFC is involved in attentional selection when irrelevant visual information must be ignored. In another example, Soutschek et al., (2013) stimulated either the intraparietal sulcus (IPS) or the pre-supplementary motor area (pre-SMA) during a task requiring participants to ignore irrelevant stimulus information. Participants decided whether a central letter surrounded by distractor letters was either a vowel or a consonant. The irrelevant distractors could either create response conflict where the irrelevant information was incongruent with required response, or perceptual conflict, where the irrelevant letters were different but required the same response (e.g. central letter and distractors were different vowels, "E" and "A"). Stimulation over IPS increased perceptual conflict whilst stimulation to pre-SMA increased response conflict. This suggests that irrelevant stimulus information interfered more when these regions were temporarily disrupted, and that therefore these regions are normally important in supressing interference from this irrelevant information. These data indicate the involvement of these regions in selective attention and also demonstrate how stimulation techniques can be used to test a causal relationship between neural activity and behaviour.

The data reviewed thus far provides strong converging evidence that frontal and parietal regions are involved in selective attention. A question remains as to *what role they play*. For example, do responses in these regions bias responses in earlier cortical regions to result in selective processing of relevant information? The combination of TMS with other techniques (e.g. electroencephalography, EEG, or fMRI), often described as

"perturb-and-measure" approaches (Paus, 2005), allows for direct assessment of how TMS affects neural processing both locally and in remote but connected brain regions (Bestmann & Feredoes, 2013). A number of studies have revealed causal influences of the frontal and/or parietal lobes on earlier processing regions by combining TMS with fMRI (e.g. Blankenburg et al., 2010; Feredoes et al., 2011; Higo et al., 2011; Miller et al., 2011; Ruff et al., 2008; Ruff et al., 2006; Ruff et al., 2009) or with EEG (e.g. Capotosto, Corbetta, Romani, & Babiloni, 2012; Morishima et al., 2009; Taylor et al., 2007; Zanto et al., 2011). For example, Zanto and colleagues (2011) used fMRI-guided, offline repetitive TMS (rTMS) to perturb function within a specific region of the PFC involved in a delayed-response task, then measured the outcome with EEG recordings. Ten minutes of rTMS to the right inferior frontal junction (IFJ) significantly reduced top-down modulation of the P1 component (a positive signal seen as early as 100ms post-stimulus onset) of the event-related potential (ERP) to colour stimuli at posterior electrodes, as well as a significant reduction in WM accuracy for colour. As P1 modulation recovered with time after rTMS, so did WM performance. The effect of rTMS-induced reduction in P1 modulation during colour processing was also found to predict the reduction in WM accuracy on an individual participant basis. Moreover, participants with stronger fMRI functional connectivity between the IFJ and visual cortices had a greater impact of stimulation over IFJ on top-down modulation. These results suggest that frontal cortex implements top–down control over perceptual areas to promote the successful establishment of task-relevant representations, potentially by modulating feature processing.

Other studies (e.g. Capotosto, Babiloni, Romani, & Corbetta, 2009, 2011) have also shown top-down modulation as a result of IPS stimulation. In these studies, disruption to the IPS during the allocation of spatial attention produced disruption of

anticipatory EEG rhythms in occipito-parietal cortex. Together these findings are consistent with a top-down role of the IPS in the endogenous allocation of attention.

EEG can inform us about the timecourse of top-down effects but provides much less information about which brain regions are affected. For this, we can use TMS in combination with fMRI. For example, in a study using combined fMRI and TMS (Higo et al., 2011), a cue instructed participants to store one of two objects in WM. fMRI revealed increased activity in visual regions specialised for processing the selected object category (e.g. houses). TMS over the frontal operculum/anterior insula was found to diminish this top-down modulation of posterior visual activity indicating a role of this region in modulating task-relevant vs. task irrelevant stimulus processing. Feredoes et al. (2011) also investigated how a region of prefrontal cortex exerts control over anatomically remote visual areas using concurrent TMS–fMRI. Participants performed a delayed recognition WM task where participants first viewed a presentation of target categories followed by an unfilled delay period (no distractor) or a distractor delay period with three distractor stimuli from the opposite category. Following this, a target probe was presented and participants had to indicate if it matched the initial target category. Participants performed the task in the scanner whilst TMS was applied to the right dlPFC. The timing of TMS coincided with the presentation of the irrelevant distractors. TMS increased the blood-oxygen level dependent (BOLD) signal in posterior visual regions. This increase was specific to the regions representing the current memory targets, and not the distractors, and was therefore taken as evidence that an important function of the dlPFC is to maintain relevant information in the face of distracting irrelevant information. In this example, and in others where TMS does not disrupt task performance, TMS was used as a causal "physiological probe" in which if stimulation modulates remote BOLD responses (or ERPs) under certain conditions, it in inferred that the two regions are functionally

connected (e.g. Bestmann et al., 2008; Blankenburg et al., 2010; Feredoes et al., 2011). These data provide evidence that control signals from frontal cortex propagate to posterior regions to help overcome the effects of irrelevant information during WM maintenance.

The role of the parietal cortex in top-down influences has also been studied with concurrent TMS-fMRI (Blankenburg et al., 2010; Heinen et al., 2011; Ruff et al., 2008; Ruff et al., 2009). For example, in Blankenburg and colleagues' (2010) investigation, TMS was applied over parietal sites during the direction of covert attention towards one hemifield which increased the BOLD signal in the contralateral early visual areas. This finding indicates a causal attention-dependent influence of parietal cortex over activity in visual areas. Together, the use of neurostimulation and neuroimaging techniques have provided evidence for a causal top-down modulation relationship between frontal/parietal cortices and specialised processing areas such as the visual cortex.

The wealth of neuropsychological, neuroimaging, and neurostimulation data on the function of frontal and parietal regions provide strong evidence that these regions play a crucial role in attentional control. I will now consider the various theoretical proposals for how these regions are thought to support flexible goal-directed behaviour. A number of theoretical models have been proposed for how the control of cognition is implemented (e.g. Baddeley, Della Sala, Robbins, & Baddeley, 1996; Badre & D'Esposito, 2009; Cohen, Dunbar, & McClelland, 1990; Dehaene, Kerszberg, & Changeux, 1998; Duncan, 2001; Koechlin & Summerfield, 2007; Miller & Cohen, 2001; Norman & Shallice, 1986; Reynolds, O'Reilly, Cohen, & Braver, 2012). The emergent consensus is a hierarchical view of cognitive processes, where a complex executive system regulates, monitors, and inhibits simpler domain-specific cognitive processes. In the following section I will discuss a selection of the most influential models of executive functions (broadly

conceived as processes that organise and control cognitive processes) that consider the way in which attentional control is realised in the brain.

Models of executive function

A prominent model proposed by Norman and Shallice (Norman & Shallice, 1980, 1986; Shallice & Burgess, 1991) states that in certain situations a "supervisory attention system" (SAS) is needed to influence selection and enable deliberate control. This model of attentional control assumes that two complementary processes operate in the selection and control of action. The basic mechanism is termed "contention scheduling", which is thought to be able to control routine activities automatically, without conscious control. According to this framework, behaviour is governed by sets of thought or action 'schemas'. A schema is a set of actions or cognitions that have become very closely associated through practice. However, in circumstances that require a higher level of cognitive control the SAS then intervenes and provides additional inhibition or activation to the appropriate schema. The operation of the SAS is thought to be necessary for behaviour in situations that involve planning and decision making, error correction, that contain novel sequences of actions, and when the overcoming of a strong habitual response is required. An impairment of the SAS then leads to difficulties in these situations. For example, in some situations, environmental triggers lead to the activation of one schema, but an alternative schema needs to be selected. In these situations, damage to the SAS will make it more likely that the previously relevant schema, triggered by environmental events, will continue to be selected, leading to behavioural rigidity. In other situations, a reduction in supervisory input may lead to the triggering of inappropriate behaviour by salient objects in the environment, leading to distractibility or an inability to filter out irrelevant details. Thus, damage to the SAS could explain excessive rigidity as

well as excessive distractibility, both of which have been reported to occur following damage to the frontal lobes (e.g. Luria, 1966; Manes et al., 2002).

The SAS model ties in closely with an earlier model of WM, defined as a short-term memory mechanism that integrates moment-to-moment perceptions across time, proposed by Baddeley and colleagues (Baddeley et al., 1996; Baddeley & Hitch, 1974). In this model, WM has three distinct components, one for verbal memories (the phonological loop), one for visual and spatial information (the visuospatial sketchpad) and a third referred to as the 'central executive' component. According to this model, the central executive coordinates the flow of attention between the components of WM and is often linked to the functioning of the frontal lobes. The central executive functions as a limited capacity system, responsible for strategy selection, planning, and the attentional control of action. The notion of a "central executive" is similar to the SAS, indeed, in an attempt to specify the subcomponents of executive control in greater detail, Baddeley and colleagues (1996) incorporated the SAS from Norman and Shallice's (1980) model of attentional control as an approximation of central executive functioning. This provided a framework for specifying the processes and capacities needed by such an attentional controller. A few basic capacities were postulated and explored (Baddeley et al., 1996): the ability to focus, to divide and to switch attention, and the ability to relate the content of WM to long-term memory. Other accounts of executive function have also referred to subcomponents of executive function (e.g. Miyake et al., 2000), but the consistent theme across these models is a view of a hierarchical and flexible system that enables goal-directed behaviour through the influence of one or more 'executive' cognitive modules on other, more specialised, ones.

More recent models have gone further, linking executive and specialised modules onto different brain regions, such as the prefrontal cortex (PFC). For example, Miller and

Cohen (2001) also drew distinctions between a central executive and other specialised systems, similar to the models of control discussed above. Based on findings from the non-human primate literature, including both neuropsychological and neuroanatomical data, Miller and Cohen (2001) proposed that the PFC is key when behaviour must be guided by internal states or intentions (i.e., top-down control, see also Fuster, 2008). In Miller and Cohen's (2001) model of prefrontal function, they argue that the capacity to support sustained activity in the face of interference is one of the distinguishing characteristics of the PFC. In their account, they discuss how the PFC is anatomically well situated to provide feedback signals and exert biasing influences on other structures throughout the brain. Other accounts of executive control consider a similar hierarchical view of the brain. For example, a computational model (Dehaene & Changeux, 2011; Dehaene et al., 1998) distinguished between a 'global workspace' and specialised modular subsystems. In this model, the global workspace interconnects multiple specialized brain areas in a coordinated fashion, similar to Miller and Cohen's (2001) model of prefrontal function.

The PFC covers a considerable portion of the human brain, and several theories of prefrontal organisation suggest that there are specialised functional regions within the PFC (e.g. Badre & D'Esposito, 2009; Botvinick, 2008; Bunge & Zelazo, 2006; Christoff & Keramatian, 2007; Koechlin & Summerfield, 2007; O'Reilly, 2010). For example, the rule abstraction model (Badre, 2008; Badre & D'Esposito, 2007, 2009) suggests a gradient with different regions in the PFC recruited according to the control demands of the task: as one moves anteriorly, prefrontal areas are recruited to support the use of progressively more complex rules. An alternative account is the information cascade model (Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007), which suggests that as one moves in a caudal to rostral direction across the lateral frontal cortex, there is a change in

the nature of the information being represented. The furthest rostral end of the hierarchy might hold information that has potential future relevance, whereas the furthest caudal region codes information about the current stimulus. These models are supported by neuroimaging and neuropsychological data demonstrating that further rostral areas process increasingly abstract representations (e.g. Azuar et al., 2014; Badre & D'Esposito, 2007; Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; Koechlin & Jubault, 2006; Nee & Brown, 2012) although see (Crittenden & Duncan, 2012; Reynolds et al., 2012).

Other accounts have been offered in which both anterior and posterior PFC represents task-relevant information, but activity is modulated according to whether information needs to be transiently updated or sustained over longer time periods (e.g. Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Dosenbach et al., 2007; Dosenbach et al., 2006; Reynolds et al., 2012). Still other proposed divisions include different involvement of dorsal and ventral regions based on stimulus modality (Goldman-Rakic, Roberts, Robbins, & Weiskrantz, 1998), or medial and lateral segregation corresponding to monitoring for conflict vs. implementing control (Botvinick, 2008), and others (e.g. O'Reilly, 2010). Corbetta and Shulman (2002) also proposed two separate networks for attention where a dorsal frontoparietal system is involved in goal-directed stimulus-response selection whilst a ventral system is important for stimulus-driven shifts of attention, working as an alerting mechanism to detect behaviourally relevant stimuli. At this point there is no consensus on these organisation schemes, and various sources of empirical evidence provide challenges to these proposals (e.g. Crittenden & Duncan, 2012).

The evolving consensus here is of a complex executive system that regulates more specialised processing areas. Specific frontal and parietal regions may form part of this executive system. In the following section I will discuss an influential theory, termed the

adaptive coding hypothesis (Duncan, 2001), that again paints a hierarchical view of the brain, but differs from the other models discussed in the emphasis placed on the adaptive response of single neurons.

The Adaptive Coding Hypothesis

The adaptive coding hypothesis (Duncan, 2001; Duncan, 2010) proposes that a network consisting of specific frontal and parietal regions are involved in processing the relevant aspects of many different types of tasks. This network is proposed to support goal-directed behaviour by adjusting its responses to code the information that is currently relevant for behaviour. Neurons in the network are not always tuned to the same specific features in the environment, but rather their response properties are highly adaptable so they shift their tuning profiles to code features or information that is currently relevant.

The adaptive coding model differs from other models discussed above in the emphasis placed on adaptive responses of single frontoparietal neurons. Rather than the recruitment of different regions for different tasks (e.g. Botvinick, 2008; Corbetta & Shulman, 2002; Dosenbach et al., 2007; Dosenbach et al., 2006; Goldman-Rakic et al., 1998; O'Reilly, 2010; Reynolds et al., 2012), this model accounts for flexible behaviour by proposing that responses within a single system adapt to represent currently needed information across tasks. Similar to the models discussed above (e.g. Dehaene et al., 1998; Miller & Cohen, 2001; Norman & Shallice, 1980), it posits a hierarchical structure of the brain in which information processing in more specialised processing areas is biased by the influence of this network (Desimone & Duncan, 1995). These specialised processing areas include those concerned with sensory inputs and the generation of motor commands. In this way, these higher regions can carry out a central function in configuring a flexible cognitive system to address what is currently relevant. This ability to adjust to the task-

relevant information, with accompanying top-down signalling to sensory, motor and other systems, is proposed to account for the flexibility of the human brain to adapt to a dynamic environment.

Single unit studies have provided evidence that neurons in both frontal and parietal cortex encode information that is relevant for the task. For example, the activity from single cells in lateral frontal cortex (e.g. Freedman, Riesenhuber, Poggio, & Miller, 2001; Kadohisa et al., 2013; Rao, Rainer, & Miller, 1997; Roy, Riesenhuber, Poggio, & Miller, 2010) and lateral parietal cortex (Andersen, Essick, & Siegel, 1985; Fitzgerald, Swaminathan, & Freedman, 2012; Gail & Andersen, 2006; Ibos, Duhamel, & Hamed, 2013; Stoet & Snyder, 2004; Swaminathan & Freedman, 2012) can discriminate a wide range of task features including task rules, cues, stimuli and responses. In Everling and colleagues' study (2002), 50% of all cells recorded in the lateral PFC (lPFC) discriminated targets (relevant) from non-targets (irrelevant), while many fewer cells made the task-irrelevant distinction between one non-target and another. Furthermore, data from non-human primate research has supported the proposal that neurons *adjust* their responses across tasks to encode what is currently relevant. For example, in a study conducted by Rao and colleagues (1997), monkeys performed a combined 'what–where' WM task, where in different phases of each trial monkeys retained either target identity or target location. Across the lPFC, many single neurons carried both identity and location information. Importantly, when the task required a switch from identity to location, this switch was reflected in the responses of individual neurons, identity information being discarded and location information taken up. These data provide evidence for adaptive coding in this area (Duncan, 2001), whereby lPFC neurons adjust to code the relevant feature information of the current phase of the task.

Another example of adaptive coding in non-human primate comes from Freedman et al (2001). Freedman and colleagues (2001) used morphing software to create stimuli that fell into two general categories: 'cats' and 'dogs'. Three species of cat and three breeds of dog were used as prototypes. The remaining stimulus set varied continuously either between 'cat' and 'dog' or between two prototypes within the same category (e.g. two types of cat, or two types of dog). Monkeys were trained to perform a categorisation task on this stimulus set where on each trial two stimuli were presented, separated by an interval. The monkeys indicated whether two sequentially presented stimuli were from the same or different categories according to an arbitrary decision-boundary, ignoring within category differences. 20% of lPFC neurons adjusted their firing rates to reflect these decision boundaries, differentiating between cats and dogs, even for those exemplars close to the decision boundary (e.g. a neuron responding strongly to cats would respond to a morph made up of 60% cat and 40% dog but not to a morph made up of 60% dog and 40% cat). Moreover, when monkeys were trained on a new, orthogonal, decision-boundary based on the same stimulus set where the cat-dog distinction was now irrelevant, the neurons altered their firing rates to reflect the new category distinctions that the monkeys had learnt. The results of this study emphasise how lPFC neurons demonstrate a remarkable flexibility to adjust their firing rates to reflect the currently task-relevant information, even across orthogonal categorical decision boundaries.

Several other studies have similarly demonstrated that prefrontal neurons can encode the behavioural meaning of visual stimuli, regardless of their physical properties (Cromer, Roy, & Miller, 2010; Freedman, Riesenhuber, Poggio, & Miller, 2002; Roy et al., 2010; Sakagami & Niki, 1994; Sakagami & Tsutsui, 1999; Watanabe, 1986). For example, two recent studies (Cromer et al., 2010; Roy et al., 2010) used a similar paradigm to that of Freedman and colleagues (2001) and showed comparable results. Roy

et al. (2010) demonstrated that 24% of lPFC neurons had a distinct firing rate in response to one category of visual stimuli over another. Similar to Freedman and colleagues' (2001) study, these neurons responded to the relevant category membership of the stimuli, rather than to their simple visual properties. They were also shown to alter their firing rate to reflect the new task when these stimuli were re-categorised in orthogonal categories. Cromer and colleagues (2010) showed that lPFC neurons responded to the relevant category membership of a different set of stimuli (sports cars vs. sedan cars), and when monkeys were re-trained to categorise a separate stimulus set (cats vs. dogs), 44% of the task-responsive neurons changed their firing rate to reflect the new task. These data also provide evidence that substantial portions of lPFC neurons are able to engage in multiple cognitive tasks, emphasising the flexibility of these neurons to alter their coding as needed for behaviour.

Similar results have been shown in the lateral parietal cortex (Freedman & Assad, 2006) where the activity of lateral parietal neurons encoded the direction of motion according to category membership. This encoding then shifted to the new category membership after the monkeys were retrained to group the same stimuli into two new categories. These data indicate that both lateral frontal and lateral parietal neurons are able to alter their responses as needed for behaviour as well as shift their responses to engage in multiple cognitive tasks.

Kadohisa et al. (2013) tracked the shift of prefrontal neurons dynamically allocated to processing task-relevant information over time. In this study, monkeys were first presented with a cue followed by one or two target objects, which could be presented to either or both visual fields. These target objects could be either relevant to the alternative cue (currently irrelevant), associated with the current cue (behaviourally relevant targets)

and/or targets that were never associated with either cue (always irrelevant). Interestingly, the results showed that when two stimuli were present in the display, prefrontal neural resources were reallocated over time. During early processing responses were dominated by the stimulus (relevant or irrelevant) in the contralateral hemisphere, but later, globally across hemispheres, activity was dominated by the behaviourally relevant object (irrespective of its location). This suggests that higher regions exert control by reallocating attentional resources, over time, to favour behaviourally relevant information. These findings again emphasise how frontal neurons have the potential to respond to different task features, and to alter their responses to favour behaviourally relevant information.

Although the electrophysiological studies reviewed above provide detailed information at the level of single-cell activity, they are limited in the breadth of brain coverage and scale of brain network under study. In the human brain, many associations have been drawn between specific subregions of frontal and parietal areas of the human brain and explicit executive functions. However, Duncan and Owen (2000) demonstrated that a wide range of tasks activate a common network of frontal and parietal regions. Duncan and Owen's (2000) meta-analysis reviewed twenty positron emission tomography (PET) and fMRI studies which implemented tasks that manipulated cognitive demand. The five categories of "demand" that these studies manipulated were response conflict, task novelty, duration of delay (WM), items to recall (WM), and perceptual difficulty. The findings showed that across these different types of demands, there was a similar recruitment of a particular constellation of frontal and parietal regions, which Duncan (2006, 2010) named 'multiple demand' or MD regions. These regions incorporate the anterior insula/frontal operculum (AI/FO), the inferior frontal sulcus (IFS), the dorsal anterior cingulate/pre-SMA (ACC/pre-SMA), and the IPS (Figure 1).

MD network   IFS   AI/FO   ACC/pre-SMA   IPS
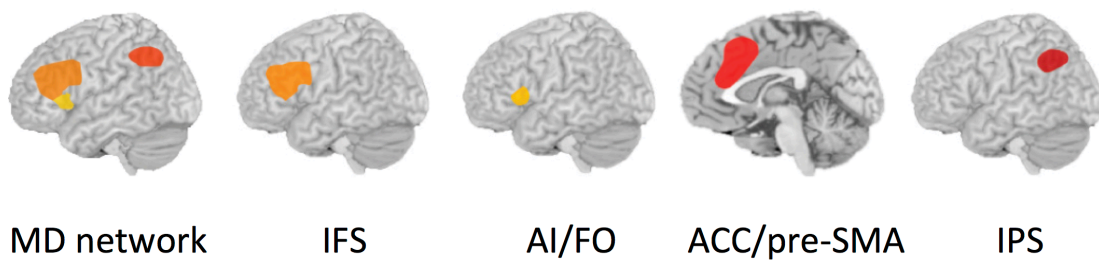
Figure 1: Illustration of the "multiple demand" (MD) regions of the human brain projected on a standard template of the left hemisphere (regions are symmetrical in the right hemisphere)

More recent studies have confirmed that the MD regions respond to a wide range of task demands (e.g. Dosenbach et al., 2006; Fedorenko et al., 2013; Niendam et al., 2012; Nyberg et al., 2003; Stiers, Mennes, & Sunaert, 2010), including at the level of single subjects (Fedorenko et al., 2013). In the study by Fedorenko and colleagues (2013), 40 participants completed 7 different tasks varying in type of cognitive demand: language, arithmetic, verbal and spatial working memory, and response selection/inhibition, while in the scanner. Similar to the tasks in Duncan and Owen's meta-analysis (2000), each task had a harder and easier version to assess increased cognitive demand. Fedorenko et al.'s (2013) study found comparable MD activity associated with increased difficulty in each separate task at both group and single-subject levels. Additionally, in individual subjects, these regions were often adjacent to regions showing a different pattern of responses that did not modulate with difficulty across the tasks. This indicates that sensitivity to task difficulty is a specific feature of the MD regions. In line with their broad response across tasks, these regions have been variously referred to as the "task positive network" (Fox et al., 2005), "frontoparietal control system" (Vincent et al., 2008), "task activation ensemble" (Seeley et al., 2007) and have been described as "flexible hubs" that adjust their connectivity patterns along with task demands (Cole et al., 2013).

The studies reviewed here provide evidence for a system that responds to increasing cognitive demands across a variety of different types of cognitive demand. This makes the MD regions candidates for adaptive coding, since a minimal prediction the adaptive coding hypothesis is that regions showing adaptive coding should be involved in a range of different tasks (Duncan, 2001; Duncan, 2010, 2013). However, there are alternative explanations for common activation between tasks; for example, common activation could reflect a very general response, such as effort, which would not necessarily be related to adaptive coding. Until recently, our capacity for examining *coding* in humans was limited. The advent of new analysis methods for fMRI that use the pattern of response across voxels to infer *what information is encoded* greatly enhance the inference possible and mean that we can test the more specific predictions of the adaptive coding hypothesis in the human brain.

Traditional univariate fMRI analyses mainly focus on mapping the magnitude of changes in the BOLD signal in various brain areas during different task conditions. These conventional methods look for voxels that show a significantly different response to experimental conditions relative to some baseline or control condition. Typically, data are spatially smoothed and activity is averaged across voxels within a region of interest. This is done to boost signal-to-noise ratio, but results in a loss of sensitivity to fine-grained spatial-pattern information, blurring out spatial patterns that might discriminate between experimental conditions (e.g. Mur, Bandettini, & Kriegeskorte, 2009).

Unlike traditional methods, multivoxel pattern analysis (MVPA) preserves the fine-grained information in fMRI data by extracting information from patterns of activity across multiple voxels (e.g. Haxby et al., 2001; Haynes & Rees, 2006; Haynes & Rees, 2005; Kamitani & Tong, 2005; Yang, Fang, & Weng, 2012). Voxels considered

22

individually might not be significantly responsive to any of the conditions of interest; however, the multi-voxel codes across many voxels have the potential to reveal patterns of activation relating to those conditions. MVPA also does not traditionally involve spatial averaging of voxel responses as with univariate-based analysis, whereby fine-grained pattern differences can go undetected unless the regional-average activation also differs. MVPA is therefore suited for detecting pattern changes even if they occur in the absence of regional-average activation changes. Consider an example response to the two novel objects shown in Figure 2 (left panels). For the hypothetical region in this example, average activation is similar for both objects. However, the *pattern* of activation over voxels is different; this region still carries information that distinguishes between the two objects at a multi-voxel scale. In this case, MVPA would be a suitable technique for examining whether information about these stimuli is carried in a particular brain region.
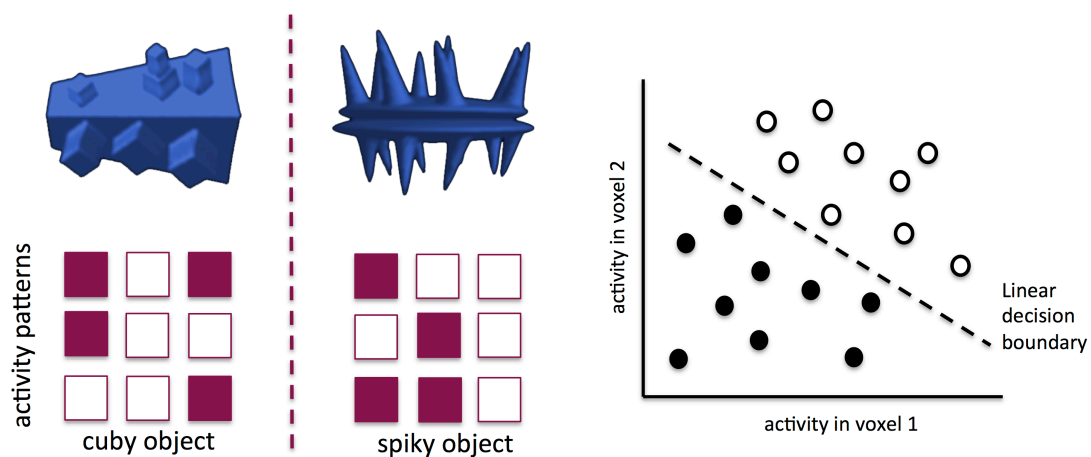
Figure 2: Left images: Hypothetical multi-voxel activity patterns for a 'cuby' and a 'spiky' object within a region of interest. In this hypothetical region, average activation for both objects is identical, but there is information that can be detected with MVPA to distinguish between the experimental conditions (voxel activation responses to conditions in colour). Right: Linear decision boundary separating condition A (e.g. cuby objects) and condition B (e.g. spiky objects)

There are many different algorithms that can be used to undertake MVPA but they all share the basic principle of interpreting the data in a multidimensional space, where each dimension (usually) corresponds to the response of an individual voxel. Figure 2 (right) illustrates an example where we consider the activity patterns of only two voxels. In order to classify these patterns we can construct a line that separates our two conditions (cuby vs. spiky object), even though the response of each voxel individually does not discriminate between the two conditions (Cox & Savoy, 2003; Haynes & Rees, 2006). Typically, the analysis would include many such voxels, so the decision line generalises to a multidimensional hyper plane.
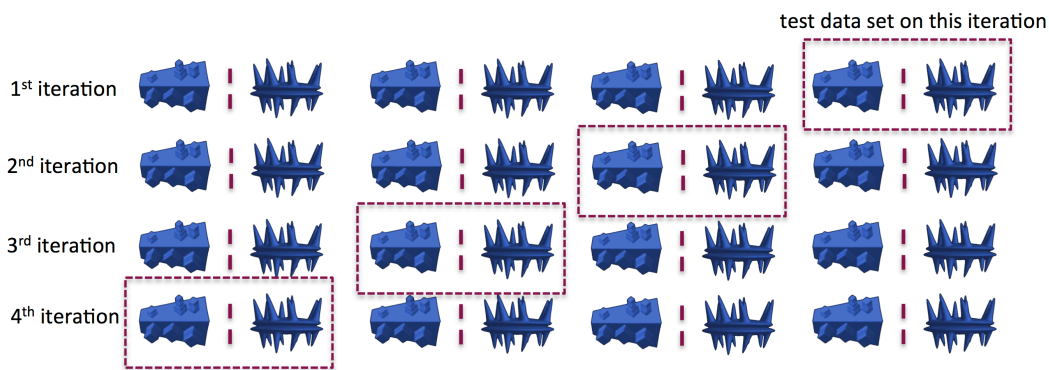
Figure 3: Example of a standard MVPA cross-validation training and testing procedure: In this example there are four experimental runs and the classifier is trained to distinguish between the patterns of activation pertaining to cuby and spiky objects. One subset of the data is left out on each iteration as the independent test data set. This procedure is repeated leaving out a different subset until all data contributes equally to test and train subsets. Performance on the different iterations is averaged together to obtain overall classifier performance.

The classifier algorithm attempts to derive a model that describes how responses to the experimental conditions are separated in this space. The classifier is trained to classify a set of patterns (referred to as samples, Figure 3) from the different experimental

conditions and is then tested with an independent set of samples (that did not contribute to the training model). The accuracy with which the classifier can predict the condition to which the test samples belong forms a measure of the information held in that region about the particular categories tested (classification accuracy).

The accuracy of the categorisation of test samples indicates how well the classifier performs in identifying differences between the samples from different conditions. If the pattern of activation in a particular region can be used to consistently discriminate between two task events, at an accuracy greater than that expected by chance (the null hypothesis), then that region is said to 'carry information' about the conditions. In order to preserve fine-grained subject-specific information, the patterns are not typically averaged across subjects. Instead, MVPA is performed in native subject space for each individual, and group analysis can be performed as a second level analysis.

MVPA is fundamentally limited by the amount of information about the neural population codes that can be provided by fMRI. Voxel resolution is one such limitation and an ongoing debate continues as to what gives rise to the pattern of voxel biases exploited by classification algorithms. Early MVPA fMRI research (Haynes & Rees, 2005; Kamitani & Tong, 2005) posited that pattern classification methods exploit the fine functional, columnar architecture of neuronal preferences. For example, a single voxel may happen to sample more of one type of cortical column (or neuronal preference e.g. orientation ~45°) relative to others, therefore the response to one condition would be slightly different between voxels even though the voxel resolution is not fine enough to resolve cortical columns directly. This has been referred to as the "hyperacuity" hypothesis. This assumption has been called into question by recent studies showing the existence of large-scale patterns of response bias for orientation (Freeman, Brouwer,

Heeger, & Merriam, 2011) and motion direction (Beckett, Peirce, Sanchez-Panchuelo, Francis, & Schluppeck, 2012), which can account for decoding these features. Op de Beeck (2010) also showed that MVPA is robust to spatial smoothing, which has been interpreted as evidence against a columnar-scale bias driving classification (Op de Beeck, 2010). In addition, the assumption that the response biases reflect neuronal response properties has been called into question by results suggesting such biases may be vascular in origin (Gardner, 2010; Shmuel, Chaimow, Raddatz, Ugurbil, & Yacoub, 2010). Despite the ongoing debate and the fact that sensitivity is limited by the measurement technique of fMRI, statistically distinct activity patterns nonetheless provide strong evidence for a difference between the underling neural activity in the region, even if we cannot be certain on what scale these patterns arise.

Now let us return to the evidence that the MD regions are involved in adaptive coding. As discussed earlier, data from the non-human primate literature has previously shown that both frontal (e.g. Freedman et al., 2001; Kadohisa et al., 2013; Rao et al., 1997; Roy et al., 2010) and parietal cortex (Andersen et al., 1985; Fitzgerald et al., 2012; Gail & Andersen, 2006; Stoet & Snyder, 2004; Swaminathan & Freedman, 2012) discriminate a range of task information. MVPA provides the potential to explore similar questions about adaptive coding in the human brain: regions demonstrating adaptive coding should code a range of different types of information under different circumstances (Duncan, 2001).

Using MVPA of fMRI data, the MD network has indeed been shown to code a range of task information demonstrating flexibility to respond to a variety of behaviourally relevant stimuli in different contexts (Bode & Haynes, 2009; Harel, Kravitz, & Baker, 2014; Haynes et al., 2007; Li, Ostwald, Giese, & Kourtzi, 2007; Nee & Brown, 2012;

Reverberi, Gorgen, & Haynes, 2011; Soon, Namburi, & Chee, 2013; Stiers et al., 2010; Waskom, Kumaran, Gordon, Rissman, & Wagner, 2014; Woolgar, Afshar, Williams, & Rich, 2015; Woolgar, Hampshire, Thompson, & Duncan, 2011; Woolgar, Thompson, Bor, & Duncan, 2011; Woolgar, Williams, & Rich, 2015). We recently conducted a meta-analysis (Appendix A, Woolgar, Jackson, & Duncan, in press) drawing on 100 published decoding analyses to examine the various types of information coding across the human brain.
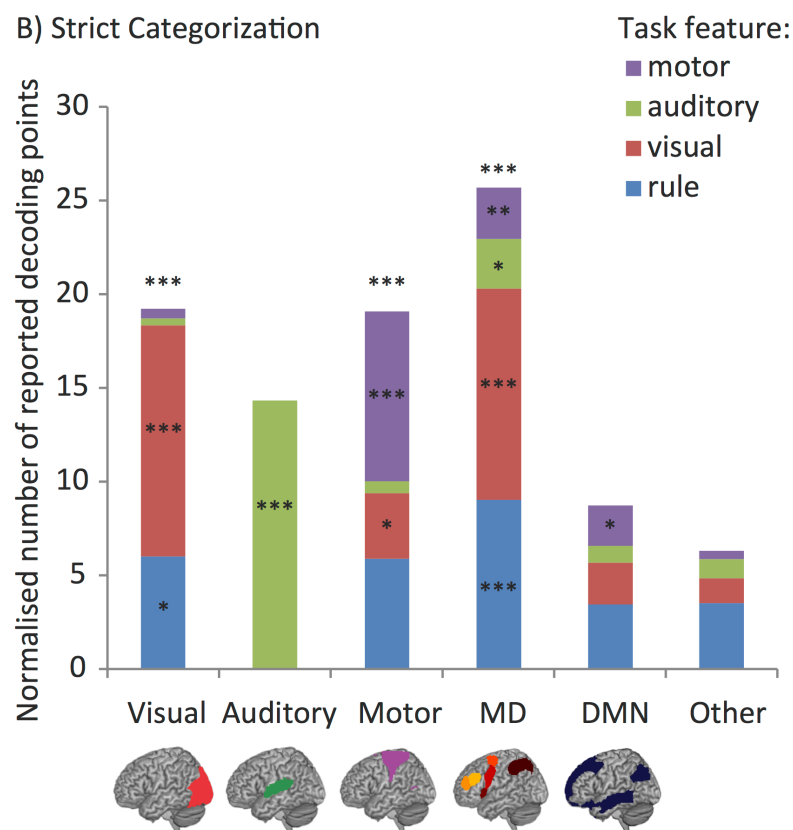
Figure 4 (taken from Woolgar et al., (in press), see Appendix A): Number of significant decoding points reported in each network, after correcting for the number of analyses examining coding of each task feature and network volume. Asterisks indicate significance of chi-squared or exact binomial goodness of fit tests examining whether there was more coding in each principal network compared to Other for all decoding points (above bars) or

for decoding points in each task feature separately (asterisks on coloured bar segments). *p<0.05, **p<0.01, *** p<0.00001.

The MD network encoded a range of task features, including visual, auditory, motor and rule information (Figure 4, taken from Woolgar et al., in press), more frequently than predicted based on network volume. This contrasted with more specialised processing areas such as the visual, auditory and motor cortices which primarily coded information from their own domain. These data confirm that the MD regions are involved in processing a range of different types of information across different tasks.

A stronger test of adaptive coding is whether individual regions in the human brain adjust responses within single tasks, for example, if task demands change. Recent MVPA studies have shown that the MD network indeed adjusts its coding of perceptual (Woolgar, Hampshire, et al., 2011; Woolgar, Williams, et al., 2015) and task rule information (Woolgar, Afshar, et al., 2015) as task demands vary. For example, in Woolgar et al.'s (2011) study, participants responded to the location of a blue square on a screen. The stimulus positions were either perceptually easy (far apart) or difficult (close together) to distinguish. Across the MD regions, position coding was significantly stronger in the more difficult relative to the easier condition. In contrast, the visual cortices showed the opposite result, with a weaker representation of the difficult relative to the easy stimuli, in line with the physical stimulus differences. These data suggest that the MD regions adjusted to emphasise the perceptual information when it was more challenging and the input from the visual cortex was weak. Woolgar and colleagues (2015) found similar results for the representation of task rules in the MD network: Coding of rule information was stronger when the rules were more complex and behaviourally confusable compared to when they were relatively simple. These studies indicate that the MD network can

indeed adjust its responses within single tasks, at least as measured at the voxel level, consistent with the adaptive coding hypothesis.

Adaptive coding in the MD network could form the basis for the operation of selective attention. Woolgar et al. (2015) found that the multi-voxel representation of objects in the MD network was stronger when objects were attended compared to when they were ignored. These data indicate that the behavioural relevance of task stimuli, in this case whether the stimuli were attended (targets) or ignored (distractors), influences coding in these regions. The behavioural relevance of a stimulus has also been found to modulate MD network coding in a category discrimination task (Erez & Duncan, 2015). In this study, participants were cued at the start of a block with the names of two target categories (e.g., shoes, butterflies). In the following four trials they were then shown a picture of an object on each trial and had to decide whether it belonged to the target category. The presented stimulus could either be a target (target category cue is shoe and subsequent presented stimulus is a shoe), consistent non-target (target category cue is shoe and subsequent presented stimulus is a sofa, and a sofa is never a target category), or an inconsistent non-target (target category cue is butterflies and subsequent presented stimulus is a shoe, which is a target on other trials). The results showed that multi-voxel patterns in the MD regions discriminated visual categories for which the distinction was behaviourally relevant. In contrast, behaviourally irrelevant category distinctions were not coded. These data show that behavioural relevance modulates category discrimination across the MD network, consistent with the view that the MD network adjusts its responses to code the relevant aspects of a task. These responses to behavioural relevance could provide the source of bias for the implementation of attention in the brain.

The MD regions are not the only areas in which attention modulates responses. Preferential responses to attended information is also seen, for example, throughout the

visual cortex, where we see a stronger response to attended versus ignored visual input (Jehee, Brady, & Tong, 2011; Moran & Desimone, 1985; Murray & He, 2006; Murray & Wojciulik, 2003; Serences, Saproo, Scolari, Ho, & Muftuler, 2009). The MD regions have extensive connectivity patterns with subcortical, sensory and motor regions (Selemon & Goldman-Rakic, 1988) and it is commonly suggested that these regions bias activity towards task-relevant information and provide control input to these other cortical and subcortical systems (Dehaene et al., 1998; Desimone & Duncan, 1995; Duncan, 2013; Miller & Cohen, 2001; Norman & Shallice, 1980). However, this is difficult to examine using fMRI data alone (see section 1.1 for evidence from combined TMS/EEG and TMS/fMRI). The MD regions may therefore play an important role in implementing the controls for selective attention.

The literature indicates that the MD network is involved in cognitive control and that it appears to work adaptively to code a range of task features. The adaptive coding hypothesis offers a possible mechanism for the way in which selective attention is achieved. However, further research is needed to test the hypothesis that these regions preferentially code task-relevant information in a range of situations, for example, not only in the case of spatial attention and for attended objects but also for attended *features* of objects. Additionally, the question of how the MD regions influence and potentially bias processing in specialised processing regions, particularly information coding in other regions, and whether this depends on the selective representation of information in this system, is a critical question that requires further investigation.

Feature-selective attention

Much of the previous research on the control of attention has focused on spatial attention. This is the process where our attention can be deployed to locations in space, for

example, where spatial attention is varied by pre-cueing the location where a target stimulus is likely to appear (e.g. Posner, 1980). Voluntary spatial attention is thought to depend on top-down mechanisms (Yantis & Serences, 2003), for example, directing attention to a particular location in the visual field has also been shown to modulate neuronal responses in corresponding part of the primate visual cortex (e.g. Luck, Chelazzi, Hillyard, & Desimone, 1997; McAdams & Maunsell, 1999; Moran & Desimone, 1985; Seidemann & Newsome, 1999). Spatial attention has dominated investigations of attention, but we also have the capacity to allocate our attention to a particular feature of an object or visual field. This type of attention can be further divided into *feature-based* and *feature-selective* attention.

Feature-based attention encompasses the ability to enhance the representation of image components throughout the visual field that are related to a particular feature. For example, it can support our ability to detect a behaviourally relevant target among distractor items, as in popular 'visual search' paradigms in visual psychophysics. In visual search experiments, targets and distractors differ by at least one feature, and target detection can be improved by enhancing the representation of image components that match the attended feature (e.g. the colour red or a vertical orientation) and by suppressing those that do not. Several psychophysical studies have demonstrated that feature-based attention improves detection or otherwise enhances behavioural performance across the visual field (e.g. Cohen & Magen, 1999; Saenz, Buraĉas, & Boynton, 2003).

Feature-based attention is typically thought to involve up-regulating processing of the parts of visual field that match a given value of a given feature (e.g. red). This overlaps with, but is distinct from, the ability to selectively attend to a particular feature dimension of an object (e.g. to make fine discriminations concerning its colour), while ignoring other features of that object (e.g. its shape). The latter is called *feature-selective* attention,

because selection affects which feature is attended to. Although the terms feature-based and feature-selective attention are often used interchangeably, they are distinct because feature-based attention does not require the selection of a certain feature of an object *whilst simultaneously ignoring* other features of the object (Chen et al., 2012). Feature-based attention requires enhancement of aspects of a visual field that match a template feature. Feature-selective attention requires enhanced processing of a task-relevant feature of an object and simultaneous suppression of a task-irrelevant feature of that same object. Feature-selective attention is particularly important in situations where two features of an object result in response conflict (e.g. if the orientation of a grating requires a left button response whilst the contrast of the grating requires a right button response).

The PFC and lateral parietal cortex (constituent parts of the MD network, see section 1.3) have been specifically implicated in feature-selective attention in the non-human primate literature (e.g. Freedman & Assad, 2006; Lauwereyns et al., 2001; Rao et al., 1997). For example, in Lauwereyns et al.'s (2001) study, macaque monkeys were trained to classify patterns of moving dots according to either their colour or their direction of motion. This required the monkeys to extract the relevant feature dimension whilst ignoring the other feature. They found that around 20% of prefrontal cells showed strong attentional modulation. These cells discriminated between stimulus colours when the monkeys attended to colour, but discriminated movement direction when the animals attended to stimulus motion. These data demonstrate adaptive coding for feature-specific information in the non-human primate brain.

In the human brain, Thompson and Duncan (2009) used an fMRI adaptation analysis to show that the MD network responds more to changes in attended stimulus features (colour/shape) than to changes in unattended stimulus features. Similarly, Li et al., (2007) used MVPA to demonstrate that the MD regions preferentially coded the

features of moving dot figures that were relevant to the task: when the task changed, so did the pattern of information coding. These findings offer suggestive evidence that adaptive neural coding in the human frontoparietal cortex may support selection at the level of object features. In this thesis, I examine this further, asking whether adaptive coding provides a mechanism for feature-selective attention through adjustment of response properties in the MD network, by prioritising processing of task-relevant object features over task-irrelevant object features.

Overview of thesis

The evidence reviewed above is a firm foundation for the hypothesis that adaptive coding in the MD network supports feature-selective attention. We know that the MD regions adjust their responses in line with task demands, encode a range of task-related information, and respond to changes in attended features. It seems reasonable then, that this selective representation in the MD regions would support task-relevant decision-making processes necessary for solving the task at hand, and could be the source of bias supporting processing of the attended feature in more specialised regions (e.g. Dehaene et al., 1998; Desimone & Duncan, 1995; Miller & Cohen, 2001).

In this thesis, I explore the role of the MD network in feature-selective attention. The main questions addressed by the following experimental chapters are: **1.** Does the MD network adjust its responses to prioritise coding of relevant over irrelevant object feature information? **2.** Do the same voxels show coding of relevant features across multiple tasks? **3.** Does the dlPFC play a causal role in filtering out irrelevant feature information? **4.** Does selection of relevant feature information and/or inhibition of irrelevant feature information in the dlPFC underlie top-down modulation of earlier processing regions? I

used behavioural, fMRI, MVPA, TMS and concurrent TMS-fMRI methods to answer these research questions in 4 experiments outlined below.

In Chapter 2, I present a study examining whether the MD network prioritises coding of task-relevant feature information over task-irrelevant feature information. I used MVPA of fMRI data acquired while participants performed a perceptually challenging categorisation task. At different times, participants were required to discriminate novel "spiky" objects across one of two orthogonal decision boundaries based on two feature dimensions, length and orientation. I tested whether MD representation of visual object features flexibly adjusted according to task-relevance (coding for length information when length was relevant, and orientation information when orientation was relevant). The results showed that the MD network coded the task-relevant feature distinctions more strongly than the equivalent task-irrelevant feature distinctions. These data demonstrate that the MD network adjusts its representation of objects to make the feature distinctions needed for the current behavioural task, providing support for adaptive coding as a mechanism for the implementation of feature-selective attention.

In Chapter 3, I examined the flexible coding in the MD system in more detail, asking whether the same voxels carried information about the relevant object features across multiple tasks. First I examined whether the MD regions coded the task-relevant features of different objects, "spikies" and "smoothies", in two task contexts. Then, I assessed the extent to which the same voxels were 're-used' to code the relevant discriminations across both task contexts and compared this to the voxel re-use predicted by chance. Finally, I compared the extent of voxel re-use between the two tasks of Chapter 3 to the extent of re-use between the tasks in Chapter 2. I replicated the finding from the first experiment (Chapter 2) that the MD network preferentially coded the task-relevant

feature distinctions, this time across different stimulus sets. I also found that irrelevant feature information was coded more strongly in the MD regions when it was *inconsistently* irrelevant (i.e. when that information had recently been relevant on a different task) compared to when it was *never relevant* to the participant's task. The data also showed that significantly more MD voxels than expected by chance were re-used to code relevant feature information across the two tasks, suggesting, at least at the level of voxels, that attentional resources can be flexibly re-allocated in different tasks. Conversely, there was no evidence to suggest that the same voxels were re-used to code relevant feature information across the two tasks in the LOC or BA 17, indicating that this flexibility may be particular to higher cortical brain regions. Despite clear predictions from the non-human primate literature (Cromer et al., 2010; Roy et al., 2010), there was no difference in the extent of voxel re-use between the data from Chapters 3 and 2.

Chapter 4 presents a study where I explored whether the dlPFC plays a causal role in filtering irrelevant feature information. I used TMS to stimulate the dlPFC during a task requiring participants to ignore irrelevant feature information and compared behavioural performance to that under three control conditions. The active TMS condition was high intensity stimulation (HIS) of the dlPFC (100% of motor threshold). The three control conditions were a sham condition (inactive coil), low intensity stimulation (LIS) to the dlPFC (40% MT) and HIS (100% MT) to a control region. The button response for the irrelevant feature was either incongruent or congruent with the required button response for the currently relevant feature. The resulting congruency difference allowed us to index the extent to which irrelevant information processing influenced behaviour.

We selected the dlPFC for stimulation as it is the largest component of the MD system, extending caudally from around the region of the IFJ, along the IFS and middle

frontal gyrus to the anterior PFC. In addition, the right dlPFC is cited in the literature as a region that plays a causal role in modulating processing in more specialised regions (e.g. Feredoes et al., 2011; Zanto et al., 2011). The prediction was that disruption to the right dlPFC would either decrease selection of relevant information and/or decrease inhibition of irrelevant feature information, which consequently would result in an increase in the magnitude of the congruency effect relative to the control conditions. This experiment was also the pilot of the paradigm to be used in a combined fMRI-TMS experiment (Chapter 5). Although there was a main effect of congruency, our behavioural data did not show an effect of disruption to the right dlPFC. A Bayes analysis showed that more evidence was needed to make a firm conclusion. I subsequently conducted two further behavioural experiments to improve the sensitivity of the paradigm.

In Chapter 5, I asked whether the right dlPFC is causally involved in the selection of task-relevant feature information and/or inhibition of task-irrelevant feature information. I employed concurrent TMS and fMRI during a task where participants were required to ignore irrelevant feature information, using the task developed in Chapter 4. Participants completed two separate sessions: in the first session, they practised the behavioural task and completed several localiser tasks in the fMRI scanner; in the second session, participants were scanned whilst completing eight runs of the main task. The second session was combined with TMS so that on every trial, participants received three pulses of either LIS (40% MT, control) or HIS (100% MT) over the right dlPFC (functionally defined). The dlPFC has been associated with both inhibition of task-irrelevant information and selection of task-relevant information. Therefore I predicted that disruption to right dlPFC would result in **a)** stronger coding of irrelevant feature information and/or **b)** weaker coding of relevant feature information. I related this to participants' behaviour, predicting that HIS compared to LIS would disrupt selection of

the relevant feature information and/or inhibition of the irrelevant feature information and therefore result in a larger congruency effect.

In line with the proposal that the dlPFC suppresses the representation of irrelevant information, we found stronger coding of irrelevant information across the frontoparietal network and other brain regions, including early visual cortices, following right dlPFC disruption. Irrelevant feature information also had more effect on participants' reaction times under HIS. HIS to the dlPFC also modulated the representation of task relevant information. However, contrary to our prediction that disruption to dlPFC would impair selection of relevant feature information, we found *stronger* coding of relevant colour information across the brain under HIS relative to LIS, possibly indicating some form of adaptive compensation. These data are in line with a causal role for dlPFC in modulating the representation of relevant and irrelevant information in the brain.

In my final Chapter (Chapter 6), I summarise the main findings of my research and consider how it fits with the current literature. I discuss limitations of my research and challenges for this field as well as future directions, before drawing it all together with general conclusions.

References

Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A., & Eberle, L. (1989). Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 74*(6), 458-462.

Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A. P., & Eberle, L. (1993). Unmasking human visual perception with the magnetic coil and its relationship to hemispheric asymmetry. *Brain research, 605*(2), 312-316.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science, 230*(4724), 456-458.

Azuar, C., Reyes, P., Slachevsky, A., Volle, E., Kinkingnéhun, S., Kouneiher, F., . . . Levy, R. (2014). Testing the model of caudo-rostral organization of cognitive control in the human with frontal lesions. *NeuroImage, 84*, 1053-1060.

Baddeley, A., Della Sala, S., Robbins, T., & Baddeley, A. (1996). Working memory and executive control [and Discussion]. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 351*(1346), 1397-1404.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation, 8*, 47-89.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences, 12*(5), 193-200.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Cognitive Neuroscience, Journal of, 19*(12), 2082-2099.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*(9), 659-669.

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., . . . Shenker, J. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Cognitive Neuroscience, Journal of, 12*(6), 988-1000.

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., . . . Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. *Cognitive Brain Research, 10*(1–2), 1-9. doi:http://dx.doi.org/10.1016/S0926-6410(00)00015-X

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia, 40*(3), 349-356.

Beckett, A., Peirce, J. W., Sanchez-Panchuelo, R.-M., Francis, S., & Schluppeck, D. (2012). Contribution of large scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early visual cortex. *NeuroImage, 63*(3), 1623-1632.

Bench, C., Frith, C., Grasby, P., Friston, K., Paulesu, E., Frackowiak, R., & Dolan, R. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologia, 31*(9), 907-922.

Bestmann, S., & Feredoes, E. (2013). Combined neurostimulation and neuroimaging in cognitive neuroscience: past, present, and future. *Annals of the New York Academy of Sciences, 1296*(1), 11-30.

Bestmann, S., Swayne, O., Blankenburg, F., Ruff, C. C., Haggard, P., Weiskopf, N., . . . Ward, N. S. (2008). Dorsal premotor cortex exerts state-dependent causal

influences on activity in contralateral primary motor and dorsal premotor cortex. *Cerebral Cortex, 18*(6), 1281-1291.

Blankenburg, F., Ruff, C. C., Bestmann, S., Bjoertomt, O., Josephs, O., Deichmann, R., & Driver, J. (2010). Studying the role of human parietal cortex in visuospatial attention with concurrent TMS–fMRI. *Cerebral Cortex, 20*(11), 2702-2711.

Bode, S., & Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *NeuroImage, 45*(2), 606-613. doi:10.1016/j.neuroimage.2008.11.031

Bornstein, R. A. (1986). Contribution of various neuropsychological measures to detection of frontal lobe impairment. *International Journal of Clinical Neuropsychology*.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences, 12*(5), 201-208.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15*(3), 118-121.

Capotosto, P., Babiloni, C., Romani, G. L., & Corbetta, M. (2009). Frontoparietal cortex controls spatial attention through modulation of anticipatory alpha rhythms. *The Journal of Neuroscience, 29*(18), 5863-5872.

Capotosto, P., Babiloni, C., Romani, G. L., & Corbetta, M. (2011). Differential contribution of right and left parietal cortex to the control of spatial attention: a simultaneous EEG–rTMS study. *Cerebral Cortex*, bhr127.

Capotosto, P., Corbetta, M., Romani, G. L., & Babiloni, C. (2012). Electrophysiological correlates of stimulus-driven reorienting deficits after interference with right

parietal cortex during a spatial attention task: A TMS-EEG study. *Journal of Cognitive Neuroscience, 24*(12), 2363-2371.

Chao, L. L., & Knight, R. T. (1995). Human prefrontal lesions increase distractibility to irrelevant sensory inputs. *Neuroreport, 6*(12), 1605-1610.

Chen, X., Hoffmann, K. P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology, 107*, 1530-1543.

Christoff, K., & Keramatian, K. (2007). Abstraction of mental representations: theoretical considerations and neuroscientific evidence. *Perspectives on Rule-Guided Behavior, SA Bunge and JD Wallis, eds.(New York: Oxford University Press)*.

Cohen, A., & Magen, H. (1999). Intra-and cross-dimensional visual search for single-feature targets. *Perception & Psychophysics, 61*(2), 291-307.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological review, 97*(3), 332.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience, 16*(9), 1348-1355. doi:10.1038/nn.3470

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201-215.

Coulthard, E. J., Nachev, P., & Husain, M. (2008). Control over conflict during movement preparation: role of posterior parietal cortex. *Neuron, 58*(1), 144-157.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI)"brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage, 19*(2), 261-270.

Crittenden, B. M., & Duncan, J. (2012). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cerebral Cortex*, bhs333.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

de Beeck, H. P. O. (2010). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage, 49*(3), 1943-1948.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences, 95*(24), 14529-14534.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Dias, R., Robbins, T., & Roberts, A. (1996). Primate analogue of the Wisconsin Card Sorting Test: effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behavioral neuroscience, 110*(5), 872.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*(3), 99-105.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., . . . Raichle, M. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences, 104*(26), 11073-11078.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron, 50*(5), 799-812. doi:10.1016/j.neuron.2006.04.031

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *NeuroImage, 14*(6), 1256-1267.

Drewe, E. (1974). The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. *Cortex, 10*(2), 159-170.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820-829.

Duncan, J. (2006). EPS Mid-Career Award 2004: brain mechanisms of attention. *The Quarterly Journal of Experimental Psychology, 59*(1), 2-27.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobes recruited by diverse cognitive demands. *Trends in Neuroscience, 23*, 475-483.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 380-390.

Erez, Y., & Duncan, J. (2015). Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *The Journal of Neuroscience, 35*(36), 12383-12393.

Eslinger, P. J., & Grattan, L. M. (1993). Frontal lobe and frontal-striatal substrates for different forms of human cognitive flexibility. *Neuropsychologia, 31*(1), 17-28.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5*(7), 671-676. doi:10.1038/nn874

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences, 108*(42), 17510-17515.

Fitzgerald, J. K., Swaminathan, S. K., & Freedman, D. J. (2012). Visual categorization and the parietal cortex. *Frontiers in integrative neuroscience, 6*(18), 1-6.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature, 443*(7107), 85-88. doi:10.1038/nature05078

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312-316.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *Journal of Neurophysiology, 88*(2), 929-941.

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *The Journal of Neuroscience, 31*(13), 4792-4804.

Fuster, J. (2008). The Prefrontal Cortex, 4th Edn: Academic Press London.

Gail, A., & Andersen, R. A. (2006). Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. *The Journal of Neuroscience, 26*(37), 9376-9384.

Gardner, J. L. (2010). Is cortical vasculature functionally organized? *NeuroImage, 49*(3), 1953-1956.

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience, 17*(3), 507-517.

Goldman-Rakic, P., Roberts, A., Robbins, T., & Weiskrantz, L. (1998). The prefrontal cortex: executive and cognitive functions. *The prefrontal cortex: executive and cognitive functions*.

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*(4), 404.

Harel, A., Kravitz, D., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the USA*, 962-971.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*, 2425-2429.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci, 7*(7), 523-534.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology, 17*(4), 323-328. doi:10.1016/j.cub.2006.11.072

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

Heide, W., Blankenburg, M., Zimmermann, E., & Kömpf, D. (1995). Cortical control of double‐step saccades: implications for spatial orientation. *Annals of neurology, 38*(5), 739-748.

Heinen, K., Ruff, C. C., Bjoertomt, O., Schenkluhn, B., Bestmann, S., Blankenburg, F., . . . Chambers, C. D. (2011). Concurrent TMS–fMRI reveals dynamic interhemispheric influences of the right parietal cortex during exogenously cued visuospatial attention. *European Journal of Neuroscience, 33*(5), 991-1000.

Higo, T., Mars, R. B., Boorman, E. D., Buch, E. R., & Rushworth, M. F. (2011). Distributed and causal influence of frontal operculum in task control. *Proceedings of the National Academy of Sciences, 108*(10), 4230-4235.

Hon, N., Epstein, R. A., Owen, A. M., & Duncan, J. (2006). Frontoparietal activity with minimal decision and control. *The Journal of Neuroscience, 26*(38), 9805-9809.

Ibos, G., Duhamel, J.-R., & Hamed, S. B. (2013). A functional hierarchy within the parietofrontal network in stimulus selection and attention control. *The Journal of Neuroscience, 33*(19), 8359-8369.

Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral neuroscience, 103*(3), 548.

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience, 31*(22), 8210-8219. doi:10.1523/jneurosci.6153-09.2011

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., . . . Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80*(1), 235-246.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679-685.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91-100.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761.

Knight, R. T., Staines, W. R., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica, 101*(2), 159-178.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*(6732), 148-151.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron, 50*(6), 963-974.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181-1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*(6), 229-235.

Koenigs, M., Barbey, A. K., Postle, B. R., & Grafman, J. (2009). Superior parietal cortex is critical for the manipulation of information in working memory. *The Journal of Neuroscience, 29*(47), 14980-14986.

Lauwereyns, J., Sakagami, M., Tsutsui, K.-I., Kobayashi, S., Koizumi, M., & Hikosaka, O. (2001). Responses to task-irrelevant visual features by primate prefrontal neurons. *Journal of Neurophysiology, 86*(4).

Lee, T. G., & D'Esposito, M. (2012). The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI–TMS study. *The Journal of Neuroscience, 32*(44), 15458-15466.

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible Coding for Categorical Decisions in the Human Brain. *Journal of Neuroscience, 27*(45), 12321-12330. doi:10.1523/jneurosci.3795-07.2007

Liu, X., Banich, M. T., Jacobson, B. L., & Tanabe, J. L. (2004). Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage, 22*(3), 1097-1106. doi:http://dx.doi.org/10.1016/j.neuroimage.2004.02.033

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of neurophysiology, 77*(1), 24-42.

Luria, A. R. (1966). Higher cortical functions in man. *New York: Basic Books*.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science, 288*(5472), 1835-1838. doi:10.1126/science.288.5472.1835

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002). Decision‐making processes following damage to the prefrontal cortex. *Brain, 125*(3), 624-639.

McAdams, C. J., & Maunsell, J. H. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *The Journal of Neuroscience, 19*(1), 431-441.

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience, 23*(1), 1-10.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of neurology, 9*(1), 90-100.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782-784.

Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., & Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nature Neuroscience, 12*(1), 85-91.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social cognitive and affective neuroscience*, nsn044.

Murray, S. O., & He, S. (2006). Contrast Invariance in the Human Lateral Occipital Complex Depends on Attention. *Current Biology, 16*(6), 606-611. doi:10.1016/j.cub.2006.02.019

Murray, S. O., & Wojciulik, E. (2003). Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience, 7*(1), 70-74. doi:10.1038/nn1161

Nee, D. E., & Brown, J. W. (2012). Rostral–caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *NeuroImage, 63*(3), 1285-1294. doi:10.1016/j.neuroimage.2012.08.034

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*(4), 313-324.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience, 12*(2), 241-268.

Norman, D. A., & Shallice, T. (1980). Attention to Action: Willed and Automatic Control of Behavior Technical Report No. 8006.

Norman, D. A., & Shallice, T. (1986). *Attention to action*: Springer.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petterson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*, 371-377.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in neurosciences, 33*(8), 355-361.

Paus, T. (2005). Inferring causality in brain images: a perturbation approach. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1457), 1109-1114.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia, 12*(3), 323-330.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology, 32*(1), 3-25.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.

Reverberi, C., Gorgen, K., & Haynes, J. D. (2011). Compositionality of Rule Representations in Human Prefrontal Cortex. *Cerebral Cortex, 22*(6), 1237-1246. doi:10.1093/cercor/bhr200

Reynolds, J. R., O'Reilly, R. C., Cohen, J. D., & Braver, T. S. (2012). The function and organization of lateral prefrontal cortex: a test of competing hypotheses. *PloS one, 7*(2), e30284.

Robinson, A. L., Heaton, R. K., Lehman, R. A., & Stilson, D. W. (1980). The utility of the Wisconsin Card Sorting Test in detecting and localizing frontal lobe lesions. *Journal of consulting and clinical psychology, 48*(5), 605.

Roca, M., Parr, A., Thompson, R., Woolgar, A., Torralva, T., Antoun, N., . . . Duncan, J. (2009). Executive function and fluid intelligence after frontal lobe lesions. *Brain*, awp269.

Rossi, A. F., Bichot, N. P., Desimone, R., & Ungerleider, L. G. (2007). Top–down attentional deficits in macaques with lesions of lateral prefrontal cortex. *The Journal of Neuroscience, 27*(42), 11306-11314.

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Ruff, C. C., Bestmann, S., Blankenburg, F., Bjoertomt, O., Josephs, O., Weiskopf, N., . . . Driver, J. (2008). Distinct causal influences of parietal versus frontal areas on human visual cortex: evidence from concurrent TMS–fMRI. *Cerebral Cortex, 18*(4), 817-827.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., . . . Driver, J. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal influences on human retinotopic visual cortex. *Current Biology, 16*(15), 1479-1488.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Weiskopf, N., & Driver, J. (2009). Hemispheric differences in frontal and parietal influences on human occipital cortex: Direct confirmation with concurrent TMS–fMRI. *Journal of Cognitive Neuroscience, 21*(6), 1146-1161.

Saenz, M., Buraĉas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision research, 43*(6), 629-637.

Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research, 97*(3), 423-436.

Sakagami, M., & Tsutsui, K.-i. (1999). The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience Research, 34*(2), 79-89. doi:http://dx.doi.org/10.1016/S0168-0102(99)00038-3

Sandrini, M., Rossini, P. M., & Miniussi, C. (2008). Lateralized contribution of prefrontal cortex in controlling task-irrelevant information during verbal and spatial working memory tasks: rTMS evidence. *Neuropsychologia, 46*(7), 2056-2063. doi:10.1016/j.neuropsychologia.2008.02.003

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., . . . Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*(9), 2349-2356.

Seidemann, E., & Newsome, W. T. (1999). Effect of spatial attention on the responses of area MT neurons. *Journal of neurophysiology, 81*(4), 1783-1794.

Selemon, L. D., & Goldman-Rakic, P. S. (1988). Common cortical and subcortical targets of the dorsolateral prefrontal and posterior parietal cortices in the rhesus monkey: evidence for a distributed neural network subserving spatially guided behavior. *The Journal of Neuroscience, 8*(11), 4049-4068.

Serences, J. T., Saproo, S., Scolari, M., Ho, T., & Muftuler, L. T. (2009). Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *NeuroImage, 44*(1), 223-231.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*(2), 727-741.

Shapiro, K., Hillstrom, A. P., & Husain, M. (2002). Control of visuotemporal attention by inferior parietal and superior temporal cortex. *Current Biology, 12*(15), 1320-1325.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*(2), 207-218.

Shmuel, A., Chaimow, D., Raddatz, G., Ugurbil, K., & Yacoub, E. (2010). Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and macroscopic blood vessels in V1 convey information on the stimulated eye. *NeuroImage, 49*(3), 1957-1964.

Soon, C. S., Namburi, P., & Chee, M. W. L. (2013). Preparatory patterns of neural activity predict visual category search speed. *NeuroImage, 66*, 215-222. doi:10.1016/j.neuroimage.2012.10.036

Soutschek, A., Taylor, P. C., Müller, H. J., & Schubert, T. (2013). Dissociable networks control conflict during perception and response selection: a transcranial magnetic stimulation study. *The Journal of Neuroscience, 33*(13), 5647-5654.

Stiers, P., Mennes, M., & Sunaert, S. (2010). Distributed task coding throughout the multiple demand network of the human frontal–insular cortex. *NeuroImage, 52*(1), 252-262. doi:http://dx.doi.org/10.1016/j.neuroimage.2010.03.078

Stoet, G., & Snyder, L. H. (2004). Single neurons in posterior parietal cortex of monkeys encode cognitive set. *Neuron, 42*(6), 1003-1012.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Stuss, D., Floden, D., Alexander, M., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39*(8), 771-786.

Swaminathan, S. K., & Freedman, D. J. (2012). Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. *Nature Neuroscience, 15*(2), 315-320. doi:10.1038/nn.3016

Taylor, P. C., Nobre, A. C., & Rushworth, M. F. (2007). FEF TMS affects visual cortical activity. *Cerebral Cortex, 17*(2), 391-399.

Thompson, R., & Duncan, J. (2009). Attentional modulation of stimulus representation in human fronto-parietal cortex. *Neuroimage, 48*(2), 436-448.

Vendrell, P., Junqué, C., Pujol, J., Jurado, M. A., Molet, J., & Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia, 33*(3), 341-352.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-80.

Waskom, M. L., Kumaran, D., Gordon, A. M., Rissman, J., & Wagner, A. D. (2014). Frontoparietal representations of task context support the flexible control of goal-directed cognition. *The Journal of Neuroscience, 34*(32), 10743-10755.

Watanabe, M. (1986). Prefrontal unit activity during delayed conditional Go/No-go discrimination in the monkey. I. Relation to the stimulus. *Brain research, 382*(1), 1-14. doi:http://dx.doi.org/10.1016/0006-8993(86)90104-6

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex, 4*(5), 470-483.

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. *Journal of Neuroscience, 31*(41), 14592-14599. doi:10.1523/jneurosci.2616-11.2011

Woolgar, A., Jackson, J., & Duncan, J. (in press). Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. *Journal of Cognitive Neuroscience*.

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., . . . Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 107*(33), 14899-14902.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage, 56*(2), 744-752. doi:10.1016/j.neuroimage.2010.04.035

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *NeuroImage, 109*, 429-437. doi:10.1016/j.neuroimage.2014.12.083

Yang, Z., Fang, F., & Weng, X. (2012). Recent developments in multivariate pattern analysis for functional MRI. *Neuroscience bulletin, 28*(4), 399-408.

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology, 13*(2), 187-193.

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*(5), 656-661.

Chapter 2

Feature-selective attention in frontoparietal cortex: Multivoxel codes adjust to prioritize task-relevant information

Jade Jackson[1]*, Anina N. Rich[1], Mark A. Williams[1], Alexandra Woolgar[1]

[1]Perception in Action Research Centre and ARC Centre of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney

*Corresponding author at: Australian Hearing Hub, Level 3, 16 University Avenue, Macquarie University, NSW 2109, Australia.

Abstract

Human cognition is characterised by astounding flexibility, enabling us to select appropriate information according to the objectives of our current task. A circuit of frontal and parietal brain regions, often referred to as the frontoparietal attention network or multiple-demand (MD) regions, are believed to play a fundamental role in this flexibility. There is evidence that these regions dynamically adjust their responses in order to selectively process information that is currently relevant for behavior, as proposed by the "adaptive coding hypothesis" (Duncan, 2001). Could this provide a neural mechanism for feature-selective attention, the process by which we preferentially process one feature of a stimulus over another? We used multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data during a perceptually challenging categorization task to investigate whether the representation of visual object features in the MD regions flexibly adjusts according to task-relevance. Participants were trained to categorize visually similar novel objects along two orthogonal stimulus dimensions (length/orientation) and performed short alternating blocks in which only one of these dimensions was relevant. We found that multi-voxel patterns of activation in the MD regions encoded the task-relevant distinctions more strongly than the task-irrelevant distinctions: The MD regions discriminated between stimuli of different lengths when length was relevant and between the same objects according to their orientation when orientation was relevant. The data suggest a flexible neural system that adjusts its representation of visual objects to preferentially encode stimulus features that are currently relevant for behavior, providing a neural mechanism for feature-selective attention.

Introduction

We live in a complex dynamic environment in which the behavioral relevance of different sensory input changes rapidly. To function successfully, we need a cognitive system that can select what is currently relevant, ignore distraction, and update its responses in accordance with events in the world. Selection of relevant information can be specific to different features of visual objects depending on the current goal. For example, if I am looking for my blue cup among other cups, color is the relevant dimension. When I find my cup and reach to pick it up other features of the cup are now relevant (e.g. orientation). Following Chen and colleagues (2012), we refer to the process of attending to and making a decision about one feature of an object, while ignoring other features of that object, as feature-selective attention. We use this nomenclature to differentiate it from feature-based attention, in which a relevant feature is used to select what object or location to attend to (e.g. attend to the red object). In feature-based attention, attention is directed towards objects and/or locations matching a cued value (e.g. red), while objects of a different color are ignored. In feature-selective attention, attention is instead directed towards a particular *stimulus dimension* (e.g., color)*,* in preference to other dimensions (e.g., shape), in order to make a judgement about the relevant feature of a stimulus (Chen et al., 2012).

The adaptive coding hypothesis (Duncan, 2001) offers a possible neural mechanism for feature-selective attention. It holds that within certain higher cortical regions, the response properties of single neurons are highly adaptable such that in any particular task context, many cells become tuned to code information that is currently relevant. Evidence of such 'adaptive coding' comes primarily from single-unit work with non-human primates in which neurons in higher cortical regions have been shown to alter

62

coding as needed for behavior (Cromer, Roy, & Miller, 2010; Freedman, 2001; Freedman & Assad, 2006; Roy, Riesenhuber, Poggio, & Miller, 2010; Sakagami & Niki, 1994; Stokes et al., 2013). For example, in a go/no-go discrimination task where the relevant feature of a cue changed between three task contexts (Sakagami & Niki, 1994), 72% of prefrontal cortex (PFC) neurons showed different responses during the cue period for each task condition. In another example, Roy et al. (2010) demonstrated that 24% of PFC neurons had a distinct firing rate in response to one category of visual stimuli over another. These neurons responded to the relevant category membership of the stimuli, rather than to their simple visual properties. Moreover, when the decision boundary changed so that these same stimuli were re-categorized into orthogonal categories, these neurons changed their firing rate to reflect the new task (see also Cromer et al. (2010)). Similar results have been found for the lateral intraparietal cortex, where neural firing rates reflect learned category boundaries and change to reflect orthogonal category boundaries on retraining (Freedman & Assad, 2006).

In the human brain, candidate regions for adaptive coding are a set of frontal and parietal brain regions known as multiple-demand (MD) regions (Duncan, 2001; Duncan, 2010). The MD regions incorporate the anterior insula/frontal operculum (AI/FO), the inferior frontal sulcus (IFS), the dorsal anterior cingulate/ pre-supplementary motor area (ACC/pre-SMA), and the intraparietal sulcus (IPS). They are characterized by their response to a wide range of task demands (Dosenbach et al., 2006; Duncan & Owen, 2000; Nyberg et al., 2003), even at the level of single participants (Fedorenko, Duncan, & Kanwisher, 2013). These regions have also been referred to as the "task positive network" (Fox et al., 2005) or "frontoparietal control system" (Vincent, Kahn, Snyder, Raichle, & Buckner, 2008), and have been described as "flexible hubs" that adjust their connectivity patterns along with task demands (Cole et al., 2013).

In addition to being active for a range of tasks, human imaging data suggests that the MD regions are capable of coding a range of task features. Evidence for this comes from MVPA of fMRI data, in which information coding is inferred if patterns of activation across voxels reliably discriminate between task events (e.g. Haxby et al., 2001; Haynes & Rees, 2005; Kamitani & Tong, 2005). Such studies suggest that the MD regions can code several different types of task-relevant information such as rules, stimuli, and motor responses (e.g. Bode & Haynes, 2009; Haynes et al., 2007; Li, Ostwald, Giese, & Kourtzi, 2007; Nee & Brown, 2012; Reverberi, Gorgen, & Haynes, 2011; Woolgar, Hampshire, Thompson, & Duncan, 2011; Woolgar, Thompson, Bor, & Duncan, 2011; Woolgar, Williams, & Rich, 2015). Moreover, in response to changes in task demands, the MD regions adjust their representation of perceptual (Woolgar, Hampshire, et al., 2011; Woolgar, Williams, et al., 2015) and rule information (Woolgar, Afshar, Williams, & Rich, 2015). For example, in Woolgar and colleagues' (2011) study, participants had to identify the spatial position of a visual stimulus. When the positions were close and overlapping, such that they were more difficult to discriminate, they were more strongly represented in MD regions, compared to when they were spaced far apart and perceptually easier, despite a weaker representation of the difficult stimuli in visual cortex. However, in this work the stimuli were always discriminated according to their spatial position, and this stimulus feature was always task relevant. Here, we examine the complementary question of whether flexibility of the MD system also underpins our capacity to attend to different features of the same object, depending on what is currently relevant. Suggestive evidence comes from a recent study in which object representations in lateral PFC (lPFC) could be decoded more strongly within a single task than between tasks (Harel, Kravitz, & Baker, 2014), raising the possibility that the same objects may be represented differently as task contexts change.

In the current study we examined the responses of the MD regions when different features of the same visual objects were made relevant. We presented a set of novel objects that varied along two dimensions (length of one of the spikes and orientation of that same spike; Figure 1). In separate blocks of trials, participants categorized the stimuli on the basis of one of the feature dimensions (length or orientation). Thus, at any one time participants were required to discriminate objects according to one (relevant) dimension and ignore the other (irrelevant) dimension. We used MVPA of fMRI data to test whether the patterns of activation in the MD regions discriminate objects according to the externally imposed decision boundary, and whether this multi-voxel categorization changes when an orthogonal decision boundary is used. If the representations in MD regions are driven by physical stimulus characteristics, the same information should be present irrespective of the task. However, the adaptive coding hypothesis predicts that neural populations in MD regions adjust their responses to adaptively code the currently relevant information, in which case we should see stronger coding for the task-relevant feature distinction than the task-irrelevant distinction. We also examined responses in regions of interest (ROIs) in the lateral occipital complex (LOC) and early visual cortex, for comparison. We found that MD coding of the relevant feature distinction was significantly stronger than discrimination along the equivalent irrelevant dimension. LOC also followed this pattern, albeit more weakly, but we did not find coding of either dimension of these highly similar stimuli in the early visual cortex. Our results suggest that the frontoparietal MD network adjusts its representation of individual objects to make the specific discrimination that is needed for the current task. We suggest that this process supports selective attention to task-relevant object features.

Materials and Methods

Participants

Twenty-six healthy adult volunteers (17 females; mean age = 24.3 years, SD= 5.27) participated. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants gave written informed consent, and the study was approved by the human research ethics committee of Macquarie University, Sydney, Australia. The participants received $50. There were twenty-seven participants initially, but one subject was excluded because he did not complete the task.

Stimuli

Stimuli were abstract novel "spiky" objects created using custom MatLab scripts (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). The stimulus set consisted of 16 objects (Figure 1) in which one spike varied along two dimensions (its length and orientation). Participants learnt to discriminate between the 16 objects across two orthogonal decision boundaries (task contexts) based on the length and orientation dimensions. The relevant visual feature of the stimuli therefore varied depending on the current decision boundary. The stimuli were aligned so that the main stem of the objects appeared at an angle of +37 degrees from vertical. During both the training and scanning sessions, the visual angle (VA) of the spiky object's length along its main axis was 8.07°. Stimulus presentation was controlled by a PC running the Psychophysics Toolbox-3 package (Brainard, 1997) in MatLab (Mathworks). Stimuli were presented in the centre of a screen and viewed through a mirror mounted on the head coil in the scanner.

Category boundary 1(length)

Category boundary 2 (orientation)
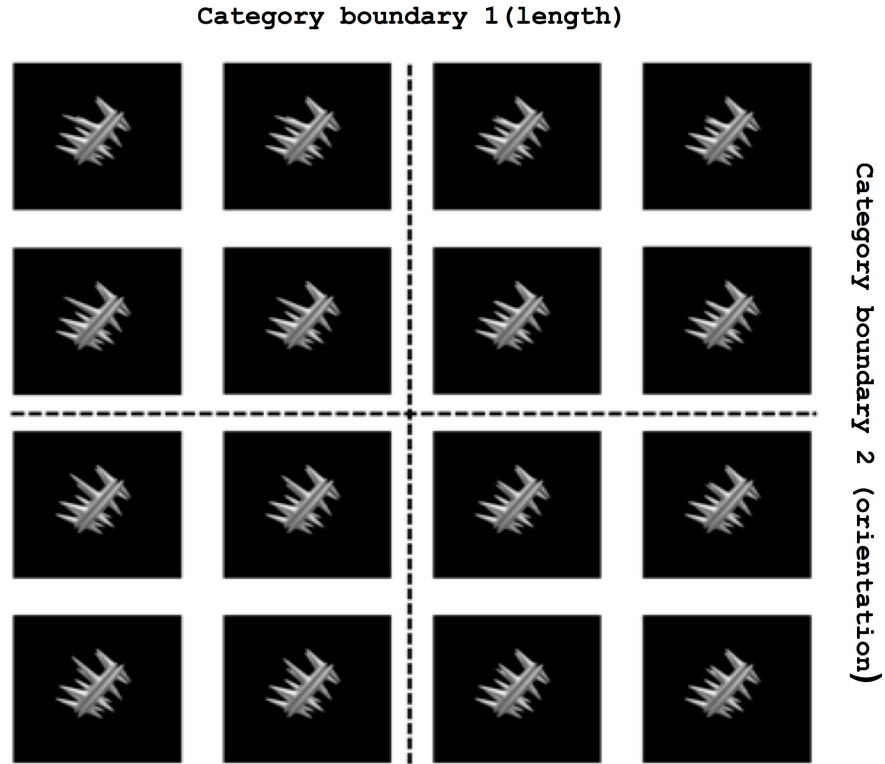
Figure 1: The stimulus set consisted of 16 'spiky' objects with eight objects on either side of two orthogonal category boundaries. Category boundaries were defined based on the length (horizontal axis) or orientation (vertical axis) of the third spike on the left side of the object. Spike length and orientation was titrated on an individual subject basis to equate the difficulty of categorization across the two dimensions.

Titrating Task Difficulty

Less than one week (1-7 days) prior to scanning, participants completed a behavioral testing session in which we titrated the discriminability of the stimuli to ensure that the length and orientation tasks were of comparable difficulty on an individual subject level. We started with a difference of 1.26° VA between the shortest and longest length of the spike, and a maximum difference of 27.02° in the angle of orientation. After 192 trials, the range of lengths or orientations in the stimulus set was adjusted if there was a significant difference in the subject's reaction times between the two tasks (p < 0.05). For

this, we increased the difficulty of the task that had the lower average reaction time. This procedure was repeated until there was no difference in reaction time between the two task contexts, as confirmed with Bayes analysis in each participant separately. In Bayes analysis, a Bayes Factor (BF) > 3 indicates strong evidence for experimental hypothesis and BF < 1/3 indicates strong evidence for the null hypothesis (Dienes, 2011; Love et al., 2015). We required BF < 1, and took this as evidence towards the null and sufficient to deem the conditions equated. The difficulty of the orientation context was increased for 11 participants (maximum angle decreased to 21.92° for 6 participants and 13.86° for 5 participants) and the difficulty of the length context was increased for 5 participants (maximum difference in VA was decreased to 1.06° for 2 participants and 0.95° for 3 participants).

Procedure

Prior to titrating the stimuli, participants completed at least 6 blocks of practice trials to learn the task. Stimuli were initially presented for 400ms until participants achieved >80% correct after which objects were presented for 216ms. Feedback (correct/incorrect) was presented after each response until participants achieved > 80% performance, after which feedback (percent correct) was only given at the end of each block. Once participants reached a high performance level (> 80% correct) in both task contexts, we then titrated the stimuli to ensure equal performance in reaction times (as described above). During titration participants only received feedback at the end of each block. Immediately prior to entering the scanner, participants completed a further 2 practice blocks of each task context to remind them of the task and to avoid initial practice effects in the scanner. These two practice blocks also introduced a response-mapping screen to be used in the scanning task, which randomly assigned the button to be pressed

for each category decision (short or long spikes in the length task and clockwise or anti-clockwise spikes in the orientation task) on a trial-by-trial basis (Figure 2). This allowed separate estimation of the blood-oxygen-level dependent (BOLD) response associated with perceptual information about each category from that associated with each button press. Participants also performed an additional two practice blocks in the scanner during the structural scan, prior to commencing the main experiment, to familiarise them with the button-response box in the scanner.
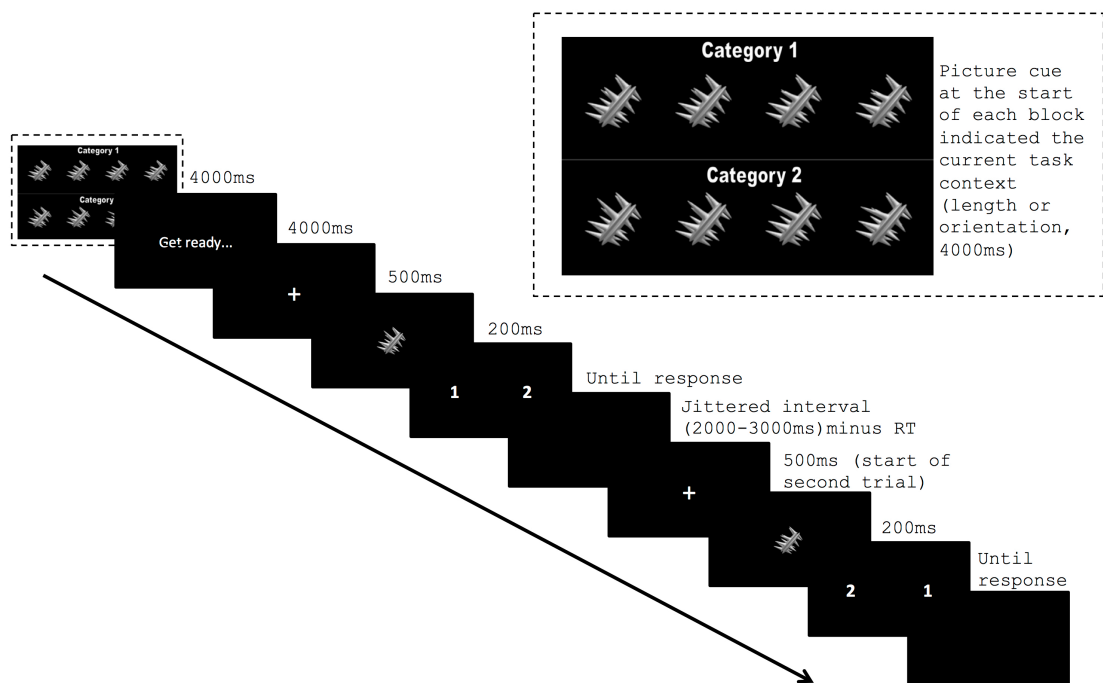
Figure 2: Stimulus categorization task: A picture cue at the start of each block indicated the current task context for categorization (orientation or length; inset shows cue display for the length task). On each trial a fixation cross was presented for 500ms followed by a 'spiky' object for 200ms. Finally a response mapping screen appeared which indicated the appropriate response button. In the example shown, the current context is length. For the first trial, the stimulus is category 1 on the length dimension and therefore the correct response was the left-button.

Participants were scanned whilst performing the categorization task shown in Figure 2. Each participant completed 4 acquisition runs (8.09min each) consisting of 4 blocks (32 trials / block) totaling 128 trials per acquisition run (2.02min / block). At the start of each block, a picture cue (4000ms) indicated the current task (length or orientation) and which attribute was category 1 and 2 (e.g. whether short spikes/long spikes were category 1 or 2; counterbalanced across participants). The order of task contexts was counterbalanced across participants as well as within-participant across runs. The picture cue depicted spiky objects from the extremes of the currently relevant dimension (see Figure 2, inset). The stimulus set was identical across the two contexts, but the relevant feature was either the length of the same spike relative to the category boundary or the orientation of a particular spike relative to the category boundary (rotated clockwise versus anti-clockwise from the boundary) in the different contexts. Thus, participants were attending to the same part of the object, but different features of that object part, in the two conditions.

On each trial, participants saw a white central fixation cross (500ms) after which the spiky object was presented at fixation for 216ms. Finally, participants saw a response mapping screen which indicated the category-to-button response mapping on this trial, and responded regarding the category membership of the stimulus. The response mapping screen randomly assigned category 1 and 2 decisions to either the left or right response button, operated by the index or middle finger of the participant's right hand. The response mapping screen was visible until a button-press was made or until the jittered time interval timed out (2000-3000ms). If a response was made before the end of the inter-trial-interval, a blank black screen was shown for the remainder of the trial time. Feedback (accuracy score) was presented at the end of each block for 6000ms after which there was

a delay of 4000ms prior to the start of the next block. At the end of each run, a blank black screen was shown for 4000ms.

Following completion of the main task during the scanning session we ran a localizer task to functionally identify the LOC as an *a priori* region-of-interest (ROI). Participants viewed central intact and scrambled versions of black and white common objects in 16.8s blocks of 16 trials (1100ms/trial), whilst fixating on a central cross. Participants had to indicate via a button response when the fixation cross changed color to remind them to fixate centrally. There were 21 blocks consisting of alternating blocks of whole objects, scrambled objects, and rest blocks (counterbalanced across participants). The EPI (acquisition time) for the localizer task was 6.25min.

Data Acquisition

The data were collected using a 3T Verio Siemens (Erlangen, Germany) Magnetic Resonance Imaging (MRI) scanner at Macquarie Medical Imaging, Macquarie University Hospital, Sydney, Australia. We used a sequential descending T2*- weighted echo planar imaging (EPI) acquisition sequence with the following parameters: acquisition time 2000 ms; echo time 30 ms; 34 oblique axial slices with a slice thickness of 3.0 mm and a 0.70 mm inter-slice gap; in plane resolution 3.0×3.0 mm; matrix 64×64; field of view 210 mm; flip angle 78°. T1-weighted MPRAGE structural images were also acquired for all participants (slice thickness 1.0 mm, resolution 1.0×1.0 mm).

Preprocessing

MRI data were preprocessed using SPM 5 (Wellcome Department of Imaging Neuroscience, www. fil.ion.ucl.ac.uk/spm) in Matlab 2011b. Functional MRI data were converted from DICOM to NIFTII format, spatially realigned to the first functional scan

and slice timing corrected, and structural images were co-registered to the mean EPI. EPIs from the main experiment were smoothed slightly (4 mm FWHM Gaussian kernel) to improve signal-to-noise ratio. LOC localizer EPIs were also smoothed (8 mm FWHM Gaussian kernel) and in all cases the data were high pass filtered (128 s). Structural scans were additionally normalized, using the segment and normalize routine of SPM5, in order to derive the individual participant normalization parameters needed for ROI definition (below).

ROIs

MD ROIs were defined using co-ordinates from a previous review of activity associated with a diverse set of cognitive demands (Duncan & Owen, 2000) using the kernel method described in Cusack, Mitchell, and Duncan (2010), as in our previous work (Woolgar, Hampshire, et al., 2011; Woolgar, Thompson, et al., 2011; Woolgar, Williams, et al., 2015). The procedure yielded a total of seven ROIs: left and right IFS (centre of mass +/−38 26 24, volume 17 cm^3); left and right AI/FO (+/−35 19 3, 3 cm^3); left and right IPS (+/−35 −58 41, 7cm^3) and ACC/ pre-SMA (0 23 39, 21 cm^3).

Left and right visual cortex ROIs were derived from the Brodmann template provided with MRIcro (Rorden & Brett, 2000) (BA 17, centre of mass −13 −81 3, 16 −79 3, volume 54 cm$_3$). All co-ordinates are given in MNI152 space (McConnell Brain Imaging Centre, Montreal Neurological Institute). MD and BA17 ROIs were deformed for each participant by applying the inverse of the participant's normalization parameters. This allowed analyses to be carried out using native space (i.e., non-normalised) EPI data.

Using the functional localiser scan data, we defined LOC for each participant as the brain area in the lateral occipital lobe that responded more strongly to whole objects than to scrambled versions of the same objects. We used the standard multiple regression

approach of SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk) to estimate values pertaining to the whole and scrambled object conditions (block design). Blocks were modelled using a box car function lasting 16s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. Whole-brain analyses (paired t-tests) compared voxelwise BOLD response in the two conditions (whole objects minus scrambled objects). The resulting map was thresholded such that there was at least one cluster with a minimum size of 20 voxels. These clusters were then imported into MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) and those active voxel clusters close to anatomical LOC coordinates from previous studies (Grill-Spector et al., 1999; Grill-Spector, Kushnir, Hendler, & Malach, 2000) were selected as the ROIs.

First-Level Model

To obtain estimated activation patterns for multivariate analysis, a General Linear Model (GLM) was estimated for each participant using the realigned, slice-time corrected and smoothed native space EPI images using SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk). We classified each stimulus as either "short" or "long" on the length dimension, and "rotated clockwise" or "rotated anti-clockwise" relative to the category boundary on the orientation dimension. Trials were modelled as events of zero duration convolved with the hemodynamic response of SPM5. Each trial contributed to the estimation of two beta values, the relevant feature (short or long for *length* context blocks, and clockwise or anti-clockwise for *orientation* context blocks) and irrelevant feature (short or long for *orientation* context blocks, and clockwise or anti-clockwise for *length* context blocks). We derived the estimates for each feature in

each block separately. The two run means were included in the model as covariates of no interest. Error trials were excluded from the analysis.

MVPA

We used MVPA to examine the representation of relevant and irrelevant stimulus features. Of central interest was whether the MD regions adapted to code length and orientation information more strongly when it was relevant for the task than when it was task-irrelevant. We also examined the same stimulus feature distinctions in the LOC, and early visual cortex (Brodmann area 17, BA 17).

We implemented MVPA using the Decoding Toolbox (Hebart, Görgen, & Haynes, 2015), which wraps the LIBSVM library (Chang & Lin, 2011). We examined coding of orientation when orientation was relevant, orientation when orientation was irrelevant, length when length was relevant, length when length was irrelevant. For each participant and ROI, a linear support vector machine was trained to decode the relevant (clockwise/anti-clockwise in orientation blocks, and short/long in length blocks) and irrelevant (clockwise/anti-clockwise in length blocks, and short/long in orientation blocks) stimulus features separately, for each task context separately. In total, there were 16 blocks for each participant: 8 with length relevant, and 8 with orientation relevant. For each classification, we used a leave-one-out 8 fold splitter whereby the classifier was trained using the data from 7 out of the 8 blocks and subsequently tested on its accuracy at classifying the unseen data from the remaining block, iterating over all possible combinations of training and testing blocks. For example, to yield a classification accuracy score for the task-relevant length distinctions (i.e., short vs. long in length blocks), we took the 8 blocks in which participants performed the length task and trained the classifier to distinguish between patterns of activation representing short and long

spikes using data from 7 out of these 8 blocks, and then tested generalization to the remaining unseen block. We repeated this iteratively for all 8 combinations of training and testing data, and then averaged the accuracies from each iteration together to give a mean accuracy score for task-relevant length coding. This was repeated for each condition, participant and ROI separately.

The mean classification accuracy for each participant in each ROI and in each condition was then entered into a second level analysis. Of central interest was whether the MD network would code task-relevant stimulus features more strongly than task-irrelevant ones. To address this, we conducted a three factor ANOVA on classifier accuracy with the factors *relevancy*, *feature*, and *MD region*. To explore any hemispheric effects, we ran an additional ANOVA with factors *relevancy, feature, MD region* and *hemisphere.*

Since a difference in coding between relevant and irrelevant conditions is only interpretable if coding in at least one condition is also significantly above chance, we also conducted one-sample t-tests against the classification accuracy expected by chance (50%) in each condition (relevant and irrelevant) separately. One-tailed significance tests were used where appropriate for inference: tests comparing classification of relevant to irrelevant feature distinctions in the MD regions are one-tailed since the direction of the effect is pre-specified, and tests comparing classification accuracy to chance are one-tailed as below chance classifications are not interpretable. All other tests are two-tailed. Alpha was adjusted for four comparisons using Bonferroni correction (0.05 divided by 4).

We also examined whether coding in the visual cortices was stronger for task-relevant than irrelevant stimulus features. For this, we used a two-factor ANOVA on classifier accuracy with factors *relevancy* (task-relevant, task-irrelevant) and *feature*

(orientation, length) for each of the visual cortex ROIs (LOC, early visual cortex) collapsed across hemisphere. Again, we also tested whether coding in the visual cortices was above chance in each condition separately (one-sample t-test against the classification accuracy expected by chance, 50%).

Finally, we ran an additional analysis in which the classifier was trained on data representing the category number decisions in one task context (category 1/category 2) and tested on the category number decisions participants made in the other task context (category 1/category 2). We included this analysis to explore whether the categorization decision was represented at the level of the stimulus (i.e. short/long) or at the level of the category number (category 1/category 2).

Results

Behavioral Results

Prior to scanning, participants practiced the task until they scored at least 80% correct in both task conditions. Task difficulty was then titrated to match reaction times between the two conditions for each participant separately (assessed with Bayes factor analysis for each participant separately, all $BF_{10} < 0.89$).

In the scanning session, participants performed with a high degree of accuracy (mean = 89.6%, SD = 9.1). Accuracy scores were assessed using Bayes Factor analysis to check for differences in performance between the two task contexts. There were no differences in accuracy score between the two conditions for any participant individually (all $BF_{10} < 0.64$). Reaction time data from the scanning session are not meaningful as the response mapping screen defined the response after the stimulus display.

Decoding Task-Relevant and Task-Irrelevant Stimulus Features

MD Regions

MVPA was used to differentiate multivoxel patterns pertaining to stimulus feature distinctions (orientation: clockwise/anti-clockwise, and length: short/long) when they were relevant to the task (orientation in orientation task blocks, and length in length task blocks) and when they were irrelevant to the task (orientation in length task blocks, and length in orientation task blocks). The resulting classification accuracy signified the strength of coding.

The adaptive coding hypothesis (Duncan, 2001) proposes that neurons dynamically adjust their responses in order to selectively code information that is currently relevant for our behavior. We asked whether this could provide a basis for feature-selective attention, with preferential coding of currently relevant visual features in the MD system. Our prediction was that the MD regions would change their representation of the visual stimuli between tasks, to reflect or emphasize the distinctions that were needed in each task context.

The results are presented in Figure 3. In line with the hypothesis, a three-way ANOVA with factors *relevancy* (task-relevant and task-irrelevant stimulus features), *region* (AI/FO, IFS, ACC/pre-SMA, and the IPS; collapsed across hemisphere where appropriate), and *feature* (orientation and length) revealed a main effect of *relevancy* ($F(1,25) = 1.13$, $p = 0.03$). The ANOVA showed no main effect of *feature* ($F(1,25)=1.01$, $p=0.32$), no main effect of *region* ($F(3,75) = 0.77$, $p = 0.51$), no significant interaction between *relevancy* and MD *region* ($F(3,75) = 2.05$, $p = 0.12$), *region* and *feature* ($F(3,75)=1.75$,$p=0.16$), or *relevancy* and *feature* ($F(3,75) = 0.21$, $p = 0.65$), and no

significant 3-way interaction ($F(3,75) = 1.41$, $p = 0.25$). These results indicate that the MD regions coded the task-relevant feature distinctions more strongly than the task-irrelevant distinctions, despite these features being actually physically identical. In our additional ANOVA, included to check for hemispheric differences, there was no main effect of *hemisphere* ($F(3,24) = 2.52$, $p = 0.13$), but there was a significant interaction between *relevancy* and *hemisphere* ($F(1,24) = 5.78$, $p = 0.02$), reflecting a stronger relevancy effect on the left.

The coding of relevant over irrelevant features is only interpretable if coding in one or more of the relevancy conditions is also significantly above chance. Therefore, we conducted one-sample t-tests against the classification accuracy expected by chance (50%) in each relevancy condition separately. We found that the MD regions coded the task-relevant stimulus features significantly (mean classification accuracy for relevant across all regions, 55.14%; one sample t-test against chance, $t(25) = 4.75$, $p < 0.001$), whereas classification of the task-irrelevant stimulus distinctions was not significantly different from chance (mean classification accuracy for irrelevant, 52.21%; $t(25) = 1.65$, $p = 0.11$). Thus, on average across the network, these regions only encoded the task-relevant stimulus distinctions. Considering each MD ROI separately, the relevant stimulus distinctions were coded in all 4 MD ROIs (ACC/pre-SMA, mean accuracy 55.64%; $t(25) = 3.01$, $p = 0.004$; IPS, mean accuracy 54.80%; $t(25) = 2.81$, $p = 0.01$; IFS, mean accuracy 57.27%; $t(25) = 5.47$, $p < 0.001$; AI/FO, mean accuracy 52.94%; $t(25) = 2.85$, $p = 0.01$; Figure 3, dark bars) while the irrelevant stimulus distinctions were only coded in the AI/FO (mean classification accuracy 53.31%; $t(25) = 2.66$, $p = 0.01$; other ps > 0.05; Figure 3, light bars).

Recall that the stimulus set was identical across conditions; each feature (length, orientation) was relevant in one condition and irrelevant in the other. The data suggest that

the MD system encoded the relevant feature in each case, that is, it adjusted such that it coded the same physical stimulus distinction (e.g. length) more strongly when it was relevant than when it was irrelevant.
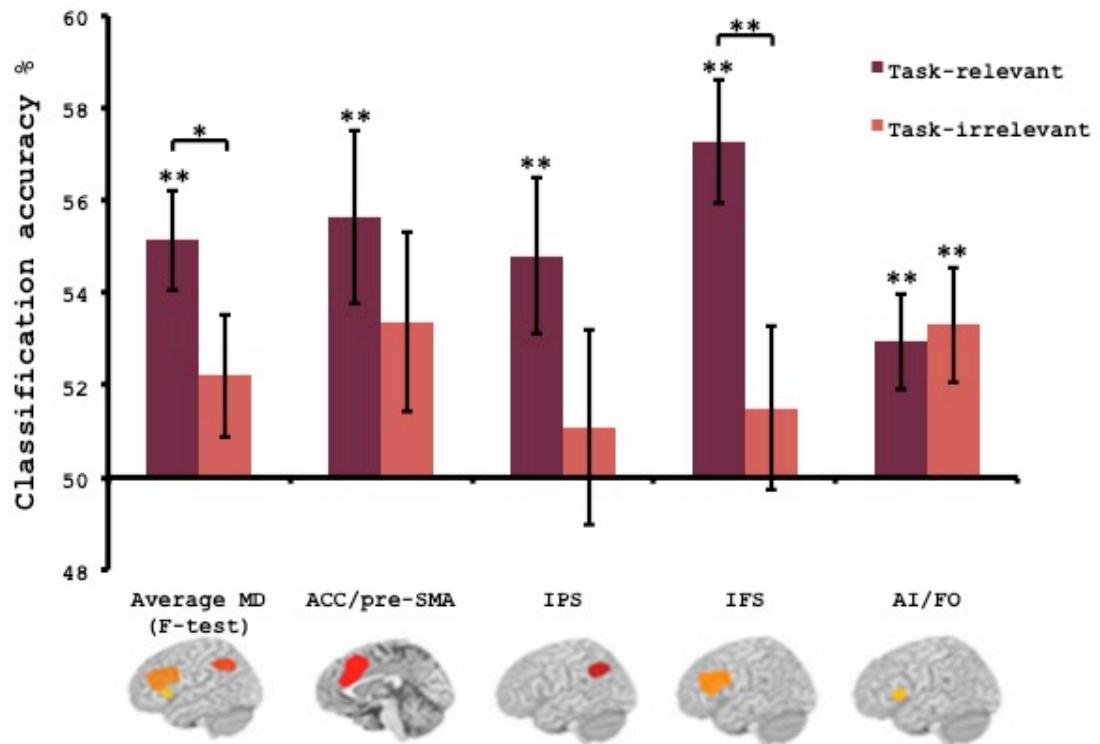
Figure 3: Coding of task-relevant and task-irrelevant stimulus distinctions in MD regions. Error bars indicate standard error. Significance marking for individual bars indicate whether coding was significantly greater than chance in each condition separately (one-sample t test against chance, 50%), significance marking between bars indicate where coding was significantly greater for relevant compared to irrelevant distinctions (main effect of relevancy / paired t-test). *$p < 0.05$, ** $p < 0.01$, alpha for analyses of individual regions corrected for four comparisons using Bonferroni correction. The MD regions coded task-relevant stimulus distinctions more strongly than the physically identical task-irrelevant distinctions.

Lateral Occipital Complex (LOC)

The LOC is known to respond strongly to object features (Grill-Spector, Kourtzi, & Kanwisher, 2001) and has previously been found to show preferential representation of attended relative to distractor objects (Woolgar, Williams, et al., 2015). It therefore seemed a likely candidate for preferential coding of relevant object features in our task. Figure 4 presents the data from the LOC, and visual inspection suggests a trend in this direction. However, an ANOVA with factors *relevancy* (relevant, irrelevant) and *feature* (orientation, length) showed no significant main effect of *relevancy* ($F(1,25) = 2.83$, p = 0.11), no main effect of *feature* ($F(1,25) = 0.32$, p = 0.57), and no *relevancy*feature* interaction ($F(1,25) = 0.08$, p = 0.78). When we compared coding to chance, the LOC carried significant information about task-relevant distinctions (mean classification accuracy 56.67%; $t(25)=3.65$, p < 0.001), but not about irrelevant distinctions (mean classification accuracy 52.63%; $t(25) = 1.39$, p = 0.15). Thus, although the LOC only discriminated objects along the stimulus dimension that was currently relevant, the interaction with relevancy did not reach significance.

Early Visual Cortex

We also tested whether information pertaining to task-relevant and task-irrelevant stimulus distinctions was coded in early visual cortex (BA 17). An ANOVA with factors *relevancy* (relevant, irrelevant) and *feature* (orientation, length) showed no main effect of *relevancy* ($F(1,25) = 0.57$, p = 0.81), no main effect of *feature* ($F(1,25) = 2.1$, p = 0.17), and no *relevancy*feature* interaction ($F(1,25) = 2.06$, p = 0.16). Thus, we found no evidence that context modulates coding of orientation and length in this region (Figure 4). Further, BA 17 did not show categorical discrimination of these visually similar objects

according to either the task-relevant (relative to chance; mean classification accuracy 52.9%; t(25) = 1.62, p = 0.14), or task-irrelevant (relative to chance; mean classification accuracy 52.3%; t(25) = 1.17, p= 0.25) stimulus features. Thus, in contrast to the MD and LOC ROIs, BA 17 did not distinguish these objects according to the task-imposed decision boundaries.
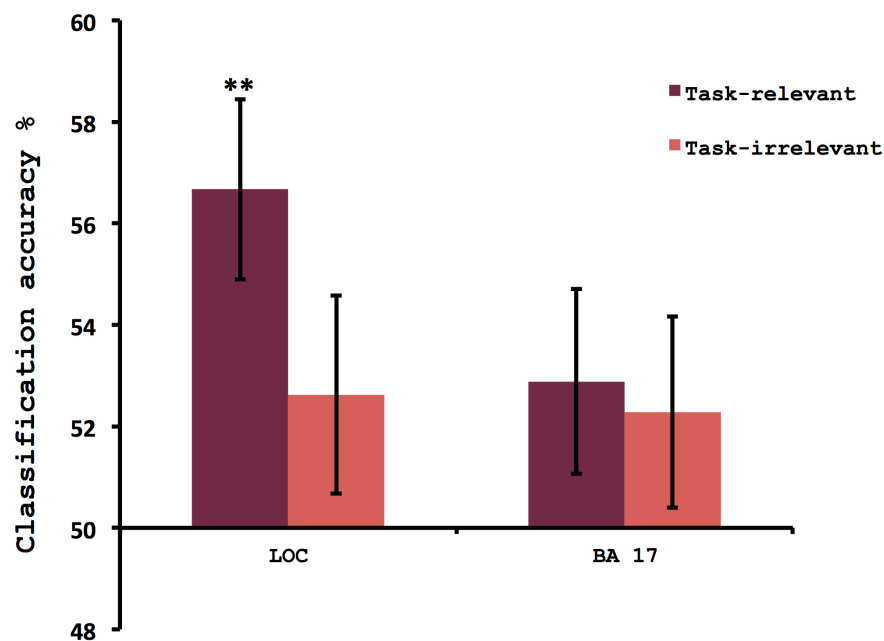
Figure 4: Coding of task-relevant and task-irrelevant stimulus distinctions in LOC and BA 17. Error bars indicate standard error. The significance markings for individual bars indicate whether coding was significantly greater that chance in each condition separately (one-sample *t* test against chance, 50%). *******p* < 0.01.

Coding of Category Placement

On each trial, participants had to categorize the object according to the relevant feature dimension (e.g. short/long on length blocks), associate that decision with the

number representing the chosen category (1 or 2, e.g. short = 1 in half the participants), and then use the response mapping screen to transform their choice into the appropriate button press response (left or right, 1 = left on half of the trials). Therefore, it is possible that as well as the categorization decision at the level of the stimulus, participants also held a category *number* in mind on each trial. To test for representation at this higher level of abstraction, we ran an additional analysis where we trained the classifier on the data representing the category number decisions in one task context and tested this on the category number decisions participants made in the other task context. We were not able to decode the category number placement of the objects in the MD system (mean classification accuracy 50.5%, $t(25) = 1.31$, $p = 0.21$). In order to interpret this null effect we calculated the Bayes Factor. Coding of category number placement revealed a Bayes Factor (BF_{10}) of 0.44. As this is less than 1 (Dienes, 2011) and approaches the level of 0.33 suggested by (Jeffreys, 1998) to represent significant evidence for the null hypothesis, we interpret this as evidence that, although the MD regions encode task-relevant stimulus distinctions, the representation is not abstracted to the level of categorical number placement.

Discussion

The adaptive coding hypothesis (Duncan, 2001) proposes that neural populations dynamically adjust their responses to selectively code information that is currently relevant for our behavior. This provides a possible mechanism for feature-selective attention, which allows information about task-relevant stimulus features to be processed in preference to irrelevant attributes. We examined the responses of the MD regions in a difficult visual object categorization task in which the relevant stimulus dimension varied on physically identical stimuli. The MD system adjusted its representation of these novel

objects to preferentially encode feature distinctions that were relevant for the task. When the task required participants to categorize the objects based on length, the MD regions coded length information and not orientation information but when the task was to categorize based on orientation, orientation was encoded in preference to length. Thus, the MD system adjusted its representation of the features of an object to encode the discrimination necessary for the current task. Consistent with the proposal that the cognitive flexibility of these regions underlies their involvement in a wide range of tasks (e.g. Cole & Schneider, 2007; Duncan, 2010; Duncan & Owen, 2000), our data suggest the coding of this adaptive system adjusts to hold the currently relevant features of a stimulus as needed for behavior.

Electrophysiological studies in non-human primates have previously shown that neurons in higher cortical regions adapt their tuning profiles to respond most strongly to the information that is currently relevant (Cromer et al., 2010; Freedman, 2001; Freedman & Assad, 2006; Roy et al., 2010; Sakagami & Niki, 1994; Stokes et al., 2013). The implementation of MVPA for fMRI has shown similar results in humans: patterns of activation in the MD regions code a range of different types of task-related information (e.g. Bode & Haynes, 2009; Haynes et al., 2007; Li et al., 2007; Nee & Brown, 2012; Reverberi et al., 2011; Woolgar, Hampshire, et al., 2011; Woolgar, Thompson, et al., 2011; Woolgar, Williams, et al., 2015), and adjust their responses when task demands vary (e.g. Li et al., 2007; Woolgar, Afshar, et al., 2015; Woolgar, Hampshire, et al., 2011; Woolgar, Williams, et al., 2015). The MD regions also encode attended objects in preference to unattended objects (Woolgar, Williams, et al., 2015) and a previous adaptation study demonstrated that these regions show greater responses to changes in attended stimulus features (color/shape) than to changes in unattended stimulus features (Thompson & Duncan, 2009). Here, we find that these regions can also flexibly adapt

their representations of single objects to emphasize task-relevant stimulus distinctions, resulting in preferential coding of attended stimulus features.

Our data align with a recent study in which objects were strongly represented in lPFC in individual task contexts but the representation did not generalize between task contexts (Harel et al., 2014). Those data suggested that the same set of objects may be represented differently as task contexts change. Here, we tested this possibility directly by specifying the specific stimulus distinctions that an adaptive system should make in each task context. We found that the MD system adjusted its representation of the set of novel objects to make this specific distinctions needed for the task. In a related study (Peelen & Caramazza, 2012), participants responded to one of two semantic dimensions of an object (how the object is used or where the object is found) in a one-back task. Results from a whole-brain searchlight revealed several regions, including the right lPFC, which showed coding of the two semantic dimensions. However, no region showed preferential coding of task-relevant over task-irrelevant dimensions. The MD regions are known to be recruited most strongly when tasks are challenging (e.g. Duncan & Owen, 2000). It may be, then, that preferential coding of task-relevant information is only observed when the task is sufficiently difficult (Woolgar, Afshar, et al., 2015), as in the current experiment.

The observation that stimulus features were coded more strongly when they were relevant than when they were irrelevant suggests that some filtering of information occurs between input (the relevant and irrelevant features were physically identical in our case) and MD representation as recorded with fMRI. In our data, we could not detect MD coding of irrelevant information in 4 of the 5 MD regions. However, recent evidence from non-human primates suggests that task irrelevant information can affect firing rates in higher cortical regions such as the frontal eye fields (Mante, Sussillo, Shenoy, & Newsome, 2013). Differences in the tasks, sensitivity of the methods and recording site,

may account for the different result. In that study, consistent with our findings, the effect of irrelevant information was weaker than that of relevant information but interestingly the size of the difference there was too small to account for the behavioral effect (Mante et al., 2013). Other work suggests a dynamic change in the responses of lateral prefrontal neurons over time (Kadohisa et al., 2013). When presented with a target and distractor object in the left and right visual field, prefrontal activity was initially dominated by the contralateral object, regardless of its relevance, but over the course of the trial prefrontal resources were quickly re-assigned such that representation of the target came to dominate in both hemispheres (Kadohisa et al., 2013).

In our data, the LOC held information about the task-relevant feature distinctions, demonstrating that it is sensitive to minimal changes in the shape of an object (e.g., to small changes in the length of one spike) when that change is relevant for behavior. The LOC did not make the task irrelevant distinctions, but the difference in coding between relevant and irrelevant conditions did not reach significance. The trend for relevant coding to be greater than irrelevant coding is in line with previous work which has emphasized a role for the LOC in responding more strongly to attended compared to unattended objects (e.g. Konen & Kastner, 2008; Murray & He, 2006; Murray & Wojciulik, 2003; Woolgar, Williams, et al., 2015; Xu & Chun, 2005). It seems likely that the magnitude of feature-selective attention effects, as in our study, would be considerably smaller than effects of attention allocated on a whole-object level. Although we could not look at it in this study, because we did not track eye movements, it is also possible similar patterns might occur in other brain regions such as the frontal eye fields, based on previous findings that they are reliably activated by attentional tasks in humans (e.g. Corbetta et al., 1998; Culham, Cavanagh, & Kanwisher, 2001; Ester, Sprague, & Serences, 2015; Jerde, Merriam, Riggall, Hedges, & Curtis, 2012) and in non-human primates (Mante et al., 2013).

We did not observe multivoxel coding of object-category information in early visual cortex (BA17), or any interaction with relevancy. Although previous investigations using MVPA with fMRI have reported preferential coding of task-relevant stimuli in this region (Jehee, Brady, & Tong, 2011; Kok, Jehee, & de Lange, 2012; Woolgar, Williams, et al., 2015), there are some marked differences between these previous studies and the current study that may account for the different results. In previous work, coding corresponded to discrimination between physically dissimilar stimuli (gratings of 55° and 145° orientation in Jehee et al., (2011), 40° and 135° in Kok et al., (2012), and different objects in Woolgar et al., (2015)). In our task, coding corresponded to discriminations across an arbitrary boundary on identical sets of stimuli. Objects close to the decision boundary were physically very similar, meaning stimulus-driven activation patterns would also be very similar and therefore difficult to classify. Moreover, the objects on either side of the category boundary were collapsed in our analyses, meaning that the classifier was required to generalize over physical differences of a similar magnitude to those it needed to discriminate between. This makes our result in the MD regions all the more striking, since the information they encoded was based on such minimal visual differences.

Successful behavior requires a flexible cognitive system. Here, we have demonstrated that the MD network adjusts its representation of visual objects to make the distinctions that are needed for the current task. In this way, visually minimal task-relevant feature distinctions are coded more strongly than the equivalent irrelevant distinctions. This study exemplifies the extent to which the MD network can flexibly emphasize different features of an object, providing a possible neural mechanism for the implementation of feature-selective attention.

Funding

Acknowledgments

References

Bode, S., & Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *NeuroImage, 45*(2), 606-613. doi:10.1016/j.neuroimage.2008.11.031

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage, 16*(2).

Chang, C. C., & Lin, C. J. (2011). LIBSVM a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 27.

Chen, X., Hoffmann, K. P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology, 107*, 1530-1543.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience, 16*(9), 1348-1355. doi:10.1038/nn.3470

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., . . . Van Essen, D. C. (1998). A common network of functional areas for attention and eye movements. *Neuron, 21*(4), 761-773.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

Culham, J. C., Cavanagh, P., & Kanwisher, N. G. (2001). Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. *Neuron, 32*(4), 737-745.

Cusack, R., Mitchell, D. J., & Duncan, J. (2010). Discrete object representation, attention switching, and task difficulty in the parietal lobe. *Journal of Cognitive Neuroscience, 22*, 32-47.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron, 50*(5), 799-812. doi:10.1016/j.neuron.2006.04.031

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobes recruited by diverse cognitive demands. *Trends in Neuroscience, 23*, 475-483.

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron, 87*(4), 893-905. doi:10.1016/j.neuron.2015.07.013

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Freedman, D. J. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science, 291*(5502), 312-316. doi:10.1126/science.291.5502.312

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature, 443*(7107), 85-88. doi:10.1038/nature05078

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. G. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 21*, 1409-1422.

Grill-Spector, K., Kushner, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*, 837-843.

Harel, A., Kravitz, D., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the USA*, 962-971.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*, 2425-2429.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology, 17*(4), 323-328. doi:10.1016/j.cub.2006.11.072

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

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8*. doi:10.3389/fninf.2014.00088

Jeffreys, H. (1998). *The theory of probability*: OUP Oxford.

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience, 31*(22), 8210-8219. doi:10.1523/jneurosci.6153-09.2011

Jerde, T. A., Merriam, E. P., Riggall, A. C., Hedges, J. H., & Curtis, C. E. (2012). Prioritized maps of space in human frontoparietal cortex. *J Neurosci, 32*(48), 17382-17390. doi:10.1523/JNEUROSCI.3810-12.2012

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., . . . Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80*(1), 235-246.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679-685.

Kok, P., Jehee, Janneke F. M., & de Lange, Floris P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron, 75*(2), 265-270. doi:10.1016/j.neuron.2012.04.034

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11*(2), 224-231. doi:10.1038/nn2036

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible Coding for Categorical Decisions in the Human Brain. *Journal of Neuroscience, 27*(45), 12321-12330. doi:10.1523/jneurosci.3795-07.2007

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E. J. (2015). JASP (Version 0.7) [Computer Software].

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature, 503*(7474), 78-84. doi:10.1038/nature12742

Murray, S. O., & He, S. (2006). Contrast Invariance in the Human Lateral Occipital Complex Depends on Attention. *Current Biology, 16*(6), 606-611. doi:10.1016/j.cub.2006.02.019

Murray, S. O., & Wojciulik, E. (2003). Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience, 7*(1), 70-74. doi:10.1038/nn1161

Nee, D. E., & Brown, J. W. (2012). Rostral–caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *NeuroImage, 63*(3), 1285-1294. doi:10.1016/j.neuroimage.2012.08.034

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petterson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*, 371-377.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate

Cortex. *Journal of Neuroscience, 26*(50), 13025-13036. doi:10.1523/jneurosci.2481-06.2006

Peelen, M. V., & Caramazza, A. (2012). Conceptual Object Representations in Human Anterior Temporal Cortex. *Journal of Neuroscience, 32*(45), 15728-15736. doi:10.1523/jneurosci.1953-12.2012

Reverberi, C., Gorgen, K., & Haynes, J. D. (2011). Compositionality of Rule Representations in Human Prefrontal Cortex. *Cerebral Cortex, 22*(6), 1237-1246. doi:10.1093/cercor/bhr200

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology, 12*, 191-200.

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research, 97*(3), 423-436.

Stokes, Mark G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron, 78*(2), 364-375. doi:10.1016/j.neuron.2013.01.039

Thompson, R., & Duncan, J. (2009). Attentional modulation of stimulus representation in human fronto-parietal cortex. *NeuroImage, 48*(2), 436-448. doi:10.1016/j.neuroimage.2009.06.066

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. *Journal of Neuroscience, 31*(41), 14592-14599. doi:10.1523/jneurosci.2616-11.2011

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage, 56*(2), 744-752. doi:10.1016/j.neuroimage.2010.04.035

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *NeuroImage, 109*, 429-437. doi:10.1016/j.neuroimage.2014.12.083

Xu, Y., & Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*(7080), 91-95. doi:10.1038/nature04262

Chapter 3

Overlapping neural codes: Frontoparietal voxels are re-used to code relevant feature information across different tasks

Jade Jackson[1], Alexandra Woolgar[1]

[1]Perception in Action Research Centre and ARC Centre of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney

Abstract

Frontoparietal cortex is thought to flexibly allocate attentional resources depending on our current goal and provide signals that configure our moment-to-moment information processing (e.g. Desimone & Duncan, 1995; Duncan, 2010). A fundamental question is how these signals are coded in neural activity and in what ways these processing mechanisms enable flexibility for the requirements of different tasks. Single-unit studies (Cromer, Roy, & Miller, 2010; Freedman, 2001; Roy, Riesenhuber, Poggio, & Miller, 2010) have shown that prefrontal neurons flexibly adjust their responses to code relevant information across multiple tasks. Additionally, these studies indicate that the extent to which these neurons are re-used across tasks ("multitask") may depend on how similar the two tasks are, with a greater proportion of neurons involved in representing more than one stimulus when the stimuli are more distinct (Cromer et al., 2010). In the human brain, the "multiple demand" (MD) system is proposed to exert control by adjusting its responses to selectively process information that is currently relevant (Duncan, 2001). Here, we used multi-voxel pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) data to examine the coding of relevant and irrelevant stimulus features in the MD network, and quantified the extent to which single voxels contributed to multiple neural codes. Participants categorised two separate groups of novel objects ("spikies" and "smoothies") based on the features of each set of objects. We found that multi-voxel patterns of activation in the MD regions encoded the task-relevant feature distinctions more strongly than the task-irrelevant distinctions, as in our previous work (Chapter 2 – Jackson et al., in press). A comparison of the two datasets revealed that irrelevant feature information was coded more strongly when it was sometimes relevant to the task (Chapter 2 – Jackson et al., in press) compared to when it was never relevant (current study). Next, we asked

whether the neural codes for each stimulus feature depended on the same or different voxels. In the MD system, we found that voxels were more likely to be re-used in multiple neural codes than we predicted based on a permutation test (i.e., by chance). This was not the case for the visual system, where the stimulus features were also encoded, but voxel-re-use was at chance. The same analysis, applied to our previous dataset (Chapter 2 – Jackson et al., in press), revealed the same pattern of results. Despite differences in the stimuli and our prediction based on the non-human primate literature, there were no significant differences in voxel re-use between the two experiments. Our data emphasise the flexibility of the MD regions, with single voxels re-used to multitask coding of relevant feature information across different tasks.

Introduction

To function successfully, we need a cognitive system that can select what is currently relevant, ignore distraction, and when we encounter increasing task difficulty, employ mechanisms enabling us to deal with an increase in cognitive demand. The ability to select between task-relevant and task-irrelevant information is a basic aspect of attentional function. This system needs to constantly update the way it responds in order to meet the requirements of our current environment. We do not fully understand how the human brain is able to swiftly adjust processing priorities in response to changing circumstances.

One influential model, the adaptive coding hypothesis, proposes that context-specific task parameters directly shape the tuning profile of higher cortical neurons (Duncan, 2001; Duncan, 2010). These neurons are not tuned to specific features in the environment but instead their response properties are thought to be highly adaptable, coding the details of stimuli, tasks and responses according to what information is

currently relevant for behaviour. This model suggests that the way that stimuli are encoded will change depending on task parameters.

Evidence for such 'adaptive coding' came originally from single-unit work with non-human primates where neural responses in higher cortical regions were found to be relatively independent of the physical attributes of stimuli, and depend instead on their behavioural significance. For example, several experiments have reported that neurons in prefrontal cortex (PFC) encode the behavioural meaning of visual stimuli, regardless of their physical properties (Cromer et al., 2010; Freedman & Assad, 2006; Freedman, Riesenhuber, Poggio, & Miller, 2001; Freedman, Riesenhuber, Poggio, & Miller, 2002; Roy et al., 2010; Sakagami & Niki, 1994; Sakagami & Tsutsui, 1999; Watanabe, 1986). This contrasts with findings showing that the responses of neurons in other association cortices, such as inferior temporal cortex, are less strongly driven by categorical membership and more responsive to visual features (Freedman & Miller, 2008; Freedman, Riesenhuber, Poggio, & Miller, 2003).

Adaptive coding is also evident in the changing response of PFC neurons during single trials (Kadohisa et al., 2013). In this study, monkeys performed a cue detection task where they were presented with one of two cues that were associated with one of two alternative targets. After a delay one or two objects were presented that could consist of a behaviourally relevant target (associated with the current cue), currently irrelevant target (associated with the alternate cue), and/or a consistently irrelevant target (never relevant to either task). The results showed that when two objects were present in the display, attentional competition was resolved by a reallocation of prefrontal neural resources over time. During early processing, different neurons responded to different items and responses were mostly dominated by the object in the contralateral hemisphere. Later, responses reflected the behaviourally relevant objects globally across hemispheres,

suggesting an adaptive re-allocation of prefrontal resources over the course of the trial. The extent of this reallocation was greater when the accompanying stimulus was never relevant to the task compared to when it currently irrelevant (but relevant for the alternative cue), suggesting that consistently irrelevant information is more easily filtered out.

Adaptive coding can also be observed in Cromer and colleagues' (2010) study where non-human primates were trained to classify stimuli according to an arbitrarily defined category boundary. In this study, individual PFC neurons displayed tuning profiles that were aligned with the task-relevant decision space. When the task changed so that the monkeys were required to classify a second group of stimuli according to a new decision boundary, 44% of these neurons changed their firing rate to reflect the new task. These data provide evidence that substantial portions of PFC neurons are able to engage in multiple cognitive tasks, emphasising the flexibility of these neurons to alter their coding as needed for behaviour. Roy et al., (2010) showed similar results in an experiment where monkeys switched between two tasks requiring them to re-categorise the same stimuli across an orthogonal decision boundary. However, in this study only 24% of the task-responsive neurons altered their firing rate to reflect the decision boundary of the new task. It appears that in some cases PFC neurons are recruited to represent different information depending on the task (termed "multitasking", Cromer et al., 2010) but in other situations, information is primarily encoded in different PFC neurons for different task contexts (Roy et al., 2010). These data reflect the striking properties of these neurons not only to adjust what they code in different circumstances, but to adjust the extent to which they act as "generalists" (coding multiple types of information) or "specialists" (coding only one type of information) according to different circumstances (Cromer et al., 2010; Roy et al., 2010). It emphasises the importance of investigating the variety of

control mechanisms, such as resource allocation, employed in higher cortical regions, in response to different task requirements.

In the human brain, candidate regions for adaptive coding are a set of frontal and parietal brain regions, often referred to as multiple-demand (MD) regions (Cole et al., 2013; Duncan, 2010, 2013; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008), elsewhere called "task positive network" (Fox et al., 2005) or "frontoparietal control system" (Vincent et al., 2008). These are a specific set of regions in the prefrontal and parietal cortex, in particular: cortex in and around the inferior frontal sulcus (IFS), anterior insula (AI), frontal operculum (AI/FO), pre-supplementary motor area and adjacent dorsal anterior cingulate (pre-SMA/ACC) and in and around the intraparietal sulcus (IPS). They are characterised by their response to a wide range of task demands (Dosenbach et al., 2006; Duncan & Owen, 2000; Nyberg et al., 2003), even at the level of single participants (Fedorenko, Duncan, & Kanwisher, 2013). They have been widely implicated in models of executive function and cognitive control (Cole & Schneider, 2007; Corbetta & Shulman, 2002).

Human imaging data has revealed that the MD regions display distinctive forms of flexibility depending on the requirements of the task. Evidence for this comes from MVPA of fMRI data where coding is revealed by regularities in fine-grained activity patterns evoked by different stimulus events (e.g. Haxby et al., 2001; Haynes & Rees, 2006; Haynes & Rees, 2005; Kamitani & Tong, 2005). For example, the MD regions have been shown to flexibly code a range of task features demonstrating flexibility to respond to a variety of behaviourally relevant aspects of a task in different contexts (e.g. Bode & Haynes, 2009; Harel, Kravitz, & Baker, 2014; Haynes et al., 2007; Li, Ostwald, Giese, & Kourtzi, 2007; Nee & Brown, 2012; Reverberi, Gorgen, & Haynes, 2011; Stiers, Mennes, & Sunaert, 2010; Woolgar, Afshar, Williams, & Rich, 2015; Woolgar, Hampshire,

Thompson, & Duncan, 2011; Woolgar, Thompson, Bor, & Duncan, 2011; Woolgar, Williams, & Rich, 2015) also refer to (Appendix A, Woolgar, Jackson, & Duncan, in press). In addition they have been shown to adjust the strength of their coding within single tasks as task demands vary, increasing their representation of perceptual (Woolgar, Hampshire, et al., 2011; Woolgar, Williams, et al., 2015) and rule task information (Woolgar, Afshar, et al., 2015) when perceptual or rule elements of the task are made more difficult.

Previously (Chapter 2 – Jackson et al. (in press)), we examined whether flexibility of the MD system could underpin our capacity to attend to different features of the same object. Participants categorised objects across an orthogonal decision boundary that created two task contexts based on separate feature dimensions. The nature of the two tasks required participants to suppress information about the irrelevant feature in order to respond accurately to the behaviourally relevant feature dimension. We refer to the mechanism employed to enable this selective processing of features as feature-selective attention (Chen, Hoffmann, Albright, & Thiele, 2012). We demonstrated that the MD regions flexibly emphasised different features of the same object as required by the current task, providing a possible mechanism for feature-selective attention. The first objective of the current study was to examine whether the result replicates in a second experiment using different stimulus sets across the two tasks.

Our second objective was to test whether coding of irrelevant information across tasks is modulated by whether the irrelevant information is currently irrelevant (relevant in the alternative task) or never relevant to the participants' task. To address this objective we drew a comparison between our previous dataset (Chapter 2 – Jackson et al. (in press)) and the current data. In Kadohisa et al.'s study, the results indicated when information was

consistently irrelevant it was easier to filter out. We therefore predicted a similar pattern of results in our data.

A third goal of this study was to investigate flexible coding in the MD system in more detail, by interrogating the extent to which multiple neural codes load on the same individual voxels. In Cromer et al.'s (2010) and Roy et al.'s (2010) studies, the firing rate of individual units discriminated between multiple different visual stimuli in different tasks. Although these studies provide detailed information at the level of single-cell activity, they are limited in the scale of brain network under study. New methods are needed to examine this process in the human brain. Accordingly, we developed a method to measure the extent to which voxels in the MD regions were 're-used' in multiple codes more often than expected by chance, as an indirect method for estimating neuronal population recruitment. In Chapter 2, the patterns across the MD regions held relevant over irrelevant information, but here we ask whether it is the *same voxels* that hold that information. We predicted, in line with the adaptive coding hypothesis (Duncan, 2001), that MD voxels would flexibly be re-used to code behaviourally relevant information across multiple tasks.

Ideally, to answer a question about whether neural resources are re-used across multiple tasks, we would exploit responses at the neural-level rather than the voxel-level. However, this is not possible in the human brain. Thus, a finding that a voxel carries different information under two task conditions may indicate that the same neural populations are active across the different conditions, consistent with the single cell data. Alternatively, it could be due to different neural populations within the voxel that are each active in one condition. It seems unlikely that there would be a bias for qualitatively different neuron populations to cluster together in single voxels, particularly given the actual voxel 'allocation' is arbitrary, but nonetheless, we refer to only 'voxel re-use' rather

than 'neural re-use'. We see this method as the first step towards testing the notion of recruitment of the same neural resources for different tasks.

Finally, these datasets allow investigation of another level of flexible control in the MD network: whether the extent to which MD regions behave as "specialists" and "generalists" is affected by task demands, as suggested in the non-human primate monkey literature (Cromer et al., 2010; Roy et al., 2010). In Cromer et al.'s (2010) and Roy et al.'s (2010) studies, less neural multitasking occurred when monkeys switched between two orthogonal decision boundaries of the same stimuli, requiring suppression of the previously relevant information (high demand), compared when task the switched between separate groups of visual objects (low demand). We asked whether the extent to which MD voxels were re-used to code behaviourally relevant information was modulated by whether the tasks required participants to switch their focus of attention between two features of the same object (Chapter 2 data, high demand) or to switch their focus of attention between features of different objects (current experiment data, low demand). This is the stronger test of neural flexibility, and the one that this study was originally designed to examine, because if there is an *interaction* in the extent to which voxels are re-used between the two experiments, this must be based on differential allocation of resources and cannot be attributed to sub-populations of neurons sampled by different voxels.

Materials and Methods

This study includes comparisons to the data presented in Chapter 2. In this Chapter, we refer to the experiment from Chapter 2 as "Experiment 1". We designed the current experiment ("Experiment 2") to be comparable to Experiment 1 to allow comparisons between the two datasets. However, there was one important difference. In

Experiment 1, the two sets of encoded stimulus features belonged to the *same* group of objects (Figure 1, top panel). Thus, in Experiment 1 participants were attending to the same object, but different features of that object, in the two conditions. As a result, the irrelevant dimension in one block was the relevant dimension in other blocks. In Experiment 2, the two sets of encoded task-relevant stimulus features belonged to *different* groups of objects, rendering the irrelevant dimension never relevant to the task (Figure 1, bottom panel).

Our first aim was to examine whether the MD regions adjust to code task-relevant feature information across different groups of objects reflecting the same type of adaptive properties as demonstrated in our previous work (Chapter 2 – Jackson et al. (in press)). Second, we compared the current data set to that from the previous Chapter to examine whether the strength of coding of *irrelevant* feature information differed depending on whether that information was sometimes relevant or never relevant in the broader experimental context. Our third aim was to examine the extent to which the same MD voxels contributed to the neural codes for the relevant information across different tasks. Finally, we assessed whether a greater proportion of voxels would be re-used to code task-relevant stimulus features when the encoded features belong to different objects (Experiment 2, low demand) compared to when encoded features belong to the same objects (Experiment 1, high demand). To allow these comparisons, the two experiments were matched as closely as possible in terms of their methods and procedure. Here I describe the details of the new experiment, Experiment 2, emphasising differences from Experiment 1 where relevant (for full details on Experiment 1, refer to Chapter 2, Jackson et al. (in press)).

Participants

Twenty-six healthy adult volunteers (17 females; mean age = 23.9 years, SD= 4.56) participated in Experiment 2. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants were selected using the Macquarie University Psychology Participant Pool (SONA) to be matched on age (BF_{10}<0.28) and gender (17 female) with the 26 participants in Experiment 1. Participants gave written informed consent and the human research ethics committee of Macquarie University (Sydney, Australia) approved the experiment. All participants received $50.

Stimuli
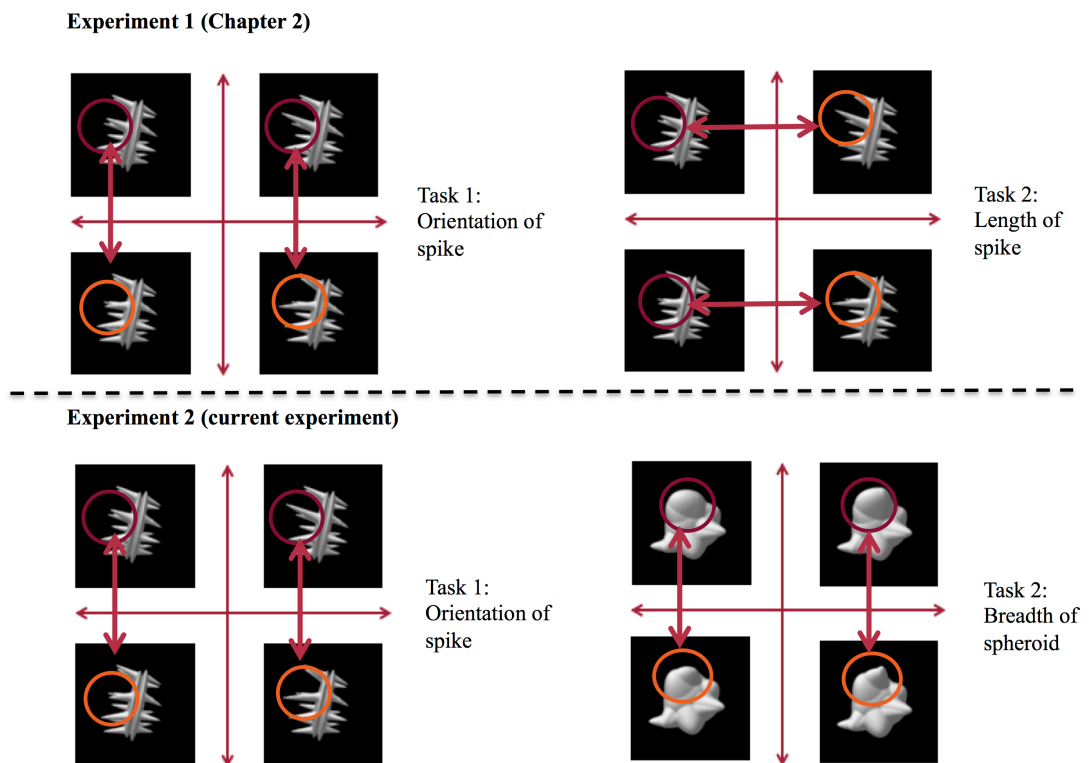
Figure 1: Experiment 1 (top panel): Participants performed two tasks in the scanner where they categorised spiky objects according to an orientation (task 1) and length (task 2) dimension. **Experiment 2 (bottom panel):** Participants performed two tasks in the scanner where they categorised spiky

objects along an orientation dimension (task 1) and smoothy objects along a breadth dimension (task 2). The objects in this figure depict the extremes of each dimension (refer to Figure 2 for all 32 objects). The red and orange circles in this figure are to illustrate the differences between objects and were not shown in the actual display

Our stimulus set consisted of abstract novel "spiky" objects and "smoothy" objects created using custom MatLab scripts (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). The stimulus set consisted of 32 objects (16 spiky and 16 smoothy objects, Figure 2). One "spike" of the spiky objects varied along two dimensions (its length and orientation) and one "spheroid" of the smoothy objects also varied along two dimensions (its breadth and height).

Participants performed two tasks. In one task, participants discriminated between the 16 spiky objects based on the orientation dimension (rotated clockwise vs. rotated anti-clockwise spikes). For the second task, subjects discriminated between the 16 smoothy objects based on the breadth dimension (wide vs. narrow spheroid). The stimuli also varied on a second dimension (for spiky objects the critical spike also varied in length, for smoothy objects the spheroid varied in height) but participants never discriminated the objects based on these irrelevant dimensions (Figure 1, bottom panel shows maximum difference between objects for the relevant dimensions). This aspect of the stimulus design was different to Experiment 1 (Chapter 2) in which only the spiky objects were used (Figure 2, left panel), and participants switched between the two dimensions of the spiky stimulus set. During both the training and scanning sessions, the visual angle (VA) of the spiky object's length along its main axis was 8.07° and for the smoothy objects it was 8.56°. A PC running the Psychophysics Toolbox-3 package (Brainard, 1997) in Matlab (Mathworks) controlled stimulus presentation. Stimuli were presented at central fixation on a screen and viewed through a mirror mounted on the head coil in the scanner.

Length dimension

Height dimension

Orientation dimension
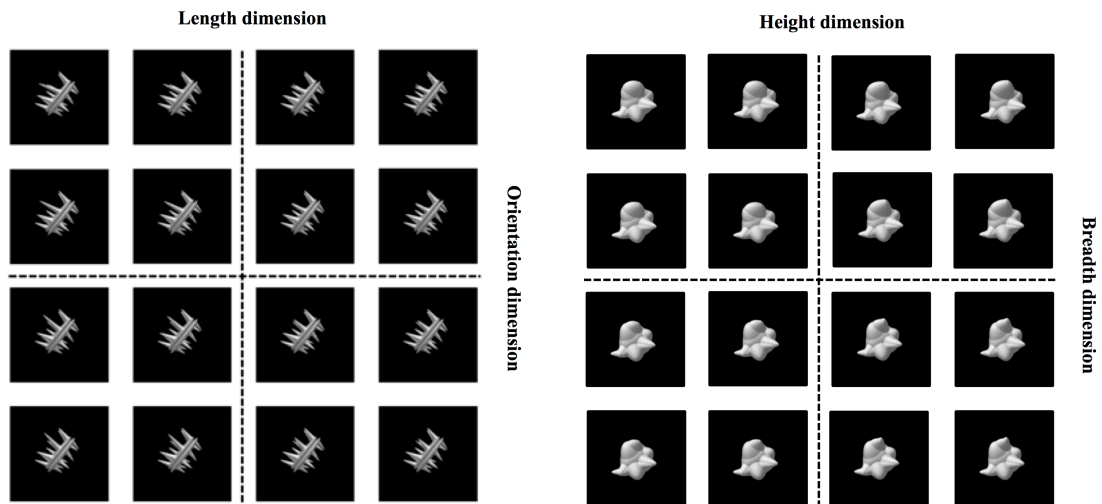
Breadth dimension

Figure 2: The stimulus set consisted of 32 objects total. One "spike" of the spiky objects varied along two dimensions (its length and orientation) and one "spheroid" of the smoothy objects also varied along two dimensions (its breadth and height). Participants categorised the spiky objects according to the orientation dimension; the length dimension was always irrelevant. For the second task, participants categorised the smoothy objects according to breadth dimension; the height dimension was always irrelevant.

Procedure

In the previous chapter, participants completed at least 6 blocks of practice trials to learn the task. Stimuli were initially presented for 400ms until participants achieved >80% correct after which objects were presented for 216ms. Feedback (correct/incorrect) was presented after each response until participants achieved >80% performance, after which feedback (percent correct) was only given at the end of each block. Once participants reached a high performance level (>80% correct) in both tasks, we titrated the stimuli to ensure no difference in reaction times. We used the spiky stimulus set that was most frequently used in Experiment 1, and matched the difficulty of the smoothy object task to this. To achieve this, we increased or decreased the difficulty of the smoothy task by using different smoothy stimulus sets varying on the maximum difference across the relevant decision boundary (breadth). This procedure was repeated until there was no difference in

reaction time between the two task contexts, as assessed with Bayes analysis in each participant separately. During titration participants only received feedback at the end of each block.

Immediately prior to entering the scanner, participants completed a further 2 practice blocks of each task context to remind them of the task and to avoid initial practice effects in the scanner. These two practice blocks also introduced a response-mapping screen to be used in the scanning task, which randomly assigned the button to be pressed for each category (clockwise or anti-clockwise spikes in the orientation task and narrow or wide spheroids in the breadth task) on a trial-by-trial basis. This allowed separate estimation of the blood-oxygen-level dependent (BOLD) response associated with perceptual information about each category from that associated with each button press. Participants also performed an additional two practice blocks in the scanner during the structural scan, prior to commencing the main experiment, to familiarise them with the button-response box in the scanner.

Participants were scanned whilst performing the categorisation task shown in Figure 3 (for comparison, see also categorisation task for Experiment 1 shown in Chapter 2, Figure 2). The procedure and task were identical to Experiment 1 except that in this experiment the two tasks were performed on different stimuli as detailed above. Each participant completed 4 acquisition runs (8.09min each) consisting of 4 blocks of 128 trials (2.02min/block). At the start of each block, a picture cue (4000ms) indicated the current task (orientation of the spikes, breadth of the spheroids). The picture cue also indicated which attribute was category 1 and 2 (e.g. whether rotated clockwise/anti-clockwise spikes were category 1 or 2; counterbalanced across participants). The order of task contexts was counterbalanced across participants as well as within-participants across runs.
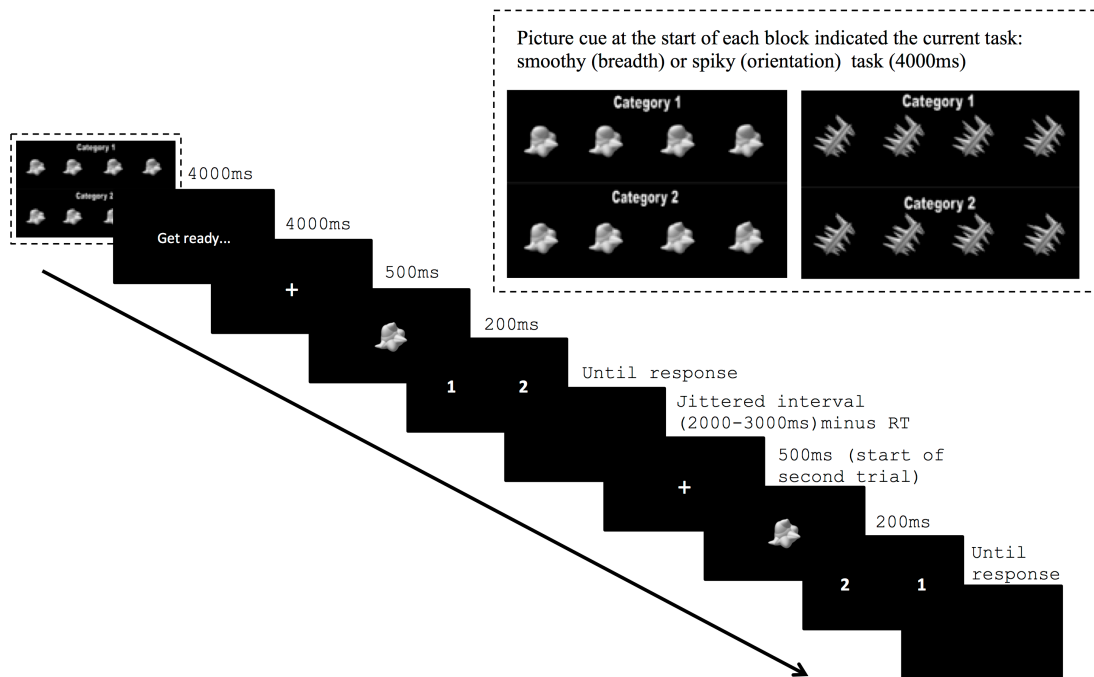
Figure 3: Stimulus categorisation task: A picture cue at the start of each block indicated the current task for categorisation: breadth (smoothy task) or orientation (spiky task). The inset shows cue display for both the orientation and breadth task. On each trial a fixation cross was presented for 500ms followed by an object for 216ms. Finally, a response mapping screen appeared which indicated the appropriate response button. In the example shown, the current task is breadth. For the first trial, the stimulus is category 1 on the breadth dimension and therefore the correct response was the left-button.

The picture cue depicted either spiky objects or smoothy objects from the extremes of the currently relevant dimension dependent on the current task context. On each trial, participants saw a white central fixation cross (500ms) after which a spiky or smoothy object was presented at fixation for 216ms. Finally, participants saw a response mapping screen which indicated the category-to-button response mapping on this trial, and responded regarding the category membership of the stimulus. The response mapping screen randomly assigned category 1 and 2 decisions to either the left or right response button, operated by the index or middle finger of the participant's right hand. The response mapping screen was visible until a button-press was made or until the jittered

time interval timed out (2000-3000ms). If a response was made before the end of the inter-trial-interval, a blank black screen was shown for the remainder of the trial time. Feedback (accuracy score) was presented at the end of each block for 6000ms after which there was a delay of 4000ms prior to the start of the next block. At the end of each run, a blank black screen was shown for 4000ms.

Following completion of the main task during the scanning session, we ran a localiser task to functionally identify the LOC as an *a priori* region-of-interest (ROI). Participants viewed centrally located intact and scrambled versions of black and white objects in 16.8s blocks of 16 trials (1100ms/trial), whilst attending to a central fixation cross. Participants indicated via a button response when the fixation cross changed colour. There were 21 blocks consisting of alternating blocks of whole objects, scrambled objects, and rest blocks (counterbalanced across participants). The EPI (acquisition time) for the localiser task was 6.25min.

Data Acquisition

Data acquisition was identical to Experiment 1. FMRI data were collected using a 3T Siemens Verio Magnetic Resonance Imaging (MRI) scanner at Macquarie University Hospital. We used a sequential descending T2*- weighted echo planar imaging (EPI) acquisition sequence with the following parameters: acquisition time 2000 ms; echo time 30 ms; 34 oblique axial slices with a slice thickness of 3.0 mm and a 0.70 mm inter-slice gap; in plane resolution 3.0×3.0 mm; matrix 64×64; field of view 210 mm; flip angle 78°. T1-weighted MPRAGE structural images were also acquired for all participants (slice thickness 1.0 mm, resolution 1.0×1.0 mm).

Preprocessing

Preprocessing was identical to Experiment 1. MRI data were preprocessed using SPM 5 (Wellcome Department of Imaging Neuroscience, www. fil.ion.ucl.ac.uk/spm) in Matlab 2011b. Functional MRI data were converted from DICOM to NIFTII format, spatially realigned to the first functional scan and slice timing corrected, and structural images were co-registered to the mean EPI. EPIs from the main experiment were smoothed slightly (4 mm FWHM Gaussian kernel) to improve signal-to-noise ratio. LOC localiser EPIs were also smoothed (8 mm FWHM Gaussian kernel) and in all cases the data were high pass filtered (128s). Structural scans were additionally normalised, using the segment and normalise routine of SPM5, in order to derive the individual participant normalisation parameters needed for ROI definition (below).

Regions of Interest

ROIs were defined in the same way as for Experiment 1. MD regions of interest (ROIs) were defined using co-ordinates from a previous review of activity associated with a diverse set of cognitive demands (Duncan & Owen, 2000) using the kernel method described in Cusack, Mitchell, and Duncan (2010), and as in previous work (Woolgar, Hampshire, et al., 2011; Woolgar, Thompson, et al., 2011; Woolgar, Williams, et al., 2015). The procedure yielded a total of seven ROIs: left and right IFS (centre of mass +/−38 26 24, volume 17 cm$_3$); left and right AI/FO (+/−35 19 3, 3 cm$_3$); left and right IPS (+/−35 −58 41, 7cm$_3$) and ACC/ pre-SMA (0 23 39, 21 cm$_3$).

Left and right visual cortex ROIs were derived from the Brodmann template provided with MRIcro (Rorden & Brett, 2000) (BA 17, centre of mass −13 −81 3, 16 −79 3, volume 54 cm$_3$). All co-ordinates are given in MNI152 space (McConnell Brain Imaging Centre, Montreal Neurological Institute). MD and BA17 ROIs were deformed for

each participant by applying the inverse of the participant's normalisation parameters. This allowed analyses to be carried out using native space (i.e., non-normalised) EPI data.

We defined LOC for each participant based on the functional localiser scan as the brain area that responded more strongly to whole objects than to scrambled versions of the same objects. We used the standard multiple regression approach of SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk) to estimate values pertaining to the whole and scrambled object conditions (block design). Blocks were modelled using a box car function lasting 16s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. Whole-brain analyses (paired t-tests) compared voxelwise BOLD response in the two conditions (whole objects minus scrambled objects). The resulting map was thresholded such that there was at least one cluster with a minimum size of 20 voxels. These clusters were then imported into MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) and clusters of activation close to anatomical LOC coordinates from previous studies (Grill-Spector et al., 1999; Grill-Spector, Kushnir, Hendler, & Malach, 2000) were selected as ROIs.

First-Level Model

To obtain estimated activation patterns for multivariate analysis, a General Linear Model (GLM) was estimated for each subject using the realigned, slice-time corrected and smoothed native space EPI images using SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk). The data were modelled as in Experiment 1. Relevant features were "rotated clockwise" or "rotated anti-clockwise" for *orientation* context blocks, and "wide" or "narrow" for *breadth* context blocks. The irrelevant features were "short" or "long" for *orientation* context blocks, and "tall" or "short" for *breadth* context blocks (height dimension). Trials were modelled as events of zero duration convolved with the hemodynamic response of SPM5. Every trial contributed

to the estimation of two beta values, the relevant feature (wide or narrow for *breadth* context blocks, and clockwise or anti-clockwise for *orientation* context blocks) and irrelevant feature (wide or narrow for *orientation* context blocks, and clockwise or anti-clockwise for *breadth* context blocks). We derived the estimates for each feature in each block separately. The two run means in each experiment were included in the model as covariates of no interest. Error trials were excluded from the analysis.

MVPA

As in the previous Chapter, we used MVPA to examine the representation of relevant and irrelevant features in the MD regions, LOC, and early visual cortex (Brodmann area 17, BA 17). We implemented MVPA using the Decoding Toolbox (Hebart, Görgen, & Haynes, 2015), which wraps the LIBSVM library (Chang & Lin, 2011). We examined coding of orientation of the spike (relevant) and length of the spike (irrelevant) in the spiky orientation task. We also examined coding of breadth of the spheroids (relevant) and the height of the spheroids (irrelevant) in the smoothy breadth task. We also drew on the data from Experiment 1 to make comparisons between the datasets. In Experiment 1 the conditions were coding of orientation when orientation was relevant, orientation when orientation was irrelevant, length when length was relevant, length when length was irrelevant.

For each participant and ROI, a linear support vector machine was trained to decode the relevant (clockwise or anti-clockwise for orientation context blocks, and wide or narrow for *breadth* context blocks) and irrelevant (short or long for *orientation* context blocks, and tall or thin for *breadth* context blocks) stimulus features for each task context separately.

In total, there were 16 blocks for each participant in each experiment: 8 with orientation relevant and 8 with breadth relevant. For each classification, we used a leave-one-out 8 fold splitter whereby the classifier was trained using the data from 7 out of the 8 blocks and subsequently tested on its accuracy at classifying the unseen data from the remaining block, iterating over all possible combinations of training and testing blocks. For example, to yield a classification accuracy score for the task-relevant breadth distinctions (i.e., wide vs. narrow in breadth blocks), we took the 8 blocks in which participants performed the breadth task and trained the classifier to distinguish between patterns of activation representing wide and narrow spheroids using data from 7 out of these 8 blocks, and then tested generalization to the remaining unseen block. The accuracies were then averaged to give a mean accuracy score for task-relevant breadth coding. The same procedure was repeated for each condition, participant and ROI separately.

The mean classification accuracy for each participant in each ROI and in each condition was then entered into a second level analysis. We conducted a three factor analysis of variance (ANOVA) on classifier accuracy with the factors *relevancy*, *object*, and *MD region*. To explore any hemispheric effects, we ran an additional ANOVA with factors *relevancy, object, MD region* and *hemisphere*. Since a difference in coding in the relevant and irrelevant conditions is only interpretable if coding in at least one condition is also significantly above chance, we also conducted one-sample t-tests against the classification accuracy expected by chance (50%) in each condition (relevant and irrelevant) separately. One-tailed significance tests were used where appropriate for inference: tests comparing classification of relevant to irrelevant feature distinctions in the MD regions are one-tailed since the direction of the effect is pre-specified, and tests comparing classification accuracy to chance are one-tailed as below chance classifications

are not interpretable. All other tests are two-tailed. Alpha was adjusted for four comparisons using Bonferroni correction (0.05 divided by 4).

Next we compared the datasets to see whether there was stronger coding of the irrelevant feature information when the irrelevant feature was relevant to the task half of the time (Experiment 1), compared to when the irrelevant feature was never relevant to the task (Experiment 2). To do this we used a mixed model ANOVA with within-subjects factor *region* (AI/FO, IFS, ACC/pre-SMA, and the IPS; collapsed across hemisphere) and between-subjects factor *experiment* (Experiment 1, Experiment 2).

As in Experiment 1, we also examined whether coding in the visual cortices was stronger for task-relevant than irrelevant stimulus features. For this, we used a two-factor ANOVA on classifier accuracy with factors *relevancy* (Task-relevant, Task-irrelevant) and *object* (Spiky, Smoothy) for each of the visual cortex ROIs (LOC, Early visual cortex) collapsed across hemisphere. Again, we also tested whether coding in the visual cortices was above chance in each condition separately (one-sample t-test against the classification accuracy expected by chance, 50%).

Voxel re-use

Voxel re-use for relevant information

We developed an extension of multi-voxel pattern analysis to examine **1)** whether voxels are re-used to code relevant information across multiple tasks; and **2)** whether the extent to which resources are re-used varies with task demands (i.e., between experiments). Although each voxel may have contributions from independent populations of neurons, it seems unlikely that there is a bias for such neuron populations to cluster together in single voxels rather than across voxels. Thus, we can look at the voxel-level re-use to investigate these two aims.

For this analysis, we identified the voxels that contributed the most signal to the stimulus discriminations using a transformation of the classifier weight outputs. A recent paper (Haufe et al., 2014) showed that multivariate classifier weight vectors alone do not provide information about the signal-of-interest because voxel weights can be independent of the underlying signal. For example, a voxel may be highly weighted (critical to the multivoxel discrimination) if it provides a good estimate of noise, which is also present in the response of other, signal-carrying voxels, even if it does not carry the signal from the underlying neural code. However, the extent to which each voxel contributes signal to the underlying neural code can be recovered by multiplying the classifier weight vector by the covariance in the data (Haufe et al., 2014). We used this approach for both Experiment 1 and Experiment 2 data, retrieving the transformed weights from the Decoding Toolbox (Hebart et al., 2015). First, for a particular participant and ROI, we trained a linear support vector machine classifier using all the data (8 blocks) of each task context separately (e.g. in Experiment 1: left vs right orientation in the 8 orientation task blocks). From this we extracted the weight assigned to each voxel by the classifier, and transformed it (Haufe et al., 2014) to recover the extent to which each voxel contributed signal to each multivoxel pattern. We then identified the top 10% of voxels with the highest signal contributing to each discrimination and calculated the proportion of these voxels that were also in the top 10% for the other relevant stimulus distinction. For example, for relevant discriminations in Experiment 1, we calculated the overlap between the voxels contributing the most signal to orientation coding in the orientation task blocks and the voxels contributing the most signal to length coding in the length task blocks. If 40 out of the 200 voxels in an ROI that contributed the highest signal to the discrimination of orientation in the orientation task were also amongst the 200 voxels that contributed the highest signal to the discrimination of length in the length task, then the proportion of re-use or overlap was 40/200 = 0.2 (20%).

For relevant discriminations in Experiment 2, we similarly calculated the overlap between the voxels contributing the most signal to orientation coding of the spiky object and the voxels contributing the most signal to breadth coding of the smoothy object. We repeated this procedure for every participant, in each ROI and each experiment separately. This measure of voxel re-use is only meaningful in regions where patterns of activation reliably discriminated between the stimuli in the first place, so we carried out this analysis only in regions and conditions where information coding was above chance in the previous analysis. As a sanity check, we also checked whether voxel re-use was at chance when information coding was at chance in Experiment 2.

Permutation tests on voxel re-use

We used a permutation test (Stelzer, Chen, & Turner, 2013) to establish the voxel re-use expected by chance. This approach accounts for within-subject factors such as vasculature that could lead to certain voxels having higher classification weights. We carried this out for each region and each experiment separately. In the first step, we exhaustively permuted the condition labels within each block (128 combinations total) for each person in each experiment and each task separately. For each permutation, we trained a classifier using the permuted data, and calculated the transformed weight vectors, in the same way as we had done for the correctly labelled data. Next, we built a group level null distribution by sampling (with replacement) from the set of 20 participants * 128 permutation results (one sample per participant per permutation, 10,000 permutations). From this, we calculated the probability p of observing the actual re-use value (from the correctly labelled data) given the group null distribution, using the Monte-Carlo approach ($k+1/(n+1)$) where k is the number of permutations in the null with equal or higher accuracy to the actual re-use value and n is the number of all permutations).

117

Statistical comparison between experiments

Finally, we examined whether the extent of MD voxel re-use varied between experiments. For this, we entered the voxel re-use values into a mixed model ANOVA with within-subjects factor *region* (AI/FO, IFS, ACC/pre-SMA, and the IPS; collapsed across hemisphere) and between-subjects factor *experiment* (Experiment 1, Experiment 2).

Results

In the following section we report the behavioural and decoding results from the current experiment (with reference to findings from Experiment 1 when relevant), followed by the comparison of voxel re-use in both experiments. For behavioural and decoding results for Experiment 1, refer to Chapter 2 (Jackson et al. (in press)).

Behavioural Results

Prior to scanning, participants practiced the task until they scored at least 80% correct in both task conditions. The stimulus set was then titrated to match reaction times between the two conditions for each participant separately (assessed with Bayes factor analysis for each participant separately: all $BF_{10} < 0.76$), with the stimulus sets matched between experiments. As BF < 1 we interpret this as evidence that the conditions were matched (Dienes, 2011).

In the scanning session, participants performed with a high degree of accuracy (94.2%, SD = 7.1). Accuracy scores were entered into a one-way ANOVA to check for differences in performance between the two task contexts. There was no main effect of feature ($F(1,27) = 0.04$, p = 0.92) , giving no evidence that at a group level participants found either condition (smoothy or spiky task) more difficult than the other. There were also no differences in accuracy score between the two conditions for any participant

individually (all $BF_{10} < 0.89$). Reaction time data from the scanning session was not analysed because the response mapping screen prevented participants from responding when they first saw the stimuli.

Decoding of Task-Relevant and Task-Irrelevant Stimulus Distinctions

MD Regions

MVPA was used to differentiate multivoxel patterns pertaining to relevant and irrelevant stimulus distinctions. The relevant feature distinctions were orientation in orientation task blocks and breadth in breadth task blocks and the irrelevant feature distinctions were length in orientation task blocks and height in breadth task blocks. The decoding results are presented in Figure 4 (refer to Chapter 2 for decoding results from Experiment 1).

Based on the adaptive coding hypothesis (Duncan, 2001) and previous results (Chapter 2), we predicted that the MD regions would code the task-relevant feature distinctions more strongly than the task-irrelevant feature distinctions. A three-way ANOVA with factors *relevancy* (task-relevant, task-irrelevant stimulus features), *region* (AI/FO, IFS, ACC/pre-SMA, and the IPS; collapsed across hemisphere where appropriate), and *object* (spikies, smoothies) revealed a main effect of *relevancy* ($F(1,25)$ = 14.5, p = 0.001), corresponding to stronger representation of the relevant compared to irrelevant stimulus dimensions. No other main effects or interactions were significant (all *p*s > 0.11). These results indicate that the MD regions coded the task-relevant feature distinctions more strongly than the task-irrelevant distinctions. We have no evidence for a difference between regions or between spiky and smoothy objects.

In our additional ANOVA, included to check for hemispheric differences, there was no main effect of *hemisphere* ($F(1,25)$ = 0.16, p = 0.69) or *relevancy* by *hemisphere*

interaction (F(1,25) = 0.20, p = 0.66). However, there was a significant interaction between *region* and *hemisphere* (F(2,50) = 6.87, p = 0.002). To explore this, we performed post hoc ANOVAs (factor: *region*) for the left and right hemispheres separately. In the left hemisphere there was a significant main effect of *region* (F(2,50) = 5.64, p = 0.006) and pairwise comparisons revealed stronger coding of information overall in the left IFS compared to the left AI/FO (p = 0.02). There was no main effect of *region* in the right hemisphere (F(2,50) = 0.28, p = 0.75).

The difference in strength of coding for relevant and irrelevant features is only interpretable if coding in one or more of the relevancy conditions is also significantly above chance. Therefore, we conducted one-sample t-tests against the classification accuracy expected by chance (50%). We found that, on average, the MD regions coded the task-relevant features significantly above chance (mean classification accuracy for relevant across all regions, 55.9%; one sample t-test against chance, t(25) = 3.93, p < 0.001), whereas classification of the task-irrelevant stimulus distinctions was numerically below chance (mean classification accuracy for irrelevant, 48.6%). Thus the MD regions only encoded the task-relevant stimulus distinctions.

To test whether this was also the case in each region individually, we considered coding in each MD ROI separately, correcting for the 4 multiple comparisons using a Bonferroni corrected alpha level (a = 0.0125). We found that the relevant stimulus distinctions were coded in 3 MD ROIs (ACC/pre-SMA, mean accuracy 56.1%; (t(25) = 2.85, p < 0.001); IPS, mean accuracy 57.3%; (t(25) = 3.74, p < 0.001); IFS, mean accuracy 56.4%; (t(25) = 3.62, p < 0.001)) with a trend in the remaining AI/FO ROI that did not reach our Bonferroni corrected significance level (mean accuracy 53.6%; (t(25) =1.81, p = 0.04)). The irrelevant feature distinctions were not coded in any of the MD regions.

Figure 4: Coding of task-relevant and task-irrelevant stimulus distinctions in MD regions. Error bars indicate standard error. Significance markings for individual bars indicate whether coding was significantly greater that chance in each condition separately (one-sample t test against chance, 50%), significance marking between bars indicate where coding was significantly greater for relevant compared to irrelevant distinctions (main effect of relevancy / paired t-test). *$p < 0.05$, ** $p < 0.01$, alpha for individual regions corrected for four comparisons using Bonferroni correction. The MD regions coded task-relevant feature distinctions more strongly than the task-irrelevant distinctions.

Comparison of irrelevant coding between experiments

Our comparable design of Experiment 1 and Experiment 2 allow us to directly test the prediction that irrelevant information that is sometimes relevant (Experiment 1) leads to stronger coding of irrelevant information compared with situations where it is never relevant (Experiment 2). To address this, we conducted a further analysis using the irrelevant feature classification from both experiments. The data are presented in Figure 5. We used a mixed model ANOVA with within-subjects factor *region* (AI/FO, IFS,

ACC/pre-SMA, and the IPS; collapsed across hemisphere) and between-subjects factor *experiment* (Experiment 1, Experiment 2). There was a significant main effect of *experiment* ($F_{(1,50)} = 4.60$, $p = 0.04$), no main effect of *region* ($F_{(3,150)} = 1.12$, $p = 0.34$) and no significant interaction ($F_{(3,150)} = 2.14$, $p = 0.09$). The results indicate that the MD regions coded the irrelevant stimulus information more strongly in Experiment 1, when this information was sometimes relevant to the task, than in Experiment 2.
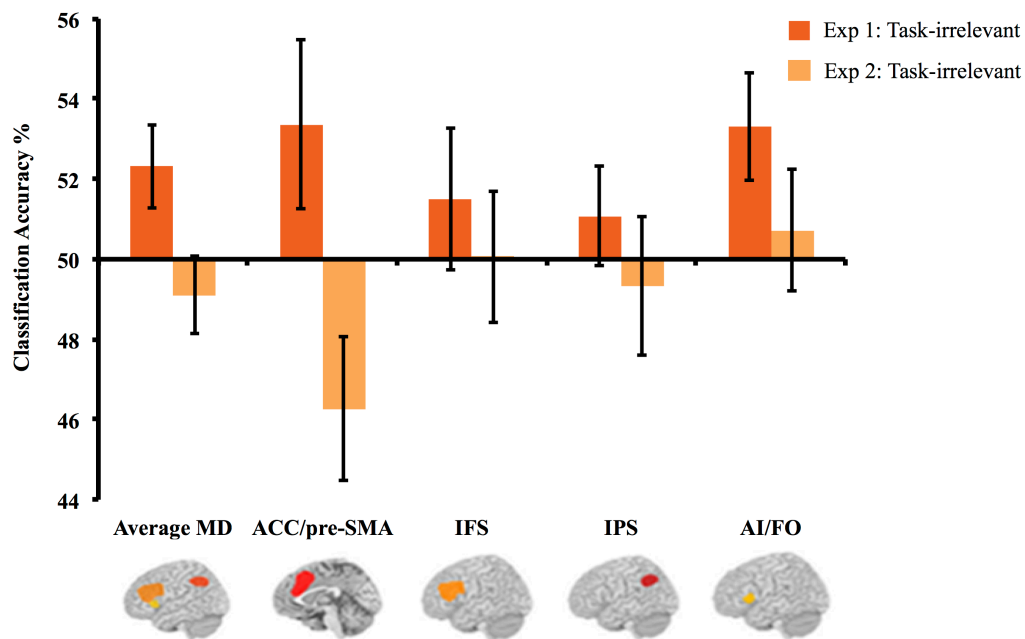
Figure 5: Coding of task-irrelevant stimulus distinctions in MD regions (Experiment 1: Left bars; Experiment 2: right bars). Error bars indicate standard error. Although coding of irrelevant information was not different from chance in either Experiment, the MD regions coded task irrelevant feature information more strongly in Experiment 1 than in Experiment 2.

Lateral Occipital Complex (LOC)

The LOC is known to respond strongly to object features (Grill-Spector, Kourtzi, & Kanwisher, 2001) and has previously been found to show preferential representation of attended relative to distractor objects (Woolgar, Williams, et al., 2015). In Experiment 1 (Chapter 2), this region only coded the relevant stimulus distinctions, but the interaction

with relevancy was not significant. In this independent set of data the trend was again in the same direction (Figure 6), but an ANOVA with factors *relevancy* (Task-relevant, Task-irrelevant) and *object* (Spikies, Smoothies) again showed no significant main effect of *relevancy* ($F(1,25) = 1.85$, $p = 0.19$). There was also no main effect of *object* ($F(1,25) = 0.44$, $p = 0.54$), and no *relevancy*object* interaction ($F(1,25) = 0.01$, $p = 0.94$). When we compared coding to chance, the LOC did not carry significant information about task relevant distinctions (mean classification accuracy = 53.0%, ($t(25) = 1.33$, $p = 0.19$)), or about irrelevant distinctions (mean classification accuracy=49.6%, ($t(25) = 0.68$, $p = 0.81$)). Thus, in Experiment 2 we did not find evidence to suggest that the LOC discriminated the objects along either dimension, or that LOC coding was modulated by relevance.
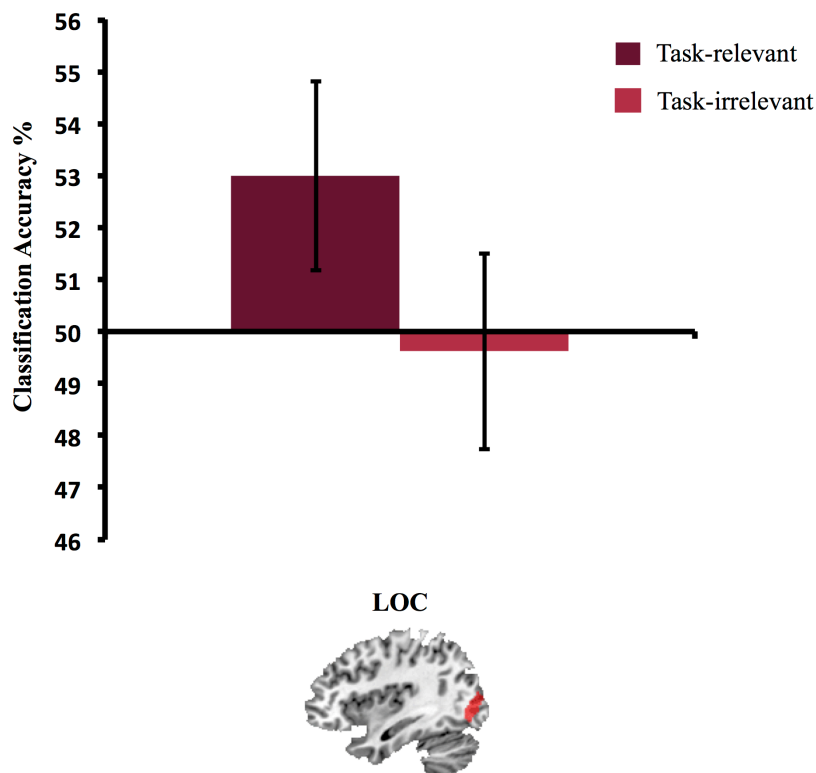
Figure 6: Coding of task-relevant and task-irrelevant stimulus distinctions in LOC. Error bars indicate standard error. Coding did not differ from chance in either condition.

123

Early Visual Cortex

Finally, we tested whether information pertaining to task-relevant and task-irrelevant stimulus distinctions was coded in early visual cortex (BA 17, Figure 7). An ANOVA with factors *relevancy* (Task-relevant, Task-irrelevant) and *object* (Spikies, Smoothies) showed no main effect of *relevancy* ($F(1,25) = 0.16$, $p = 0.69$), no main effect of *object* ($F(1,25) = 0.08$, $p = 0.78$), and no *relevancy*object* interaction ($F(1,25) = 1.02$, $p = 0.32$). Thus, we found no evidence that context modulates coding of feature information in this region. However, BA 17 did show above chance classification of these objects according to both the task-relevant (relative to chance, mean classification accuracy 56.2%; ($t(25) = 2.35$, $p = 0.002$)) and the task-irrelevant (relative to chance; mean classification accuracy 55.8%; ($t(25) = 2.95$, $p = 0.006$)) stimulus features. This indicates that although BA 17 was sensitive to small differences in the appearance of novel objects, coding in this region was not modulated by behavioural relevance.

Voxel re-use

We developed an extension of multi-voxel analysis to investigate whether MD voxels would code for relevant information across multiple tasks (multitasking). We also compared the extent of multitasking between our two datasets (Experiment 1 and Experiment 2).

Permutation tests

We ran permutation tests to compare the voxel re-use we observed to that expected by chance. We carried out this analysis for all the ROIs that showed significant coding of the stimulus information (whether relevant or irrelevant). These were the MD regions, LOC (Experiment 1), and BA 17 (Experiment 2), for the relevant stimulus dimensions, and BA 17 (Experiment 2) for the irrelevant stimulus dimension.
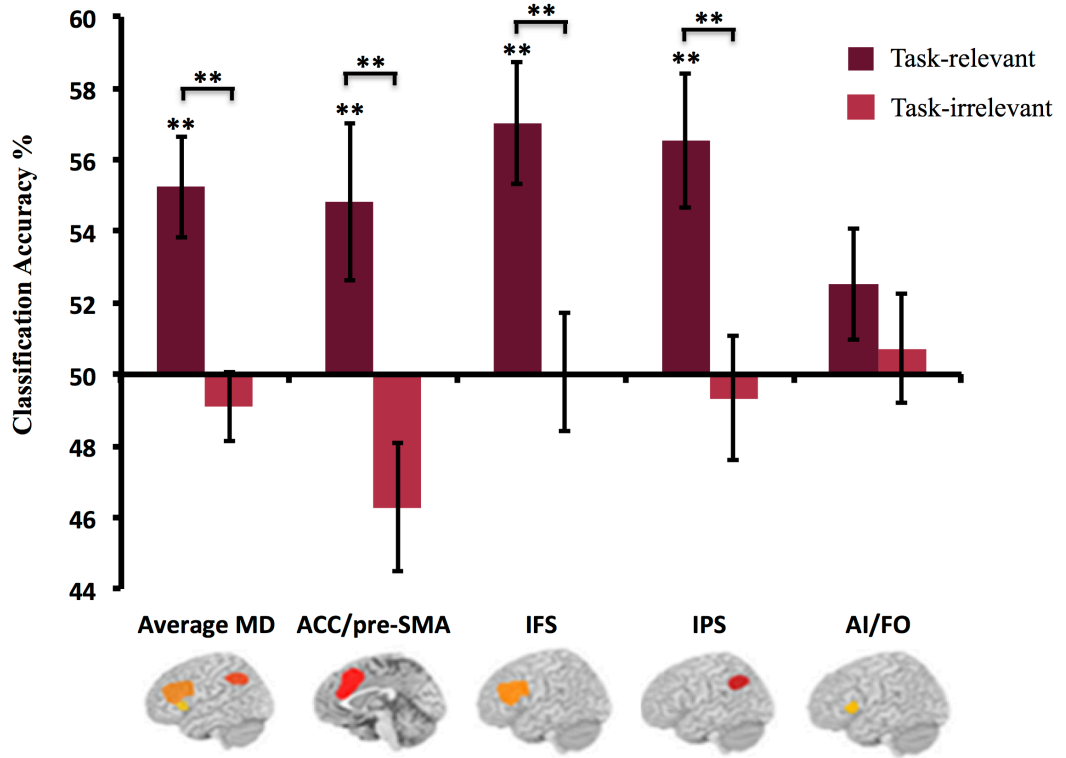
124
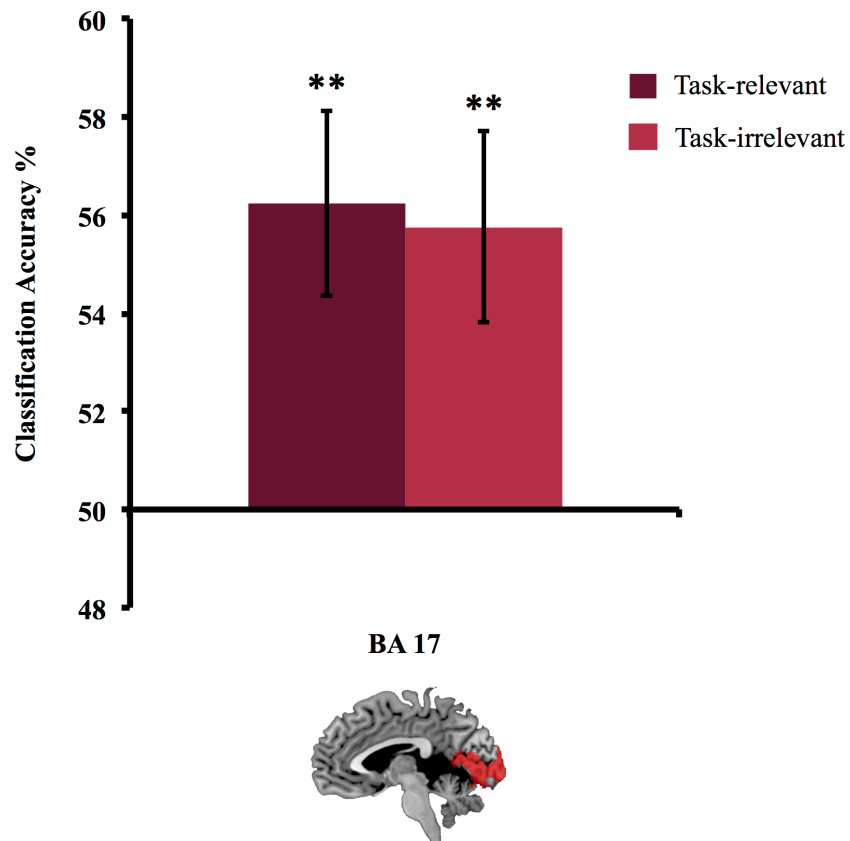
Figure 7: Coding of task-relevant and task-irrelevant stimulus distinctions in BA 17. Error bars indicate standard error. Significance markings for individual bars indicate whether coding was significantly greater that chance in each condition separately (one-sample *t* test against chance, 50%). ** *p* < 0.01.

The results for the MD network are shown in Figure 8. Collapsing over region, we found that in both Experiment 1 and 2, the MD network displayed a higher proportion of voxel re-use for the relevant dimensions than what would be expected by chance (22.6% Experiment 1; p < 0.0001, 23.9% Experiment 2; p < 0.01). Considering the regions separately, voxel re-use was above chance in the IFS in Experiment 1 (28.1%, p < 0.0001), and the IPS in both experiments (Experiment 1: 22.2%, p < 0.01, Experiment 2: 25.1%, p < 0.01). Conversely, LOC (Experiment 1) and BA 17 (Experiment 2) did not multitask coding of the relevant (p > 0.09) or the irrelevant information (BA 17, Experiment 2; p > 0.25) across the two tasks.

As a sanity check we tested whether voxel re-use was at chance when information coding was at chance in Experiment 2. Voxel re-use was at chance for irrelevant information coding across the MD regions (all ps > 0.12) as well as for relevant (p = 0.72) and irrelevant (p = 0.85) information coding in LOC.

MD regions: Comparing the proportion of voxel re-use between Experiment 1 and Experiment 2

Our novel approach allowed us to examine whether the extent to which frontoparietal voxels were re-used between tasks was dependent on the demands of the task (Experiment 1 vs. Experiment 2, Figure 8). We entered these data into a mixed model ANOVA with within-subjects factor *region* (AI/FO, IFS, ACC/pre-SMA, and the IPS; collapsed across hemisphere) and between-subjects factor *experiment* (Experiment 1, Experiment 2).
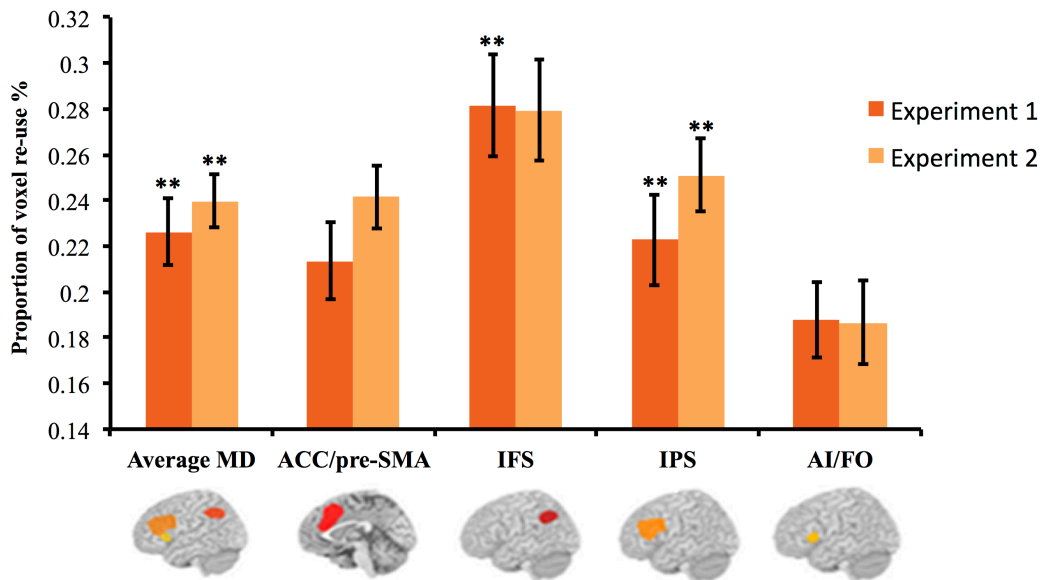
Figure 8: Proportion of voxel re-use in MD regions (for relevant dimensions) for Experiment 1 (left bars) and Experiment 2 (right bars), ** *p* < 0.01 (chance established with permutation tests, Stelzer, Chen, & Turner, 2013). There was strong evidence to suggest that voxel re-use does not differ with task demands (BF = 0.28, Experiment 1 vs. Experiment 2)

Based on the findings from the single-unit literature (Cromer et al., 2010; Roy et al., 2010) we predicted that there would be a higher proportion of voxel re-use in Experiment 2 (low demand) than in Experiment 1 (high demand). However, the ANOVA showed no significant main effect of *experiment* ($F(1,50) = 0.28$, $p = 0.60$) and no significant interaction between *experiment* and *region* ($F(3,150) = 0.64$, $p = 0.59$). There was a significant main effect of *region* ($F(3,150) = 13.4$, $p < 0.001$), and post-hoc pairwise comparisons showed that there was a higher proportion of voxel re-use in the IFS than in the three other MD regions (all $ps < 0.01$) and a lesser extent of multitasking voxels in the AI/FO than in the three other MD regions (all $ps < 0.01$). We entered the voxel re-use data into a Bayes analysis and found strong evidence for the null hypothesis that voxel re-use did not differ between experiments ($BF_{10} = 0.28$).

Discussion

In the human brain, a network of frontal and parietal brain regions referred to as the MD regions, are thought to exert control by flexibly processing information that is currently relevant (Duncan, 2001; Duncan, 2010, 2013). To explore the underlying mechanisms that promote adaptive flexibility in these regions, the current study had several aims. First, we aimed to replicate our findings that the MD regions preferentially encoded the task-relevant over task-irrelevant features of visual objects. Second, we compared our two data sets to examine whether MD coding of irrelevant information differs according to whether this information is sometimes relevant or never relevant. Third, we quantified the extent to which MD resources are re-used to code relevant feature information across tasks. Finally, we investigated whether the extent to which MD resources are re-used to code different information across tasks (whether they act as

'generalists' or 'specialists', Cromer et al., 2010; Roy et al., 2010) varies with task demands.

We examined the responses of the MD regions in a difficult visual object categorisation task in which participants discriminated the features of two objects. Unknown to the participant, an irrelevant dimension of those objects varied but was never relevant to the task. We compared the results of this experiment to our previous work where participants categorised a single set of stimuli along two orthogonal dimensions and the MD regions adjusted their coding to reflect the currently relevant dimension (Chapter 2, Jackson et al. (in press)). We found that the MD system again preferentially coded the relevant feature distinctions. In addition, we asked whether coding of the irrelevant dimension would vary depending on whether the participants were required to attend to the irrelevant dimension at other times in the experiment. We found stronger coding for the irrelevant information in Experiment 1 (when the irrelevant dimension was sometimes relevant to the task) compared to Experiment 2 (when the irrelevant information was never relevant). Consistent with the proposal that the cognitive flexibility of these regions underlies their involvement in a wide range of tasks (e.g. Cole & Schneider, 2007; Duncan, 2010; Duncan & Owen, 2000), these data suggest that this system adjusts to prioritise processing of the currently relevant features of a stimulus.

Electrophysiology studies in non-human primates have previously shown that neurons in higher cortical regions adapt their tuning profiles to respond to the information that is currently relevant (Cromer et al., 2010; Freedman & Assad, 2006; Freedman et al., 2001; Freedman et al., 2002; Roy et al., 2010; Sakagami & Niki, 1994; Sakagami & Tsutsui, 1999; Stokes et al., 2013). In the human brain, the MD regions are considered to be candidate regions for this type of control because they have been shown to code a range of different types of task information (e.g. Bode & Haynes, 2009; Haynes et al., 2007; Li

et al., 2007; Nee & Brown, 2012; Reverberi et al., 2011; Woolgar, Hampshire, et al., 2011; Woolgar et al., in press; Woolgar, Thompson, et al., 2011; Woolgar, Williams, et al., 2015), and adjust their responses when task demands vary (Li et al., 2007; Woolgar, Afshar, et al., 2015; Woolgar, Hampshire, et al., 2011; Woolgar, Williams, et al., 2015). In our previous work (Chapter 2, Jackson et al. (in press)) we have shown that these regions adapt their representations of single objects to emphasise task-relevant stimulus distinctions, resulting in preferential coding of attended stimulus features. Here we show that these regions code the task-relevant stimulus distinctions across two different groups of objects, and that, again, coding of the attended stimulus features is stronger than coding of the irrelevant stimulus features. In Experiment 2, the relevant and irrelevant stimulus dimensions were not identical, leaving open the possibility that the relevant stimulus dimensions were encoded preferentially due to differences in the physical stimuli, rather than relevance *per se*. We also chose not to swap the relevant and irrelevant dimensions over participants because, unlike with typical univariate analyses, confounds at the individual subject level do not 'average out' at the group level in decoding analyses (Todd, Nystrom, & Cohen, 2013; Woolgar, Golland, & Bode, 2014). These design choices mean that the original result in the previous chapter is more conclusive in demonstrating that the relevancy of the stimulus-dimension, rather than its physical properties, drives decoding in the MD system. However, these design choices were necessary to allow a direct comparison between coding of irrelevant information in situations where this irrelevant information is sometimes versus never relevant in other blocks, and in order to manipulate the similarity of the stimuli in the two tasks so that we could investigate the effect of stimulus similarity on voxel re-use.

Influential theories propose that cognitive control is exerted from higher cortical regions by biased processing toward task-relevant information (e.g. Desimone & Duncan,

1995; Duncan, 2001; Miller & Cohen, 2001). In both Experiment 1 and 2, MD representation of task-irrelevant stimulus features was weak, which may reflect a processing bias away from irrelevant information. When we compare data across Experiment 1 and 2, however, we can see that these regions code more irrelevant information if it has been relevant recently (i.e. in the wider task context) than if it was never relevant to the task. This may be because irrelevant information that is never relevant to the participant's task is easier to filter out than information that will potentially be used at another time. This concept is supported by evidence from Kadohisa et al.'s (2013) single-unit work which showed that when prefrontal resources were reallocated from the distractor to the behaviourally relevant target, the extent of this reallocation was greater if the accompanying stimulus was consistently irrelevant compared to when it was currently irrelevant (relevant for the alternative cue). This supports the idea that information pertaining to the irrelevant distinctions in these regions is filtered out, and that this may occur more efficiently when irrelevant information is consistently irrelevant.

In Experiment 1 (refer to Chapter 2 for full decoding results), the LOC held information about the task-relevant features but the difference in coding between relevant and irrelevant conditions did not reach significance. In Experiment 2, we observed the same trend, but multi-voxel coding of stimulus information in this region was not significantly above chance. Previous research has shown that the LOC responds to attended objects more strongly than to unattended objects (e.g. Konen & Kastner, 2008; Murray & He, 2006; Murray & Wojciulik, 2003; Woolgar, Williams, et al., 2015; Xu & Chun, 2005) so we might expect that the LOC would code relevant information more strongly than irrelevant information in both experiments. However, in both experiments we examined coding of minimal visual changes within object, which are likely to have considerably smaller effects than those of attention allocated on a whole-object level. This

makes the significant coding of relevant information in the MD regions all the more striking, since the information they encoded was based on minimal visual differences.

BA 17 held information about both the task-relevant and the task-irrelevant feature distinctions to a similar extent, indicating that it does not discriminate object features according to behavioural relevance. This is in line with previous work that has shown that multiple features of objects are encoded in early visual areas even if they are task-irrelevant (Emmanouil, Burton, & Ro, 2013), alongside the idea that these regions are primarily stimulus-driven. Other authors have reported attentional influences on early visual cortex (e.g. Jehee, Brady, & Tong, 2011; Kok, Jehee, & de Lange, 2012; Woolgar, Williams, et al., 2015) but we did not find any evidence for an influence of feature-selective attention on the discrimination between our physically similar stimuli in either experiment. The results of Experiment 2 differ from our findings from Experiment 1 (Chapter 2) both relevant and irrelevant feature information were coded significantly in Experiment 2, whereas we did not observe coding of either in Experiment 1. The two main differences between the experiments were the different stimuli (smoothies were only included in Experiment 2), and the requirement to supress the irrelevant dimension of the stimuli, which was unique to Experiment 1. However, decoding was always within stimulus sets (i.e. between 2 sets of spikies, or between 2 sets of smoothies), and there was no evidence that coding in Experiment 2 was driven by the smoothies, so it is not obvious how the different stimulus sets could account for the difference. The requirement to suppress one aspect of the stimuli may account for the different results between experiments. However, to account for the data it would have to be a general suppression of coding rather than suppression of the irrelevant information specifically.

In this experiment, we examined another level of flexibility: namely whether MD regions can effectively multitask (voxels are re-used) coding of relevant information

across different tasks, and whether their behaviour as "specialists" or "generalists" is affected by task demands. To do this we developed an extension of multi-voxel analysis and identified the voxels that contributed the most signal to the stimulus discriminations using a transformation of the classifier weight outputs. We found that voxel re-use in the MD regions was significantly above chance, while voxel re-use in LOC and BA17 was not. Voxel re-use could potentially reflect the flexibility of underlying neural populations to dynamically adjust their responses to reflect relevant feature information across multiple tasks. These results fit within the predictions made by the adaptive coding hypothesis (Duncan, 2001) and replicates the findings from the non-human primate literature in the human brain at the level of voxel responses (Cromer et al., 2010; Freedman, 2001; Roy et al., 2010)

To establish chance for voxel re-use we used a permutation test (Stelzer et al., 2013) which accounts for within-subject factors such as vasculature that could lead to certain voxels having higher classification weights. Clearly, however, voxel re-use is an indirect measure of the extent to which individual neurons are re-used (20,000 to 30,000 neurons per mm^3 of cortex (Logothetis, 2008)), and it is possible that in our data, the same neurons were not re-used, even when voxels were, because those voxels happened to sample two independent populations of neurons each responding to different tasks. Accordingly, we draw conclusions only at the level of voxels, not neurons. However, it seems unlikely that such an explanation could completely account for our results, because the key independent neural populations (coding for the arbitrary categorisations imposed by the task), would have to happen to concentrate within single voxels more frequently than they are distributed across voxels, and this would have to be consistent across the MD regions, across participants and across the two experiments. Future research

combining electrophysiological recordings with fMRI in non-human primates could further elucidate these possibilities.

Based on the single-unit literature (Cromer et al., 2010; Roy et al., 2010), we predicted that there would be a higher proportion of MD voxel re-use for coding stimulus information across tasks with different objects (Experiment 2, low demand) than across tasks with the same stimulus set (Experiment 1, high demand). Our results showed that on average, in the MD network, in Experiment 1, 22.6% of the top 10% of voxels were re-used to code the two relevant stimulus dimensions in the two tasks in Experiment 1, and 23.9% of the top 10% of voxels were re-used in Experiment 2. However, although the numerical trend was in the predicted direction there was no significant difference in voxel re-use between the two experiments. In fact, a Bayes Factor analysis indicated strong evidence for the null hypothesis that the extent of multitasking does not vary in accordance with the demands of the tasks. One possibility is that the extent of re-use at the single neuron level does not vary in humans, or at least not within the MD network, and therefore voxel re-use also did not vary. This would mean that, contrary to our prediction, these regions re-use the same proportion of resources regardless of the circumstances. For example, they may use some optimal division of labour regardless of similarities in the information being encoded. However, there are a number of alternative explanations for our null result. It may be that the demands of our two tasks were not sufficiently different (i.e. that our manipulation was not strong enough) to induce a difference that our methods would be sensitive to. It could also be the case that changes did occur on a sub-voxel scale but we were not able to detect them. There are limitations to what we can infer about resource re-use using this method, but if one is careful about the inference we draw, then this and similar methods (e.g. van Kemenade et al., 2014) are promising approaches to examine the distribution of neural resources between tasks in the human brain.

Successful behaviour requires an adaptive cognitive system that can implement various mechanisms in order to process information in our environment flexibly and efficiently. Here, we have replicated our finding that the frontoparietal cortex prioritises feature information that is needed for the current task. Task-relevant feature distinctions were coded more strongly in the MD regions than the irrelevant distinctions. The irrelevant feature distinctions themselves were not reflected in the pattern information across this network even though they were coded in early visual areas. We also investigated the distribution of neural resources between tasks. Our data show that MD voxels are re-used to code information across multiple tasks, more often than predicted by chance, while voxels in visual regions are not. There was evidence that voxel re-use did not depend on task demands, at least as far as were manipulated them. The new methods developed here have potential for investigating the distribution of neural resources in the human brain. This study provides evidence that the MD network emphasises task relevant features of different objects and can flexibly re-use its resources in order to do so, providing a mechanism for the implementation of feature-selective attention.

Funding

Acknowledgments

the stimuli. We also thank Tijl Grootswagers, Tom Carlson and Susan Wardle for helpful discussions of the experimental design and analysis.

References

Bode, S., & Haynes, J.-D. (2009). Decoding sequential stages of task preparation in the human brain. *NeuroImage, 45*(2), 606-613. doi:10.1016/j.neuroimage.2008.11.031

Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage, 16*(2).

Chang, C. C., & Lin, C. J. (2011). LIBSVM a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 27.

Chen, X., Hoffmann, K. P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology, 107*, 1530-1543.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience, 16*(9), 1348-1355. doi:10.1038/nn.3470

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

Cusack, R., Mitchell, D. J., & Duncan, J. (2010). Discrete object representation, attention switching, and task difficulty in the parietal lobe. *Journal of Cognitive Neuroscience, 22*, 32-47.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron, 50*(5), 799-812. doi:10.1016/j.neuron.2006.04.031

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobes recruited by diverse cognitive demands. *Trends in Neuroscience, 23*, 475-483.

Emmanouil, T. A., Burton, P., & Ro, T. (2013). Unconscious processing of unattended features in human visual cortex. *Journal of Cognitive Neuroscience, 25*, 329-337.

Fedorenko, E., Duncan, J., & Kanwisher, N. G. (2013). Broad domain-generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the USA, 110*(41), 16616-16621.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Freedman, D. J. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science, 291*(5502), 312-316. doi:10.1126/science.291.5502.312

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature, 443*(7107), 85-88. doi:10.1038/nature05078

Freedman, D. J., & Miller, E. K. (2008). Neural mechanisms of visual categorization: insights from neurophysiology. *Neuroscience & Biobehavioral Reviews, 32*(2), 311-329.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312-316.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: neurophysiology and behavior. *Journal of Neurophysiology, 88*(2), 929-941.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience, 23*(12), 5235-5246.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. G. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 21*, 1409-1422.

Grill-Spector, K., Kushner, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*, 837-843.

Harel, A., Kravitz, D., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences of the USA*, 962-971.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage, 87*, 96-110.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*, 2425-2429.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci, 7*(7), 523-534.

Haynes, J.-D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology, 17*(4), 323-328. doi:10.1016/j.cub.2006.11.072

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

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8*. doi:10.3389/fninf.2014.00088

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience, 31*(22), 8210-8219. doi:10.1523/jneurosci.6153-09.2011

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., . . . Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80*(1), 235-246.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679-685.

Kok, P., Jehee, Janneke F. M., & de Lange, Floris P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron, 75*(2), 265-270. doi:10.1016/j.neuron.2012.04.034

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience, 11*(2), 224-231. doi:10.1038/nn2036

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible Coding for Categorical Decisions in the Human Brain. *Journal of Neuroscience, 27*(45), 12321-12330. doi:10.1523/jneurosci.3795-07.2007

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878. doi:10.1038/nature06976

Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Reviews Neuroscience, 24*, 167-202.

Murray, S. O., & He, S. (2006). Contrast Invariance in the Human Lateral Occipital Complex Depends on Attention. *Current Biology, 16*(6), 606-611. doi:10.1016/j.cub.2006.02.019

Murray, S. O., & Wojciulik, E. (2003). Attention increases neural selectivity in the human lateral occipital complex. *Nature Neuroscience, 7*(1), 70-74. doi:10.1038/nn1161

Nee, D. E., & Brown, J. W. (2012). Rostral–caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *NeuroImage, 63*(3), 1285-1294. doi:10.1016/j.neuroimage.2012.08.034

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petterson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*, 371-377.

Reverberi, C., Gorgen, K., & Haynes, J. D. (2011). Compositionality of Rule Representations in Human Prefrontal Cortex. *Cerebral Cortex, 22*(6), 1237-1246. doi:10.1093/cercor/bhr200

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology, 12*, 191-200.

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research, 97*(3), 423-436.

Sakagami, M., & Tsutsui, K.-i. (1999). The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience Research, 34*(2), 79-89. doi:http://dx.doi.org/10.1016/S0168-0102(99)00038-3

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage, 65*, 69-82. doi:10.1016/j.neuroimage.2012.09.063

Stiers, P., Mennes, M., & Sunaert, S. (2010). Distributed task coding throughout the multiple demand network of the human frontal–insular cortex. *NeuroImage, 52*(1), 252-262. doi:http://dx.doi.org/10.1016/j.neuroimage.2010.03.078

Stokes, Mark G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron, 78*(2), 364-375. doi:10.1016/j.neuron.2013.01.039

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage, 77*, 157-165. doi:10.1016/j.neuroimage.2013.03.039

van Kemenade, B. M., Seymour, K., Wacker, E., Spitzer, B., Blankenburg, F., & Sterzer, P. (2014). Tactile and visual motion direction processing in hMT+/V5. *NeuroImage, 84*, 420-427. doi:http://dx.doi.org/10.1016/j.neuroimage.2013.09.004

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Watanabe, M. (1986). Prefrontal unit activity during delayed conditional Go/No-go discrimination in the monkey. I. Relation to the stimulus. *Brain research, 382*(1), 1-14. doi:http://dx.doi.org/10.1016/0006-8993(86)90104-6

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

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. doi:10.1016/j.neuroimage.2014.04.059

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. *Journal of Neuroscience, 31*(41), 14592-14599. doi:10.1523/jneurosci.2616-11.2011

Woolgar, A., Jackson, J., & Duncan, J. (in press). Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. *Journal of Cognitive Neuroscience*.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage, 56*(2), 744-752. doi:10.1016/j.neuroimage.2010.04.035

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *NeuroImage, 109*, 429-437. doi:10.1016/j.neuroimage.2014.12.083

Xu, Y., & Chun, M. M. (2005). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature, 440*(7080), 91-95. doi:10.1038/nature04262

Chapter 4

Next, I turned to the causal role of the MD regions in feature-selective attention. For this I used a similar paradigm, in which participants attend to one of two features of an object. Since I wanted to measure the interference by irrelevant features on behaviour, I used more obvious features categories (colour and shape) than the very subtle categories of my previous work (e.g. orientation and length). I reasoned that I would be more likely to be able to detect the effect on behaviour if the irrelevant categories were easy to distinguish without being closely attended. I carried out substantial piloting and development work to refine this paradigm for use in a subsequent concurrent TMS-fMRI experiment, as detailed in this Chapter. This chapter contains the results of a TMS experiment I carried out at Macquarie (Sydney) and two behavioural experiments I carried out at the University of Reading (UK).

Exploring the causal role of the right dorsolateral prefrontal cortex in feature-selective attention

Jade Jackson[1], Anina N. Rich[1], Paul Sowman[1], Alexandra Woolgar[1]

[1]Perception in Action Research Centre and ARC Centre of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney

Abstract

Our brains receive more information from the world than we can process at once. To function successfully, we must flexibly select what is relevant and filter out what is irrelevant. The dorsolateral prefrontal cortex (dlPFC) is proposed to be fundamental to this process, acting as a selective gating or filtering mechanism (e.g. Shimamura, 2000). It is part of a network of frontal and parietal 'multiple demand', or MD, regions that are proposed to be critical to goal-directed behaviour, supporting selective processing of task-relevant information across the brain (Duncan, 2010, 2013). This network is often contrasted with the default mode network (DMN) which is found to deactivate in response to demanding cognitive tasks (Buckner, Andrews-Hanna, & Schacter, 2008). Here, we used transcranial magnetic stimulation (TMS) to stimulate right dlPFC whilst participants performed a task that measured intrusion by irrelevant feature information. Participants paid attention to one feature of novel objects (e.g. colour) and ignored another (e.g. form). Crucially, the response for the irrelevant feature could either be congruent or incongruent with the required response for the relevant feature. Participants completed this task in four separate TMS sessions consisting of high-intensity (100% motor threshold (MT)) stimulation of dlPFC, low intensity (40% MT) stimulation of dlPFC, high intensity stimulation of a control region targeting the DMN, and a sham coil over dlPFC. We predicted that when dlPFC processing was disrupted during the task we would see an increase in the magnitude of the congruency effect relative to our control conditions. Although we found a main effect of congruency, where participants were less accurate on incongruent than congruent trials, we did not find evidence to support our prediction. A Bayes analysis indicated that more evidence was needed to either reject or accept the null hypothesis. Collecting more data was beyond the scope of this initial study, which served primarily as a pilot for a subsequent combined TMS-functional magnetic resonance

imaging (fMRI) experiment (Chapter 5). Instead, we tested a number of modifications in two further experiments to increase the sensitivity of the behavioural measure to interference from irrelevant features. We used the final design of the main task for the subsequent combined TMS-fMRI study (Chapter 5). The outcome of this set of experiments is a sensitive task for measuring the extent to which participants are able to filter out irrelevant information coming from an attended object.

Introduction

The ability to prioritise task-relevant over task-irrelevant information is important in navigating every day life. We can only encode a limited amount of information in working memory (WM) meaning that we must select some information for maintenance while filtering out other information. Prominent theorists (Desimone & Duncan, 1995; Miller & Cohen, 2001) postulate that this type of control is exerted from higher cortical regions of the brain, such as the dorsolateral prefrontal cortex (dlPFC). This region is thought to exert influence over more specialised processing areas by biasing processing toward task-relevant information and away from distracting, irrelevant information (Desimone & Duncan, 1995; Duncan, 2001). Selection of relevant information can operate over spatial locations (spatial attention), sensory features (feature based attention), objects (object based attention) or at the level of single features of objects ('feature-selective attention', Chen, Hoffmann, Albright, & Thiele, 2012). In this experiment we focus on feature-selective attention.

The prefrontal cortex (PFC) has traditionally been associated with executive control (e.g. Curtis & D'Esposito, 2003; Fuster, 2008; Miller & D'Esposito, 2005; Miller & Cohen, 2001). This is supported by findings from electrophysiological studies showing that lateral PFC neurons respond flexibly to task-relevant information (Cromer, Roy, &

Miller, 2010; Everling, Tinsley, Gaffan, & Duncan, 2002; Freedman, 2001; Kadohisa et al., 2013; Lauwereyns et al., 2001; Rao, Rainer, & Miller, 1997; Roy, Riesenhuber, Poggio, & Miller, 2010; Sakagami & Niki, 1994; Sakagami & Tsutsui, 1999; Watanabe, 1986). In the human brain, regions of the PFC are found to be involved in tasks where filtering of irrelevant information is required (Banich, Milham, Atchley, Cohen, Webb, Wszalek, Kramer, Liang, Wright, et al., 2000; Banich, Milham, Atchley, Cohen, Webb, Wszalek, Kramer, Liang, Barad, et al., 2000; Liu, Banich, Jacobson, & Tanabe, 2004). Neurostimulation data has additionally shown that disruption to the dlPFC only impairs performance in a WM task when irrelevant information is present (Sandrini, Rossini, & Miniussi, 2008). This suggests that prefrontal regions may act to amplify neural representations of task-relevant information and/or to inhibit representations of what is irrelevant (e.g. "dynamic filtering theory" Shimamura, 2000). There is still debate, however, whether amplification, inhibition, or a combination of both mechanisms occur (e.g. Aron, 2007; Kanwisher & Wojciulik, 2000). Regardless of the mechanism, the combination of neuroimaging and neurostimulation techniques have demonstrated causal top-down influences from prefrontal regions to earlier processing areas (e.g. Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Lee & D'Esposito, 2012; Miller, Vytlacil, Fegen, Pradhan, & D'Esposito, 2011; Morishima et al., 2009; Zanto, Rubens, Thangavel, & Gazzaley, 2011). Together the literature indicates a crucial role for the dlPFC in executive function, providing top-down feedback via the selection of task-relevant, and/or inhibition of task-irrelevant, information.

The dlPFC is part of a circuit of frontal and parietal brain regions, often referred to as multiple-demand (MD) regions or the frontoparietal control system, that are consistently shown to play a fundamental role in attentional mechanisms (Cole & Schneider, 2007; Corbetta & Shulman, 2002; Duncan, 2001; Duncan, 2010; Vincent,

Kahn, Snyder, Raichle, & Buckner, 2008). For example, these regions respond to a wide range of task demands (Dosenbach et al., 2006; Duncan & Owen, 2000; Nyberg et al., 2003), even at the level of single participants (Fedorenko, Duncan, & Kanwisher, 2013). The MD regions also adjust their representation of perceptual (Woolgar, Hampshire, Thompson, & Duncan, 2011; Woolgar, Williams, & Rich, 2015) and rule information (Woolgar, Afshar, Williams, & Rich, 2015) in response to changes in task demands.

In contrast to the active role of the MD network, a set of brain regions known as the default mode network (DMN) are usually associated with the opposite response profile than that exhibited by the MD system (Buckner et al., 2008; Raichle & Snyder, 2007). This network incorporates sections of the medial temporal lobe, parts of the medial prefrontal cortex, the posterior cingulate cortex, and the precuneus, and is usually associated with the task negative or resting state (Fedorenko et al., 2013; Fox et al., 2005; Mazoyer et al., 2001; Raichle et al., 2001; Shulman et al., 1997). For instance, Shulman and colleagues conducted a meta-analysis comparing data between active and passive tasks. They found a consistent set of brain regions (DMN) that were more active under passive conditions than goal-directed task conditions. The DMN can be detected using task-free connectivity fMRI (Greicius, Krasnow, Reiss, & Menon, 2003) and is also associated with past thinking and episodic memory processing (Buckner & Carroll, 2007; Buckner et al., 2005). Given the previous literature, it seems likely that regions of the DMN compared to the MD network would perform different roles in a feature-selective attention task.

To test for a casual role in feature-selective attention, we can disrupt processing in a region thought to be critical in controlling attention, and compare this to disruption in a region proposed to play a different role (part of the DMN). We can do this by employing TMS, a neurostimulation technique in which magnetism is used to induce an electric field

in the brain. TMS delivers short pulses that penetrate the skull and disrupt neural processing in a non-invasive, temporary way (Walsh & Cowey, 2000). This "virtual lesion" technique allows us to investigate the causal relationship between neural activity and behavioural performance. The distinctive roles of the MD and the DMN systems indicated by previous literature give rise to different predictions for the effect of TMS stimulation on regions of either network during a task requiring inhibition of the irrelevant feature information.

Here, we developed a paradigm where subjects were required to select the relevant feature of an object whilst ignoring the irrelevant feature information in order to successfully perform the task. We presented stimuli that varied on two dimensions: colour and form. Participants categorised one dimension (e.g., colour) while ignoring the other (e.g., form) in one block of trials, and switched to the other dimension on alternate blocks. The button response for the irrelevant feature was either congruent or incongruent with the required response for the relevant feature. To test the hypothesis that the dlPFC is causally involved in feature-selective attention, we employed TMS whilst participants performed this task in four separate sessions. Neurostimulation studies often include only one control to compare against their main experimental condition (Davis, Gold, Pascual-Leone, & Bracewell, 2013). Here, we employed a more rigorous design by comparing our main experimental condition (stimulation to the dlPFC) to three separate control conditions.

Our main experimental condition was high intensity stimulation (100% motor threshold (MT)) to the right dlPFC. Our first control was low intensity stimulation (40% MT) of the dlPFC, to reflect the control used in concurrent TMS-fMRI paradigms (Chapter 5). This type of control is practical for concurrent TMS-fMRI experiments, as it does not require removing participants from the scanner to reposition the coil, as other control types would. It may also control for any non-specific effects of TMS (e.g,

psychological effects, scalp tactile sensation). However, this control cannot tell us whether the outcome of stimulation reflects a function specific to our region-of-interest. We therefore included two other controls.

Although many studies opt to use the vertex of the skull as a control region for TMS, it often falls between cerebral hemispheres and is therefore only a control for the sensation of stimulation; rather than actual stimulation of brain tissue. Thus, to control for changes due to stimulation of brain tissue in a region not predicted to be critical for the task, as well as the scalp sensation of stimulation, we selected a prefrontal brain region close, but superior and anterior, to our dlPFC target region (Brodmann area 45). This region is part of the DMN network, and provides an interesting control because the DMN often shows the opposite activation response to the MD network (e.g. Fedorenko et al., 2013). Stimulation of a brain region that plays a different role for a particular task should not lead to the same performance deficits as a region that is critical for a particular task. Finally, we also used a sham coil, which is a coil that visually resembles the actual coil and produces a sound during a TMS pulse, but does not produce similar scalp sensations or superficial muscle twitching in the same manner as the active coil. This condition acted as a control for the acoustic artefact of dlPFC stimulation. Additionally, we included no-TMS trials (trials with no TMS stimulation) as a baseline for performance on each session. We predicted that high intensity stimulation of dlPFC would affect performance resulting in increased magnitude of the congruency effect relative to performance in our control conditions.

This study was designed to investigate the role of the dlPFC in feature-selective attention as measured through behaviour and to develop an appropriate task for the following combined TMS-fMRI experiment (Chapter 5). Chapter 5 investigates how disruption to dlPFC function modulates coding of task-relevant and task-irrelevant feature

information in the rest of the frontoparietal network and earlier processing regions. Here we report the methods and results of the TMS behavioural experiment and review the findings. We did not find evidence to support our initial prediction, and a Bayes analysis suggested that this may be due to a lack of power. As this experiment served primarily as a pilot for the subsequent chapter (Chapter 5) we then report two behavioural experiments where we tested a number of modifications to increase the sensitivity of the behavioural measure to interference from irrelevant features.

Experiment 1 (TMS)

Materials and Methods

Participants

There were initially sixteen adult participants, but only ten (7 female; mean age = 25.2 years, SD = 5.09) completed all sessions and are therefore included in this study. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants gave written informed consent, passed relevant safety screening for TMS, and Macquarie University Research Ethics Committee approved the study. All of the participants for the TMS behavioural experiment were recruited from our two previous fMRI experiments (Chapter 2 and Chapter 3), so that the structural MR images could be used to define the stimulation sites. All participants received $120.00.

Stimuli and Task

Stimuli were abstract novel "spiky" and "cuby" objects created using custom MatLab scripts (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). There were 4 objects in the stimulus set: a blue and a green version of each of the spiky and cuby stimuli, obscured by Gaussian noise (Figure 1). We used a PC running the Psychophysics

Toolbox-3 package (Brainard, 1997) in MatLab (Mathworks) to present the stimuli. Objects were presented at central fixation on a screen. Participants performed two separate tasks that alternated in blocks. In the form task, participants discriminated between spiky and cuby objects, ignoring colour. For the colour task, they discriminated between green (RGB 95 171 96) and blue (RGB 95 114 172) objects, ignoring form.
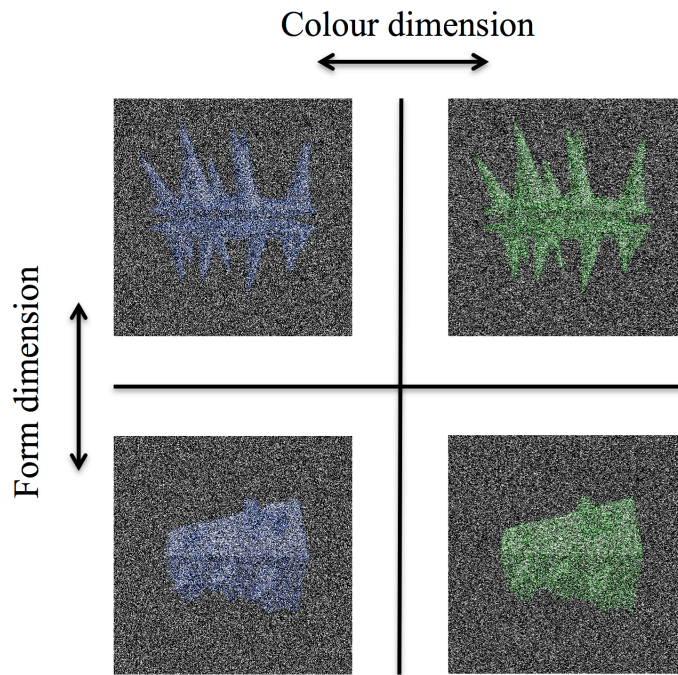
Figure 1: The stimulus set for the feature-selective congruency measure consisted of four novel objects differing on the basis of two feature dimensions: colour (green/blue) and form ("spiky"/"cuby").

The current task context was indicated by a cue at the start of the block (colour or form, Figure 2). Participants responded by pressing the left or right response button ('z' or 'm'; on the keyboard) depending on the colour or form of the stimulus. For example, if participants were performing the colour task and the presented object was a green spiky, participants would press the button associated with 'green'. Importantly, this button response could either be congruent or incongruent with the response for the irrelevant feature in the alternate context. In this example, in the form context the button for "spiky"

could also be left (congruent) or it could be right (incongruent). The mapping of which colour/form was associated with which button response (stimulus-response mapping) was randomised over participants and swapped within participants half way through each session. This stimulus response-mapping switch was included because the experiment served in part as a pilot for the subsequent combined TMS-fMRI experiment (Chapter 5). It is an important aspect of the TMS-fMRI design as it allows separation of the BOLD response associated with each stimulus feature from that associated with each button press.
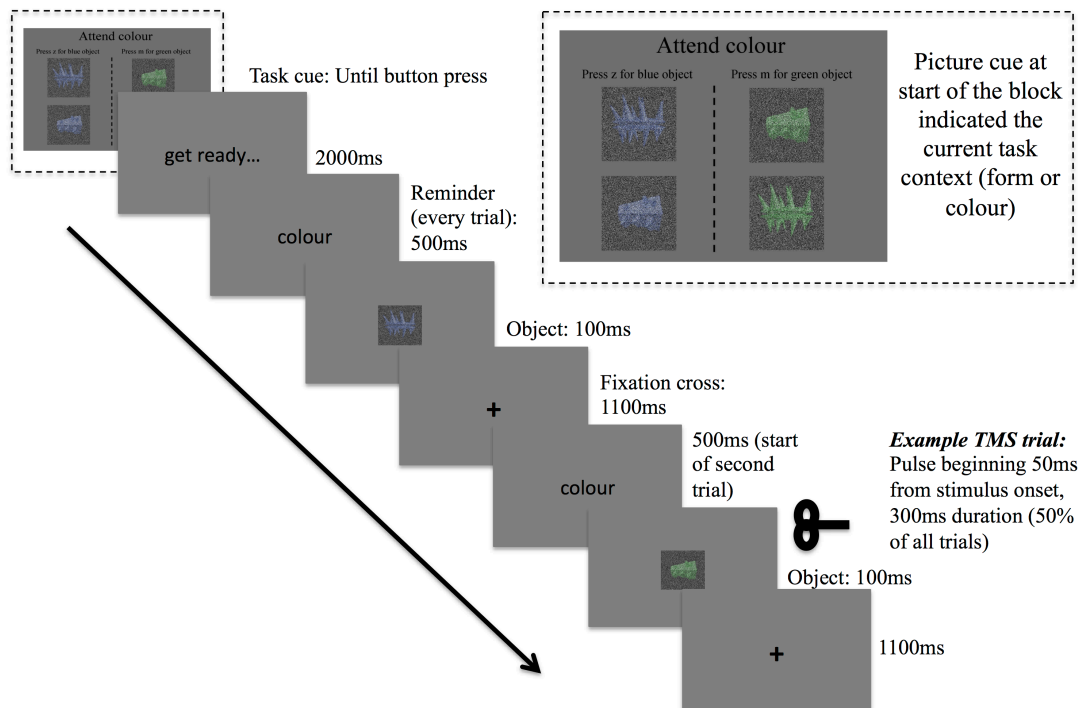
Figure 2: Feature-selective congruency measure: A picture cue at the start of each block indicated the current task context (inset shows cue display for colour task). On each trial, a cue reminded participants of the current task context (500ms) followed by the object to categorise (100ms), followed by the fixation cross (1100ms). On a TMS trial (50% of all trials), participants would receive three pulses of TMS starting 50ms after stimulus onset.

TMS

We used a transcranial magnetic stimulator (Magstim model 200, Magstim, Whitland, UK), with a focal figure-of-eight stimulating coil (90-mm outer diameter) for stimulation. The Matlab (Mathworks) script running the experimental task on a PC remotely triggered the onset of the TMS pulses to occur 50ms after stimulus presentation. Our protocol contained a total of 16 TMS stimulation blocks per session (384 pulses in total per session at 10 Hz), complying with published safety limits for TMS stimulation (Rossi, Hallett, Rossini, & Pascual-Leone, 2009). Each session lasted 1 hour in total including practise and set-up.

Region Selection

Right dlPFC
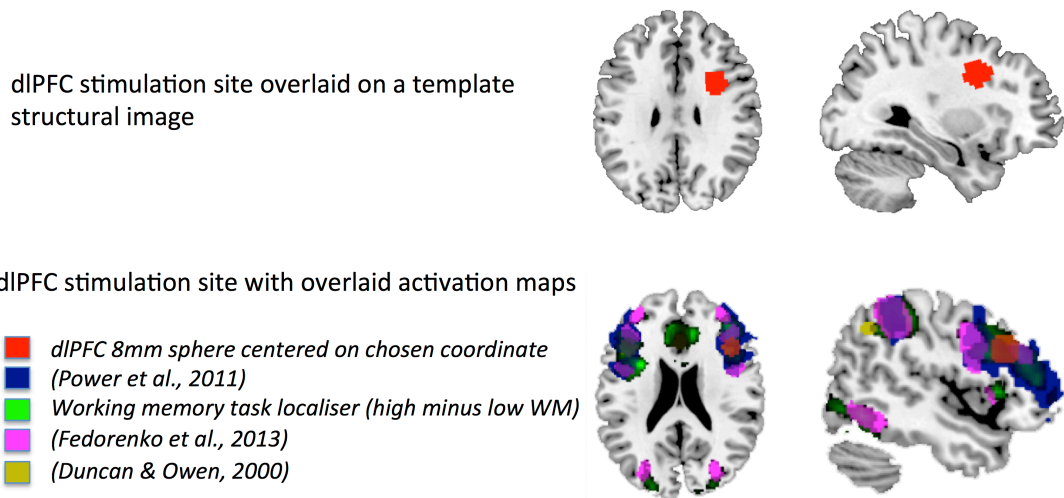
Figure 3: Selected dlPFC region for stimulation: The top panel shows the right dlPFC stimulation site and the bottom panel shows the dlPFC stimulation site with the overlaid activation maps used to guide selection. The coordinates for the right dlPFC stimulation site were transformed into native space for each participant.

The location of our main stimulation site was a region in the right dlPFC. The target region was selected by overlaying activation and resting-state connectivity maps

from previous research that identified the MD network (Duncan & Owen, 2000; Fedorenko et al., 2013; Power et al., 2011) with functional localiser data collected for our participants in a previous scanning session (Figure 3, bottom panel, description of localiser below). We focused on the right dlPFC because TMS to this region has previously been shown to affect activity in visual brain regions during a task requiring selective attention (e.g. Feredoes et al., 2011). Additionally right dlPFC is activated in conditions with a high level of conflict requiring more control compared to conditions with a lower level of conflict (e.g. Egner & Hirsch, 2005). This, combined with the practical reason of the dlPFC being accessible during TMS-fMRI (Chapter 5), made it the ideal stimulation site.

We targeted the MNI152 coordinates [44 31 28] (McConnell Brain Imaging Centre, Montreal Neurological Institute, Figure 3, top panel) as the site of stimulation based as the intersection of the functional and resting-state connectivity maps. We also ensured that the stimulation location was lateral and within the zone identified by Opitz et al. (2016) as most likely to affect the frontoparietal control network as opposed to the DMN. The right dlPFC coordinate was deformed for each participant by applying the inverse of the participant's normalisation parameters. This allowed us to target the site of stimulation in native (i.e., non-normalised) space for each participant.

Control region

The control region we selected is part of the DMN, a network often contrasted against the frontoparietal network as it is found to deactivate in response to demanding cognitive tasks (Buckner et al., 2008). This region was selected by overlaying activation maps from a previous study defining the DMN (Power et al., 2011) and our previous fMRI localiser data (Figure 4, bottom panel). The contrasts used to define the DMN region were the opposite contrasts to that used to determine our dlPFC site (task negative).

Out of the resultant clusters, we selected the region that was superior and anterior to the region selected for dlPFC stimulation. We also ensured that this stimulation location was within the target zone identified by Opitz et al. (2016) as most likely to affect the DMN rather than the frontoparietal network. We selected an 8mm sphere and targeted the central MNI152 coordinates [16 47 46], Figure 4, top panel). The DMN coordinate was deformed for each participant as for the right dlPFC coordinate.

DMN stimulation site overlaid on a template structural image

DMN (task negative) stimulation site with overlaid activation maps

- DMN 8mm sphere centered on chosen coordinate
- (Power et al., 2011)
- Working memory task localiser (low minus high WM)
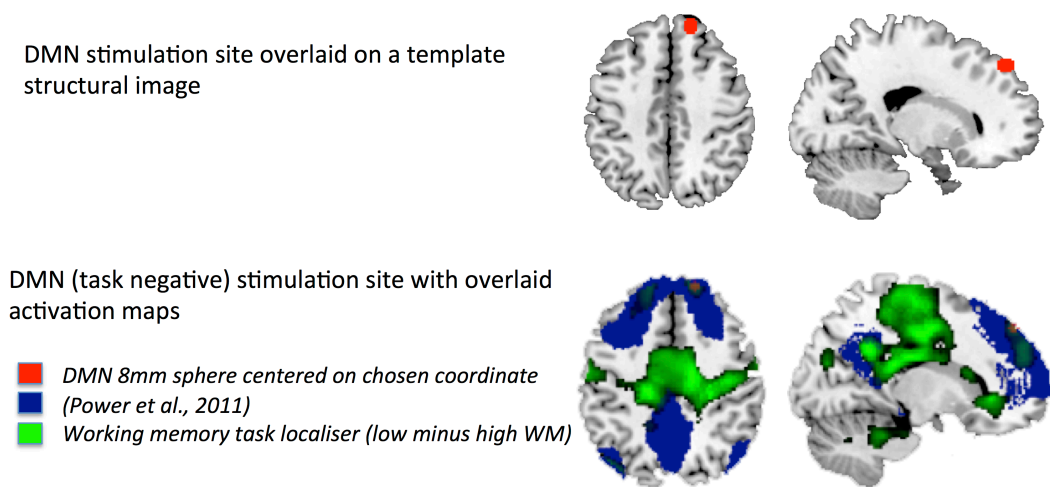
Figure 4: Selected DMN region for stimulation: The top panel shows the DMN stimulation site and the bottom panel shows the DMN stimulation site with the overlaid activation maps used to guide selection. The central coordinate for the DMN stimulation site was transformed into native space for each participant.

Localiser task

We drew on data from an fMRI localiser task that participants completed on a previous occasion (run in conjunction with the experiments outlined in Chapter 2 and 3). Participants performed a spatial WM task designed to identify the frontoparietal regions responsive to increased WM demand (Fedorenko et al., 2013). On each trial, participants were presented with a black fixation cross (500ms), followed by a series of four 3*4 grids

(1000ms) in each of which one (low WM) or two (high WM) squares were blue (Figure 5). Participants had to remember the spatial locations of all the blue squares in the grid. Finally participants were presented with a two-alternative forced choice screen depicting two grids (max of 3750ms). One showed the summation of the blue squares in the preceding grids (correct), and the other showed the same grid with one incorrect square (incorrect). Participants indicated the correct grid by pressing the left or right button, at which point they were immediately shown feedback (a green tick for correct and a red cross for incorrect, 250ms). If the participant responded in under 3750ms, a fixation cross was shown until the next trial began (4000ms after choice screen). Participants performed blocks of low WM, high WM and rest (fixation cross) lasting 16s each (order counterbalanced over participants). The EPI time was 8.42m.
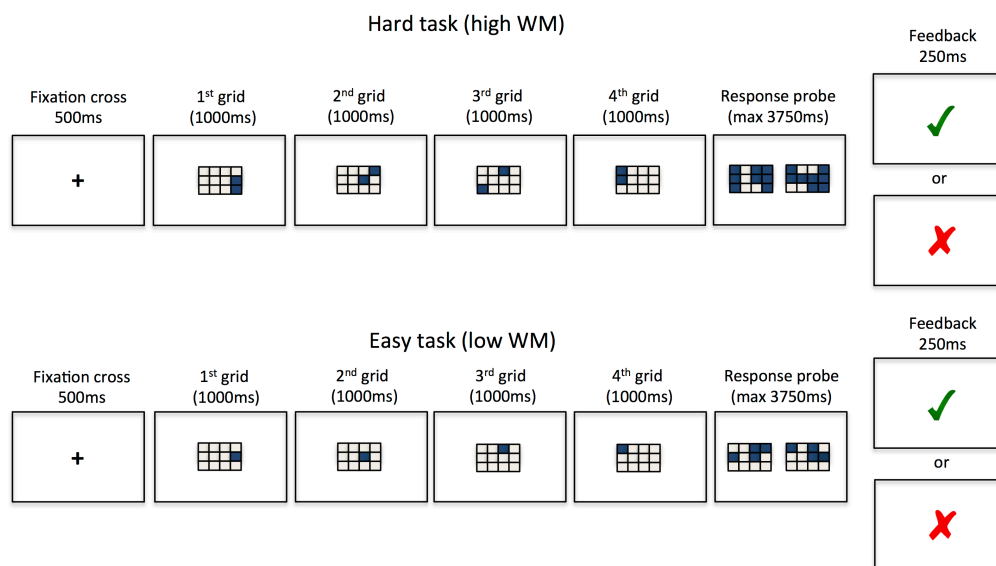
Figure 5: Spatial working memory task used to localise MD and DMN regions: On each trial participants were presented with a fixation cross followed by a series of 4 grids in which one (low WM) or two (high WM) squares were blue. Participants were then presented with a two-alternative forced choice screen depicting two grids and were required to press the left or right button to select their choice. They received feedback at the end of the trial.

MRI data were preprocessed using SPM 5 (Wellcome Department of Imaging Neuroscience, www. fil.ion.ucl.ac.uk/spm) in Matlab 2011b. Functional MRI data were converted from DICOM to NIFTII format, spatially realigned to the first functional scan and slice timing corrected, and structural images were co-registered to the mean EPI. EPIs were smoothed (8 mm FWHM Gaussian kernel) and in all cases the data were high pass filtered (128s). We used the standard multiple regression approach of SPM5 to estimate values pertaining to the high WM and low WM conditions (block design). Blocks were modelled using a box car function lasting 16s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. Mass univariate whole-brain analyses (paired t-tests) compared group-level BOLD response across conditions (high WM minus low WM to define the MD regions and low WM minus high WM to define the DMN regions).

Neuronavigation

We used a neuronavigational system (Visor 2, Berlin, Germany) that uses MR anatomical information and predefined stimulation target coordinates on each individual to allow visual guidance of the coil. The navigational system displays the estimated electric field projected onto the cortical structures based on the current coil position on the participant's head. We registered a head tracker with four reflecting spheres and a coil tracker with six reflecting spheres into the navigational system. The standard Visor 2 routines were used to determine the coil position on the participant's head, which was then marked on the scalp. This assisted efficient re-location of the location following the practice component of the task (start of session and half way through each session). During the main task, the coil was maintained in position over the ROI using the navigational system. In the dlPFC stimulation sessions, the coil was oriented with the handle pointing posteriorly and anteriorly with respect to the participant's head and

roughly parallel to the midline, to increase the overlap with the frontoparietal network as opposed to the DMN (Opitz et al., 2016). During the task, we used a chinrest to maintain the participant's position.

Procedure

Defining motor threshold

At the start of the first session for each participant, we measured his or her individual resting motor threshold. This is the minimal intensity at which a single pulse to the motor cortex reliably produced a visible twitch in the resting abductor pollicis brevis of the hand in 3 out of 5 pulses. The motor threshold offers a means of normalising stimulation intensity across participants, and was used to determine the intensity of TMS stimulation during the main task. The average motor threshold across participants was 58% of maximum stimulator output. The intensity of stimulation in the low intensity session was 40% of the participant's motor threshold, and in the high intensity stimulation sessions was 100% of the participant's motor threshold.

Overall procedure

Participants completed four sessions in total. Each session followed the same protocol apart from the site and intensity of the stimulation. The four sessions were: a) high intensity stimulation (100% of MT) of dlPFC; b) low intensity stimulation (40% of MT) of dlPFC; c) high intensity stimulation (100% of MT) of the control region; and d) sham coil at 100% of MT over dlPFC. We counterbalanced the order of the sessions across participants.

In every session, prior to starting the main experiment, participants performed 6 practice blocks of the task. Each block started with the task cue (4000ms) indicating the current task context (form or colour). On each trial participants would first see a cue reminding them of the current task (form or colour, 500ms) followed by the object

(100ms) and then a black fixation cross (1100ms) (Figure 2). During the first two practice blocks, participants received feedback following each trial as well as feedback (percent correct) at the end of the block. For the last four blocks of practice participants only received feedback at the end of the block.

Following the practice, participants completed 8 blocks of the main task where they only received feedback at the end of the block. In this main task, half of the trials were a TMS stimulation trial (participants received three pulses of stimulation) and half were a no TMS trial (no stimulation) (pseudo-random order). On stimulation trials, participants received three pulses of stimulation (10Hz) beginning 50ms after stimulus onset. The TMS coil was held against their head throughout the stimulation blocks and the stimulation location was maintained using the online neuronavigational system described above.

Following this, participants then completed a further 6 blocks of practice in which they learnt the new stimulus response-mapping (described above). Participants then completed the last 8 blocks of the main task with stimulation on 50% of trials according to the current stimulation condition and the new stimulus-response mapping.

Results

In line with the prominent proposal that the dlPFC plays a critical role in selective attention (e.g. Duncan, 2001; Miller & Cohen, 2001), we predicted that interruption of dlPFC function with TMS would affect performance on our feature-selection task, which requires selection of relevant information and/or inhibition of irrelevant feature information. Therefore, we expected the location of the stimulation, as well as whether the trial was a TMS trial or not, to interact with the congruency effect. The experiment has a complex design, resulting in four factors.

We analysed RT (correct trials only) and accuracy (percent correct) data separately with four-way ANOVAs, with the factors *Session* (High intensity dlPFC stimulation, Low intensity dlPFC stimulation, High intensity control region stimulation, Sham), *Stimulation* (TMS trial, No TMS trial), *Task* (Colour task, Form task), and *Congruency* (Congruent, Incongruent). As a note, the factor *Stimulation* for the sham session included sham and no sham trials instead of TMS and no TMS trials. Greenhouse-Geisser correction for non-sphericity (assessed with Mauchly's test of sphericity) were applied when appropriate.

RT data

RT data are presented in Figure 6. Difference score RT data (incongruent-congruent trials) are presented in Figure 7 to illustrate the magnitude of the congruency effect in each condition. Our main question was whether interrupting dlPFC processing would increase the magnitude of the congruency effect. The four-way ANOVA revealed a main effect of *Task* ($F(1,9) = 7.86$, $p = 0.02$), where participants were faster in the colour task (312ms) than in the form task (327ms). There was also a main effect of *Congruency* ($F(1,9) = 9.21$, $p = 0.02$), significant two-way interactions between *Stimulation* and *Congruency* ($F(1,9) = 6.01$, $p = 0.04$) and *Session* and *Congruency* ($F(3,27) = 3.12$, $p = 0.04$), and a significant three-way interaction between *Stimulation, Session* and *Congruency* ($F(3,27) = 4.91$, $p = 0.01$). There were no other significant main effects or interactions (all ps > 0.15).
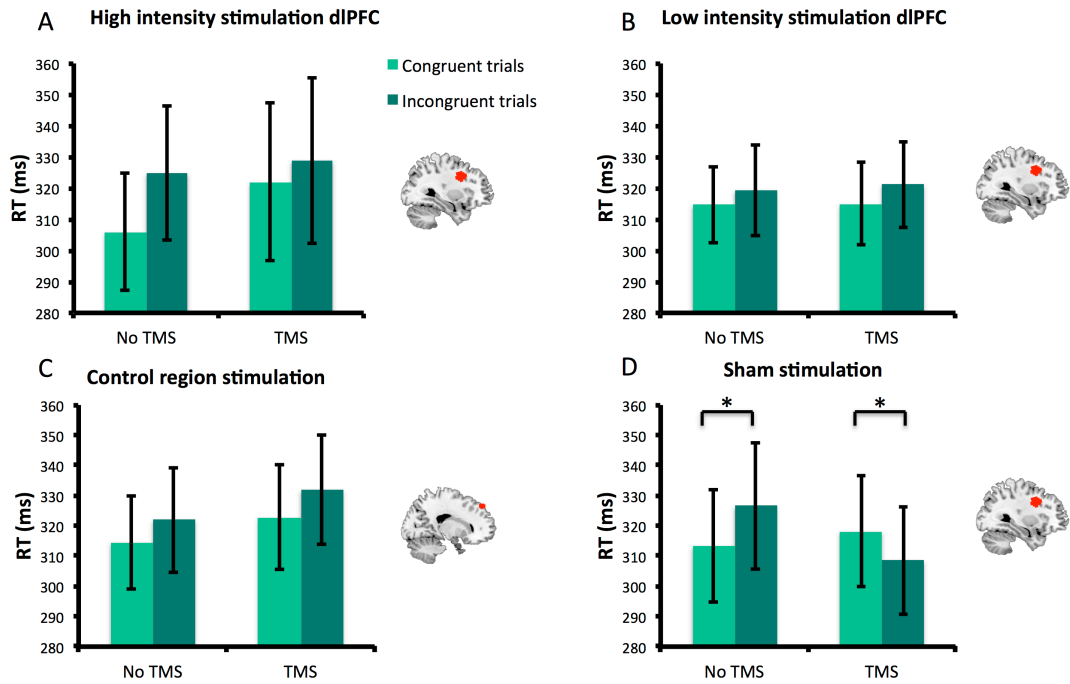
Figure 6: Experiment 1 (TMS). Reaction time data (ms) presented separately by session and stimulation condition. A: High intensity stimulation dlPFC; B: Low intensity stimulation dlPFC; C: Control region stimulation; D: Sham stimulation. Error bars indicate standard error. In both the high intensity dlPFC and control region stimulation, participants were faster in congruent than incongruent trials. Participants were faster in congruent than incongruent trials in no-stimulation sham trials, but were faster in incongruent than congruent trials in stimulation sham trials. Note: y-axis does not start at zero.
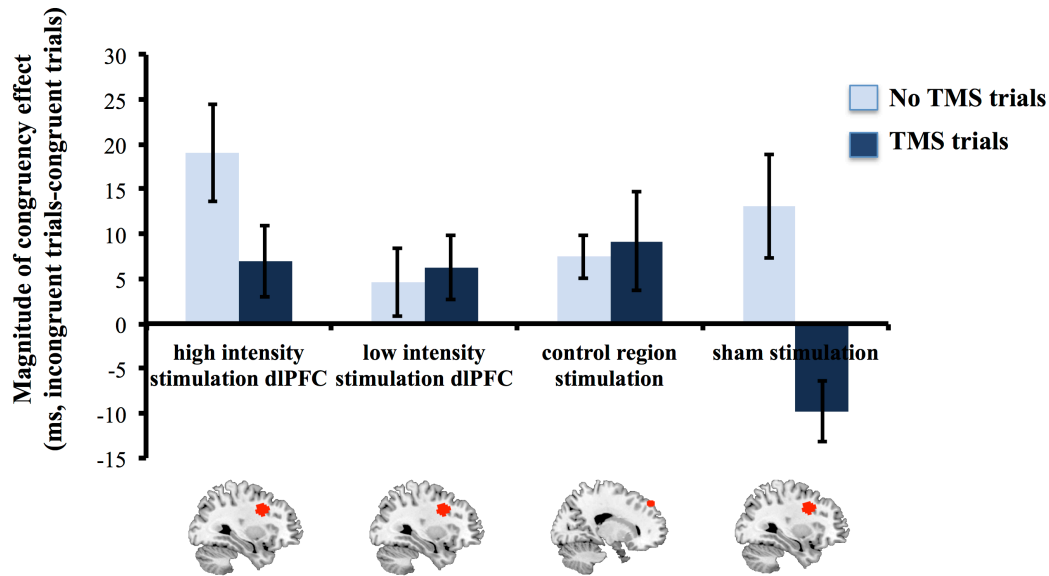
Figure 7: Experiment 1 (TMS). Difference scores for RT data: light bars are trials without TMS stimulation, dark bars are trials with TMS stimulation. The y-axis indicates the magnitude of the congruency effect (incongruent-congruent trials). A higher score indicates a larger congruency effect. The bars are split by session. Error bars indicate standard error.

To explore the interaction between *Stimulation, Session* and *Congruency*, we conducted an ANOVA on the data from each *Session* separately (High intensity dlPFC stimulation, Low intensity dlPFC stimulation, High intensity control region stimulation, Sham). These ANOVAs had factors *Stimulation* and *Congruency*.

If TMS of the dlPFC interferes with selection of the relevant feature and/or filtering out of the irrelevant feature, we should see an interaction between *Congruency* and *Stimulation*, with a larger congruency effect on TMS relative to no-TMS trials. The ANOVA on the high intensity dlPFC stimulation data revealed a main effect of *Congruency* ($F(1,9) = 11.4$, $p < 0.01$), where participants were significantly slower on incongruent trials (327ms) than congruent trials (314ms) and no main effect of *Stimulation* ($F(1,9) = 1.72$, $p = 0.23$). The interaction of *Congruency* and *Stimulation* did not reach significance ($F(1,9) = 4.61$, $p = 0.06$). If we look at Figure 6, panel A, it is clear that the

trend is actually in the opposite direction to our prediction: TMS, if anything, reduced the congruency effect. Therefore, although there was a significant congruency effect overall in this session, we have no evidence that it is larger for TMS relative to no-TMS trials.

For the three control sessions, we predicted only a main effect of *Congruency* (with RTs slowed on incongruent relative to congruent trials) only, as TMS should have no impact. The result for the low intensity dlPFC stimulation session showed that the main effect of *Congruency* was on the cusp of significance (F(1,9) = 4.76, p = 0.052), in the predicted direction. There was no main effect of *Stimulation* (F(1,9) = 0.09, p = 0.77) and no interaction of *Congruency* and *Stimulation* (F(1,9) = 0.09, p = 0.75). The data for the control region (DMN) revealed a main effect of *Congruency* (F(1,9) = 6.01, p < 0.01), where participants were significantly slower on incongruent trials than congruent trials. There was again no main effect of *Stimulation* (F(1,9) = 0.81, p = 0.39) or interaction of *Congruency* and *Stimulation* (F(1,9) = 0.11, p = 0.74). The sham condition showed no main effect of *Congruency* (F(1,9) = 0.22, p = 0.64) or *Stimulation* (F(1,9) = 1.71, p = 0.23), but did show an interaction between *Congruency* and *Stimulation* (F(1,9) = 12.9, p < 0.01). Post-hoc comparisons of congruent and incongruent trials separately for Sham trials and No sham trials revealed the source of this interaction. On No sham trials, we saw the predicted congruency effect: participants were slower on incongruent trials (326ms) than on congruent trials (313ms, t(9) = 2.26, p = 0.04). However, on Sham trials, participants were significantly *faster* on incongruent trials (308ms) than on congruent trials (318ms, t(9) = 2.86, p = 0.02).

Overall, then, we did not see support for our hypothesis from the RT data. The three-way *Session*Congruency*Stimulation* interaction seems to be driven by a larger *Congruency*Stimulation* interaction in the sham session relative to the other sessions.

Accuracy data

Accuracy data (percent correct) are presented in Figure 8. Difference score accuracy data (congruent-incongruent trials) are presented in Figure 9 to illustrate the magnitude of the congruency effect in each condition. As for reaction time data, we predicted a 3-way *Session*Stimulation*Congruency* interaction, in which accuracy would be higher for congruent relative to incongruent trials, and this congruency effect would be larger for TMS relative to no TMS trials in the high dlPFC session only.
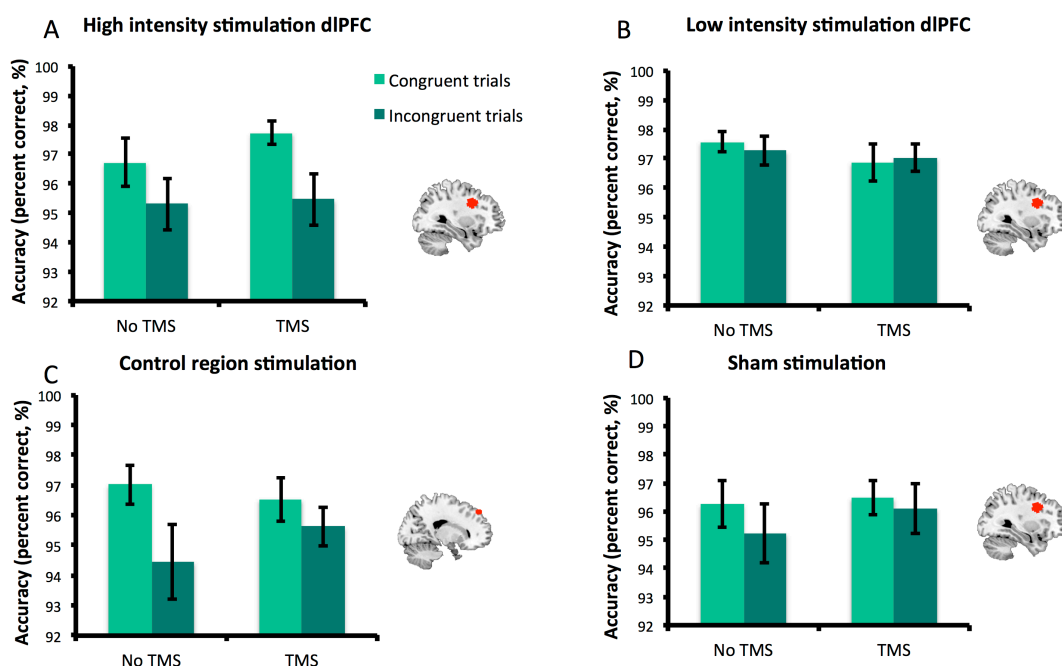
Figure 8: Experiment 1 (TMS). Accuracy data (percent correct %): The data are presented separately by session and stimulation condition condition. A: High intensity stimulation dlPFC; B: Low intensity stimulation dlPFC; C: Control region stimulation; D: Sham stimulation). Error bars indicate standard error. There was a significant main effect of congruency where participants were more accurate on congruent than incongruent trials. Note: y-axis does not start at zero.

The ANOVA revealed a main effect of *Congruency* (F(1,9) = 8.12, p = 0.02) where participants were more accurate on congruent trials compared to incongruent trials. There was no main effect of *Session* (F(3,27) = 1.61, p = 0.21), *Stimulation* (F(1,9) = 0.34, p = 0.55), or *Task* (F(1,9) = 3.83, p = 0.09). Visual inspection of the data (Figure 8, Figure

9) suggested a trend for a larger congruency effect in the high stimulation dlPFC and control region sessions, but there were no significant two-way, three-way and no four-way interactions (all ps> 0.14). Thus, although we found the predicted effect of congruency overall, we have no evidence from the accuracy data to suggest that TMS impacted the magnitude of the congruency effect.
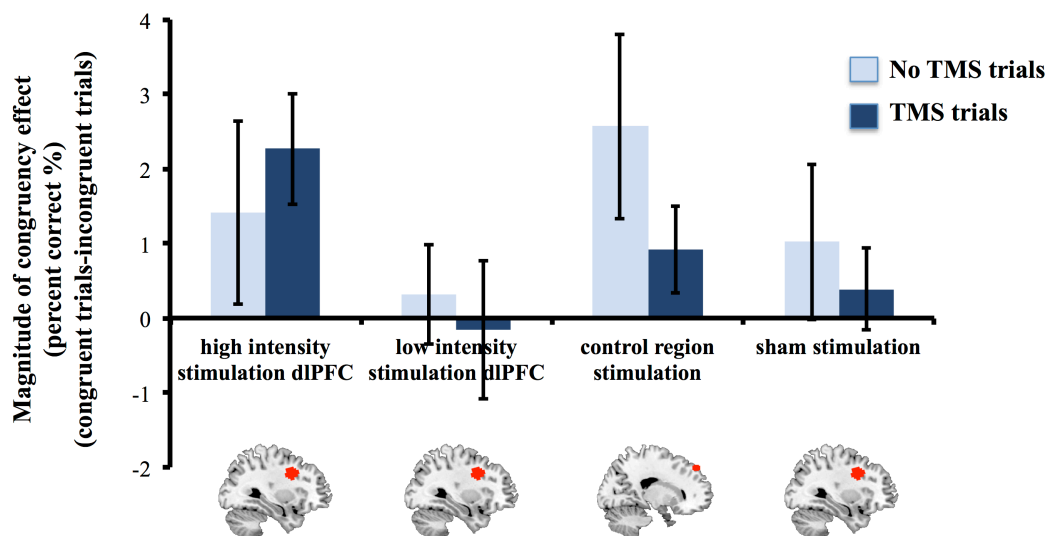
Figure 9: Experiment 1 (TMS). Difference scores for accuracy data: The dark-coloured bars are TMS stimulation trials and the lighter coloured bars are trials without TMS stimulation. The y-axis indicates the magnitude of the congruency effect (congruent-incongruent trials). The bars are split by session. Error bars indicate standard error.

Bayes Analysis

The accuracy data clearly show a congruency effect, but do not show a significant effect of stimulation site or intensity. If this could be interpreted as *evidence for no effect* (i.e., support for the null hypothesis), this would be evidence against the dlPFC being involved in selection of relevant features and/or filtering irrelevant information (assuming one accepts that we have a valid measure of this with our paradigm). Alternatively, it is

possible that our lack of effect reflect a lack of sufficient power or sensitivity. We therefore entered the accuracy difference scores (congruent-incongruent trials) into a Bayes analysis to differentiate these alternatives. We tested evidence for the null hypothesis that *Session* and *Stimulation* have no effect (BF > than 3: strong evidence for experimental hypothesis; BF < 1/3: strong evidence for the null hypothesis; BF 1/3-3: the experiment is not sensitive enough; (Dienes, 2011; Love et al., 2015)). The BF_{10} was 0.91, suggesting we did not have enough enough evidence to distinguish between our hypothesis and the null, although it is moving in the direction of support for the null hypothesis. Unfortunately, collecting more data was beyond the scope of this thesis.

Discussion

The right dlPFC is part of a circuit of frontal and parietal brain regions, often referred to as multiple-demand (MD) regions or the frontoparietal attention network. This network is consistently shown to play a fundamental role in cognitive control (Cole & Schneider, 2007; Duncan, 2001; Duncan, 2010; Vincent et al., 2008). Here we examined whether disruption to the dlPFC influences the magnitude of a congruency effect in a task requiring selection of relevant and/or inhibition of irrelevant feature information. Both RT and accuracy data showed a congruency effect (poorer performance on incongruent relative to congruent trials), but we did not find the predicted interactions with TMS stimulation and stimulation site. A Bayes analysis on accuracy data suggests this may be due to a lack of power; we return to this below.

One of the features of this experiment was the inclusion of three different types of control. The inclusion of a control site accounts for general effects of stimulation on the brain. Additionally, if this region is proposed to be involved in different processes compared to the main stimulation site, it can provide an effective comparison for the involvement in the cognitive process under investigation. Here, we found comparable

results in our main site of stimulation and our control site. The RT data revealed a congruency effect in both the high stimulation dlPFC stimulation session and the control region stimulation session, but this was not influenced by whether participants received stimulation or not on a particular trial. One explanation for this is that the high intensity TMS on stimulation trials may have caused a flow-over effect onto other trials within the same session. Alternatively, it may be because these regions do not act differently in respect to filtering irrelevant feature information, or it could be due to a lack of sensitivity of our behavioural measure. A recent study showed that very slight changes in coil orientation and stimulation zone of dlPFC can alter the network that is affected (Opitz et al., 2016). Although we took precautions in choosing our two stimulation coordinates (DMN and frontoparietal network) as within the correct region and coil orientiation-invariant zone for each network respectively (Opitz et al., 2016), it is still possible that we missed the target network. Another possibility is that both the frontoparietal and DMN support feature-selective attention but under different mechanisms. For example, it may be the case that deactivation of DMN is essential to performing a difficult task. However, neither of these regions showed a larger congruency effect than the other control regions, making this explanation unlikely.

Our second control, low intensity TMS, is an ideal control for concurrent TMS-fMRI experiments as it allows TMS conditions (high and low) to be presented on a trial-by-trial basis. If it is to be considered an 'ineffective' level of stimulation, it would be reassuring to see that the low-TMS does not modulate congruency effects. However, in this experiment, the effect of low intensity TMS on RT was difficult to interpret, because there was no congruency effect in either TMS or no TMS trials. Finally, in the sham condition, we observed a surprising effect where no "stimulation" trials had a standard congruency effect whilst sham "stimulation" showed the opposite (participants were faster

in incongruent trials than congruent trials). The only difference between stimulation and no stimulation trials in the sham session is a sound that mimics the acoustic artefact of actual stimulation. It is therefore difficult to explain this result. However, if this is a general effect of the noise associated with TMS, it at least acts against our prediction that TMS will increase the congruency effect (making our test more conservative). Overall, the use of three different controls raises important questions about the way in which we interpretat active site data when we compare it to different choices of control, but the data were not able to adjudicate between them.

In the accuracy data, we found clear evidence for a difference in performance on incongruent and congruent trials, indicating that our paradigm can track the influence of irrelevant feature information on performance. However, there was no evidence that the magnitude of the congruency effect (in accuracy) was affected by TMS, regardless of stimulation type. These results are surprising considering the evidence supporting a role for the dlPFC in filtering irrelevant information, but as our Bayes analyses suggest, presumably reflect a lack of power. The right dlPFC is part of a network of frontoparietal regions and previous work studying patients with frontal lobe injury has also shown compensatory over-activation of this network when part of the network is damaged (Woolgar, Bor, & Duncan, 2013). Although it is beyond the scope of this experiment to assess, it may be the case that disruption to dlPFC was compensated for by the rest of the MD network. The combination of TMS with other neuroimaging techniques may provide an approach to examine this and enhanced understanding of the mechanisms of action of TMS. In addition, TMS stimulation did not affect participants' accuracy in the task overall. This may be because participants' performance was close to ceiling (~95%) and it has been suggested that TMS effects on accuracy at this level are difficult to observe (Manenti, Cotelli, Calabria, Maioli, & Miniussi, 2010). Thus, our results form only a

preliminary step in testing whether this region is involved in filtering irrelevant feature information.

This experiment was carried out in part as a pilot for the subsequent TMS-fMRI experiment where we tested whether disruption to this region affected brain responses in the rest of the MD network and in more specialised processing areas (Chapter 5). Given the clear need for a robust behavioural measure, we next made modifications to the paradigm with the goal of increasing sensitivity. These behavioural experiments are reported below.

Experiment 2

Introduction

There were several possible explanations for the lack of a reliable effect of stimulation in Experiment 1. First, accuracy was close to ceiling (~95%) and TMS effects on accuracy at this level may be difficult to observe (Manenti et al., 2010). Second, it was possible that the task was not difficult enough to recruit the MD regions. These areas are characterised by their response to a wide range of task demands, and these are typically all challenging tasks (Dosenbach et al., 2006; Duncan & Owen, 2000; Nyberg et al., 2003). Indeed, the MD regions have been shown to adjust their representation of perceptual information in response to task demand, such that in easier conditions they do not hold relevant information (Woolgar et al., 2011; Woolgar, Williams, et al., 2015). Therefore, in the following experiment, we increased the difficulty of the task.

Materials and Methods

The design of the task in Experiment 2 is identical to Experiment 1 apart from the modifications explained below.

Participants

Eight healthy adult volunteers (5 females; mean age = 22.8 years, SD = 4.08) participated in this experiment, which was conducted at the University of Reading, UK. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants gave written informed consent, and Reading University Research Ethics Committee approved the study. All participants received 2 course credits for participating.

Stimuli
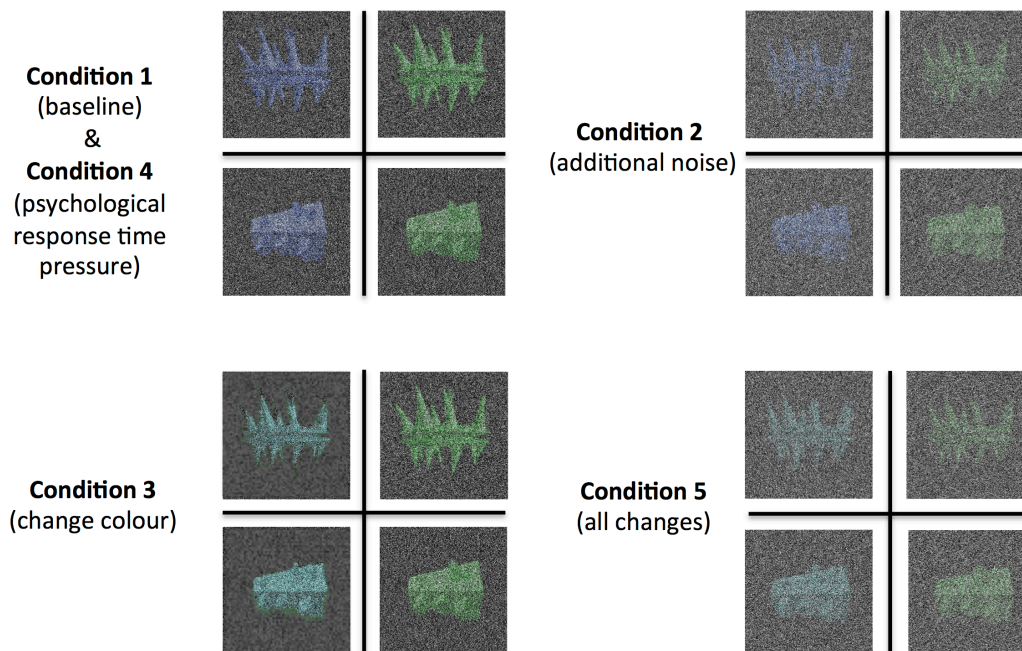
Figure 10: Experiment 2. We used three additional versions (Condition 2, 3 and 5) of the stimulus set used in Experiment 1 (TMS). The stimulus set for Conditions 1 and 4 were identical to the stimulus set used in Experiment 1.

There were four different stimulus sets each consisting of 4 objects (Figure 10). For Condition 1 (baseline) and 4 (psychological response time pressure, see procedure below), the objects consisted of the same set as used in Experiment 1 (TMS). For Condition 2 (additional noise), we used Adobe Photoshop CC 2015 to edit the stimuli and add another layer of Gaussian noise, with the aim of increasing the perceptual difficulty of

the task; otherwise the stimuli were identical to Condition 1. For Condition 3 (change colour), we used Adobe Photoshop CC 2015 to edit the colour of the stimuli, with the aim of increasing the difficulty of distinguishing between the colour of the objects (blue object RGB values: 98, 179, 180), otherwise the stimuli were identical to Condition 1. For Condition 5 (all changes), we edited the stimuli to add another layer of noise to match Condition 2 *and* changed the RGB values of the blue objects to match Condition 3.

Procedure

Participants completed the five different versions (conditions) of the task in a single testing session. Participants completed four blocks (two blocks of each task: colour and form) of each condition (20 blocks total). There were 32 trials per block.

For Condition 1 (baseline), 2 (additional noise), and 3 (change colour), the task that participants completed was identical to Experiment 1 (TMS). For Condition 4 (psychological response time pressure) we used the stimulus set from Condition 1 (baseline). However, in this condition, participants were informed that their responses would only be recorded within 500ms of the object presentation. This modification to the task was added to increase the perceived difficulty, encouraging participants to respond as fast as possible following stimulus presentation. In this condition, participants were encouraged to respond during the white cross period following object presentation, which would change to black after 500ms. If participants did not respond within 500ms, their response would be marked as incorrect and reflected in their feedback. Responses following this 500ms time were still recorded for data analysis. 500ms was chosen because it was long enough to be achievable (based on the results of Experiment 1) but short enough to apply psychological time pressure. For Condition 5 (all changes), all three modifications were added from Condition 2, 3 and 4. The order in which participants performed conditions 1, 2 and 3 was counterbalanced across participants whilst Condition

4 was always presented fourth and Condition 5 was always presented last. The last two conditions were not counterbalanced because the addition of the response time limit would have affected performance on subsequent conditions.

At the start of the session, participants completed 2 practice blocks of the task in the first condition. For example, if the first condition for this participant was the baseline condition, then the practice blocks would consist of baseline stimuli. During the practice blocks, participants received feedback following each trial as well as feedback (percent correct) at the end of the block. For the main task participants did not receive feedback following each trial but received feedback at the end of the block. The task for Conditions 1, 2, and 3 was identical to Experiment 1 (TMS). For Conditions 4 and 5, the task was the same except for the addition of the psychological response time pressure.

Results

RT and accuracy data were submitted to separate three-way ANOVAs with the factors *Condition* (Baseline, Colour change, Additional noise, Psychological response time pressure, All change), *Task* (Colour, Form) and *Congruency* (Congruent, Incongruent). Here, we were interested in examining whether the modifications to the paradigm (Conditions 1-5) increased the difficulty of the task. Greenhouse-Geisser correction for non-sphericity (assessed with Mauchly's test of sphericity) were applied when appropriate.

RT data

RT data are presented in Figure 11. A three-way ANOVA with factors *Condition* (Baseline, Colour change, Additional noise, Psychological response time pressure, All change), *Task* (Colour, Form) and *Congruency* (Congruent, Incongruent) revealed a main effect of *Condition* ($F(4,28) = 5.62$, $p < 0.01$), no main effect of *Congruency* ($F(1,7) = 0.17$, $p = 0.71$) and no main effect of *Task* ($F(1,7) = 1.2$, $p = 0.32$). There were no

significant two-way interactions (all ps > 0.08). However, there was a significant

interaction between *Condition, Task* and *Congruency* (F(4,28) = 3.31, p = 0.04).

Figure 11: Experiment 2. RT data: The two graphs are separated by task (Panel A: Colour, Panel B: Form). Within each graph the data are separated by condition. For each condition the light bars are the congruent trials and the dark bars are the incongruent trials. Error bars indicate standard error. Post-hoc ANOVAs following a three-way interaction revealed a main effect of condition in the colour task (response time condition was faster than the baseline, change colour, and all change, conditions). The form task showed faster RTs in the response time condition than in the all change and additional noise condition. The all change condition and change colour conditions also showed faster RTs than the additional noise condition.

To explore the three-way interaction, we conducted a simple main effects analysis. We split the data by *Task* as we were interested in whether we had a larger main effect of *Congruency* in certain *Conditions*. Additionally for the purpose of this experiment, we were interested in seeing if there was a main effect of condition, to assess whether the conditions differed in difficulty. The two ANOVAs therefore had the factors of *Congruency* and *Condition.* The colour task ANOVA showed a main effect of *Condition* (F(4,28) = 3.36, p = 0.02), no main effect of *Congruency* (F(1,7) = 1.81, p = 0.22) and no

interaction between *Condition and Congruency* (F(4,28) = 1.62, p = 0.19). Pairwise comparisons on the *Condition* main effect showed that RTs on the *psychological response time pressure* condition (307ms), were faster than the *baseline* (400ms, p = 0.03), *change colour* (366ms, p = 0.04), and the *all change* (340ms, p = 0.04), conditions. There were no further significant differences between the conditions (p > 0.07 all). Therefore, the *psychological response time pressure* condition resulted in faster RTs for the colour task, but there was no evidence for a congruency effect on RT overall or a difference in congruency effect between tasks.

The form task ANOVA showed a main effect of *Condition* (F(4,28) = 8.73, p < 0.01). The main effect of *Congruency* did not quite reach significance (F(1,7) = 5.12, p = 0.06) and there was no interaction between *Congruency* and *Condition* (F(4,28) = 1.31, p = 0.28). Pairwise comparisons revealed that RTs were significantly faster for the *psychological response time pressure* condition (332ms) than in the *all change* condition (353ms, p < 0.01), which in turn had significantly faster RTs than to the *additional noise* condition (413ms, p < 0.001). The *psychological response time pressure* condition and the *change colour* condition (340ms) also had significantly faster RTs than the *additional noise* condition (ps < 0.001). There were no further significant differences between the conditions (p > 0.08 all). Therefore, the *response time* condition resulted in faster RTs than some of the other conditions, but there was no evidence for a difference in congruency effect.

Accuracy data

Accuracy data are presented in Figure 12. Visual inspection suggested a larger congruency effect in the *all change* condition. A three-way ANOVA with factors *Condition* (Baseline, Colour change, Additional noise, Psychological response time pressure, All change), *Task* (Colour, Form) and *Congruency* (Congruent, Incongruent)

revealed a main effect of *Condition* (F(4,28) = 5.72, p < 0.01). The ANOVA also showed a main effect of *Congruency* (F(1,7) = 13.6, p < 0.01) modulated by a significant interaction between *Task* and *Congruency* (F(1,7) = 11.1, p = 0.01). There was no main effect of *Task* (F(1,7) = 1.45, p = 2.72) and no other significant interactions (all ps > 0.11).

Pairwise comparisons following up the main effect of *Condition* (which did not interact with any other factors) revealed that the *all change* condition had significantly lower accuracy scores (85.5%) than the *baseline* condition (92.2%, p < 0.01), *additional noise* condition (92.3%, p = 0.02) and the *colour change* condition (93.4%, p < 0.01). There were no further significant differences between the conditions (p > 0.07 all).
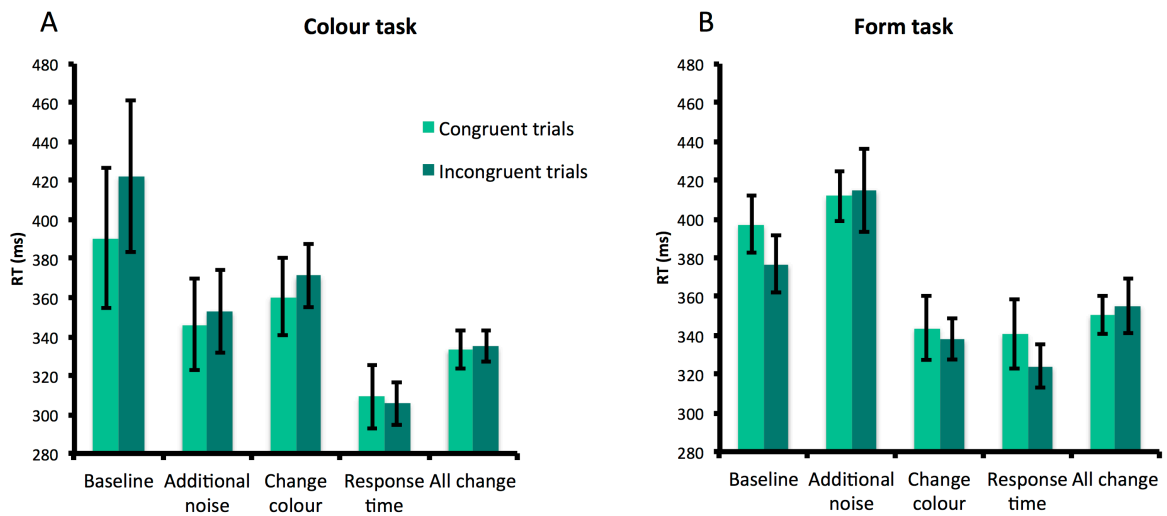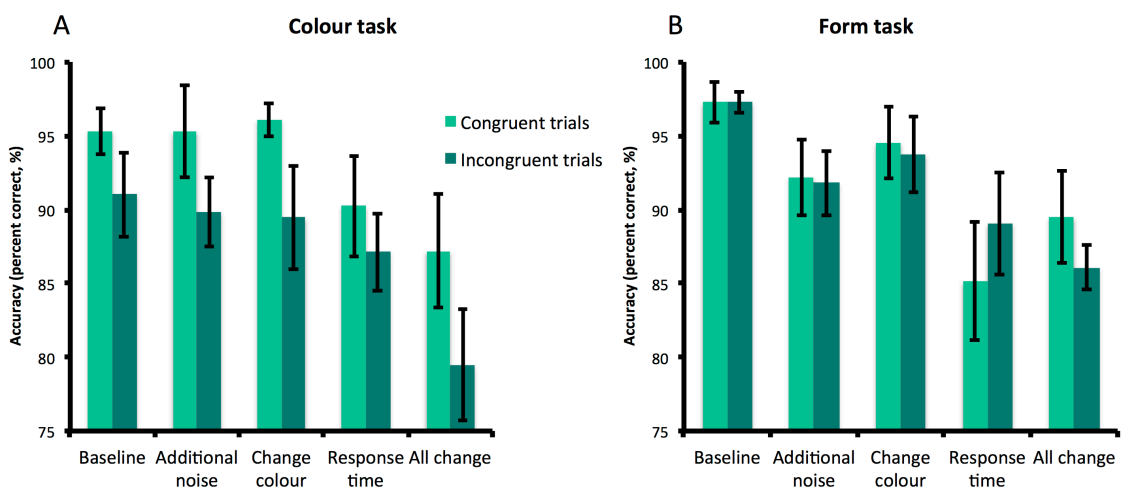

Figure 12: Experiment 2: Accuracy data: The two graphs are separated by task (Panel A: Colour, Panel B: Form). Within each graph the data are separated by condition. For each condition the light bars are the congruent trials and the dark bars are the incongruent trials. Error bars indicate standard error. The data showed a main effect of condition where the change all condition had lower accuracy scores than the baseline, additional noise and change colour conditions. An interaction between Task and Congruency showed that the Congruency effect was larger for the colour task. Participants had significantly lower accuracy scores in incongruent trials compared to congruent trials for the colour task but there was no effect of congruency in the form task. Note: y-axis does not start at zero.

To explore the interaction between *Task* and *Congruency,* we conducted *post hoc* paired t-tests for each task (colour and form) separately. In the colour task, participants had significantly lower accuracy scores (87.4%) on the incongruent trials than on the congruent trials (92.8%, t(7) = -4.39, p < 0.01). In the form task, there was no difference in accuracy scores between incongruent (91.6%) and congruent trials (91.7%, t(7) = -0.15, p = 0.89). Therefore, there was an effect of congruency, but only in the colour task, and this was not modulated by the different conditions. Although performance on the *all change* condition was poorer overall, the trend towards a larger congruency effect in this condition was not enough to drive a significant two or three way interaction.

Discussion

The purpose of this experiment was to increase the difficulty of the task and to increase the robustness of the congruency effect. The accuracy data indicated that the condition where all aspects of the task were changed (all change) was the most difficult, and relative to the other condition with psychological response time pressure (response time condition), RTs were also slower on this condition. This condition also showed the largest numerical congruency effect in accuracy. However, the statistical analysis showed that the congruency effect was specific to the colour task.

The purpose of this paradigm was to utilise it for a combined TMS-fMRI project. If the congruency effect only occurs in one task context, this would result in only half of the data from our TMS-fMRI paradigm being used for further analysis. We therefore ran a further experiment to match the congruency effect between the two task contexts.

Experiment 3

Introduction

In Experiment 2, we found evidence for a congruency effect in the colour task but not in the form task indicating that it was easier to ignore the irrelevant dimension in the form task than the colour task. Additionally, participants found the condition where we made the task both perceptually more challenging in the colour domain, degraded the stimuli, and introduced a psychological response time pressure the most difficult, and this condition had the largest numerical congruency effect. We therefore changed the object forms to increase the difficulty of the form discrimination, reasoning that by making it harder to process form; we would increase our sensitivity to any interference from colour. Woolgar et al. (2015) found that participants were faster to respond to "spikies" than they were to "cubies" and "smoothies", suggesting that spikies may be more distinct that the other objects. Therefore we altered the stimulus set so that participants would need to distinguish between "cubies" and "smoothies" (rather than cubies and spikies) in the form task. We conducted this experiment with the original colours from the TMS paradigm (Exp 3: Condition 1) and the new more perceptually challenging colours used in conditions 3 and 5 of Experiment 2 (Exp 3: Condition 2). We included Condition 1 with the original colours to test whether the colour dimension interfered more when the colours were perceptually distinct.

Materials and Methods

The design of the task here is identical to Experiment 1 and 2 apart from a few modifications as explained below.

Participants

A group of seven healthy adult volunteers (6 females; mean age = 21.5 years, SD = 4.56) participated in this experiment. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants gave written informed consent, and Reading University Research Ethics Committee approved the study. All participants received 2 course credits for participation.

Stimuli

Condition 1 (more distinct colours)

Condition 2 (less distinct colours)
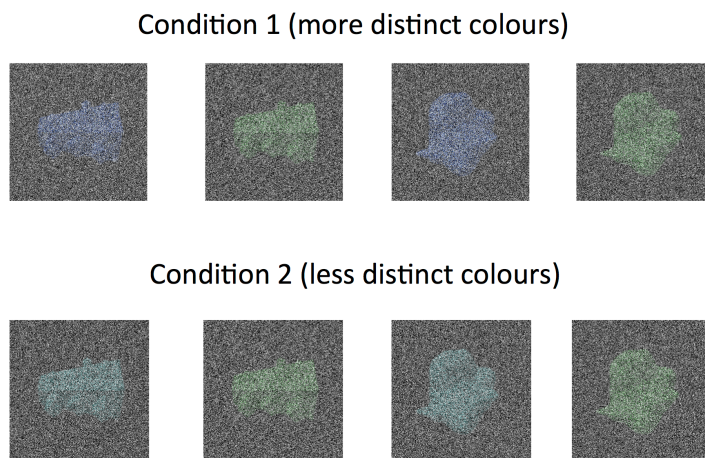
Figure 13: Stimulus sets used for Experiment 3. This stimulus set consisted of "cuby" and "smoothy" objects. Condition 1 stimuli consisted of the same colours as used in Experiment 1 and Experiment 2 with additional noise (Experiment 2: Condition 2). Condition 2 stimuli consisted of the less perceptually distinct colours from Experiment 2 with additional noise (Condition 5).

The stimulus set was similar to Experiment 1 and 2. However, instead of using "spiky" and "cuby" objects, we used "cuby" and "smoothy" objects (Figure 13) (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). The form task required participants to discriminate between cuby and smoothy objects. For Condition 1 of this experiment, we used the stimuli from Experiment 2, Condition 2 (*additional noise)*, in which the colour of the stimuli matched the colour of the original stimuli from Experiment 1 but had an extra layer of Gaussian noise (Figure 13, top panel). For Condition 2 in this experiment we used

the stimuli from Experiment 2, Condition 5 (*all changes*) which also had the additional

layer of noise and the colours were less distinct (Figure 13, bottom panel).
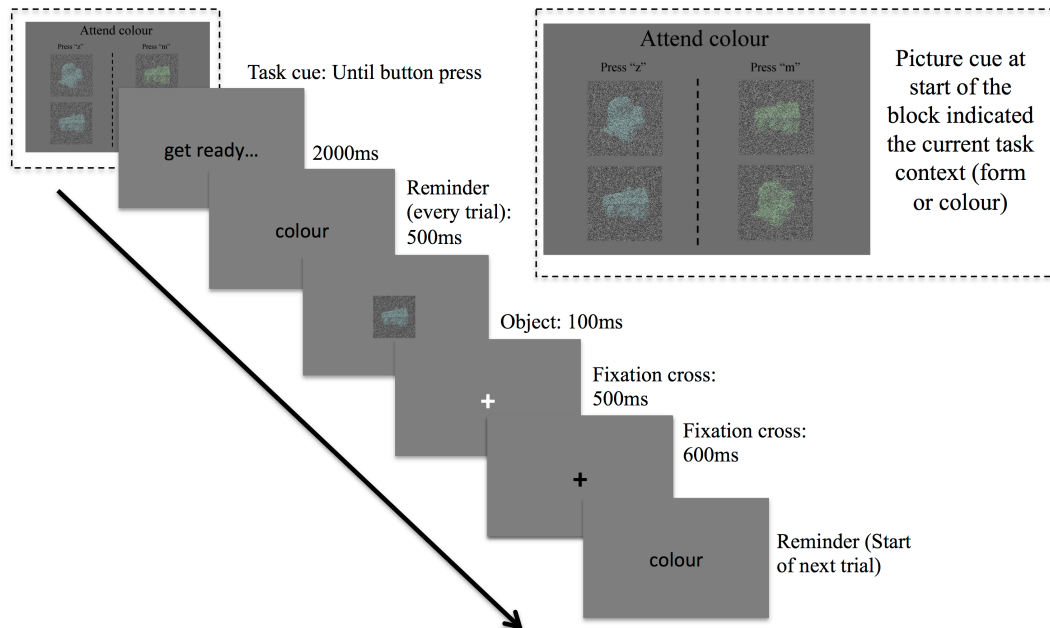
Figure 14: Feature-selective congruency measure in Experiment 3. A picture cue at the start of each block indicated the current task context (inset shows cue display for colour task for Condition 2 of this experiment). On each trial a cue reminded participants of the current task context (500ms) followed by the object to categorise (100ms) followed by the white fixation cross (500ms) and the black fixation cross (600ms). The white fixation cross induced psychological time pressure on responses.

Procedure

The task was identical to the task of Conditions 4 (*response time*) and 5 (*all change*) of Experiment 2 (Figure 14), in which participants were told they had only 500ms to make their responses. The two versions of the task were identical apart from the stimulus set. The order that participants completed the two conditions (Condition 1: more distinct colours, Condition 2: less distinct colours) was counterbalanced across participants. Participants completed two blocks of practice trials that used the stimulus set from the first condition they would carry out. Following this they completed 10 blocks of each condition (20 blocks total).

Results

The goal of Experiment 3 was to develop a task in which the congruency effect was not dependent on the task context (form or colour). As we were not interested in whether there was a difference between Condition 1 and Condition 2, but only whether there was a usable congruency effect in either (or both) Conditions, we ran separate two-way ANOVAs for Condition 1 and 2 including factors *Task* (Colour, Form) and *Congruency* (Incongruent, Congruent). We did this for both RT and accuracy data separately.

RT data

The results for Condition 1 (more distinct colours) are shown in Figure 15A. A two-way ANOVA factors *Task* (Colour, Form) and *Congruency* (Incongruent, Congruent) showed no main effect of *Congruency* ($F(1,6) = 1.71$, $p = 0.24$) or *Task* ($F(1,6) = 0.34$, $p = 0.58$) and no significant interaction between *Congruency* and *Task* ($F(1,6) = 1.32$, $p = 0.29$).
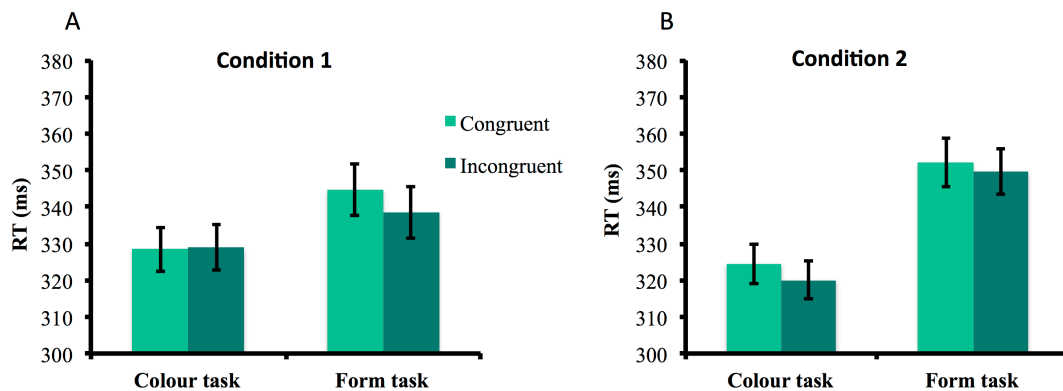
Figure 15: Experiment 3. Reaction time data: The data here are separated by condition and task (Panel A: Condition 1, Panel B: Condition 2). Error bars indicate standard error. Condition 1: more distinct colours; Condition 2: less distinct colours. The RT data showed no differences between any of the conditions. Note: y-axis does not start at zero

182

The results for Condition 2 (less distinct colours) are shown in Figure 15B. The two-way ANOVA again showed no main effect of *Congruency* ($F_{(1,6)} = 1.21$, $p = 0.35$) or *Task* ($F_{(1,6)} = 2.99$, $p = 0.14$) and no significant interaction between *Congruency* and *Task* ($F_{(1,6)} = 0.11$, $p = 0.76$). Therefore, there were again no effects of Congruency on RT in this Experiment.

Accuracy data

The accuracy data for Condition 1 are shown in Figure 16A. As for RT, we conducted a two-way ANOVA with factors *Task* (Colour, Form) and *Congruency* (Incongruent, Congruent). This ANOVA showed no main effect of *Congruency* ($F_{(1,6)} = 2.47$, $p = 0.17$) or *Task* ($F_{(1,6)} = 1.34$, $p = 0.29$) and no significant interaction between *Congruency* and *Task* ($F_{(1,6)} = 0.79$, $p = 0.41$).
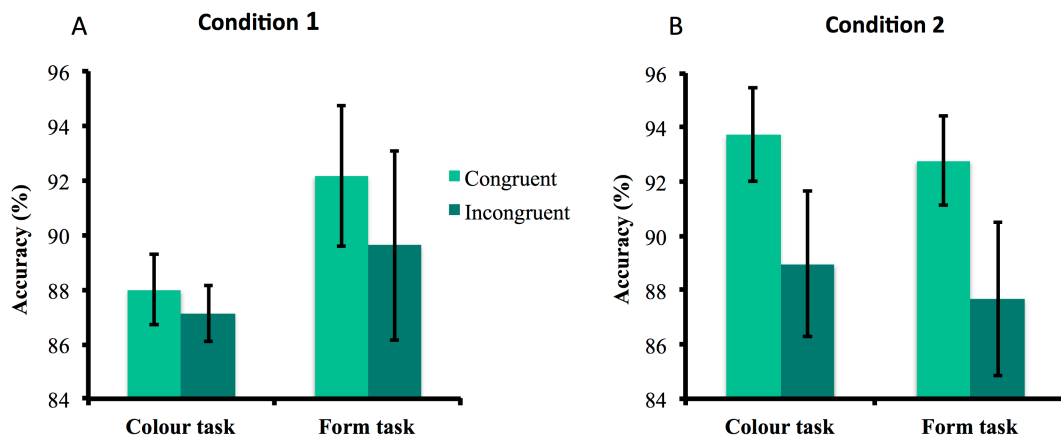
Figure 16: Experiment 3. Accuracy data: The data here are separated by condition and task (Panel A: Condition 1, Panel B: Condition 2). Error bars indicate standard error. Condition 1: more distinct colours. Condition 2: less distinct colours. Condition 2 showed a main effect of congruency in which participants were more accurate in congruent compared to incongruent trials. Note: y-axis does not start at zero.

The accuracy data for Condition 2 are shown in Figure 16B. We again conducted a two-way ANOVA with factors *Task* (Colour, Form) and *Congruency* (Incongruent, Congruent). The ANOVA showed a main effect of *Congruency* ($F(1,6) = 6.91$, $p = 0.03$) with participants having higher accuracy scores in congruent trials (93.2%) compared to incongruent trials (88.3%). The ANOVA showed no main effect of *Task* ($F(1,6) = 4.23$, $p = 0.09$) and no significant interaction between *Congruency* and *Task* ($F(1,6) = 0.02$, $p = 0.92$). Therefore, there was a significant congruency effect in Condition 2 which did not differ between the two tasks.

Discussion

For this experiment we used the version of the paradigm that participants found most difficult from Experiment 2. We also changed the objects to increase the difficulty of discriminating form in the form task. Although there was no congruency effect evident in the RT data, the accuracy data revealed a main effect of congruency that did not depend on task context in Condition 2 (less distinct colours). Therefore we selected Condition 2 of Experiment 3 as the final paradigm for the concurrent TMS-fMRI project.

Conclusions

In Experiment 1, we examined whether disruption to the dlPFC impaired participants' performance in a task requiring them to select for relevant information and/or inhibit irrelevant feature information. The dlPFC has been proposed to play a critical role in this type of control (Duncan, 2001; Miller & D'Esposito, 2005; Miller & Cohen, 2001; Shimamura, 2000) and evidence from previous neurostimulation data supports this view (e.g. Sandrini et al., 2008; Zanto et al., 2011). We did not find evidence to support this suggested role of the dlPFC but this may be due to a number of factors. Aside from a small sample size, participants' accuracy scores were close to ceiling, which may decrease our potential for finding an effect. Therefore in Experiment 2, we modified the stimulus

set and the design of the task to increase its difficulty. We found that Condition 5 of Experiment 2, where we had modified the colour of the stimuli, added noise, and added psychological response time pressure, was the most difficult for participants, reflected in the accuracy data. However, our congruency effect here appeared to be specific to the colour task and was not found in the form task. In Experiment 3, we increased the difficulty of the form task by changing the objects that participants discriminated between. In this experiment we found a congruency effect in our accuracy data and no evidence to suggest that this was influenced by the task that participants performed. This was therefore selected as the final design of the paradigm for the subsequent TMS-fMRI experiment (Chapter 5). Overall, these experiments gave us a thorough understanding of the processes involved in using TMS, including the importance of considering appropriate controls, and developed a sensitive task that could be used to measure feature-selective attention in the following study.

References

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The neuroscientist, 13*(3), 214-228.

Banich, M. T., Milham, M. P., Atchley, R., Cohen, N. J., Webb, A., Wszalek, T., . . . Shenker, J. (2000). fMRI studies of Stroop tasks reveal unique roles of anterior and posterior brain systems in attentional selection. *Cognitive Neuroscience, Journal of, 12*(6), 988-1000.

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., . . . Brown, C. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. *Cognitive Brain Research, 10*(1–2), 1-9. doi:http://dx.doi.org/10.1016/S0926-6410(00)00015-X

Buckner, R. L., Andrews‑Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences, 1124*(1), 1-38.

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49-57.

Buckner, R. L., Snyder, A. Z., Shannon, B. J., LaRossa, G., Sachs, R., Fotenos, A. F., . . . Morris, J. C. (2005). Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *The Journal of Neuroscience, 25*(34), 7709-7717.

Chen, X., Hoffmann, K. P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology, 107*, 1530-1543.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*(9), 415-423.

Davis, N. J., Gold, E., Pascual‑Leone, A., & Bracewell, R. M. (2013). Challenges of proper placebo control for non‑invasive brain stimulation in clinical and experimental applications. *European Journal of Neuroscience, 38*(7), 2973-2977.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron, 50*(5), 799-812. doi:10.1016/j.neuron.2006.04.031

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobes recruited by diverse cognitive demands. *Trends in Neuroscience, 23*, 475-483.

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8*(12), 1784-1790.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5*(7), 671-676. doi:10.1038/nn874

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013a). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621.

Fedorenko, E., Duncan, J., & Kanwisher, N. G. (2013b). Broad domain-generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the USA, 110*(41), 16616-16621.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences, 108*(42), 17510-17515.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America, 102*(27), 9673-9678.

Freedman, D. J. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science, 291*(5502), 312-316. doi:10.1126/science.291.5502.312

Fuster, J. (2008). The Prefrontal Cortex, 4th Edn: Academic Press London.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences, 100*(1), 253-258.

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., . . . Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80*(1), 235-246.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91-100.

Lauwereyns, J., Sakagami, M., Tsutsui, K.-I., Kobayashi, S., Koizumi, M., & Hikosaka, O. (2001). Responses to task-irrelevant visual features by primate prefrontal neurons. *Journal of Neurophysiology, 86*(4).

Lee, T. G., & D'Esposito, M. (2012). The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI–TMS study. *The Journal of Neuroscience, 32*(44), 15458-15466.

Liu, X., Banich, M. T., Jacobson, B. L., & Tanabe, J. L. (2004). Common and distinct neural substrates of attentional control in an integrated Simon and spatial Stroop task as assessed by event-related fMRI. *NeuroImage, 22*(3), 1097-1106. doi:http://dx.doi.org/10.1016/j.neuroimage.2004.02.033

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E. J. (2015). JASP (Version 0.7) [Computer Software].

Manenti, R., Cotelli, M., Calabria, M., Maioli, C., & Miniussi, C. (2010). The role of the dorsolateral prefrontal cortex in retrieval from long-term memory depends on strategies: a repetitive transcranial magnetic stimulation study. *Neuroscience, 166*(2), 501-507.

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houde, O., . . . Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive

functions sustain the conscious resting state in man. *Brain research bulletin, 54*(3), 287-298.

Miller, B. T., & D'Esposito, M. (2005). Searching for "the Top" in Top-Down Control. *Neuron, 48*(4), 535-538. doi:http://dx.doi.org/10.1016/j.neuron.2005.11.002

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience, 23*(1), 1-10.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., & Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nature Neuroscience, 12*(1), 85-91.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petterson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*, 371-377.

Opitz, A., Fox, M. D., Craddock, R. C., Colcombe, S., & Milham, M. P. (2016). An integrated framework for targeting functional networks via transcranial magnetic stimulation. *NeuroImage, 127*, 86-96.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., . . . Schlaggar, B. L. (2011). Functional network organization of the human brain. *Neuron, 72*(4), 665-678.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676-682.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *NeuroImage, 37*(4), 1083-1090.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology, 120*(12), 2008-2039. doi:10.1016/j.clinph.2009.08.016

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Sakagami, M., & Niki, H. (1994). Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: Relation to relevant task conditions. *Experimental Brain Research, 97*(3), 423-436.

Sakagami, M., & Tsutsui, K.-i. (1999). The hierarchical organization of decision making in the primate prefrontal cortex. *Neuroscience Research, 34*(2), 79-89. doi:http://dx.doi.org/10.1016/S0168-0102(99)00038-3

Sandrini, M., Rossini, P. M., & Miniussi, C. (2008). Lateralized contribution of prefrontal cortex in controlling task-irrelevant information during verbal and spatial working

memory tasks: rTMS evidence. *Neuropsychologia, 46*(7), 2056-2063. doi:10.1016/j.neuropsychologia.2008.02.003

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*(2), 207-218.

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of Cognitive Neuroscience, 9*(5), 648-663.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-80.

Watanabe, M. (1986). Prefrontal unit activity during delayed conditional Go/No-go discrimination in the monkey. I. Relation to the stimulus. *Brain research, 382*(1), 1-14. doi:http://dx.doi.org/10.1016/0006-8993(86)90104-6

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

Woolgar, A., Bor, D., & Duncan, J. (2013). Global increase in task-related fronto-parietal activity after focal frontal lobe lesion. *Journal of Cognitive Neuroscience, 25*(9), 1542-1552.

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. *Journal of Neuroscience, 31*(41), 14592-14599. doi:10.1523/jneurosci.2616-11.2011

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *NeuroImage, 109*, 429-437. doi:10.1016/j.neuroimage.2014.12.083

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*(5), 656-661.

Chapter 5

A concurrent TMS-fMRI study investigating feature-selective top-down signals from the dorsolateral prefrontal cortex

Jade Jackson[1], Eva Feredoes[2], Anina N. Rich[1], Alexandra Woolgar[1]

[1]Perception in Action Research Centre and ARC Centre of Excellence in Cognition and its Disorders, Department of Cognitive Science, Macquarie University, Sydney

[2]School of Psychology and Clinical Language Sciences, University of Reading, Reading

This chapter contains the results of a TMS-fMRI experiment that I conducted at the University of Reading (UK). For this I was awarded a CCD student exchange award in order to travel to Reading, UK, to work with an expert in concurrent TMS-fMRI as this technique is not available at Macquarie University.

Abstract

A critical aspect of successful goal-directed behaviour is the ability to select between task-relevant and task-irrelevant information. A circuit of frontal and parietal multiple-demand (MD) brain regions are suggested to play a fundamental role in this type of control (Duncan, 2010). Neurons in these regions are proposed to exhibit substantial flexibility, adapting their response properties to code the information required for the current task ("adaptive coding hypothesis", Duncan, 2001; Duncan, 2013). A prominent region in this network is the dorsolateral prefrontal cortex (dlPFC), often conceptualised in terms of attention-mediated filtering (e.g. Desimone & Duncan, 1995; Miller & Cohen, 2001; Shimamura, 2000). This region has been associated with both inhibition of irrelevant information and selection of task-relevant representations (e.g. Kanwisher & Wojciulik, 2000; Knight, Staines, Swick, & Chao, 1999). In view of this we devised an experiment to provide causal evidence for one or both of these mechanisms. Our task measured intrusion by irrelevant feature information. Participants attended to one feature of a novel object (e.g., its colour) whilst simultaneously ignoring another feature (e.g., form) of the same object. We employed transcranial magnetic stimulation (TMS) during functional magnetic resonance imaging (fMRI) in order to investigate the mechanisms underlying feature-selective attention in the right dlPFC. We using multivoxel pattern analysis (MVPA) to compare what task information was coded across the brain when the right dlPFC was disrupted under high intensity stimulation (HIS, 110% of motor threshold) compared to low intensity stimulation (LIS, 40% of motor threshold), and related this to participant's behavioural data. We predicted that if the dlPFC normally inhibits irrelevant information, then disrupting dlPFC function would result in stronger coding of irrelevant feature information in other brain areas. Conversely, we predicted that if the dlPFC normally enhances coding of relevant information, then disrupting dlPFC

would result in weaker coding of relevant feature information. Our results showed that TMS had a significant effect on behaviour, with irrelevant feature information impacting participants' reaction times more strongly under HIS. This behavioural effect was mirrored in the neural data: under HIS we found stronger coding of irrelevant information across the MD network and visual brain regions. These results provide causal evidence for a functional role of dlPFC in inhibition of irrelevant feature information. Conversely, HIS had either no effect (for object form) or resulted in stronger coding (for object colour) of task relevant information. We speculate that this adaptive response could reflect compensation for disruption of prefrontal function. These data are in line with top-down bias accounts of prefrontal function, supporting the role of the dlPFC in modulating processing of relevant and irrelevant feature information.

Introduction

The complex environment we live in makes it necessary to distinguish relevant from irrelevant information constantly and reliably. This type of attention can even be specific to selecting a relevant feature of an object, referred to as feature-selective attention (Chen, Hoffmann, Albright, & Thiele, 2012). Our ability to attend to specific features of objects whilst filtering out irrelevant information requires a sophisticated system that can coordinate processing across the human brain towards our goals. The details of this system, and the way in which it influences processing across the rest of the brain, is currently an area of active research.

A circuit of frontal and parietal brain regions, often referred to as multiple-demand (MD) regions (Cole & Schneider, 2007; Corbetta & Shulman, 2002; Duncan, 2001; Duncan, 2010; Kanwisher & Wojciulik, 2000; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008), are suggested to play a fundamental role in selective attention. These

regions extend over a specific set of regions in the prefrontal and parietal cortex, specifically cortex in and around the dorsolateral prefrontal cortex (dlPFC), anterior insula (AI), frontal operculum (AI/FO), pre-supplementary motor area and adjacent dorsal anterior cingulate (ACC/pre-SMA) and in and around the intraparietal sulcus (IPS). Duncan (2001) proposed that neurons within these regions dynamically adjust their responses in order to selectively process information that is currently relevant ("adaptive coding hypothesis"). In our previous work (Chapter 2 and Chapter 3) we showed that these regions held a stronger representation of task-relevant features compared to task-irrelevant feature information. This selective representation may provide a source of bias, modulating responses in earlier processing areas via back-projections (also known as "top-down" control) (Corbetta & Shulman, 2002; Desimone & Duncan, 1995; Duncan, Humphreys, & Ward, 1997; Frith, 2001; Knight et al., 1999; Miller & D'Esposito, 2005; Miller & Cohen, 2001; Ruff & Driver, 2006; Serences & Yantis, 2006).

It is increasingly recognised that regions within this frontoparietal network have the potential to influence processing in other brain regions. The dlPFC, in particular, is often cited in the literature as maintaining top-down signals that guide neural activity according to behavioural relevance (e.g. Curtis & D'Esposito, 2003; Knight et al., 1999; Miller & D'Esposito, 2005; Miller & Cohen, 2001; Shimamura, 2000). Electrophysiology studies in non-human primates have provided evidence that lateral prefrontal neurons respond flexibly to task-relevant information, adaptively coding information that is currently relevant (Cromer, Roy, & Miller, 2010; Everling, Tinsley, Gaffan, & Duncan, 2002; Freedman, Riesenhuber, Poggio, & Miller, 2001; Kadohisa et al., 2013; Rao, Rainer, & Miller, 1997; Roy, Riesenhuber, Poggio, & Miller, 2010). This biased processing of task-relevant information is then suggested to influence processing across the brain (Miller & Cohen, 2001), as prefrontal regions send projections to much of the

cerebral cortex (Pandya & Barnes, 1987; Ungerleider, Gaffan, & Pelak, 1989). For example, in electrophysiological studies, prefrontal regions have been associated with top-down influences on inferior temporal cortex (e.g. Miller, Erickson, & Desimone, 1996) and superior colliculus (Johnston & Everling, 2006). Moreover, studies on non-human primates with induced prefrontal lesions have provided *causal* evidence for interactions with early cortical regions (Chafee & Goldman-Rakic, 2000; Fuster, Bauer, & Jervey, 1985; Monosov, Sheinberg, & Thompson, 2011; Rossi, Bichot, Desimone, & Ungerleider, 2007; Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999). This evidence has provided the foundation for several models of prefrontal function (e.g. Curtis & D'Esposito, 2003; Duncan, 2001; Miller & Cohen, 2001; Shimamura, 2000). However, the method by which the PFC influences other regions to achieve selective processing requires further research.

Competition-based models of attention (Desimone, 1998) suggest that task-relevant excitatory signals to selective visual neurons outcompete task-irrelevant inputs thus sharpening the focus of attention (also refer to Desimone & Duncan, 1995; Kastner & Ungerleider, 2000; Miller & Cohen, 2001; Pessoa, Kastner, & Ungerleider, 2003). Other accounts suggest that top-down modulation is a result of enhancement of task-relevant information and simultaneous inhibition of task-irrelevant inputs (e.g. Knight & Stuss, 2002; Shimamura, 2000; Smith & Jonides, 1999). For example, a dynamic filtering account of prefrontal function (Shimamura, 2000) maintains that prefrontal regions, with extensive projections to and from many cortical and subcortical regions, orchestrate signals by means of inhibiting some areas whilst maintaining activation of others. It is often difficult to disentangle these two mechanisms (enhancement of task-relevant information and suppression of task-irrelevant information), particularly if the enhancement of relevant information by top-down projection has a subsequent effect on

task-irrelevant information due to local competition (Desimone, 1998; Desimone & Duncan, 1995). For example, in a study that employed functional magnetic resonance imaging (fMRI), participants were presented with stimuli composed of moving and stationary dots (O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997). When participants paid attention to moving dots, and ignored stationary ones, activity in motion-specific regions was selectively increased. This is suggestive of top-down attentional facilitation, but these data cannot disentangle whether attentional modulation is a result of enhancement of task-relevant information or suppression of irrelevant information or a combination of both.

One way in which we can further investigate these two potential mechanisms is to examine causal interactions between brain regions by perturbing the function of one region and recording the effect in other regions. This is now possible in the human brain with the cutting-edge combination of transcranial magnetic stimulation (TMS) and neuroimaging e.g., fMRI or electroencephalography (EEG). The combination of TMS with neuroimaging can help to test causal brain-behaviour relations not only at the stimulated target site, but also for interconnected brain regions (Bestmann, Baudewig, Siebner, Rothwell, & Frahm, 2005; Bestmann & Feredoes, 2013; Driver, Blankenburg, Bestmann, Vanduffel, & Ruff, 2009; Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011). With the combination of these techniques, prefrontal regions have been causally linked with top-down modulation of earlier cortical responses (e.g. Feredoes et al., 2011; Higo, Mars, Boorman, Buch, & Rushworth, 2011; Lee & D'Esposito, 2012; Miller, Vytlacil, Fegen, Pradhan, & D'Esposito, 2011; Morishima et al., 2009; Zanto, Rubens, Thangavel, & Gazzaley, 2011). For example, Zanto et al. (2011) combined EEG with offline TMS targeting the inferior frontal junction (IFJ). They found that EEG signatures from posterior electrodes that distinguished between task-relevant and task-irrelevant stimuli were

diminished following TMS to the IFJ. For task-relevant stimuli, the response amplitude following TMS to IFJ was decreased, whilst for task-irrelevant stimuli the response amplitude increased. These data indicate a causal role for this region in the selection of task-relevant information as well as the inhibition of task-irrelevant information.

Studies using concurrent TMS with fMRI have also demonstrated top-down modulatory effects from prefrontal regions. For example, Feredoes et al. (2011) used fMRI with online TMS to right dlPFC and found that when distractors were presented during a delay period, there was increased activity in posterior visual regions during dlPFC stimulation. An increase in activation was observed in fusiform face area (FFA) under dlPFC-TMS during memory for faces with house distractors whilst parahippocampal place area (PPA) responses were increased by TMS during house memory for face distractors. This increase in activity was specific to the regions representing the current memory targets but not the distractors and was therefore taken as evidence that an important function of the dlPFC is to maintain relevant information in the face of distracting irrelevant information.

The literature indicates that dlPFC appears to play an important role in causal top-down influences. However, to our knowledge, previous work has mainly focused on the *magnitude* of EEG, or fMRI responses, using univariate analyses. This allows inference about the change in activation in different brain regions, but does not tell us about changes in *information coding*. This is an important distinction because activation does not always reflect the storage or representation of information (D'Esposito & Postle, 2015) and therefore the observed changes in overall activation may not reflect changes in the representation of information. If we want to understand the role of the dlPFC in

supporting the preferential representation of attended information in the brain (Duncan, 2001), we need to use an analysis technique that taps information coding more directly.

Multi-voxel pattern analysis (MVPA) allows us to assess *what information is represented* in the brain by taking into account the full spatial pattern of brain activity (Haynes & Rees, 2006). We know there are top-down effects on information coding in visual cortex (e.g. Jehee, Brady, & Tong, 2011), and with TMS, we now have the opportunity to address whether this is causal result of dlPFC function. The combination of TMS with fMRI-MVPA can allow us to examine the change in information representation across the human brain in response to disruption of the right dlPFC.

As we previously demonstrated, MVPA allows us to separate the representation of relevant and irrelevant object features (Chapters 2 and 3). The MD cortex appears to flexibly represent this information, providing a potential source of bias to other brain regions such as the early visual cortex. Our first question, therefore, was whether disrupting dlPFC function with TMS would affect information coding in the rest of the MD system, and in the visual cortex. Second, an open question concerns whether top-down influences are primarily exerted by the enhancement of task-relevant information, and/or via the inhibition task-irrelevant information (e.g. Aron, 2007; Kanwisher & Wojciulik, 2000). It is entirely possible that both of these mechanisms occur in combination, whereby relevant information is enhanced and irrelevant information is simultaneously suppressed (Knight et al., 1999; Shimamura, 2000). Another possibility is that irrelevant information is not directly suppressed by top-down influences but rather driven out by local competition from relevant inputs (Desimone, 1998; Desimone & Duncan, 1995; Miller & Cohen, 2001). In combination with TMS, we have the

opportunity to examine the causal effect of dlPFC activity on the representation of relevant and irrelevant object features.

In this study we used online TMS in combination with fMRI-MVPA methods to causally investigate these two mechanisms (enhancement and inhibition) in right dlPFC. We focused on the right dlPFC because stimulation to this region has previously been shown to affect activity in visual brain regions during a task requiring selective attention (Feredoes et al., 2011). We used the following logic: if the right dlPFC usually enhances coding of the attended object feature then interrupting this mechanism with TMS stimulation should result in a *decrease* in coding of the attended object feature elsewhere in the brain. Conversely if this region usually filters out distracting information, stimulation should result in an *increase* in coding of the unattended object feature. Both mechanisms are possible and could happen concurrently; with this method we investigated whether either one, or both, are subserved by dlPFC.

Participants performed a task in which they attended to one feature of a novel object (e.g. its colour) whilst ignoring an irrelevant feature (e.g., its form). The response for the irrelevant feature (form in colour blocks; colour in form blocks) was either congruent or incongruent with the required response for the relevant feature. We randomly interleaved high intensity (effective) TMS (110% motor threshold, MT) with low intensity (ineffective) TMS (40% MT) on any given trial. Our first prediction was that disruption of the two mechanisms outlined would be reflected in the behavioural data. Specifically, we expected that irrelevant feature information would have a larger effect on performance in the high intensity stimulation (HIS) compared to the low intensity stimulation (LIS) condition, resulting in a larger congruency effect. Our second prediction was that HIS would affect information coding in the brain. We predicted that, relative to LIS, HIS to

right dlPFC would result in **a)** stronger multi-voxel coding of *irrelevant* feature information (release from suppression); and/or **b)** weaker coding of *relevant* feature information (disrupted upregulation of attended information). We anticipated that this change in information coding would be seen in the rest of the frontoparietal network, which is proposed to work together as a system (Duncan, 2010), as well as in visual cortex (LOC and early visual cortex), reflecting top-down modulation (e.g. Desimone & Duncan, 1995).

Materials and Methods

Participants

A group of twenty healthy adult volunteers (15 females; mean age = 21.6 years, SD= 3.36) participated in this experiment. All participants were right-handed with normal or corrected-to-normal vision and no history of neurological or psychiatric disorder. Participants gave written informed consent and passed relevant screening for TMS and MRI. The University of Reading Research Ethics Committee approved the study. All participants received £30.00. Initially thirty-one participants signed up for the experiment. However, four participants were excluded for not passing screening requirements, and seven participants were excluded for not completing both session 1 and session 2 of the experiment, leaving us with 20 datasets for analysis.

Stimuli

Stimuli were abstract novel "smoothy" and "cuby" created using custom MatLab scripts (Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006). The stimulus set consisted of 4 objects (Figure 1). Participants learnt to discriminate between a cuby and a smoothy object in the form task. For the colour task they discriminated between a blue (RGB

values: 98 179 18) and a green object (RGB values: 95 171 96). The relevant visual

feature of the stimuli varied depending on the current decision boundary (form or colour).

We controlled stimulus presentation with a PC running the Psychophysics Toolbox-3

package (Brainard, 1997) in MatLab (Mathworks). Stimuli were presented at central

fixation on a screen and viewed through a mirror mounted on the head coil in the scanner.

Procedure

Participants took part in two separate sessions 2-8 days apart. In session 1, we first

familiarised participants with the sensation of TMS and measured their resting motor

threshold (MT). Following this, participants completed a structural scan and three

functional localiser tasks in the scanner in order to determine individual stimulation sites

and regions of interest (ROIs) for analysis. In session 2, participants underwent concurrent

TMS-fMRI whilst completing the main experimental task.

Session 1

For each participant, we determined the minimum intensity at which a single pulse

through the TMS coil, positioned over the hand area of the primary motor cortex, reliably

produced a visible twitch in the abductor pollicis brevis when at rest, in 5 of 10 successive

pulses. The MT of each participant determined the intensity of TMS stimulation for that

participant in session 2. The average MT recorded across participants was 58% of

maximal stimulator output. The intensity of stimulation in session 2 was pseudo-randomly

varied over trials at either 110% or 40% of the individual participant's MT.

After acquiring the individual MT, participants were given instructions about the

structural scan and tasks that they would complete in the scanner. These tasks were a task

localiser for the right dlPFC, a lateral occipital complex (LOC) localiser, and a fixation

localiser. Participant practised the right dlPFC task localiser before they entered the scanner. Participants were instructed on the other localisers but were not required to practise the tasks, as they were very simple.
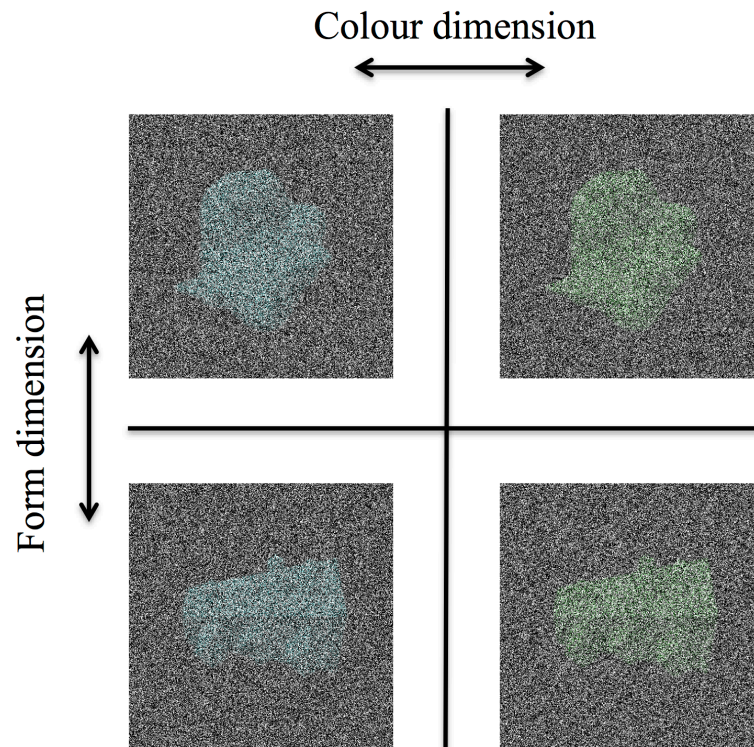
Figure 1: The stimulus set consisted of four novel objects differing on the basis of two feature dimensions; colour (blue/green) and form ("smoothy"/"cuby").

Right dlPFC task localiser

This localiser was designed to activate the dlPFC using a modified version of the main experimental task (see description below). Since the task localiser was performed in a separate session to the TMS-fMRI session, it was entirely independent data and therefore valid to use for determining a target ROI for TMS. We aimed to find a region of dlPFC that responded more strongly on incongruent relative to congruent trials. We

reasoned that this activity would reflect the additional processes required in upregulating relevant and/or suppressing irrelevant information on incongruent trials.

In the localiser, congruent and incongruent trials were blocked to give maximal power for analysis. The congruent condition blocks consisted of the objects where the response button for the irrelevant dimension was congruent with the required response for the relevant dimension (e.g. a blue cuby would be congruent if both 'cuby' and 'blue' were associated with the 'left' response). The incongruent condition blocks consisted of the objects where the response for the irrelevant dimension was incongruent with the required response for the relevant dimension (a green cuby would be incongruent if 'cuby' was associated with the 'left' response but 'green' was associated with the 'right' response). A rest block was also included where participants viewed a black cross at fixation for 16s. The order of task (colour/form) and congruent/incongruent blocks was counterbalanced across participants.

Participants completed at least 6 practice blocks outside of the scanner until they achieved >70% performance. For the first 2 blocks participants received feedback on every trial, following this they received feedback only at the end of the block (percent correct). At the start of each block a picture cue (3000ms) indicated the current task (form or colour). The picture cue displayed all four objects and separated them according to the feature dimension that participants would discriminate between in the upcoming block (Figure 2, inset). The stimulus set was thus identical across the two task contexts, but the currently relevant dimension differed. The picture cue indicated the correct button responses for the relevant dimension.

On each trial participants would first see a cue that reminded them of the current task (form or colour, 500ms) followed by an object displayed at fixation for 100ms. A

white cross would then be displayed for 500ms after the object offset followed by a black cross for 1000ms. The white cross would be replaced by the black cross before 500ms if participants responded within this time period and the remaining time that the white cross would have been displayed for was added on to the duration of the black cross period. Participants were told that their responses were only recorded during the white cross period. This was done to introduce time pressure in order to increase difficulty. Responses were also recorded if they occurred during the black cross period (total recorded response period of 1500ms) but this was not reflected in participant's feedback scores.
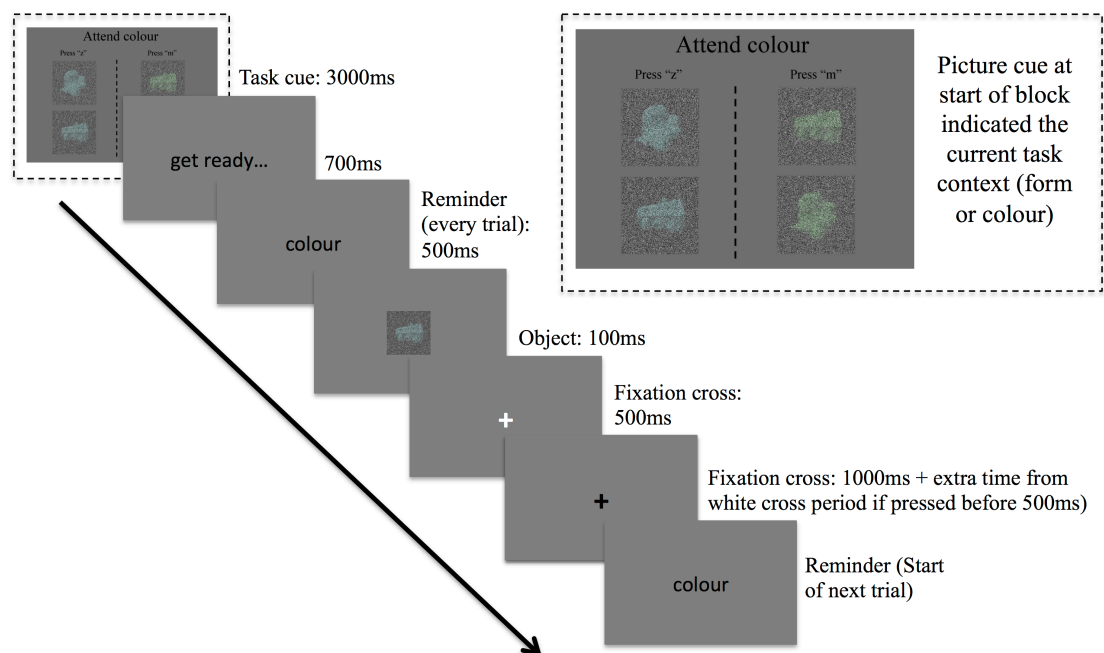
Figure 2: dlPFC task localiser: A picture cue at the start of each block indicated the current task context (inset shows cue display for the colour task). On each trial a cue reminded participants of the current task context (500ms) followed by the object to categorise (100ms) followed by the white fixation cross (500ms) and the black fixation cross (1000ms). Following the picture cue the task was blocked so that participant completed 8 trials of the congruent condition followed by 8 trials of the incongruent condition followed by a 16s rest period (black fixation cross).

In the scanner, each participant completed 2 runs (6.00m each) of the task localiser. Within each run participants completed 18 blocks of each condition; congruent (16.8s, 8 trials), incongruent (16.8s, 8 trials) and rest (16.8s). Participants received feedback (percent correct) at the end of every set of congruent and incongruent blocks before the rest period. In the second run the button response mapping (the button associated with each object) was switched in order to mimic the procedure of the main task. This was an essential feature of the design of the main task in order to dissociate activity associated with participant's motor response from that associated with the stimulus features. The button response mapping order was counterbalanced across participants.

Lateral Occipital Complex (LOC) localiser

After the task localiser, we ran a second localiser to functionally identify object-sensitive cortex in the LOC. Participants viewed centrally located intact and scrambled versions of black and white objects in 16.8s blocks of 16 trials (1100ms/trial), whilst attending to a central fixation cross. Participants indicated via a button response when the fixation cross changed from black to blue. There were 21 blocks consisting of alternating blocks of whole objects, scrambled objects, and rest blocks (order counterbalanced across participants). The EPI (acquisition) time for this localiser task was 6.25min.

Fixation localiser

Following the LOC localiser we ran a fixation localiser task (Figure 3) to identify the cortical region stimulated by visual information at fixation (covering the same area of central visual field as the objects in the task localiser and main task). Participants viewed a small white circle at the centre of the screen that was visible across all experimental blocks. Rest blocks (16s) consisted only of the circle presentation. In the stimulus blocks, participants viewed either a flickering checkerboard presented only at fixation (the spatial

extent of which matched the objects in the main task) or a checkerboard filling the entire field of view leaving only a blank grey screen in the place of the fixation checkerboard.

Fixation localiser task
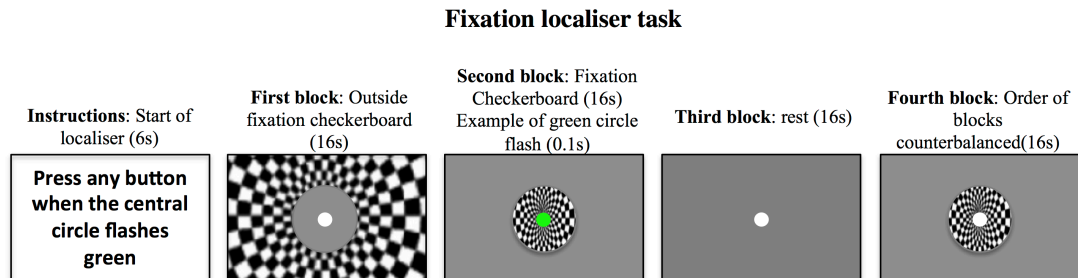
Figure 3: Fixation localiser: A white central circle was present throughout all blocks. In blocks participants viewed either a flickering checkerboard presented only at fixation, a checkerboard filling the entire field of view leaving only a blank grey screen in the place of the fixation checkerboard or just the white central circle (rest). When the central white circle flashed green (0.1s, random intervals) participants were required to respond with a button press. In the example above there are four blocks of the task depicted, the task consisted of 24 blocks total (8 blocks of each condition). The block order was counterbalanced.

Participants were required to press a button when the central circle on the screen flashed green. This task was included to encourage attention and fixation, however, as we did not include eye-tracking in this experiment, we cannot be certain that participants did not move their eyes from fixation during the localiser task. However, it seems unlikely that this would have occurred, as there was no advantage to looking away from the central circle. The EPI time for the fixation localiser was 8.00m.

Session 2

In session 2 we used the Brainsight infrared frameless stereotaxy neuronavigational system (Rogue Research Inc.) that relies on acquired MR anatomical

information (acquired in session 1) to navigate to each participant's individual target stimulation location (see below) and marked the stimulation site on the scalp. Following this, participants practiced the main task outside the scanner and then completed 8 runs of the task in the scanner with simultaneous online TMS stimulation on every trial.

Target region selection

For each participant, we defined the target region as follows. First, we contrasted univariate activation on *incongruent minus congruent* blocks in our dlPFC task localiser (contrast 1), and *congruent and incongruent minus rest* (contrast 2; see Univariate contrasts section below for contrast details). The results of these contrasts were compared to a sphere of radius 8mm centered on MNI co-ordinates [44 31 28], which was the target location in our previous work (Chapter 4), deformed into individual subject native space. The comparison sphere location was chosen on the basis of published functional activation and connectivity data (Cole et al., 2013; Duncan & Owen, 2000; Fedorenko, Duncan, & Kanwisher, 2013) and localiser data collected in our previous experiment (refer to Chapter 4 for further details, see Figure 4 for overlaid activation).

If the peak activation from contrast 1 (*incongruent minus congruent*) was within 8mm of the 8mm sphere in native space then the central coordinate of peak activation was used for that participant as their stimulation target. This was the case for 12 out of 20 participants. If contrast 1 showed no activation within range of our sphere then we compared activation for contrast 2 (*congruent and incongruent minus rest*) to the sphere applying the same procedure. The resultant cluster was used for 5 participants. If there were no clusters of activation from these contrasts near to our pre-defined sphere, then the central point of the sphere was chosen as the target. This was the case for 3 participants.

Right dlPFC sphere overlaid on a template
structural image

Right dlPFC sphere with overlaid activation maps

■ *dlPFC 8mm sphere centered on chosen coordinate*
■ *(Power, Cohen et al. 2011)*
■ *Working memory task localiser*
■ *(Fedorenko, Duncan et al. 2013)*
■ *(Duncan and Owen, 2000)*
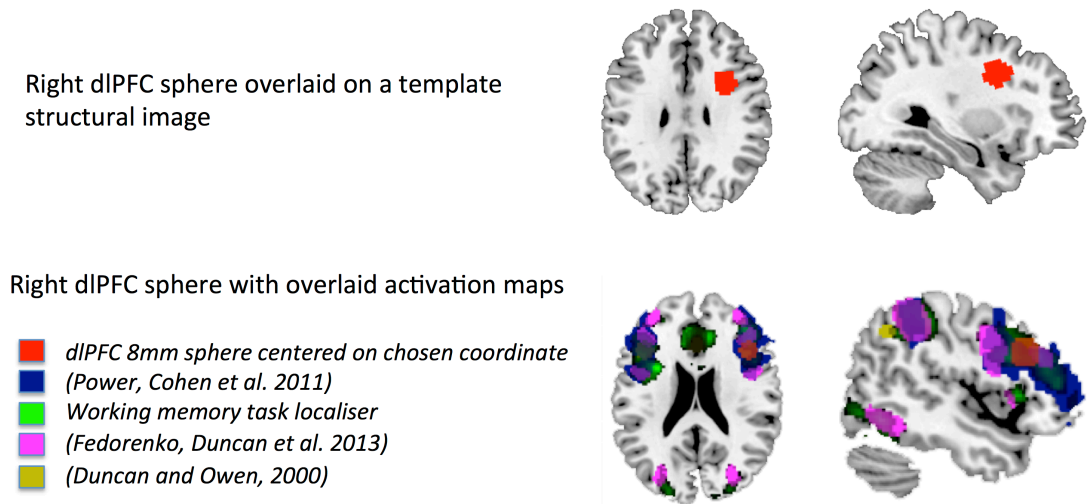
Figure 4: The top panel shows a sphere (radius = 8mm) centered on MNI152 coordinates [44 31 28]. This sphere was used for guidance in selecting our stimulation coordinate from task localiser data. The bottom panel shows overlaid activation maps used for selection of this sphere (see Chapter 4). The location of the sphere was transformed into native space for each participant.

Neuronavigation

At the start of session 2, we used the neuronavigational system with the predefined stimulation target coordinates of each individual (functionally defined, see below) to guide coil placement. A tracker with three reflecting spheres, attached to the participant's head, and a pointer with four reflecting spheres was registered into the Brainsight system. The standard Brainsight neuronavigation routines were used to determine the coil position on the participant's head. The target location was marked on the participant's scalp with a non-permanent pen for later targeting with TMS inside the MRI scanner. The position of the TMS coil was adjusted in the scanner so that it fit within the head coil and was comfortable for participants. The stimulating coil was oriented with the handle pointing posteriorly with respect to the participant's head, and roughly parallel to the midline.

TMS-fMRI

Prior to entering the scanner the participants practiced the main task for at least 6 blocks without TMS (about 20 minutes). The main task was similar to the task localiser that the participants completed in session 1. However, now the congruent and incongruent conditions were not blocked. Instead these two conditions occurred *pseudo*-randomly on a trial-by-trial basis. In addition, there was no rest block and the timings were slowed down to prolong the duration between the TMS pulses (Figure 5 shows the task design). During practice, participants received feedback following each trial for the first two blocks as well as feedback at the end of the block. For the last four blocks of practice participants only received feedback (percent correct) at the end of the block. Participants repeated the practice trials until they scored >70% correct.

Participants completed 8 runs of the task in the scanner with concurrent TMS. Runs lasted 6.3 minutes and there was a minimum of 1 minute break between runs. In each run participants completed one block of the colour task and one block of the form task. The block started with a picture cue (4000ms) indicating the current task context and response mapping. On each trial participants first saw a cue reminding them of the current task (form or colour, 500ms) followed by the object (100ms). Participants received a train of three pulses; the first pulse at 75ms after stimulus onset. The following two pulses were separated by 75ms. The train of pulses was delivered at 110% or 40% of participant's individual MT, with intensity varying pseudo-randomly over trials. A white cross appeared for 500ms after the object offset followed by a black cross for a variable time period of between 3500-4000ms. The white cross was replaced with the black cross before 500ms if participants responded within this time period, and in this case the remaining time that the white cross would have been displayed was added on to the duration of the black cross.
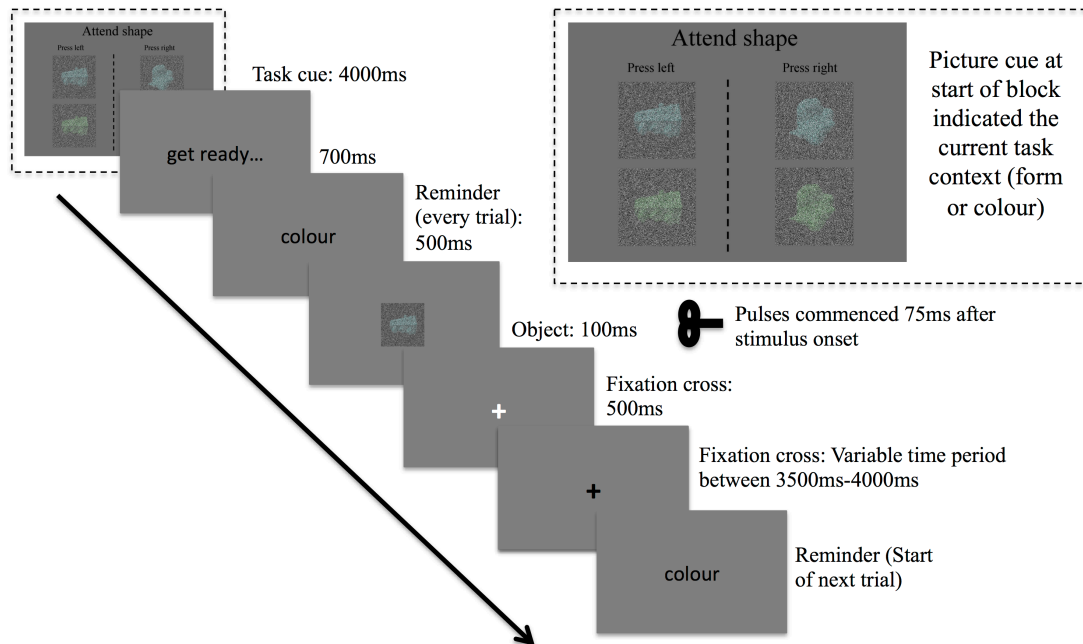
Figure 5: Concurrent TMS-fMRI task: A picture cue at the start of each block indicated the current task context (inset shows cue display for form task). On each trial a cue reminded participants of the current task (500ms) followed by the object to categorise (100ms: first pulse (high or low intensity) 75ms after stimulus onset), followed by the white fixation cross (500ms) and the black fixation cross (3500ms-4000ms). In this example, participants are cued to attend to the form of the objects. The first object is a blue cuby, which is associated with the left button response. Participants were asked to respond as fast as possible (in white cross time period) with the correct button response.

Participants received feedback (percent correct) at the end of each block. After the first four runs, the button-response mapping was swapped (e.g. if the button response for a green smoothy was the left button, this would now be the right button response). TMS timing was carefully controlled to coincide with the readout period of slice acquisition, so that the artifact caused by each pulse would affect only one slice and would occur at a different slice of each volume.

TMS

The MR-compatible TMS figure-8 stimulating coil (MRI B90 II. MagVenture, Farum, Denmark) was held firmly in position by a custom made MR compatible non-ferromagnetic coil holder with several degrees of freedom in each direction inside the MR head coil. The cable of the TMS coil passed through the back of the scanner and out through a wave guide on the scanner wall. It connected to the TMS machine located in the MR control room.

A MatLab (Mathworks) script running on an experimental PC connected to the TMS machine remotely controlled and triggered the onset of the TMS pulses to be synchronized to the fMRI scanning sequence. TMS pulses were delivered to coincide with the readout phase of a slice acquisition. In this way, the TMS artefact affected only the one slice on which TMS was applied (Bestmann, Baudewig, & Frahm, 2003). This slice was later discarded (see Preprocessing, below). The same TMS machine and coil were used in session 1 for determining individual MT and in session 2 for the main experiment. The main experiment consisted of a total of 16 blocks (8 runs) with a total of 1536 pulses across all of the runs complying with published safety limits for TMS stimulation (Rossi, Hallett, Rossini, & Pascual-Leone, 2009).

fMRI

Data acquisition

The data were collected for both session 1 and 2 using a Siemens Magnetom Trio 3T whole body Magnetic Resonance Imaging (MRI) scanner at Reading University, Reading, UK.

Session 1

We used a sequential ascending T2*- weighted echo planar imaging (EPI) acquisition sequence with the following parameters: acquisition time 2080 ms; echo time 30 ms; 60 oblique axial slices with a slice thickness of 3.0 mm and a 0.70 mm inter-slice gap; in plane resolution 3.0×3.0 mm; matrix 64×64; field of view 256 mm; flip angle 78°. T1-weighted MPRAGE structural images were also acquired for all participants (slice thickness 1.0 mm, resolution 1.0×1.0 mm).

Session 2

We used a sequential ascending T2*- weighted echo planar imaging (EPI) acquisition sequence with the following parameters: acquisition time 2450ms; echo time 30 ms; 35 oblique axial slices with a slice thickness of 3.0 mm and a 0.70 mm inter-slice gap; in plane resolution 3.0×3.0 mm; matrix 64×64; field of view 256 mm; flip angle 90°; 50% phase oversampling in the phase encoding direction to shift any Nyquist ghost artefact, due to the presence of the TMS coil, to outside the volume of interest.

Preprocessing

Session 1

MRI data were preprocessed using SPM 5 (Wellcome Department of Imaging Neuroscience, www. fil.ion.ucl.ac.uk/spm) in MatLab 2013b. Functional MRI data were converted from DICOM to NIFTI format, spatially realigned to the first functional scan and slice timing corrected, and structural images were co-registered to the mean EPI. EPIs were smoothed slightly (4mm FWHM Gaussian kernel) and in all cases the data were high pass filtered (128s). Structural scans were additionally normalized, using the segment and normalise routine of SPM5, in order to derive the individual participant normalization parameters needed for TMS target definition.

Session 2

After we had reconstructed the TMS-fMRI data, we removed the slices that were affected by TMS pulses (one slice per pulse). Slices with a signal magnitude of >1.5 SD from the run mean were replaced by the mean of the same slices from the preceding and proceeding volumes (following Feredoes et al., 2011; Ruff et al., 2008; Ruff et al., 2006). Aside from this initial step for session 2 data, preprocessing followed the same steps as session 1. EPIs from session 2 were smoothed slightly (4 mm FWHM Gaussian kernel) to improve signal-to-noise ratio for multi-variate analyses and were smoothed separately with a larger smoothing kernel (following standard practice) for univariate analyses (8 mm FWHM Gaussian kernel).

Univariate contrasts

Right dlPFC Task localiser

The right dlPFC was functionally defined on an individual level on the basis of: **a)** our main effect of interest (additional activation associated with conflict or intrusion from the irrelevant dimension); or **b)** as activation pertaining to performing the overall task compared to rest. We used the standard multiple regression approach of SPM5 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk) to estimate values corresponding to the congruent and incongruent task conditions as well as the rest conditions (block design). Blocks were modelled using a box car function lasting 16.8s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. Whole-brain analyses (paired t-tests) compared blood-oxygen-level-dependent (BOLD) responses across the following conditions: [incongruent – congruent] (congruency effect), and [(congruent + incongruent) – rest] (task effect). The resulting map was thresholded such that there was at least one cluster with a minimum

216

size of 20 voxels. The peak of one of these clusters was chosen as the target stimulation site as explained above.

LOC

For each participant, we defined object-selective LOC as the occipital brain area that responded more strongly to whole objects than to scrambled versions of the same objects. We again used the standard multiple regression approach of SPM5. Here, we estimated values pertaining to the whole and scrambled object conditions. Blocks were modelled using a box car function lasting 16s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. Paired t-tests compared voxelwise BOLD response in the two conditions (whole objects minus scrambled objects). The resulting map was thresholded such that there was at least one cluster with a minimum size of 20 voxels in lateral occipital cortex. These clusters were then imported into MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) and clusters of activation close to LOC coordinates from previous studies (Grill-Spector et al., 1999; Grill-Spector, Kushnir, Hendler, & Malach, 2000) were selected as ROIs.

Fixation

We defined an ROI based on increased activation in response to visual stimulation at fixation where the objects in the main task were presented compared to visual stimulation outside of this area. Here, we estimated activity pertaining to the fixation checkerboard and outside-fixation checkerboard conditions. Blocks were again modelled using a box car function lasting 16s convolved with the hemodynamic response of SPM5. The run mean was included in the model as a covariate of no interest. We contrasted BOLD responses with paired t-tests in the two conditions using our visual localiser data (flickering checkerboard at main task object location minus flickering checkerboard

outside object location). The resulting map was thresholded such that there was at least one cluster with a minimum size of 20 voxels within early visual cortex.

Comparison of overall activation for HIS contrasted with LIS

Here we examined activity differences under HIS compared to LIS using a mass-univariate (whole brain) approach. We also examined activity differences under HIS compared to LIS in our pre-defined MD regions (see MD network definition below). A General Linear Model (GLM) was estimated for each participant using the realigned, slice-time corrected and smoothed normalised EPI images (8 mm FWHM Gaussian kernel) from session 2 using SPM5. We modelled HIS and LIS trials separately and contrasted BOLD responses for HIS and LIS at the second (random effects across subjects) level with paired t-tests at each voxel.

MD network definition

MD regions of interest (ROIs) for decoding and univariate analyses were defined using co-ordinates from a previous review of activity associated with a diverse set of cognitive demands (Duncan & Owen, 2000) using the kernel method described in Cusack, Mitchell, and Duncan (2010), as in our previous work (Chapter 2, Chapter 3, see also Woolgar, Hampshire, Thompson, & Duncan, 2011; Woolgar, Thompson, Bor, & Duncan, 2011; Woolgar, Williams, & Rich, 2015). The procedure yielded a total of seven ROIs: left and right IFS (centre of mass $+/-38\ 26\ 24$, volume 17 cm^3); left and right AI/FO ($+/-35\ 19\ 3$, 3 cm^3); left and right IPS ($+/-35\ -58\ 41$, 7cm^3) and ACC/ pre-SMA (0 23 39, 21 cm^3).

MVPA

First-Level Model for TMS-fMRI task

To obtain estimated activation patterns for multivariate pattern analysis, a GLM was estimated for each participant using the realigned, slice-time corrected and smoothed native space EPI images from session 2 using SPM5. We modelled stimuli according to their two different feature dimensions, form (cuby or smoothy) and colour (green or blue), separately for HIS and LIS trials. To account for trial by trial variation in reaction time (Todd, Nystrom, & Cohen, 2013), trials were modelled as events lasting from stimulus onset until response (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008; Henson, 2007; Woolgar, Golland, & Bode, 2014) convolved with the hemodynamic response of SPM5. Every trial contributed to the estimation of two beta values, the relevant feature (*cuby* or *smoothy* in the form task, and *green* or *blue* in the colour task) and the irrelevant feature (*cuby* or *smoothy* in the colour task, and *green* or *blue* in the form task), for HIS and LIS trials separately (a total of 8 regressors per block).

ROI Analysis

We implemented MVPA using the Decoding Toolbox (Hebart, Görgen, & Haynes, 2015) which wraps the LIBSVM library (Chang & Lin, 2011). We used MVPA to examine the representation of relevant and irrelevant stimulus features in the HIS and LIS conditions separately. For this we examined coding of colour when colour was relevant (colour task), colour when colour was irrelevant (form task), form when form was relevant (form task), form when form was irrelevant (colour task). This was conducted for the HIS and LIS trials separately. We examined these stimulus feature distinctions in each MD region (including the stimulated right dlPFC), LOC, and our fixation ROI.

For each participant and ROI, a linear support vector machine was trained to decode the relevant (green/blue in colour blocks, and cuby/smoothy in form blocks) and irrelevant (green/blue in form blocks, and cuby/smoothy in colour blocks) stimulus features. In total, there were 16 blocks for each participant: 8 with colour relevant, and 8 with form relevant. Half of the trials in these 8 blocks contributed to the HIS classification, and half contributed to the LIS classification.

For each classification, we used a leave-one-out 8 fold splitter whereby the classifier was trained using the data from 7 out of the 8 blocks and subsequently tested on its accuracy at classifying the unseen data from the remaining block. For example, to yield a classification accuracy score for the task-relevant colour distinctions in the HIS condition, we took the 8 blocks in which participants performed the colour task and used the classifier to distinguish between patterns of activation representing green and blue objects in HIS trials in 7 out of these 8 blocks, and then tested generalization to the remaining unseen block. This procedure was repeated iterating over all possible combinations of training and testing blocks. The accuracies were then averaged over iterations to give a mean accuracy score for task-relevant colour coding in the HIS condition. This was repeated for each condition, participant and ROI separately.

This cross-validation decoding approach, which is becoming standard in the field (e.g. Mahmoudi, Takerkart, Regragui, Boussaoud, & Brovelli, 2012; Mur, Bandettini, & Kriegeskorte, 2009), gives unbiased estimates of the true classification accuracy for each condition, participant and ROI. Therefore, parametric statistics can be used to analyse this data at the second level (random effects). Accordingly, we entered the decoding scores into two separate ANOVAs to examine changes in coding of relevant and irrelevant feature information separately. We did this to address our two *a priori* predictions that

disruptive/effective HIS compared to ineffective LIS to right dlPFC would result in **a)** stronger coding of *irrelevant* feature information; and/or **b)** weaker coding of *relevant* feature information. We conducted these ANOVAs for the MD regions (all regions except the stimulated right dlPFC), right dlPFC, LOC and fixation separately.

Since a difference in coding between HIS and LIS is only interpretable if coding in at least one condition is also significantly above chance, we also conducted one-sample t-tests against the classification accuracy expected by chance (50%) in each condition separately. *Tests* comparing classification accuracy to chance are one-tailed as below chance classifications are not interpretable. Alpha was adjusted for six comparisons for the 6 MD regions using Bonferroni correction (0.05 divided by 6).

Searchlight Analysis

In order to identify any additional brain regions coding feature-relevant and feature-irrelevant information under HIS and LIS, whole brain pattern classification was carried out using a roaming spotlight (Kriegeskorte, Goebel, & Bandettini, 2006). For each participant, data were extracted from a spherical ROI (radius 5 mm) centered in turn on each voxel in the brain. A linear support vector machine was trained and tested as before, using data from each sphere, and the classification accuracy value for that sphere was assigned to the central voxel. This yielded whole brain classification accuracy maps for each individual for each of the effects of interest outlined above.

To combine data across individuals, classification accuracy maps were normalized and were subsequently smoothed using a 8mm FWHM Gaussian kernel. Classification accuracy for each condition was compared to chance at the group level using a one-sample t-test against chance (50%). Our prediction was that there would be stronger coding of *irrelevant* feature information in the brain following disruption to the right dlPFC, and

weaker coding of *relevant* feature information. To examine this across the whole brain, we conducted paired t-tests that compared coding of relevant information between the HIS and LIS condition, and coding of irrelevant information between the two stimulation conditions. The resultant maps were thresholded at $p < 0.05$ with family wise error (FWE) correction for multiple comparisons at the cluster level.

Results

Behavioural Results

In line with prominent theorists who have propagated the view that the dlPFC plays a critical role in selective attention (Duncan, 2001; Miller & Cohen, 2001), we predicted that disruption of the right dlPFC would affect participants' ability to select the relevant feature information and/or to inhibit irrelevant feature information. On congruent trials, the response for the irrelevant feature is congruent with the response for the relevant feature and may assist in selecting the correct response. In contrast, on incongruent trials, the irrelevant feature calls for the opposite response and therefore interferes with selection of the correct button response. The difference between performance on incongruent and congruent trials − the 'congruency effect' − is thus a measure of the extent to which irrelevant information affects performance. If the dlPFC normally supports the selection of the relevant information, and/or the inhibition of the irrelevant information, we would expected to see a larger congruency effect when the dlPFC is disrupted in the HIS condition in comparison to the (ineffective) LIS condition.

RT data

Reaction time data are presented in Figure 6 (panels A and B). We analysed RT (correct trials only) data with a three-way repeated measures ANOVA, with factors *TMS*

(HIS, LIS), *Task* (Colour task, Form task), and *Congruency* (Congruent, Incongruent). We predicted an interaction between TMS and Congruency, in which the congruency effect would be larger for the HIS condition.
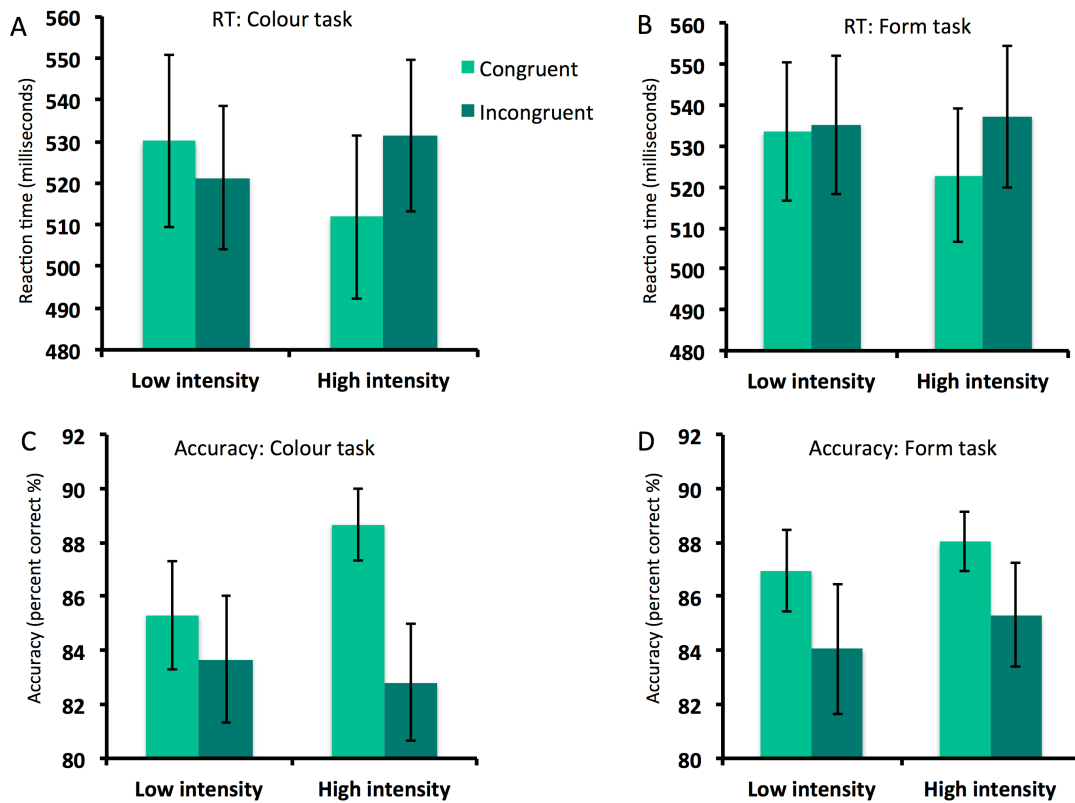
Figure 6: Reaction time (top panels) and accuracy (bottom panels), for the colour task (left panels) and form task (right panels), on congruent and incongruent trials under HIS and LIS. RT data showed that participants were faster in congruent trials than incongruent trials under HIS. Accuracy data showed a main effect of congruency only. Error bars indicate standard error.

In line with our prediction, there was a significant interaction between *TMS* and *Congruency* ($F(1,19) = 4.78$, $p = 0.04$). Post-hoc paired t-tests showed that in the HIS condition, participants were significantly slower in incongruent trials (535ms) than in congruent trials (518ms, $t(19) = 4.78$, $p < 0.0001$), but this did not occur in the LIS condition (incongruent: 532ms, congruent: 530ms, $t(19) = 0.36$, $p = 0.73$). No other main

effects or interactions were significant (all ps > 0.11). This demonstrates that HIS to dlPFC increased the magnitude of the congruency effect in RT relative to LIS.

Accuracy data

Accuracy data (percent correct %) are presented in Figure 6 (panels C and D). As for RT, accuracy data were entered into a three-way ANOVA with factors *TMS* (HIS, LIS), *Task* (Colour task, Form task), and *Congruency* (Congruent, Incongruent). This ANOVA did not show the predicted *Task*Congruency* interaction ($F(1,19) = 0.97$, $p = 0.34$). Instead, there was a significant main effect of *Congruency* ($F(1,19) = 11.58$, $p = 0.003$), reflecting more accurate performance on congruent (87.5%) relative to incongruent (84.4%) trials overall. No other main effects or interactions were significant (all ps > 0.23).

Univariate data: HIS vs. LIS

Our experiment was designed to examine information coding using MVPA, but we also carried out a whole-brain mass-univariate analysis to examine whether overall activation levels changed following HIS to right dlPFC. Although the univariate effects of disruptive TMS can be complex (Sack et al., 2007), we had hoped this analysis would be an additional source of information to which networks were affected by TMS stimulation. However, we did not find evidence for a change in overall activation in the two intensity conditions. No clusters survived FWE correction for the univariate contrasts of [HIS – LIS] or [LIS – HIS]. We also conducted univariate ROI analyses in our pre-defined MD regions; but, no clusters survived Bonferroni correction for the univariate contrasts of [HIS – LIS] or [LIS – HIS] (all ps > 0.12).

A univariate effect is not a pre-requisite for our main analyses: MVPA is known to be a more sensitive technique as it preserves fine-grained information from patterns of activity across multiple voxels. Therefore information can often be decoded from activity patterns even when activity itself is low (e.g. Harrison & Tong, 2009). The following analyses used MVPA to examine whether information coding was affected by TMS.

ROI based pattern classification

We conducted two separate ANOVAs for each ROI to examine changes in coding of relevant and irrelevant feature information separately. We did this to address our two *a priori* predictions that HIS compared to LIS would result in **a)** stronger coding of *irrelevant* feature information; and/or **b)** weaker coding of *relevant* feature information. The right dlPFC was the target for TMS stimulation and so we analysed it separately from the rest of the MD system.

MD regions

Decoding irrelevant information
Decoding results for irrelevant information are presented in Figure 7 (Panels A and B). Classification accuracies for irrelevant feature information were entered into a three-way ANOVA with factors *TMS* (HIS, LIS), *Feature* (Colour, Form) and *Region* (ACC/pre-SMA, left AI/FO, right AI/FO, left IPS, right IPS, left dlPFC). We reasoned that if the dlPFC was normally involved in inhibiting the coding of irrelevant feature information, then disrupting processing in this region should result in an increase in coding of irrelevant information. Therefore, we predicted a main effect of *TMS,* with more coding in HIS relative to LIS. The ANOVA indeed revealed a main effect of *TMS* ($F(1,19) = 17$, $p = 0.001$), reflecting stronger coding of irrelevant feature information in the HIS condition (59.3%) compared to the LIS condition (49.9%).

225
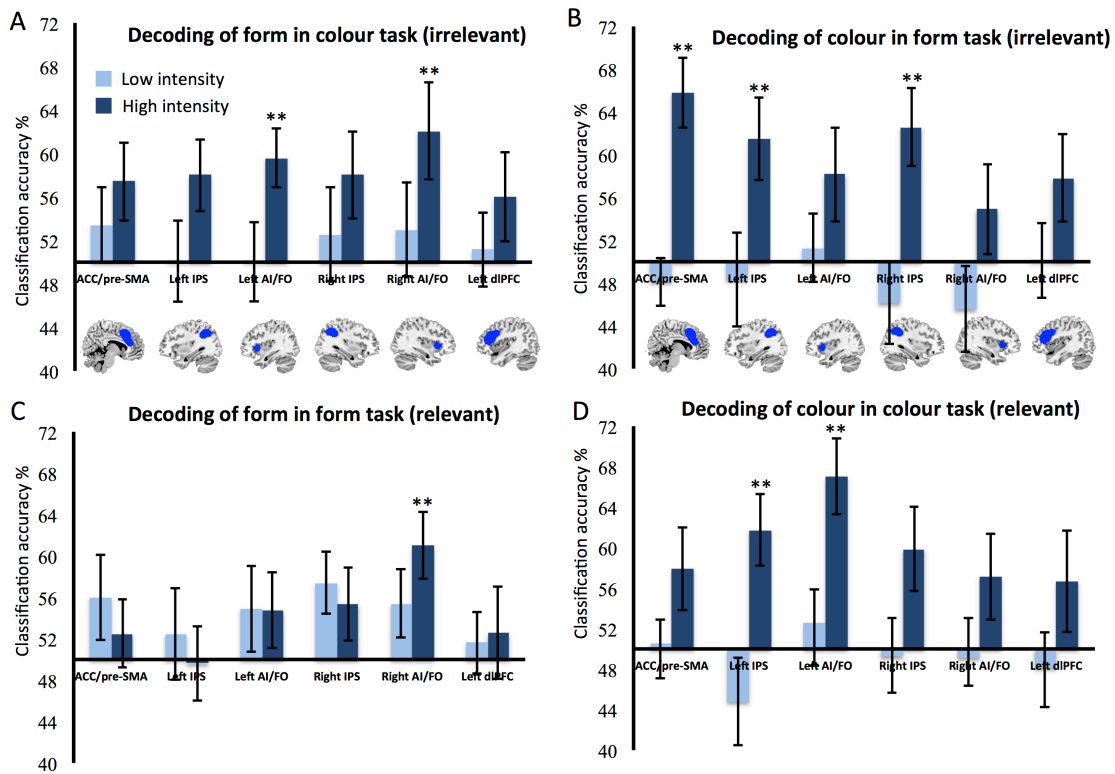
Figure 7: Decoding results in MD regions (not including right dlPFC): Stronger coding of *irrelevant* features across MD regions in HIS over LIS (ANOVA). Stronger coding of *relevant* colour information across MDs in HIS over LIS (post-hoc paired t tests). Error bars indicate standard error. The significance markings for individual bars indicate whether coding was significantly greater than chance in each condition separately (one sample t-test against chance, 50%). **p<0.008 (i.e., 0.05 / 6 to correct for multiple comparisons).

There was no main effect of *Feature* (F(1,19) = 0.21, p = 0.66) or *Region* (F(5,95) = 0.49, p = 0.78), and no interactions between *TMS* and *Feature* (F(1,19) = 1.27, p = 0.27), *TMS* and *Region* (F(5,95) = 0.31, p = 0.91) or *TMS*Feature*Region* (F(5,95) = 0.91, p = 0.48). The data therefore indicated a network-wide effect in which TMS increased coding of irrelevant information.

The stronger coding of irrelevant features in HIS over LIS is only interpretable if coding in one or more of the conditions is also significantly above chance. Therefore, we conducted one-sample t-tests against the classification accuracy expected by chance (50%) for HIS and LIS conditions separately. We found that the MD regions coded the irrelevant feature information significantly under HIS ($t(19) = 5.19$, $p < 0.01$) whereas classification of this information under LIS was at chance ($t(19) = 0.05$, $p = 0.45$) (Figure 8, Panel A). Thus, the results support our prediction that the right dlPFC is involved in inhibiting irrelevant feature information, and disruption to this region increases the extent to which irrelevant information is available through the MD system.

Decoding relevant information

Decoding results for relevant information are presented in Figure 7 (Panels C and D). Classification accuracies for relevant feature information were entered into a three-way ANOVA as described above. Our prediction here was that if the dlPFC is involved in supporting coding of relevant information, the disrupting processing in this region should result in a decrease in coding of relevant information. Therefore, we predicted a main effect of *TMS,* with more coding in LIS relative to HIS.

The ANOVA revealed a main effect of *TMS* ($F(1,19) = 10.94$, $p < 0.01$) but the effect was in the opposite direction to our prediction, with stronger coding of task relevant information in HIS (57.2%) compared to LIS (51.9%), as for irrelevant information. The main effect was modulated by a significant interaction between *TMS* and *Feature* ($F(1,19) = 5.84$, $p = 0.03$). Post-hoc paired t-tests showed stronger coding of relevant feature information in the HIS condition (60.1%) compared to the LIS condition for colour (49%, $t(19) = 3.44$, $p < 0.01$), but no difference in coding of relevant information between the HIS condition (54.3%) and the LIS condition for form (54.6%, $t(19) = 0.15$, $p = 0.46$). No other main effects or interactions were significant (all ps > 0.17).

227
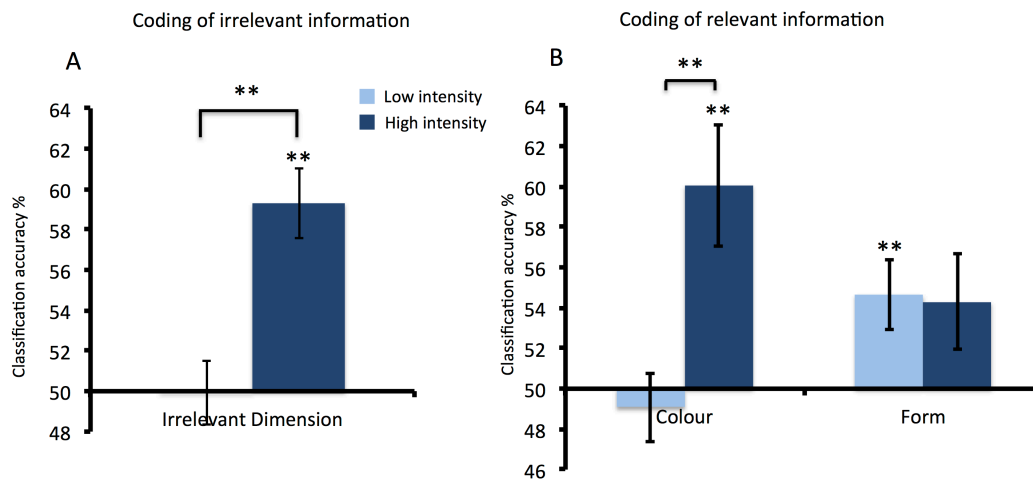
Figure 8: Coding of irrelevant (panel A) and relevant (panel B) feature information in MD regions (collapsed) under LIS and HIS. Error bars indicate standard error. The significance markings for individual bars indicate whether coding was significantly greater than chance in each condition separately (one sample t-test against chance, 50%). Markings between bars indicate where coding was significantly greater under HIS compared to LIS **p<0.01.

We conducted one-sample t-tests against the classification accuracy expected by chance (50%) for HIS and LIS conditions and relevant colour and form separately. We found that the MD regions coded the relevant colour information significantly under HIS (mean accuracy across all regions 60.1%; t(19) = 3.26, p < 0.01) whereas classification of this information under LIS was at chance (mean accuracy across all regions 49.1%; t(19) = 0.52, p = 0.31) (Figure 8, Panel B). Classification of form relevant information did not quite reach significance under HIS (54.3%; t(19) = 1.77, p = 0.06) but was significant under LIS (54.6%; t(19) = 2.64, p < 0.01).

Right dlPFC

Decoding results for the stimulated right dlPFC region are presented in Figure 9. To assess changes in coding in this region, we conducted a two-way ANOVA including

factors *TMS* (HIS, LIS) and *Feature* (Colour, Form) for the irrelevant and relevant information separately. For relevant information, there were no significant effects of *TMS* or *Feature* (all main effects and interaction p > 0.55). One sample t-tests showed no significant coding of irrelevant information under LIS or HIS (all ps > 0.15). For irrelevant information, there were again no significant effects of *TMS* or *Feature* (all main effects and interaction p > 0.13). One sample t-tests showed no significant coding of relevant or irrelevant information under LIS or HIS (all ps > 0.17).
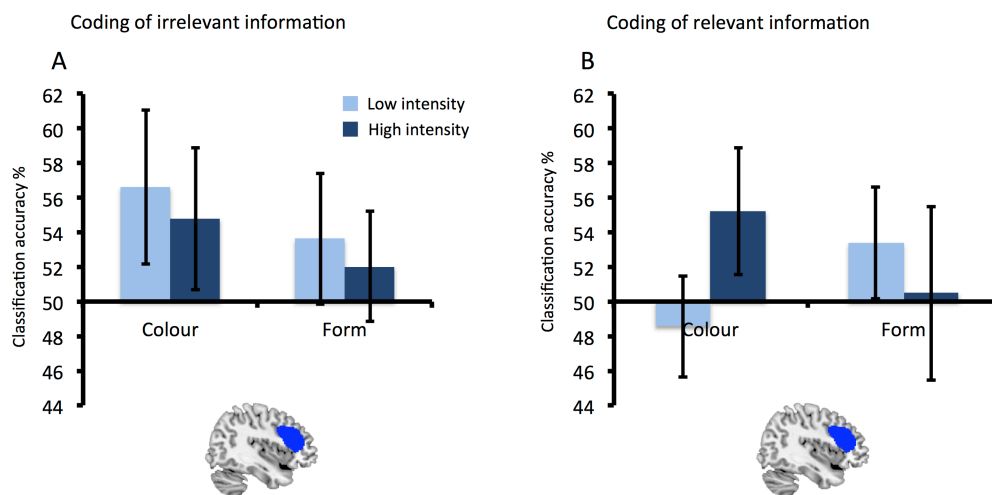
Figure 9: Coding of irrelevant (panel A) and relevant (panel B) feature information in right dlPFC (stimulated region) in HIS and LIS: Error bars indicate standard error.

We did not find evidence to suggest that information coding changes under HIS compared to LIS in right dlPFC for either relevant or irrelevant information. This pattern of results is in contrast with the results from analysing the rest of the MD frontoparietal network, which showed a clear change in the strength of coding as a result of disruption.

Lateral Occipital Complex (LOC)

Decoding irrelevant information

Decoding results for irrelevant information in LOC are presented in Figure 10, Panel A. The ANOVA was conducted as for the right dlPFC. We predicted that LOC would show stronger coding of irrelevant information with HIS than with LIS to dlPFC.
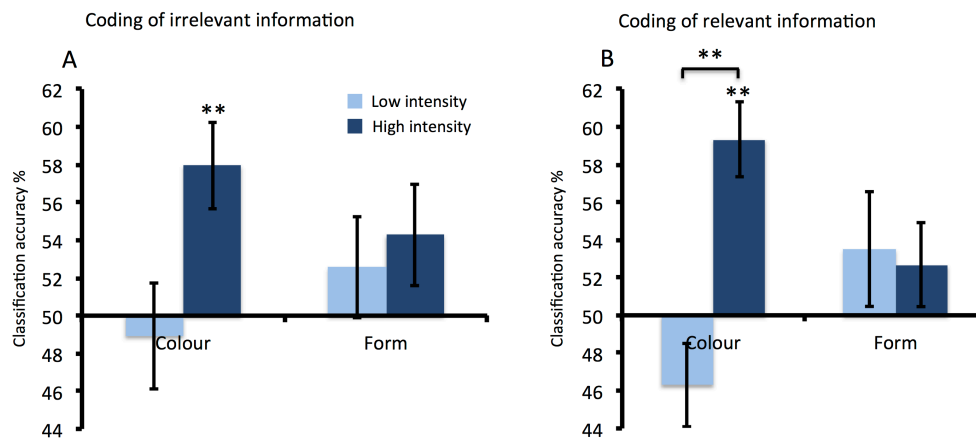
Figure 10: Coding of irrelevant (panel A) and relevant (panel B) feature information in LOC in HIS and LIS: Paired t-tests revealed stronger coding of relevant colour information (left panel) in HIS compared to LIS. Error bars indicate standard error. The significance markings for individual bars indicate whether coding was significantly greater than chance in each condition separately (one sample t-test against chance, 50%). **p<0.01.

There was a trend in this direction (Figure 10, right) that did not quite reach significance (main effect of *TMS* F(1,19) = 4.41, p = 0.051). There was no main effect of *Feature* (F(1,19) = 0.01, p = 0.99) and no interaction between *TMS* and *Feature* (F(1,19) = 1.78, p = 0.21). One sample t-tests showed that classification of the irrelevant colour information was significantly above chance under HIS (57.9%, t(19) = 3.45, p < 0.01). No other t-tests were significant (all ps > 0.07).

230

Decoding relevant information

Decoding results for relevant information in LOC are presented in Figure 10, Panel B. Classification accuracies were entered into a two-way ANOVA as above. We predicted that LOC would show weaker coding of relevant information under HIS relative to LIS to dlPFC. The ANOVA revealed a main effect of *TMS* ($F(1,19) = 6.92$, $p = 0.02$) modulated by a significant interaction between *TMS* and *Feature* ($F(1,19) = 9.71$, $p < 0.01$). Post hoc paired t-tests showed that for colour, there was stronger coding of relevant colour feature information in the HIS condition (59.3%) compared to the LIS condition (49.3%, $t(19) = 4.83$, $p < 0.01$), opposite to our prediction and in line with the results from the MD network. There was no difference in coding of relevant form information between the HIS condition (52.7%) and the LIS condition (53.5%, $t(19) = 0.23$, $p = 0.82$). There was no main effect of *Feature* ($F(1,19) = 0.01$, $p = 0.91$). One sample t-tests showed that classification of the relevant colour information was significantly above chance under HIS (59.3%, $t(19) = 4.71$, $p < 0.01$). No other t-tests were significant (all ps > 0.07). In this case, LOC followed a similar pattern to that of the MD network, coding more strongly for relevant colour information under HIS.

Visual fixation ROI

Decoding irrelevant information

Decoding results for irrelevant information in the ROI at fixation are presented in Figure 11, Panel A. The ANOVA was conducted as above. For this ROI we expected to see stronger coding of irrelevant feature information under HIS relative to LIS, reflecting disruption to an inhibiting dlPFC mechanism. However, there were no significant effects of *TMS* or *Feature* (all main effects and interaction, $p > 0.23$). One sample t-tests showed

that classification of irrelevant colour information was significantly above chance under HIS (57.4%, t(19) = 3.01, p < 0.01). No other t-tests were significant (all ps > 0.06).

Decoding relevant information

Decoding results for relevant information in the fixation ROI are presented in Figure 11, Panel B. The ANOVA was conducted as above. The ANOVA showed no main effect of *TMS* (F(1,19) = 0.04, p = 0.84) or *Feature* (F(1,19) = 3.93, p = 0.06), and no interaction between *TMS* and *Feature* (F(1,19) = 2.36, p = 0.14).
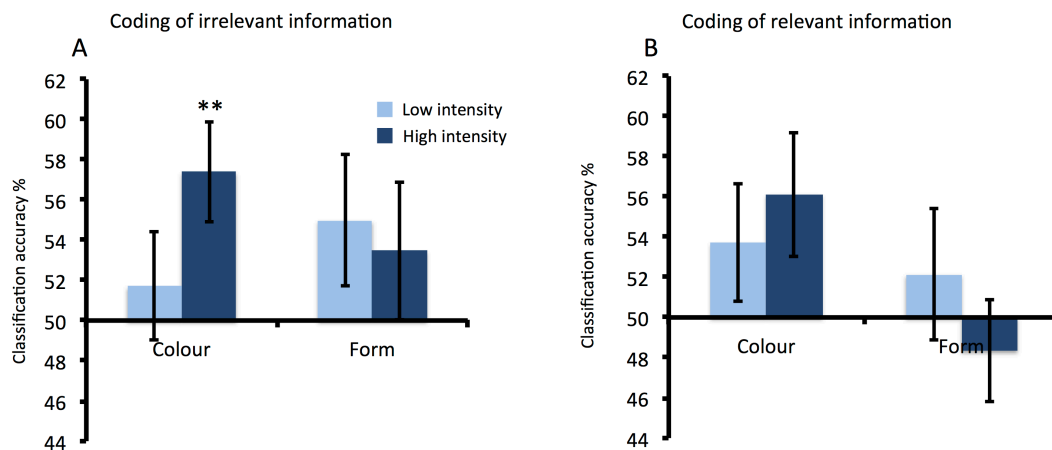
Figure 11: Coding of irrelevant (panel A) and relevant (panel B) feature information in the visual fixation ROI in HIS and LIS. Error bars indicate standard error. The significance markings for individual bars indicate whether coding was significantly greater than chance in each condition separately (one sample t-test against chance, 50%). **p<0.01.

Coding of relevant colour information under HIS was above chance but did not pass correction for multiple comparisons (p = 0.03). Coding was not significantly above chance in any other condition (all ps > 0.07).

Whole brain pattern classification

Overall, the results from our ROIs showed that there was stronger coding of irrelevant features across the MD network following disruption to right dlPFC, providing

support for an inhibitory mechanism. Coding of irrelevant information also tended to be stronger in the fixation ROI and LOC under HIS, but the difference from LIS was not significant. Interestingly, against our prediction that disruption to dlPFC would result in weaker coding of task-relevant information across the system, we saw either no effect (form task), or the opposite result (colour task) for the frontoparietal regions and LOC.

We checked for additional regions showing multivoxel coding of relevant and irrelevant feature information by performing the decoding analysis across the whole-brain using a roaming searchlight (Kriegeskorte et al., 2006). Since the searchlight analysis is performed on the same data as the ROI analyses, corresponding results are not surprising (although, since searchlights are typically smaller than ROIs, the comparison may inform the spatial scale of the multi-voxel patterns). In addition, a searchlight analysis is typically less sensitive than a ROI-based approach due to the need to correct for multiple comparisons. However, the advantage of this approach is that is it free from *a priori* spatial hypotheses, meaning we can potentially identify additional regions missed by the ROI approach.

We designed the searchlight analyses to mimic the key components of our ROI analysis. For irrelevant information, we performed searchlights comparing coding to chance in the HIS and LIS condition separately. Then, we conducted paired t-tests to directly compare coding between the two stimulation conditions. For relevant information, since our ROI analyses had suggested different results in the two tasks, we carried out the same analysis, comparing coding to chance in HIS and LIS, but kept colour and form data separate. We again conducted paired t-tests to directly compare coding between the HIS and LIS condition.
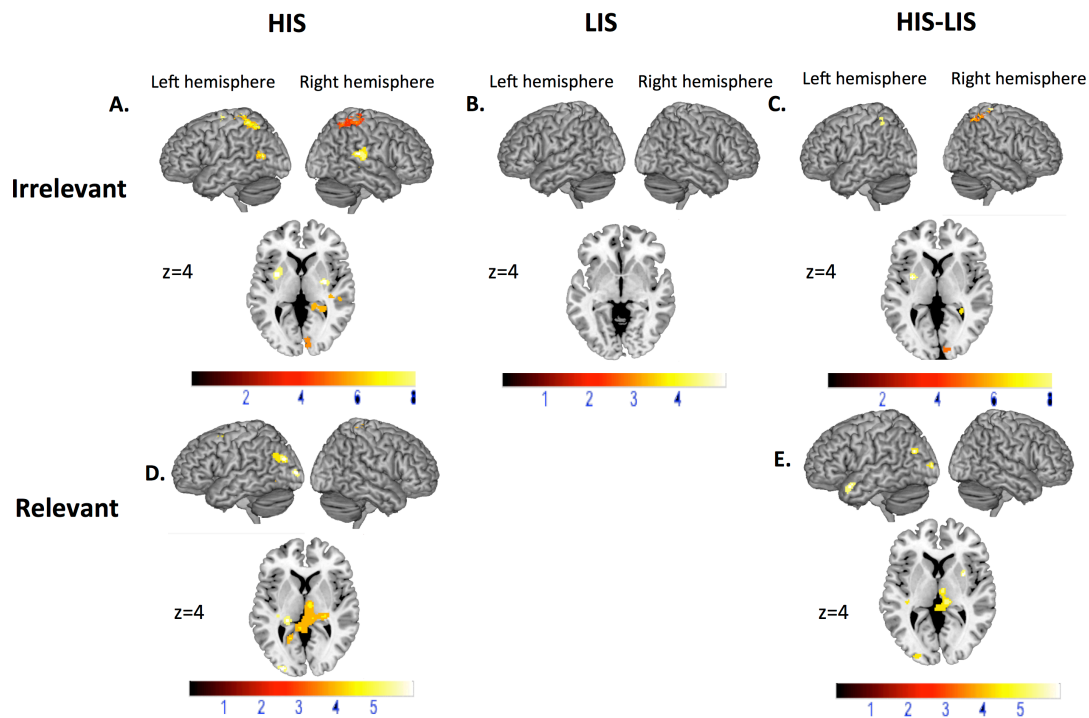
Figure 12: Feature coding assessed with a roaming searchlight. Whole-brain maps show voxels where patterns of activation in the local neighborhood (5 mm sphere) discriminated **(A)** irrelevant feature information under HIS; **(B)** irrelevant feature information under LIS; **(C)** irrelevant feature information in HIS minus LIS; **(D)** relevant colour information under HIS; **(E)** relevant colour information in HIS minus LIS. Results are thresholded clusterwise at p <0.05 with FWE for multiple comparisons and an additional extent threshold of 40 voxels. Coordinates of peak decoding are given in Tables 1 and 2.

For irrelevant coding under LIS, the only cluster of significant coding was in the cerebellum (Table 1, Figure 12, panel B). By contrast, under HIS, significant coding of irrelevant information was seen across the brain including several visual cortical regions, temporal regions, precuneus, postcentral gyrus, orbitofrontal cortex, putamen, supplementary motor area, posterior cingulate cortex, and cingulate gyrus, as well as within our ROI ACC-pre/SMA (Table 1 and Figure 12, panel A). The direct comparison

234

of these maps revealed stronger coding of irrelevant feature information under HIS compared to LIS in several occipital regions, the precentral gyrus, putamen and within our left IPS ROI (Table 1 and Figure 12, panel C). There were no significant clusters in the reverse comparison (LIS − HIS).

Coding of relevant colour feature information under HIS confirmed our ROI analysis revealing clusters in lateral occipital regions but additionally in several other regions including superior anterior prefrontal, postcentral gyrus, early visual cortex, hippocampus, fusiform gyrus, middle temporal areas, and frontal eye fields (Table 2 and Figure 12, panel D). There were no significant clusters under LIS. Stronger coding of relevant colour information under HIS compared to LIS was observed in occipital lobes (lateral occipital, fusiform gyrus, lingual gyrus), subcortical regions (hippocampus, putamen, thalamus, cerebellum), and superior temporal pole (Table 2 and Figure 12, panel E). There were no significant clusters in the reverse comparison (LIS − HIS). For coding of relevant form information, there were no significant clusters under any contrast.

Contrast	Lobe	Cluster	Hemisphere	Peak coordinates			Brodmann area	Cluster size	t
				x	y	z			
Coding of irrelevant feature information under LIS	Subcortical	Cerebellum	left	-12	-30	-26	30	49	5.8
Coding of irrelevant feature information under HIS	Frontal	ACC/pre-SMA	right	4	24	56	8	174	7.1
		Medial orbitofrontal	right	6	30	-12	11	48	6.6
		Supplementary motor area	right	2	-2	72	6	206	6.1
	Parietal	Precuneus	right	6	-42	60	5	2136	11.8
		Postcentral gyrus	left	-22	-38	70	3	634	7.5
	Visual	Lingual gyrus	right	12	-58	-2	18	1391	8.1
		Calcarine sulcus	right	10	-90	4	17	76	5.1
	Temporal	Middle temporal gyrus	right	52	-36	-4	21	340	7.8
		Superior temporal	left	-50	-42	14	42	74	5.7
		Middle temporal gyrus	left	-48	-66	12	37	113	7.7
	Subcortical	Cingulate gyrus	left	-2	-14	46	23	59	5.4
		Posterior cingulate gyrus	left	-6	-52	34	23	49	6.1
		Putamen	right	30	-8	4	48	261	6.5
		Putamen	left	-32	2	6	48	184	6.3
Coding of irrelevant feature information in HIS minus LIS	Parietal	Precentral gyrus	right	18	-30	70	4	1074	8.54
		IPS	left	-44	-48	52	40	79	5.7
	Visual	Lingual gyrus	right	28	-46	-4	37	231	6.82
		Superior occipital gyrus	left	-20	-76	38	19	58	6.8
		Calcarine sulcus	right	16	-92	6	17	46	5.2
	Subcortical	Putamen	left	-32	0	2	48	57	6.1

Table 1: Peak coordinates for whole-brain searchlights: irrelevant feature information in HIS minus LIS, irrelevant feature information under HIS and irrelevant feature information under LIS.

Contrast	Lobe	Cluster	Hemisphere	Peak coordinates			Brodmann area	Cluster size	t
				x	y	z			
Coding of relevant colour information under HIS	Frontal	Superior anterior prefrontal	left	-4	54	28	10	1754	7.64
		frontal eye fields	left	-32	4	56	8	157	6.9
	Parietal	Postcentral gyrus	right	12	-40	70	3	206	7.6
	Visual	Middle occipital	left	-44	-78	26	39	155	7.1
		Lingual gyrus	left	-6	-78	-10	17	371	6.6
		Lateral mid occipital	left	-28	-94	6	18	75	6.4
	Temporal	Fusiform gyrus	right	26	-40	-16	37	261	6.2
		Middle temporal gyrus	left	-52	-22	-2	21	114	5.6
	Subcortical	Hippocampus	left	-38	-22	-10	20	442	6.3
Coding of relevant colour information in HIS minus LIS	Visual	Lateral mid occipital	left	-24	-94	8	18	77	7
		Lateral mid occipital	left	-44	-72	28	39	64	6.9
		Lingual gryus	left	-6	-80	-10	17	144	6.5
	Temporal	Superior temporal pole	left	-48	12	-20	38	97	5.8
		Fusifrom gyrus	left	-34	-60	-6	37	147	6.7
	Subcortical	Hippocampus	left	-36	-24	-8	20	224	6.8
		Thalamus	right	6	-14	6		606	7.4
		Putamen	right	32	6	4	48	58	5.8
		Cerebellum	right	30	-54	-20	37	133	5.7

Table 2: Peak coordinates for whole-brain searchlights. Relevant colour information in HIS minus LIS and relevant colour information under HIS

Discussion

The dlPFC is thought to be crucial for goal-directive behaviour, providing the source of top-down signals that modulate processing in earlier cortical regions (Desimone & Duncan, 1995; Miller & Cohen, 2001). It is part of a circuit of frontal and parietal brain regions, referred to as multiple-demand (MD) regions (Duncan, 2010), or the frontoparietal attentional network (Fox et al., 2005; Vincent et al., 2008), that are thought to play a fundamental role in attentional and executive control. Non-human primate research has shown that these brain regions preferentially code task relevant information (Cromer et al., 2010; Everling et al., 2002; Freedman et al., 2001; Kadohisa et al., 2013; Rao et al., 1997; Roy et al., 2010), as has human neuroimaging (e.g. Chapter 2, Chapter 3, also Woolgar, Williams, et al., 2015). These regions are also shown to have a causal effect in more specialised brain regions such as the visual cortex (e.g. Feredoes et al., 2011; Higo et al., 2011; Lee & D'Esposito, 2012; Miller et al., 2011; Zanto et al., 2011). In addition, top-down effects, such as explicitly allocating attention, can affect information coding in visual cortex (e.g. Jehee et al., 2011). However, to our knowledge no work has explicitly linked disruption to MD function with information coding in a causal framework.

Here we examined whether disrupting right dlPFC with TMS affected the strength of coding of irrelevant and relevant feature information in the brain. We predicted that, relative to our control low intensity stimulation (LIS), disruptive high intensity stimulation (HIS) would result in **a)** increased coding of the irrelevant feature information; and/or **b)** decreased coding of the relevant feature information in the rest of the brain. Understanding the role of dlPFC in this type of top-down control will further our understanding of how the processing of task relevant information comes to be prioritised in the brain.

Following HIS to the dlPFC, we observed an *increase* in coding of *irrelevant* information in the rest of the MD system, in line with our prediction. One of the strengths

of this paradigm is that the causal effect of TMS to dlPFC can be used to examine top-down effects on coding in visual brain regions. This is because any task-mediated changes in non-stimulated regions must be due to the change caused by TMS to right dlPFC. The results in LOC and the fixation ROI showed the same trend as the MD regions, both coding irrelevant information under HIS but not under LIS. Although the difference in the visual ROIs was not significant, the searchlight revealed three clusters in the visual cortex (calcarine sulcus, lingual gyrus, superior occipital gyrus) where there was significantly more coding of irrelevant information under HIS. These data are in line with the suggestion that TMS caused irrelevant information to be released from suppression normally exerted by the dlPFC. Our searchlight results suggest that this was a widespread effect encompassing much of the frontal and parietal cortex and even early visual cortex. In our behavioural data, we observed an increase in the behavioural congruency effect under HIS: participants were faster to respond when the irrelevant feature directed them to the correct response, compared to when it directed them to the incorrect response, but only under HIS. These data therefore support the interpretation that TMS to the dlPFC disrupted participants' ability to ignore the irrelevant information which was mirrored in the decoding results. The data provide causal evidence for a right dlPFC filtering mechanism; disruption to this mechanism caused a stronger representation of irrelevant feature information across many regions of the brain, and had a corresponding effect on behaviour. This also provides support for predictions of the adaptive coding and biased competition models (Desimone, 1998; Desimone & Duncan, 1995; Duncan, 2001), which suggest that selective responses in higher regions bias brain responses in earlier cortical areas.

For task-relevant information, TMS had significantly different effects on the coding of form and colour, so we discuss them separately here. If dlPFC normally

facilitates the coding of relevant information, then disrupting dlPFC function should disrupt that facilitation. For both features, we therefore predicted a *decrease* in *relevant* information coding after HIS but not LIS TMS. However, for form we did not see any effect of TMS on relevant information coding, and for colour we actually saw the opposite effect. For colour, we found *stronger* coding of relevant colour information in HIS compared to LIS. This effect was seen in the rest of the MD system (with no significant differences between different MD regions), the LOC, and several other temporal, occipital and subcortical regions. One possible explanation is that the unexpected *increase* in relevant feature coding reflects a compensation mechanism. This seems particularly likely in the rest of the MD network since these regions are thought to be adaptive (Duncan, 2001).

Previous research has shown that the MD regions increase their coding of task relevant information when the system is challenged. For example, we have previously shown that the MD regions represent task-relevant visual objects, but only when the stimuli were degraded with Gaussian noise such that the physical input, and the representation of the stimuli in visual cortices, was weak (Woolgar, Williams, et al., 2015). Similar results have been reported when visual stimuli are highly confusable (Woolgar, Hampshire, et al., 2011), or when stimulus-response mapping rules are complex (Woolgar, Afshar, Williams, & Rich, 2015); in all these cases coding was weak or absent when the task was easy, but increased significantly when the task was more difficult. We speculate that this is what is driving our relevant coding result: HIS to the dlPFC activated a similar mechanism in the rest of the MD system: the input became more akin to a more difficult visual stimulus (Harris, Clifford, & Miniussi, 2008), triggering an adaptive increase in coding of task-relevant information in the rest of the system.

Previous research has indicated that TMS disruption to a brain region can result in potential compensation mechanisms in other brain areas, particularly in the contralateral hemisphere (Andoh & Paus, 2011; Lee & D'Esposito, 2012; O'Shea, Johansen-Berg, Trief, Göbel, & Rushworth, 2007), providing general support for this type of interpretation of our data. For example, O'Shea et al. (2007) applied 1Hz offline rTMS over the left premotor cortex and used fMRI to record activity from the right premotor cortex. They found an increase in activity in the right premotor cortex and connected premotor areas following disruption to left premotor cortex. Subsequent TMS of right premotor cortex disrupted performance, confirming that the pattern of functional reorganisation of the right premotor cortex made a causal contribution in preserving behaviour after neuronal interference.

Compensation for disruption is also often reported in patient data, where increased activity in the intact hemisphere may reflect a compensatory response for the loss of function in damaged frontal lobe tissue (Buckner, Corbetta, Schatz, Raichle, & Petersen, 1996; Corbetta, Kincade, Lewis, Snyder, & Sapir, 2005; Voytek et al., 2010). Most relevant here is prior research showing that the frontoparietal network over-activates in response to damage to part of the network (Woolgar, Bor, & Duncan, 2013). Note that in our data there was no overall increase in activity in the MD regions, but rather an increase in information coding. Further work is needed to understand how changes in information coding relate to the type of compensatory mechanisms inferred based on patient and univariate changes. In our case, we also observed an increase in coding of relevant information elsewhere in the brain, particularly in visual cortices. This could be a direct result of TMS stimulation of right dlPFC or may have been mediated by the increase in coding of relevant information in the MD system.

The results showed an increase in coding for irrelevant information as well as relevant colour information in the frontoparietal cortices and visual cortical areas. One interpretation, therefore, is that HIS simply causes a general increase in coding of information across the brain. Two observations caution against this. First, the searchlight revealed that although the effect was widespread, it was far from encompassing the whole brain. Similarly, in the ROI analysis we did not observe the same effect in the stimulated region. This suggests that the results were not due to a very general mechanism or artefact of the analysis. Second, we did not observe stronger coding of task-relevant form information, but only colour information. It is difficult to reconcile the interpretation that HIS increases information coding in general with this task-sensitive effect. Nonetheless, further work is needed to explore the relationship between the TMS disruption and pattern information. In our data we have clear evidence that HIS to dlPFC affects coding of both relevant and irrelevant information, and no evidence for a differential effect.

It was interesting that the increased behavioural congruency effect in the HIS condition appeared to be driven as much, if not more, by an increased facilitation in the congruent condition as by an increased interference in the incongruent condition. It would have been interesting to examine coding of relevant and irrelevant information in congruent and incongruent conditions separately to explore this further. However, it was not possible to do this with the present data because decoding of relevant or irrelevant information in congruent/incongruent conditions was confounded with the other dimension. For example, decoding of irrelevant colour information (blue vs. green) in congruent trials always corresponded with decoding of relevant form information (cuby vs. smoothy). This may be an important focus for further studies.

In our data, we failed to find any effect of TMS on information coding in the right dlPFC itself. We predicted a decrease in information coding in this region, for example, due to increased neuronal noise induced by HIS (Walsh & Cowey, 2000, although see, Harris et al., 2008; Ruff, Driver, & Bestmann, 2009; Sandrini, Umiltà, & Rusconi, 2011). In fact, we did not observe significant coding of either type of information under LIS or HIS. The right dlPFC ROI was large (we used this definition to account for any variability in the target site between subjects) and this may have washed out the effect of stimulation, for example if it affected only a small portion of the ROI. Another possibility is that the BOLD response directly at the site of stimulation is not sensitive to the direct effect of stimulation; for example, if its primary effect is on the axons (e.g. Siebner, Hartwigsen, Kassuba, & Rothwell, 2009), while the BOLD response primarily indexes post-synaptic potentials (e.g. Logothetis, 2008). This would imply that only the indirect effect of stimulation (i.e. the effect on other brain regions) would be observable with BOLD signal.

Another limitation with TMS stimulation is that factors such as minimal changes in coil orientation might alter whether the frontoparietal network or default mode network is affected (Opitz, Fox, Craddock, Colcombe, & Milham, 2016). This makes null results difficult to interpret in behavioural TMS studies, but a promising aspect of concurrent TMS-fMRI is that it has the potential to reveal the stimulated network (Bestmann et al., 2008). In our case, we have evidence from both the ROI and searchlight analyses that we successfully influenced information processing in the MD network.

In this experiment, we used a control condition of LIS as this allowed the two TMS conditions to be presented on a trial-by-trial basis. This makes the upcoming level of stimulation unpredictable to the participant, and is preferable because fMRI scanner drift makes it preferable to always compare events happening close together in time. Other

controls such as sham or different stimulation site would have involved removing the participant from the scanner and therefore splitting the session in time, or carrying out an additional scanning session. LIS controls for non-specific effects of TMS (noise artefact, scalp sensation, etc.) and targets the same ROI as the main stimulation site, and is therefore a more conservative control than a no-TMS comparison. However, the use of this control cannot specifically disentangle the function of the dlPFC compared to the function of other brain regions as would be possible with a control stimulation site. We cannot rule out that similar results could be seen with disruption to another region of the MD network, or indeed another brain region altogether. Future research could use a control region to examine whether the effect we have seen here is specific to the right dlPFC. It would also be critical for determining whether the compensatory interpretation holds regardless of which MD region is disrupted.

A final, important consideration is whether the differences in RT could have driven artefactual differences in decoding. If conditions on either side of classification differ in difficulty, for example if one has longer RTs, this could drive decoding performance (Todd et al., 2013; Woolgar et al., 2014). In our case, however, differences in RTs for the congruent and incongruent conditions could not have contributed to decoding because all object features were equally represented in both the congruent and incongruent conditions. In addition, to account for trial by trial differences in RT (Todd et al., 2013), trials were modelled as epochs lasting from stimulus onset until response in the GLM to removes the effect of RT from the estimate of activity (Woolgar, Williams, et al., 2015), as in our previous work (e.g. Woolgar, Afshar, et al., 2015; Woolgar, Hampshire, et al., 2011; Woolgar, Thompson, et al., 2011).

This experiment is one of the first to combine MVPA with concurrent TMS-fMRI. The combination of these approaches provides an opportunity to link information coding and behavioural effects in a causal framework. We found that perturbing the right dlPFC with TMS affected information coding in the rest of the MD network and in the visual cortex. The data suggest a complex role of the dlPFC in representing relevant and irrelevant information. In line with a role for the dlPFC in suppressing irrelevant information (e.g. Shimamura, 2000), coding of irrelevant information was stronger in the rest of the (non-stimulated) MD network and in some regions of the visual cortex under HIS compared to LIS. However, against our prediction that disruption to a selection mechanism would result in *weaker* coding of relevant information across the brain, the representation of task-relevant colour information also increased under HIS in these areas. We speculate that this may reflect an adaptive or compensatory response to the challenge imposed on the MD system, but cannot rule out a more general effect of HIS to dlPFC in which coding of all task features is increased in task-related brain regions. Under either interpretation the data provide causal evidence that the right dlPFC is involved in modulating the representation of relevant and irrelevant feature information in the brain. More broadly, the results demonstrate interactions between the MD network and specialised cortex, providing evidence in support of major theories of executive control and attention, and demonstrate methods that could be used to address the predictions of these theories in more detail in the future.

Funding

Acknowledgments

References

Andoh, J., & Paus, T. (2011). Combining functional neuroimaging with off-line brain stimulation: modulation of task-related activity in language areas. *Journal of Cognitive Neuroscience, 23*(2), 349-361.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The neuroscientist, 13*(3), 214-228.

Bestmann, S., Baudewig, J., & Frahm, J. (2003). On the synchronization of transcranial magnetic stimulation and functional echo‐planar imaging. *Journal of Magnetic Resonance Imaging, 17*(3), 309-316.

Bestmann, S., Baudewig, J., Siebner, H. R., Rothwell, J. C., & Frahm, J. (2005). BOLD MRI responses to repetitive TMS over human dorsal premotor cortex. *NeuroImage, 28*(1), 22-29.

Bestmann, S., & Feredoes, E. (2013). Combined neurostimulation and neuroimaging in cognitive neuroscience: past, present, and future. *Annals of the New York Academy of Sciences, 1296*(1), 11-30.

Bestmann, S., Ruff, C. C., Blankenburg, F., Weiskopf, N., Driver, J., & Rothwell, J. C. (2008). Mapping causal interregional influences with concurrent TMS–fMRI. *Experimental Brain Research, 191*(4), 383-402.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage, 16*(2).

Buckner, R. L., Corbetta, M., Schatz, J., Raichle, M. E., & Petersen, S. E. (1996). Preserved speech abilities and compensation following prefrontal damage. *Proceedings of the National Academy of Sciences, 93*(3), 1249-1253.

Chafee, M. V., & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of Neurophysiology, 83*(3), 1550-1566.

Chang, C. C., & Lin, C. J. (2011). LIBSVM a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*(3), 27.

Chen, X., Hoffmann, K. P., Albright, T. D., & Thiele, A. (2012). Effect of feature-selective attention on neuronal responses in macaque area MT. *Journal of Neurophysiology, 107*, 1530-1543.

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience, 16*(9), 1348-1355. doi:10.1038/nn.3470

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., & Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nature Neuroscience, 8*(11), 1603-1610.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7*(9), 415-423.

Cusack, R., Mitchell, D. J., & Duncan, J. (2010). Discrete object representation, attention switching, and task difficulty in the parietal lobe. *Journal of Cognitive Neuroscience, 22*, 32-47.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology, 66*, 115.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353*(1373), 1245-1255.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18,* 193-222.

Driver, J., Blankenburg, F., Bestmann, S., Vanduffel, W., & Ruff, C. C. (2009). Concurrent brain-stimulation and neuroimaging for studies of cognition. *Trends in Cognitive Sciences, 13*(7), 319-327.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2,* 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50.

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

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobes recruited by diverse cognitive demands. *Trends in Neuroscience, 23,* 475-483.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5*(7), 671-676. doi:10.1038/nn874

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences, 110*(41), 16616-16621.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences, 108*(42), 17510-17515.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, 102*(27), 9673-9678. doi:10.1073/pnas.0504136102

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*(5502), 312-316.

Frith, C. (2001). A framework for studying the neural basis of attention. *Neuropsychologia, 39*, 1367–1371.

Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain research, 330*(2), 299-307.

Grill-Spector, K., Kushner, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24*, 187-203.

Grill-Spector, K., Kushnir, T., Hendler, T., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience, 3*, 837-843.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage, 43*(3), 509-520.

Harris, J. A., Clifford, C. W., & Miniussi, C. (2008). The functional effect of transcranial magnetic stimulation: signal suppression or neural noise generation? *Journal of Cognitive Neuroscience, 20*(4), 734-740.

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

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci, 7*(7), 523-534.

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8*. doi:10.3389/fninf.2014.00088

Henson, R. (2007). Efficient experimental design for fMRI. *Statistical parametric mapping: The analysis of functional brain images*, 193-210.

Higo, T., Mars, R. B., Boorman, E. D., Buch, E. R., & Rushworth, M. F. (2011). Distributed and causal influence of frontal operculum in task control. *Proceedings of the National Academy of Sciences, 108*(10), 4230-4235.

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention Improves Encoding of Task-Relevant Features in the Human Visual Cortex. *Journal of Neuroscience, 31*(22), 8210-8219. doi:10.1523/jneurosci.6153-09.2011

Johnston, K., & Everling, S. (2006). Monkey dorsolateral prefrontal cortex sends task-selective signals directly to the superior colliculus. *The Journal of Neuroscience, 26*(48), 12471-12478.

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., . . . Duncan, J. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80*(1), 235-246.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91-100.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience, 23*(1), 315-341. doi:doi:10.1146/annurev.neuro.23.1.315

Knight, R. T., Staines, W. R., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica, 101*(2), 159-178.

Knight, R. T., & Stuss, D. T. (2002). Prefrontal cortex: The present and the future. *Principles of frontal lobe function*, 573-597.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*(10), 3863-3868.

Lee, T. G., & D'Esposito, M. (2012). The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI–TMS study. *The Journal of Neuroscience, 32*(44), 15458-15466.

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878. doi:10.1038/nature06976

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel pattern analysis for fMRI data: A review. *Computational and mathematical methods in medicine, 2012*.

Miller, B. T., & D'Esposito, M. (2005). Searching for "the Top" in Top-Down Control. *Neuron, 48*(4), 535-538. doi:http://dx.doi.org/10.1016/j.neuron.2005.11.002

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience, 23*(1), 1-10.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience, 16*(16), 5154-5167.

Monosov, I. E., Sheinberg, D. L., & Thompson, K. G. (2011). The effects of prefrontal cortex inactivation on object responses of single neurons in the inferotemporal cortex during visual search. *The Journal of Neuroscience, 31*(44), 15956-15961.

Morishima, Y., Akaishi, R., Yamada, Y., Okuda, J., Toma, K., & Sakai, K. (2009). Task-specific signal transmission from prefrontal cortex in visual selective attention. *Nature Neuroscience, 12*(1), 85-91.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social cognitive and affective neuroscience*, nsn044.

O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT–MST. *Neuron, 18*(4), 591-598.

O'Shea, J., Johansen-Berg, H., Trief, D., Göbel, S., & Rushworth, M. F. (2007). Functionally specific reorganization in human premotor cortex. *Neuron, 54*(3), 479-490.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *Journal of Neuroscience, 26*(50), 13025-13036. doi:10.1523/jneurosci.2481-06.2006

Opitz, A., Fox, M. D., Craddock, R. C., Colcombe, S., & Milham, M. P. (2016). An integrated framework for targeting functional networks via transcranial magnetic stimulation. *NeuroImage, 127*, 86-96.

Pandya, D. N., & Barnes, C. L. (1987). Architecture and connections of the frontal lobe.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *The Journal of Neuroscience, 23*(10), 3990-3998.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.

Rossi, A. F., Bichot, N. P., Desimone, R., & Ungerleider, L. G. (2007). Top–down attentional deficits in macaques with lesions of lateral prefrontal cortex. *The Journal of Neuroscience, 27*(42), 11306-11314.

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology, 120*(12), 2008-2039. doi:10.1016/j.clinph.2009.08.016

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Ruff, C. C., Bestmann, S., Blankenburg, F., Bjoertomt, O., Josephs, O., Weiskopf, N., . . . Driver, J. (2008). Distinct causal influences of parietal versus frontal areas on human visual cortex: evidence from concurrent TMS–fMRI. *Cerebral Cortex, 18*(4), 817-827.

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., . . . Driver, J. (2006). Concurrent TMS-fMRI and psychophysics reveal frontal

influences on human retinotopic visual cortex. *Current Biology, 16*(15), 1479-1488.

Ruff, C. C., & Driver, J. (2006). Attentional preparation for a lateralized visual distractor: behavioral and fMRI evidence. *Journal of Cognitive Neuroscience, 18*(4), 522-538.

Ruff, C. C., Driver, J., & Bestmann, S. (2009). Combining TMS and fMRI: from 'virtual lesions' to functional-network accounts of cognition. *Cortex, 45*(9), 1043-1049.

Sack, A. T., Kohler, A., Bestmann, S., Linden, D. E., Dechent, P., Goebel, R., & Baudewig, J. (2007). Imaging the brain activity changes underlying impaired visuospatial judgments: simultaneous FMRI, TMS, and behavioral studies. *Cerebral Cortex, 17*(12), 2841-2852.

Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neuroscience & Biobehavioral Reviews, 35*(3), 516-536.

Serences, J. T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences, 10*(1), 38-45. doi:http://dx.doi.org/10.1016/j.tics.2005.11.008

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*(2), 207-218.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex, 45*(9), 1035-1042.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657-1661.

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *NeuroImage, 77*, 157-165. doi:10.1016/j.neuroimage.2013.03.039

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature, 401*(6754), 699-703.

Ungerleider, L., Gaffan, D., & Pelak, V. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Experimental Brain Research, 76*(3), 473-484.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Voytek, B., Davis, M., Yago, E., Barceló, F., Vogel, E. K., & Knight, R. T. (2010). Dynamic neuroplasticity after human prefrontal cortex damage. *Neuron, 68*(3), 401-408.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-80.

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

Woolgar, A., Bor, D., & Duncan, J. (2013). Global increase in task-related fronto-parietal activity after focal frontal lobe lesion. *Journal of Cognitive Neuroscience, 25*(9), 1542-1552.

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. doi:10.1016/j.neuroimage.2014.04.059

Woolgar, A., Hampshire, A., Thompson, R., & Duncan, J. (2011). Adaptive Coding of Task-Relevant Information in Human Frontoparietal Cortex. *Journal of Neuroscience, 31*(41), 14592-14599. doi:10.1523/jneurosci.2616-11.2011

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multi-voxel coding of stimuli, rules, and responses in human frontoparietal cortex. *NeuroImage, 56*(2), 744-752. doi:10.1016/j.neuroimage.2010.04.035

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multi-voxel representation of novel objects in frontal, parietal and visual cortices. *NeuroImage, 109*, 429-437. doi:10.1016/j.neuroimage.2014.12.083

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*(5), 656-661.

Chapter 6

Discussion

The focus of this thesis has been on the critical ability to select between relevant and irrelevant information in a single object. I explored the key question of the way in which the brain might implement feature-selective attention. In particular, I focused on the multiple-demand (MD) regions as candidate regions (Cole & Schneider, 2007; Corbetta & Shulman, 2002; Duncan, 2001; Duncan, 2010; Kanwisher & Wojciulik, 2000; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008) and the adaptive coding hypothesis as a viable underlying mechanism (Duncan, 2001). The driving motivation of this thesis was to investigate whether adaptive coding provides a mechanism for feature-selective attention in the MD network, and whether this in turn modulates responses across the rest of the brain. In this chapter, I first review the findings from each study and then discuss implications and outstanding questions within the broader theoretical context. Finally I summarise the overall findings from this thesis.

Overview of findings

Chapter 2: Feature-selective attention in frontoparietal cortex: Multivoxel codes adjust to prioritize task-relevant information

In this chapter I presented a functional magnetic resonance imaging (fMRI) experiment using multivoxel pattern analyses (MVPA) studying the implementation of feature-selective attention in the MD network. Participants performed a task where they

classified novel objects along two orthogonal stimulus dimensions (length and orientation) in alternating blocks. I examined information in the MD network pertaining to the object feature when *relevant* (e.g. orientation feature information in the orientation task) versus *irrelevant* (e.g. orientation feature information in the length task) to the task. Lateral occipital complex (LOC) and Brodmann area 17 (BA 17) were included as regions of interest for comparison.

The MD network discriminated between the objects according to the currently relevant feature dimension: there was stronger representation of equivalent stimulus distinctions when they were relevant compared to when they were irrelevant. LOC showed a qualitatively similar pattern of results, coding object features only when they were relevant for the current task, although this coding was not significantly stronger than coding of the irrelevant feature. BA 17 did not appear to discriminate the object features in either condition. These data provide a potential mechanism for feature-selective attention in the MD network, in that regions of this network adjust their responses to code the feature information that is currently relevant.

Chapter 3: Overlapping neural codes: Frontoparietal voxels are re-used to code relevant feature information across different tasks

This chapter extends the work from Chapter 2. In Chapter 2, I had shown that multi-voxel patterns in the MD network preferentially code relevant over irrelevant feature information. In Chapter 3, I first investigated whether this effect replicated with a new stimulus set. Second, I developed a new method to test whether the MD voxels with the highest signal contributing to the classification of relevant features in one task also provided the highest signal contributing to the classification of the relevant features in another task. The design of this experiment was similar to the previous chapter with the

goal of drawing a comparison between the two datasets. However, in contrast to the experiment from Chapter 2, where participants discriminated along orthogonal dimensions of the same stimulus set ('spikies'), in this experiment, participants discriminated along dimensions of two different groups of objects ('smoothies' and 'spikies'). In comparing the two datasets I refer to the experiment presented in Chapter 2 as Experiment 1, and the new experiment as Experiment 2.

The results of the new data in Chapter 3 replicated the finding from Chapter 2: there was stronger coding of task-relevant object features than of the task-irrelevant feature information across the MD network. I then identified the top 10% of voxels with the highest signal contributing to the two relevant feature discriminations in each experiment. In both experiments, a higher proportion of voxels than that expected by chance contributed to discriminating the relevant feature dimension of the two tasks (termed 'voxel re-use').

Based on non-human primate work (Cromer, Roy, & Miller, 2010; Roy, Riesenhuber, Poggio, & Miller, 2010), I predicted that voxel re-use would be higher in Experiment 2, when the two stimulus sets were independent, than in Experiment 1. However, a comparison of the two datasets revealed no difference between the two datasets in the extent of voxel re-use. These data suggest that, at least at the level of voxels, MD resources used to discriminate relevant feature information in one task can be re-used to code relevant feature information in another task context, but there was no evidence for a meta-level of flexibility in which this allocation of resources varied according to the broader demands of the experiment. The finding that individual voxels in the MD regions contributed to multiple multi-voxel codes more frequently than expected by chance, provides further support for adaptive coding in the MD network.

Chapter 4: Exploring the causal role of the right dorsolateral prefrontal cortex in feature-selective attention

In this chapter, I conducted three experiments with the aims of developing a robust paradigm for a subsequent transcranial magnetic stimulation (TMS)-fMRI study (Chapter 5), developing expertise in TMS methods, and examining the effect of different control methods: low intensity stimulation (LIS), control site, and sham. In the first experiment I examined whether the right dorsolateral prefrontal cortex (dlPFC) was causally involved in a task requiring selection of task-relevant over task-irrelevant feature information.

Participants categorised novel coloured objects according to colour or form. The required response for the irrelevant feature was either incongruent or congruent with the required response for the relevant feature. I employed online TMS whilst participants performed the task in four separate sessions (high intensity stimulation (HIS) to right dlPFC, LIS to right dlPFC, HIS to a control region and sham stimulation). I predicted that when right dlPFC function was disrupted during the feature-selective congruency task this would result in an increase in the magnitude of the congruency effect relative to the control conditions. Although there was a main effect of congruency, indicating that participants were less accurate on incongruent than on congruent trials, there was no evidence that the size of the congruency effect changed with TMS stimulation. A Bayes analysis suggested we had insufficient power, but it was beyond the scope of this study to test more participants. Instead, I tested a number of modifications in two further behavioural experiments to increase the sensitivity of the measure of interference from irrelevant features; to optimise the design for my subsequent combined TMS-fMRI experiment (Chapter 5).

Chapter 5: A concurrent TMS-fMRI study investigating feature-selective top-down signals from the dorsolateral prefrontal cortex

This chapter presented a concurrent TMS-fMRI experiment where I investigated the role of the right dlPFC in feature-selective attention. Participants attended to one feature of novel objects (e.g., colour) whilst simultaneously ignoring another feature (e.g., form). TMS stimulation randomly alternated between 110% or 40% of an individual participant's motor threshold on any given trial. I used MVPA of the fMRI data to compare what object feature information was coded in the brain when the right dlPFC was disrupted under HIS compared to LIS. The role of the dlPFC in feature-selective attention could involve (a) inhibition of irrelevant information and/or (b) selection and maintenance of task-relevant representations. I predicted that disrupting (HIS compared to LIS) the inhibition of irrelevant information would result in *stronger* coding of *irrelevant* feature information in other brain areas, whilst disrupting the selection of task relevant information would result in *weaker* coding of *relevant* feature information in other brain areas.

TMS had a significant effect on behaviour. There was a larger congruency effect under HIS compared to LIS, which indicated that irrelevant information had a larger effect on behaviour when dlPFC function was perturbed by TMS. This behavioural effect was reflected in the neural data: under HIS I found stronger coding of irrelevant information across the MD frontoparietal network and visual brain regions. This can be interpreted as evidence for a functional role of the right dlPFC in inhibiting irrelevant information. However, contrary to our second prediction, I found either no effect (for object form) or *stronger* coding (for object colour) of task-relevant features under HIS. I speculated that this could reflect some form of compensation for disruption to dlPFC function, but from this pattern of data I cannot completely rule out a general effect of dlPFC disruption which

causes an overall increase in information coding regardless of relevancy. Under either interpretation, these data provide causal evidence for the role of right dlPFC in modulating the coding of relevant and irrelevant information in the brain. A searchlight analysis also revealed significant changes in information coding in parts of the visual cortex, in line with a role, for the dlPFC in providing top-down information to the visual cortex as predicted by various models of prefrontal function.

Implications, outstanding questions, and future directions
Adaptive coding as a framework for feature-selective attention

The experiments in this thesis were designed to test the predictions laid out by the adaptive coding hypothesis (Duncan, 2001) in the context of feature-selective attention. The first prediction of the adaptive coding hypothesis is that neurons will adjust their responses to selectively encode information that is currently relevant. I tested this prediction in tasks that required selection at the level of single features of an object, in candidate (MD) regions in the human brain. If responses in the MD regions are adaptive then they should adjust to code relevant *feature* information of a single object over the irrelevant feature information. The mechanism for this adaptive response, as applied to feature-selective attention, is set out in Figure 1.
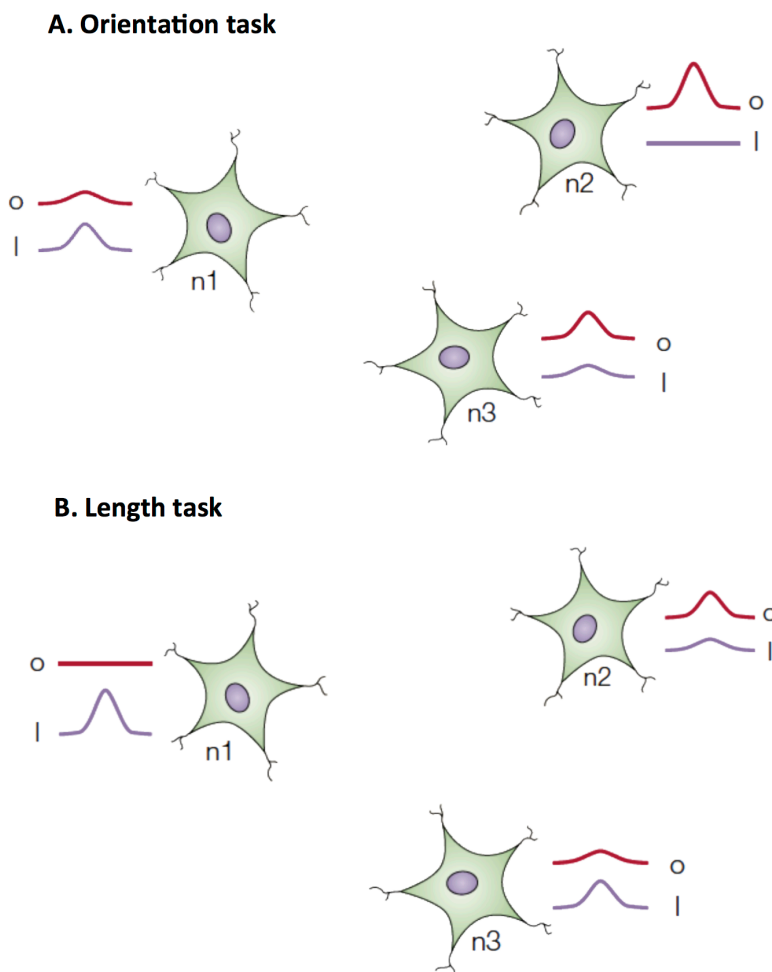
Figure 1: Adaptive coding model, figure modified with permission from Duncan (2001). The figure depicts the selectivity of three hypothetical neurons to two object features (orientation, o, and length, l). Selectivity is indicated by tuning curves next to each neuron where a sharper curve reflects greater sensitivity for the indicated feature dimension. Crucially, neurons may vary in selectivity for either orientation or length information but this selectivity is modulated by the current task (orientation or length). As an example, neuron 2 (n2), displays a greater selectivity along the orientation dimension than along the length dimension but this response is still modulated by task context (stronger response to the preferred orientation in orientation task, weaker in length task, and vice versa for length). The response of each individual neuron is modulated by context, and the result across the population is that the task relevant dimension is represented preferentially.

The results from Chapter 2 showed that multi-voxel patterns of activation in the MD regions did indeed encode the same feature distinctions more strongly when they were task-relevant than when they were task-irrelevant. Moreover, the data from Chapter 3 indicated that this effect reflects the modulation of individual voxels, with the same voxels used to code the relevant feature distinctions across both tasks more frequently than predicted by chance. Figure 2 illustrates how adaptive responses to relevant feature information could lead to more distinctive multi-voxel patterns representing the task-relevant over the task-irrelevant feature distinctions.

A further prediction laid out by the adaptive coding hypothesis is that responses in the MD regions provide a form of top-down modulation to earlier cortical regions, biasing processing in favour of what is behaviourally relevant. In Chapter 5 I tested whether disruption to right dlPFC causally modulates coding of both relevant and irrelevant feature information across the system. It was apparent that disruption to this region not only modulated coding of feature information in early visual cortices but also in the rest of the MD network. This provides support for the prediction that these regions modulate feature-selective responses across the brain. All in all the data in this thesis suggest a flexible system, in line with the adaptive coding hypothesis, that can adjust its responses to code relevant feature information. This adjustment may then modulate responses across the brain.

Orientation task

A

length   orientation

Length task

B

length   orientation

Decoding of relevant information

Orientation task: decoding orientation

Short 45 Long 45   Short 45 Long 45

Short, 45   Short 90 Long 90   Short 90 Long 90   **Short, 90**

Short 45 Long 45   Short 45 Long 45

Long, 45   Short 90 Long 90   Short 90 Long 90   **Long, 90**

Decoding of irrelevant information

Length task: decoding orientation

Short 45 Long 45   Short 45 Long 45

Short, 45   Short 90 Long 90   Short 90 Long 90   **Short, 90**

Short 45 Long 45   Short 45 Long 45

Long, 45   Short 90 Long 90   Short 90 Long 90   **Long, 90**

Length task

C

length   orientation

Orientation task

D

length   orientation

Decoding of relevant information

Length task: decoding length

Short 45 Long 45   Short 45 Long 45

Short, 45°   Short 90 Long 90   Short 90 Long 90   **Long, 45°**

Short 45 Long 45   Short 45 Long 45

Short, 90   Short 90 Long 90   Short 90 Long 90   **Long, 90**

Decoding of irrelevant information

Orientation task: decoding length

Short 45 Long 45   Short 45 Long 45

Short, 45°   Short 90 Long 90   Short 90 Long 90   **Long, 45°**

Short 45 Long 45   Short 45 Long 45

Short, 90   Short 90 Long 90   Short 90 Long 90   **Long, 90**

266

Figure 2: Example of adaptive coding at a multi-voxel level. The graphs indicate the selectivity of a neuron of type n2 (refer to figure 1). This neuron is tuned to a particular orientation and length, but its response is modulated by task context. Underneath the graphs in each panel is the hypothetical response of a voxel dominated by neurons of this type. Each square represents a voxel, and the preferential selectivity (for length and orientation dimensions) of that voxel is indicated above it e.g. the upper left voxel is tuned for short objects and for objects at 45°. In each quadrant there are four voxels, each of which is selective for a different combination of features. Next to each quadrant is the currently presented stimulus (e.g. short 45° stimulus). The colour shading in each voxel is a result of the voxels underlying selectivity, the features of the current stimulus, and the modulation by task context. For example, in Panel A for the orientation task, when a short 45° stimulus is presented, the voxel which has a preference for 'long' and '45°' will receive a score of 10 (the value is arbitrary but is assigned to illustrate a relative response). This is because it is this voxel's preferred orientation (score of 10) but non-preferred length (score of 0) to give a total of 10. For this same voxel the response would be lower for the identical stimulus in the length task (Panel B) as it would have adjusted its response to show greater selectivity for its preferred length and less selectivity for its preferred orientation (score of 8). Panels A and B show the multivoxel patterns that would be used to decode orientation in the orientation (Panel A) and length (Panel B) task. As can be observed, the multivoxel patterns for orientation (compare left and right of the decision boundary) are highly distinct in Panel A, but more confusable in Panel B. The same result is found for decoding length (Panels C and D). This is an example of how adaptive coding may result in stronger decoding of relevant compared to irrelevant feature information. The same result is found if the neuron is modelled on n1 or n3 from Figure 1.

Functional specialisation within the MD network

Several alternative models of prefrontal cortex suggest dissociations between different regions (see Chapter 1, e.g. Badre, 2008; Badre & D'Esposito, 2009; Botvinick, 2008; Bunge & Zelazo, 2006; Christoff & Keramatian, 2007; Corbetta & Shulman, 2002; Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Dosenbach et al., 2007; Dosenbach et al., 2006; Goldman-Rakic, Roberts, Robbins, & Weiskrantz, 1998; Koechlin, Ody, & Kouneiher, 2003; Koechlin & Summerfield, 2007; O'Reilly, 2010; Reynolds, O'Reilly, Cohen, & Braver, 2012). Although the experiments in my thesis were not designed to directly test any of these models, it was noticeable that there were almost no suggestions of dissociation of functions within the MD network. In my experiments, I saw similar recruitment of the IFS, IPS, AI/FO and ACC/pre-SMA across Chapters 2, 3 and 5; the only suggestion for a difference was in Chapter 3 in which voxel re-use was significantly higher in IFS (and lower in AI/FO).

There are also a few cases where the data seem to directly contradict the predictions of these models. For instance, one proposal is that the prefrontal cortex (PFC) follows a rostrocaudal gradient with different regions recruited according to the demands of the task ("Rule Abstraction model", Badre, 2008; Badre & D'Esposito, 2007, 2009). In this proposal, the cortex is functionally organised so that more abstract goals are represented along a rostrocaudal hierarchy. For example, a low level rule (e.g. press the 'left' button for rectangle and press the 'right' button for a square) is suggested to recruit the dorsal pre-motor cortex. Recruitment of anterior regions requires a higher level of rule abstraction (e.g. if the stimulus is red, use rule 1 (left for rectangle, right for square) otherwise, use rule 2 (right for rectangle, left for square)). In my paradigms, rules were

restricted to one level of abstraction (e.g. short = category 1), so should not have recruited regions anterior to the dorsal pre-motor area. However, I consistently found that task-relevant stimulus features were coded in the IFS (Chapter 2, 3 and 5, see alsoWoolgar, Afshar, Williams, & Rich, 2015). It is also not clear that the Rule Abstraction model would predict coding of stimulus features (as opposed to rules) in these regions at all. It may be, however, that functional dissociations between these regions only become apparent under specific conditions, perhaps determined by the nature of task.

My results support the proposal that the MD regions work together as a system, at least under certain conditions (in this case, feature-selective attention to novel objects). The findings fit well with the framework of the adaptive coding hypothesis whereby responses are flexible and adjust according to the currently relevant information. Thus, the weight of evidence from this thesis support the function of the MD regions as a network, being best understood through the general principle that coding in these regions is flexible and that they adapt to code relevant over irrelevant feature information.

Enhancement of relevant/inhibition of irrelevant?

My experiments have shown that the MD network plays an important role in feature-selective attention. There is a long-standing debate in the field regarding whether the effects of attention on neural activity is exerted by enhancing the representation of the relevant information or by inhibiting the representation of the irrelevant information, or both (e.g. Aron, 2007; Kanwisher & Wojciulik, 2000; Knight, Staines, Swick, & Chao, 1999; Shimamura, 2000). Miller and Cohen (2001) suggest that attention biases competing inputs in favour of task-relevant information in the PFC, which in turn has a consequence on whether irrelevant information is further represented in the system. In this scenario, task-irrelevant inputs are not actively suppressed but are competed against by the selection

of task-relevant inputs. A competition-based model of visual attention was similarly proposed by Desimone (1998), who suggested that excitatory signals to selective visual neurons would result in inhibition of task-irrelevant visual neurons, sharpening the focus of attention (also refer to Desimone & Duncan, 1995; Kastner & Ungerleider, 2000; Pessoa, Kastner, & Ungerleider, 2003). Although my data do not resolve this debate, the results provide an interesting piece of the puzzle, which I will return to shortly.

Evidence for competition-based models of attention stems primarily from non-human primate research where PFC neurons have been found to maintain task-relevant information in delayed-response tasks (e.g. Fuster, 1973; Fuster & Alexander, 1971; Kubota & Niki, 1971). PFC neurons also adjust their responses to code task-relevant information across multiple tasks (e.g. Cromer et al., 2010; Freedman, 2001; Rao, Rainer, & Miller, 1997; Roy et al., 2010). This adjustment for task-relevant information provides a potential source of feedback, which could bias representations in earlier sensory processing areas in favour of task-relevant inputs (Miller & Cohen, 2001). These studies provide support for the notion that attention exerts influence by enhancing the representation of relevant information.

In favour of the alternate mechanism of inhibition of irrelevant information, the PFC has been suggested to provide inhibitory regulation of neural activity (e.g. Knight & Stuss, 2002; Shimamura, 2000; Smith & Jonides, 1999), increasing the signal-to-noise ratio across the brain via a prefrontal-thalamic inhibitory system (Knight et al., 1999). An example of this was shown by reversible cooling of the prefrontal-thalamic system in the cat brain which increased the amplitude of evoked responses in primary sensory cortex (Skinner, 1984; Yingling & Skinner, 1977). Non-human primates studies have also provided support for an inhibitory mechanism of PFC (Bartus & Levere, 1977; Malmo,

1942). For example in Bartus et al.'s (1977) study, monkeys with dlPFC lesions were significantly impaired during a working memory (WM) task where irrelevant information was during a trial compared to monkeys with undamaged brains. In the human brain, several studies employing visual WM tasks have also emphasised an inhibitory role of the PFC (e.g. Chao & Knight, 1998; Clapp, Rubens, & Gazzaley, 2009; Postle, 2005). Chao and Knight (1998) showed that evoked responses to distracting stimuli in the auditory cortex during a delay period were greater for patients with PFC lesions than for control subjects. This was interpreted as evidence for the PFC as contributing to inhibition of irrelevant information. Inhibitory deficits have also been found in patients with prefrontal damage in cognitive tasks requiring suppression of prior learned material (Mangels, Gershberg, Shimamura, & Knight, 1996; Shimamura, Jurica, Mangels, Gershberg, & Knight, 1995). Thus, data from both non-human and human studies suggest a role for inhibitory processes in selective attention, which may also rely on regions that form part of the MD network (Kanwisher & Wojciulik, 2000).

There is a clear indication for both a mechanism that supports selection and maintenance of task-relevant inputs, as well as a mechanism that promotes inhibition of task-irrelevant inputs. However, it has been argued that evidence for an inhibition mechanism is less convincing (e.g. Aron, 2007; Shimamura, 2000). One reason for this is because it is often difficult to disentangle the two mechanisms. To illustrate, neuropsychological data has found that frontal patients show increased interference effects in the Stroop task (Perret, 1974; Stuss, Floden, Alexander, Levine, & Katz, 2001; Vendrell et al., 1995). The Stroop task (Stroop, 1935) requires participants to attend to one stimulus dimension whilst ignoring another. Increased interference by the irrelevant dimension in this task is often taken as evidence for an inhibition mechanism as disrupting inhibition of irrelevant information should result in a larger effect of the irrelevant information (e.g.

Dimitrov et al., 2003; Vendrell et al., 1995). However, in this case, it could also be that inadequate selection and maintenance of task-relevant information makes it more likely for irrelevant inputs to be preserved thus causing the observed interference effect. This does not exclude either possibility but emphasises the difficulty with the interpretation of underlying neural mechanisms.

In consideration of the literature is possible that both of these mechanisms occur, whereby relevant information is enhanced and irrelevant information is simultaneously suppressed (e.g. D'Esposito & Postle, 2015; Knight et al., 1999; Shimamura, 2000) to optimise performance. Another possibility is that irrelevant information is not actively suppressed but rather driven out by competition from relevant inputs (e.g. Desimone, 1998; Desimone & Duncan, 1995; Miller & Cohen, 2001). However, in many cases it is difficult to distinguish these two mechanisms. For example, Chapter 2 and Chapter 3 of this thesis present evidence that the MD regions adjust to code information that is currently relevant. Whilst this makes it possible for adaptive MD regions to underlie selection of task-relevant information, it does not eliminate the possibility that irrelevant information is also actively inhibited.

In Chapter 2, participants performed a task where they classified objects along two orthogonal stimulus dimensions in alternating blocks. Thus in this experiment, the irrelevant information was sometimes relevant to the task (in the alternating task block). In comparison, in Chapter 3, the irrelevant feature was never relevant to the participant's task as they alternated between performing two tasks on different groups of objects. A direct comparison revealed that there was stronger MD coding of irrelevant information in Chapter 2 (when the irrelevant information was sometimes relevant) than in Chapter 3 (when it was never relevant). This may provide evidence for inhibition of irrelevant

information in this network: when irrelevant information is never relevant, there is less coding of this information across the network, perhaps reflecting the ease with which one can inhibit it. There was no suggestion that coding of *relevant* information differed between experiments, making it unlikely that this difference was due to an increase in suppression from local competition with task relevant input. However, in both experiments, coding of irrelevant features was at chance across the MD system, making it difficult to interpret this finding.

In Chapter 4 I used TMS to address the causal effect of right dlPFC on behaviour. However, even if I had shown a behavioural interference effect of disruption to the right dlPFC, it would not have been possible to disentangle the two possible underlying mechanisms. Interference could have been caused by disruption to selection of relevant over task-irrelevant inputs, or by disruption to inhibition of irrelevant feature information. This highlights the difficulty in dissociating the underlying mechanisms involved in feature-selective attention.

As discussed in Chapter 5, the combination of TMS-fMRI, or TMS-electroencephalography (EEG), has the potential to further elucidate the relationship between selection and inhibition (e.g. Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Zanto, Rubens, Thangavel, & Gazzaley, 2011). In Chapter 5, I employed concurrent TMS with fMRI-MVPA, which can reveal changes in the informational content in the brain following disruption to a particular region. These data showed that disruption to a region of the MD network increases the strength of coding of irrelevant feature information across the network and other brain regions. Moreover, this occurred in the *absence* of a decrease in coding of relevant information, ruling out an explanation based on local competition. These data in combination with the behavioural interference effect provide

evidence for a role of the right dlPFC in active inhibition of irrelevant feature information: when right dlPFC function is disrupted, irrelevant information has a larger effect on behaviour and is coded more strongly in the rest of the MD network and the visual cortex than when dlPFC is intact.

The data in Chapter 5 showed an increase in both irrelevant and relevant information coding. This points to a disruption of an active inhibition mechanism as any passive inhibition due to upregulation of the relevant information should still have been in play (since relevant information was strongly represented in the MD system and elsewhere). However, an alternative is that, as I observed both an increase in the strength of relevant, and irrelevant information coding under HIS, the strength of coding of *all* information regardless of relevancy increases as a result of TMS stimulation. I suggest that this is unlikely in the case of these data as the stimulated target region, and many other brain regions, showed no effect of HIS, and the increase in strength of coding for relevant information under HIS was specific to colour information. Moreover, following the logic of TMS, the interpretation would have to be that the dlPFC usually suppresses the representation of task information in general (such that both relevant and irrelevant information is released from suppression under TMS). Given the previous literature, this interpretation seems unlikely. Nonetheless, it will be important to conduct further work using these kinds of paradigms to elucidate the top-down mechanisms supporting flexible behaviour (e.g. which regions modulate coding in other regions, what is the nature of the information that is modulated, is it dependent on task load, etc) and to examine the effect of TMS stimulation on information coding in general.

General limitations

An inherent problem in examining underlying neural processes, such as the mechanisms supporting flexible behaviour, lies in the measurement techniques that we have available. To begin with, it is important to remember that fMRI is an indirect measure, indexing changes in the blood-oxygen-level-dependent (BOLD) response rather than neuronal responses directly. Thus, BOLD responses measured with fMRI are only an indirect reflection of the underlying neural responses (e.g. Aron, 2007; Logothetis, 2008b). The accurate interpretation of the BOLD signal depends on characterising the nature of the underlying neural activity that gives rise to the haemodynamic response, and the way in which these two aspects are linked (Arthurs & Boniface, 2002). There is evidence suggesting that the BOLD signal correlates strongly with the underlying local field potential (LFP) as opposed to spike rates (Goense & Logothetis, 2008; Logothetis, 2008b; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). The LFP is a mass neural signal that captures a multitude of neural processes, but the exact composition of LFP remains under investigation (Berens, Logothetis, & Tolias, 2010; Magri, Schridde, Murayama, Panzeri, & Logothetis, 2012). This means that modulations in fMRI signal reflect the pooled signal of a very large number of neurons with the precise underlying machineries unknown (Scannell & Young, 1999). We can only improve our interpretation of imaging data when we have a clearer understanding of how neuronal activity influences the fMRI signal.

The problem of interpretation is not limited to fMRI. Other techniques such as electrical measurements of brain activity also have a complex relationship with network activity (Logothetis, 2008a). Using a combination of available techniques will both help us to clarify underlying brain function and potential mechanisms, and allow us to

investigate the relationship of our measurement to neural activity, for example by the simultaneous acquisition of electrophysiological recordings and fMRI data in non-human primates.

Despite fMRI being an indirect measure of neural activity, the spatiotemporal response patterns reflected in fMRI signals contain detailed information that can be revealed using MVPA (e.g. Hebart, Görgen, & Haynes, 2015; Yang et al., 2014). For example, MVPA has been used to decode visual, auditory, motor and task rule information across different brain networks (Woolgar, Jackson, & Duncan, in press, refer to Appendix A). In this thesis, MVPA was used to examine the representation of task-relevant and task-irrelevant object features in specific regions of the brain. However, although MVPA is highly sensitive to whether or not decodable information is present, an open debate remains as to the nature and content of the underlying signal that is driving the classifier.

Hubel and Wiesel's seminal work showed that orientation selective cells (Haxby et al., 2001), and likewise, ocular dominance selective cells (Hubel & Wiesel, 1963; 1972) are clustered into a columnar-like structure where groups of cells with similar response properties are arranged vertical to the cortical surface. An estimate of the size of ocular dominance columns have been measured from post-mortem samples of patients as approximately 1mm (Hubel & Wiesel, 1969). Converging evidence for this size comes from high-field, high resolution fMRI techniques of both ocular dominance columns (Adams, Sincich, & Horton, 2007) and orientation columns (Cheng, Waggoner, & Tanaka, 2001; Yacoub, Shmuel, Logothetis, & Uğurbil, 2007). In some of the first studies that employed MVPA, Haynes and Rees (2005) and Kamitani and Tong (2005) showed that it was possible to decode the orientation of visual gratings using MVPA. Both papers

proposed that the biased sampling of the orientation columnar structures at a smaller scale than voxels is what drives the successful classification. This is often referred to as the "hyperacuity" hypothesis (de Beeck, 2010a). The hyperacuity hypothesis works on the assumption that each voxel samples a mixture of different columns. The columns sampled are not uniform and this can lead to a slightly bias for a given measurement (e.g. orientation) in different voxels (e.g. de Beeck, 2010b). The argument then is that, as MVPA considers the activity of a large number of voxels together, small biases across voxels could lead to distinctive responses for different conditions, allowing underlying neural selectivity to be inferred.

The "hyperacuity" hypothesis and the source of decoded information remains a subject of significant debate (Boynton, 2005; Haynes & Rees, 2005; Kamitani & Tong, 2005). Some have argued that without high-field, high resolution scanning, the resolution required to measure activity from columnar structures is not feasible as a voxel still contains many tens of thousands of neurons (Beckett, Peirce, Sanchez-Panchuelo, Francis, & Schluppeck, 2012; Chaimow, Yacoub, Ugurbil, & Shmuel, 2011). Another possibility is that decoded information relies on large-scale biases (e.g. Alink, Krugliak, Walther, & Kriegeskorte, 2013; Carlson, 2014; de Beeck, 2010a; Freeman, Brouwer, Heeger, & Merriam, 2011; Freeman, Heeger, & Merriam, 2013; Kamitani & Sawahata, 2010; Kriegeskorte, Cusack, & Bandettini, 2010). Potential large-scale biases for orientation in the human brain include (but are not limited to), broad preferences for radial orientations (e.g. Sasaki et al., 2006) and preferences across the visual field for cardinal orientations (e.g. Furmanski & Engel, 2000).

Other researchers have claimed that the biases sampled are vascular in origin (Gardner, 2010; Shmuel, Chaimow, Raddatz, Ugurbil, & Yacoub, 2010). Additionally, de

Beeck (2010a) has shown that MVPA is robust to spatial smoothing, which he interprets as evidence against a columnar-scale bias driving classification, as smoothing should be detrimental to picking up small-scale functional organisation. However, the precise function and origin of large-scale biases that allow classification are still unknown (e.g. Freeman et al., 2011; Freeman et al., 2013; Maloney, 2015). Recent data presents an alternative account suggesting that orientation decoding relies on edge-related activity, and that this edge-related activity masquerades as a large-scale radial bias (Carlson, 2014; Wardle, Ritchie, Seymour, & Carlson, 2015). At this stage, all presented accounts depicting the source of decodable information (for orientation) remain viable. Thus, overall, MVPA can tell us *what* can be decoded in the brain but an open question remains as to *why* it is decodable. In the case of the data presented in this thesis, the source of the decoding (hyperacuity, large-scale biases, or some as yet un-identified cause) is less important than the *changes* in decoding between the physically identical conditions, which reflects that responses are adapting. It seems unlikely that we could find evidence of such adaptation if the patterns did not ultimately reflect some change in responses at the neural level.

As is clear, we do not fully understand the distribution of neural responses underlying voxel-wise responses. Thus in Chapter 3, I refer to the data in terms of voxel re-use rather than neural re-use. In this chapter, MD voxels providing the strongest signal in the discrimination of relevant object features in one task were "re-used" in the sense that they also provided the strongest signal to the discrimination of the relevant object features in another task. I took this as evidence that these regions flexibly re-use resources to encode information across multiple tasks, as previously shown in the non-human primate literature (Cromer et al., 2010; Freedman, 2001; Roy et al., 2010). However, as we do not yet fully understand the pattern and distribution of neural responses underlying

voxel-wise responses, it is possible that within the voxels sampled, there are separate populations of neurons, each of which held the relevant feature in one (and only one) task. On the other hand, it is possible, as I inferred in this Chapter, that the voxels sampled represented an overall bias where the underlying neural populations showed preference to relevant feature dimensions of both tasks. We cannot discriminate these possibilities from the current data.

Combining single-unit with fMRI measurements in non-human primates may help us elucidate these relationships in the future. In general, not only in the case of the data from Chapter 3, but in the case of all multi-voxel data, it is important to bear in mind that classifiers will exploit whatever information is present in the patterns of activity and that the interpretation of the source of information should be made with caution. Thus, is critical to design well-controlled experiments and avoid confounds. For example, in Chapter 2, we decoded physically identical objects in the task-relevant and irrelevant conditions, removing any low-level potential drivers for the classifiers. We also presented a response-mapping screen in order to dissociate the motor response (left vs. right) from the decision (short spike vs. long spike). Thus, the only potential driver for classification to be successful is our critical comparison of whether the stimulus was behaviourally relevant or irrelevant. Well-controlled designs allow us to infer the likely source of decodable information.

The interpretation of MVPA of fMRI data also depends critically on the use of appropriate statistical analysis (e.g. Allefeld, Görgen, & Haynes, 2015; Combrisson & Jerbi, 2015; Hebart et al., 2015; Nichols & Holmes, 2002; Noirhomme et al., 2014; Stelzer, Chen, & Turner, 2013). In this thesis, standard parametric statistics were applied for second-level decoding analyses. Parametric statistics were appropriate in this case as

the classifier was trained across sets of data and tested on the independent data sets, and cross-validation provides an unbiased estimate of the classification accuracy. Compared to other available methods (e.g. two-step permutation test (Stelzer et al., 2013)), they are the most appropriate for making group-level inferences. However, methods for decoding analyses are under ongoing development and it will be important to consider the appropriateness and available inference as statistical theory is developed and more sophisticated methods become available.

A further discussion in the MVPA literature is the extent to which the information that is decoded is actually used by the brain (de-Wit, Alexander, Ekroll, & Wagemans, 2016). In their recent paper, de-Wit and colleagues (2016) discuss how the ability to decode something from fMRI data indicates that the relevant information is available to our classifier, but it does not mean that these patterns are used as information in the brain. The authors draw on examples of patients who have poor spatial perception even though it is possible to construct retinotopic maps in their early visual cortex. These maps can inform the "observer" (in this case, a researcher using pattern analysis) as to where objects are in relation to each other, but are apparently not useful for the patient's behaviour. de-Wit et al. (2016) argue that if the strength of coding can be related to performance on a task, then this provides evidence that the underlying patterns were also relevant for behaviour. They use Williams et al.'s (2007) study as a key demonstration of the strength of this approach. In this study, the authors computed spatial pattern correlations between objects pairs and found higher correlations for the same category objects pairs (e.g. two spiky objects) over different category object pairs (e.g. spiky and smoothy objects). This was the case for both retinotopic cortex and LOC. Crucially, in LOC, this was only true for correct trials (and not incorrect trials). These data suggest that the strength of coding in LOC can be related to behavioural performance, indicating that the brain uses this

information. In this thesis, one recurrent finding is that the MD regions code task-relevant features more strongly than task-irrelevant. An important extension would be to see whether the strength of this coding predicts behavioural success, which could provide support for the hypothesis that these regions have a causal role for behaviour.

Although a correlation of decodable information in a region and behavioural performance is useful, de-Wit et al. (2016) also argue that even this does not provide the full picture without indication of *how* the decoded information is subsequently used elsewhere in the brain. Therefore, the combination of fMRI with other techniques, such as TMS, is useful. It has the potential to inform us how information coding in one brain region is causally related to information coding in other brain regions.

A TMS-induced change in behaviour can be used to inform models of causal relations between specific brain regions and individual cognitive functions, and the combination of TMS with neuroimaging techniques can allow us to investigate the causal relationship between neural processing in different brain regions. In Chapter 5, I used concurrent TMS-fMRI with MVPA to investigate the causal relationship between activity in the dlPFC, information coding elsewhere in the brain, and participant behaviour. However, despite the intriguing possibilities afforded by TMS, its precise mechanisms of action remain unclear (Sandrini, Umiltà, & Rusconi, 2011). For example, TMS has been shown to elicit both enhanced and suppressed activity at the cortical level (Allen, Pasley, Duong, & Freeman, 2007; Moliadze, Zhao, Eysel, & Funke, 2003). It is even shown that the same stimulation parameters can result in either activation, suppression, or both depending on the brain regions stimulated (Paus, 2005). The effect of stimulation is dependent on a number of factors including the stimulation parameters themselves (e.g. frequency, intensity and duration, Pasley, Allen, & Freeman, 2009; Sandrini et al., 2011)

and the choice of design (e.g. coil placement, region/zone selection, Opitz, Fox, Craddock, Colcombe, & Milham, 2016). In addition, the following can change the effect of stimulation (and the list is not exhaustive): excitatory vs. inhibitory neurons, the structure of stimulated neural circuit and the modifications of synaptic connections (Pasley et al., 2009). This variance creates a challenge for using and interpreting the results of neural stimulation techniques, despite the promise in combining methodologies to demonstrate causal relationships.

In Chapter 4 and 5, I used online TMS where participants received three pulses of stimulation, a parameter commonly used in disruption studies (Hallett, 2007; Walsh & Pascual-Leone, 2003). In Chapter 5, this TMS stimulation caused an increase in the behavioural congruency effect, which is an index of the extent to which irrelevant information affects behaviour, and I therefore interpreted my results in terms of disruption to the dlPFC. However, it is not entirely clear whether the effects of stimulation are predominantly excitatory or inhibitory (Parkin, Ekhtiari, & Walsh, 2015). Sandrini et al. (2011) discuss how lengthening the duration of online TMS will cause more disruption by temporal summation of the effects of stimulation. However, online TMS can also vary in behavioural outcome depending on the activation state of the stimulated region (Sandrini et al., 2011; Siebner, Hartwigsen, Kassuba, & Rothwell, 2009; Silvanto & Muggleton, 2008). For example, stimulation during the time period when a region is involved in a task usually causes disruption, whilst a number of studies report facilitation if stimulation is delivered early, prior to the commencement of involvement in the task (e.g. Grosbras & Paus, 2003; Töpper, Mottaghy, Brügmann, Noth, & Huber, 1998). In my case, I selected a target region that was predicted to be involved in the task on theoretical grounds (previous meta-analytic work) and based on individual subject activation in the task localiser, timed the first TMS pulse to arrive 75ms after stimulus onset, and observed a disruption in

behaviour. I am therefore confident that this design caused a disruptive effect, and have interpreted my results accordingly.

Another important consideration in the design of a TMS paradigm is the use of the control or placebo condition. For example, sham TMS approaches are widely used as a control condition, but this only controls for the sensory side effects (e.g. acoustic artefact) of TMS. Sham does not control completely for psychological variables because it does not mimic the somatosensory effect of TMS, and therefore participants may be aware of the different condition. In addition, sham cannot inform us of whether stimulation of a particular brain area has specific behavioural or physiological consequences. For this we need a control stimulation site (Duecker & Sack, 2015). However, there are limitations associated with this control as well. For example, this is likely to have a different sensation to the main target stimulation, and the interconnection of the brain may result in unexpected interactions. Using a lower stimulation of the critical area is a third method that controls for these factors but has the problem of unpredictability in terms of whether LIS could have a different effect (e.g., activation instead of inhibition) to the HIS. Chapter 4 of this thesis presented a paradigm with three controls to TMS stimulation. I aimed to highlight the importance of the choice of control by comparing the controls, but unfortunately the sample size was too small to draw any firm conclusions. In future, similar set-ups with more than one control condition can further elucidate the complex relationship of TMS effect with behavioural outcome.

Recent studies that combine TMS with neuroimaging have shown that TMS may affect remote brain areas interconnected with the stimulated brain region (Ruff, Driver, & Bestmann, 2009). Thus, although a behavioural effect of TMS indicates causal involvement of a targeted brain area in performance of the current task, it does not indicate

whether task-related performance is directly connected to the function of this region or reflects the influence of this region upon interconnected areas that are also involved. In Chapter 5, I used LIS as a control for the main effect of TMS. This was primarily for practical reasons as it allowed the two TMS conditions to be presented on a trial-by-trial basis. Although it would be ideal to have a control stimulation site, it comes at the cost of having to remove participants from the scanner in order to readjust the coil. Future research using concurrent TMS-fMRI-MVPA could consider using multiple controls including stimulation at different time points to investigate state-dependent effects, a control region, as well as LIS allowing a stronger inference to be made on the pattern of results. Nonetheless, the findings from Chapter 5 indicate that investigating the pattern of information coding following concurrent TMS disruption has the potential to be a powerful explanatory tool.

Conclusions

Now let us return to adaptive coding in the MD network and the implementation of feature-selective attention. Despite the limitations of the techniques that I have expounded in the above section, the data presented in this thesis provide strong evidence for adaptive responses in the MD regions as providing a mechanism for feature-selective attention. As mentioned previously, the adaptive coding hypothesis predicts that responses in these regions will adjust to task-relevant information, and this adjustment is further proposed to be a potential source of bias to earlier cortical regions (Duncan, 2001). Indeed in Chapter 2, these regions were shown to code identical object features more strongly when they were task-relevant compared to when they were task-irrelevant. Chapter 3 showed again that the MD regions code task-relevant features more strongly than task-irrelevant ones, and further demonstrated that that the same voxels are re-used for the relevant feature

discrimination across multiple tasks. Although in Chapter 4 I did not find evidence for a causal role of dlPFC in feature-selective attention, these results were observed in Chapter 5 where disruption to this region modulated both behaviour and coding of feature information across the MD network and earlier cortical regions. Therefore this thesis provides evidence for the predictions stated by the adaptive coding hypothesis, in the context of feature-selective attention, in that the MD regions adjust their responses to code relevant feature information and that this affects coding in earlier cortical regions.

In conclusion, the findings presented in this thesis advance knowledge on the neural processes underlying feature-selective attention in the MD network. Specifically, the findings provide evidence that adaptive coding provides a mechanism for feature-selective attention in the MD network. Together the experiments advance the understanding of flexible mechanisms employed in the frontoparietal cortices in the context of feature-selective attention.

References

Adams, D. L., Sincich, L. C., & Horton, J. C. (2007). Complete pattern of ocular dominance columns in human primary visual cortex. *The Journal of Neuroscience, 27*(39), 10391-10403.

Alink, A., Krugliak, A., Walther, A., & Kriegeskorte, N. (2013). fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Frontiers in psychology, 4*.

Allefeld, C., Görgen, K., & Haynes, J.-D. (2015). Valid population inference for information-based imaging: Information prevalence inference. *arXiv preprint arXiv:1512.00810.*

Allen, E. A., Pasley, B. N., Duong, T., & Freeman, R. D. (2007). Transcranial magnetic stimulation elicits coupled neural and hemodynamic consequences. *Science, 317*(5846), 1918-1921.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The neuroscientist, 13*(3), 214-228.

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in neurosciences, 25*(1), 27-31.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro–caudal organization of the frontal lobes. *Trends in Cognitive Sciences, 12*(5), 193-200.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Cognitive Neuroscience, Journal of, 19*(12), 2082-2099.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10*(9), 659-669.

Bartus, R. T., & Levere, T. (1977). Frontal decortication in rhesus monkeys: a test of the interference hypothesis. *Brain research, 119*(1), 233-248.

Beckett, A., Peirce, J. W., Sanchez-Panchuelo, R.-M., Francis, S., & Schluppeck, D. (2012). Contribution of large scale biases in decoding of direction-of-motion from high-resolution fMRI data in human early visual cortex. *NeuroImage, 63*(3), 1623-1632.

Berens, P., Logothetis, N. K., & Tolias, A. S. (2010). Local field potentials, BOLD and spiking activity–relationships and physiological mechanisms. *Nat. Precedings*.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences, 12*(5), 201-208.

Boynton, G. M. (2005). Imaging orientation selectivity: decoding conscious perception in V1. *Nature Neuroscience, 8*(5), 541-542.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15*(3), 118-121.

Carlson, T. A. (2014). Orientation decoding in human visual cortex: new insights from an unbiased perspective. *The Journal of Neuroscience, 34*(24), 8373-8383.

Chaimow, D., Yacoub, E., Ugurbil, K., & Shmuel, A. (2011). Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. *NeuroImage, 56*(2), 627-642.

Chao, L. L., & Knight, R. T. (1998). Contribution of human prefrontal cortex to delay performance. *Cognitive Neuroscience, Journal of, 10*(2), 167-177.

Cheng, K., Waggoner, R. A., & Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron, 32*(2), 359-374.

Christoff, K., & Keramatian, K. (2007). Abstraction of mental representations: theoretical considerations and neuroscientific evidence. *Perspectives on Rule-Guided Behavior, SA Bunge and JD Wallis, eds.(New York: Oxford University Press)*.

Clapp, W. C., Rubens, M. T., & Gazzaley, A. (2009). Mechanisms of working memory disruption by external interference. *Cerebral Cortex*, bhp150.

Cole, M. W., & Schneider, W. (2007). The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage, 37*(1), 343-360. doi:10.1016/j.neuroimage.2007.03.071

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods, 250*, 126-136.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Cromer, J. A., Roy, J. E., & Miller, E. K. (2010). Representation of Multiple, Independent Categories in the Primate Prefrontal Cortex. *Neuron, 66*(5), 796-807. doi:10.1016/j.neuron.2010.05.005

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual review of psychology, 66*, 115.

de Beeck, H. P. O. (2010a). Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage, 49*(3), 1943-1948.

de Beeck, H. P. O. (2010b). Probing the mysterious underpinnings of multi-voxel fMRI analyses. *NeuroImage, 50*(2), 567-571.

de-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic bulletin & review*, 1-14.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 353*(1373), 1245-1255.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Dimitrov, M., Nakic, M., Elpern-Waxman, J., Granetz, J., O'Grady, J., Phipps, M., . . . Grafman, J. (2003). Inhibitory attentional control in patients with frontal lobe damage. *Brain and Cognition, 52*(2), 258-270.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*(3), 99-105.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., . . . Raichle, M. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences, 104*(26), 11073-11078.

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., . . . Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron, 50*(5), 799-812. doi:10.1016/j.neuron.2006.04.031

Duecker, F., & Sack, A. T. (2015). Rethinking the role of sham TMS. *Frontiers in psychology, 6*.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2*, 820-829.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14*(4), 172-179. doi:10.1016/j.tics.2010.01.004

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences, 108*(42), 17510-17515.

Freedman, D. J. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science, 291*(5502), 312-316. doi:10.1126/science.291.5502.312

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *The Journal of Neuroscience, 31*(13), 4792-4804.

Freeman, J., Heeger, D. J., & Merriam, E. P. (2013). Coarse-scale biases for spirals and orientation in human visual cortex. *The Journal of Neuroscience, 33*(50), 19695-19703.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience, 3*(6), 535-536.

Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: neuronal correlates of transient memory. *Journal of Neurophysiology*.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science, 173*(3997), 652-654.

Gardner, J. L. (2010). Is cortical vasculature functionally organized? *NeuroImage, 49*(3), 1953-1956.

Goense, J. B., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology, 18*(9), 631-640.

Goldman-Rakic, P., Roberts, A., Robbins, T., & Weiskrantz, L. (1998). The prefrontal cortex: executive and cognitive functions. *The prefrontal cortex: executive and cognitive functions*.

Grosbras, M. H., & Paus, T. (2003). Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *European Journal of Neuroscience, 18*(11), 3121-3126.

Hallett, M. (2007). Transcranial magnetic stimulation: a primer. *Neuron, 55*(2), 187-199.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*, 2425-2429.

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

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8*. doi:10.3389/fninf.2014.00088

Hubel, D., & Wiesel, T. (1963). Shape and arrangement of columns in cat's striate cortex. *The Journal of physiology, 165*(3), 559.

Hubel, D., & Wiesel, T. (1969). Anatomical demonstration of columns in the monkey striate cortex. *Nature, 221*(5182), 747-750.

Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculo‑cortical fibers in the macaque monkey. *Journal of Comparative Neurology, 146*(4), 421-450.

Kamitani, Y., & Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *NeuroImage, 49*(3), 1949-1952.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*, 679-685.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience, 1*(2), 91-100.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of Visual Attention in the Human Cortex. *Annual Review of Neuroscience, 23*(1), 315-341. doi:doi:10.1146/annurev.neuro.23.1.315

Knight, R. T., Staines, W. R., Swick, D., & Chao, L. L. (1999). Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica, 101*(2), 159-178.

Knight, R. T., & Stuss, D. T. (2002). Prefrontal cortex: The present and the future. *Principles of frontal lobe function*, 573-597.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181-1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11*(6), 229-235.

Kriegeskorte, N., Cusack, R., & Bandettini, P. (2010). How does an fMRI voxel sample the neuronal activity pattern: compact-kernel or complex spatiotemporal filter? *NeuroImage, 49*(3), 1965-1976.

Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology*.

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

Logothetis, N. K. (2008b). What we can do and what we cannot do with fMRI. *Nature, 453*(7197), 869-878. doi:10.1038/nature06976

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412*(6843), 150-157.

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., & Logothetis, N. K. (2012). The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *The Journal of Neuroscience, 32*(4), 1395-1407.

Malmo, R. B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. *Journal of Neurophysiology*.

Maloney, R. T. (2015). The basis of orientation decoding in human primary visual cortex: fine-or coarse-scale biases? *Journal of Neurophysiology, 113*(1), 1-3.

Mangels, J. A., Gershberg, F. B., Shimamura, A. P., & Knight, R. T. (1996). Impaired retrieval from remote memory in patients with frontal lobe damage. *Neuropsychology, 10*(1), 32.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202.

Moliadze, V., Zhao, Y., Eysel, U., & Funke, K. (2003). Effect of transcranial magnetic stimulation on single‐unit activity in the cat primary visual cortex. *The Journal of physiology, 553*(2), 665-679.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping, 15*(1), 1-25.

Noirhomme, Q., Lesenfants, D., Gomez, F., Soddu, A., Schrouff, J., Garraux, G., . . . Laureys, S. (2014). Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. *NeuroImage: Clinical, 4*, 687-694.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in neurosciences, 33*(8), 355-361.

Opitz, A., Fox, M. D., Craddock, R. C., Colcombe, S., & Milham, M. P. (2016). An integrated framework for targeting functional networks via transcranial magnetic stimulation. *NeuroImage, 127*, 86-96.

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: a primer. *Neuron, 87*(5), 932-945.

Pasley, B. N., Allen, E. A., & Freeman, R. D. (2009). State-dependent variability of neuronal responses to transcranial magnetic stimulation of the visual cortex. *Neuron, 62*(2), 291-303.

Paus, T. (2005). Inferring causality in brain images: a perturbation approach. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360*(1457), 1109-1114.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia, 12*(3), 323-330.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: from modulation of sensory processing to top-down control. *The Journal of Neuroscience, 23*(10), 3990-3998.

Postle, B. R. (2005). Delay-period activity in the prefrontal cortex: one function is sensory gating. *Cognitive Neuroscience, Journal of, 17*(11), 1679-1690.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276*(5313), 821-824.

Reynolds, J. R., O'Reilly, R. C., Cohen, J. D., & Braver, T. S. (2012). The function and organization of lateral prefrontal cortex: a test of competing hypotheses. *PloS one, 7*(2), e30284.

Roy, J. E., Riesenhuber, M., Poggio, T., & Miller, E. K. (2010). Prefrontal Cortex Activity during Flexible Categorization. *Journal of Neuroscience, 30*(25), 8519-8528. doi:10.1523/jneurosci.4837-09.2010

Ruff, C. C., Driver, J., & Bestmann, S. (2009). Combining TMS and fMRI: from 'virtual lesions' to functional-network accounts of cognition. *Cortex, 45*(9), 1043-1049.

Sandrini, M., Umiltà, C., & Rusconi, E. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: a new synthesis of methodological issues. *Neuroscience & Biobehavioral Reviews, 35*(3), 516-536.

Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. (2006). The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron, 51*(5), 661-670.

Scannell, J. W., & Young, M. P. (1999). Neuronal population activity and functional imaging. *Proceedings of the Royal Society of London B: Biological Sciences, 266*(1422), 875-881.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*(2), 207-218.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience, 7*(2), 144-152.

Shmuel, A., Chaimow, D., Raddatz, G., Ugurbil, K., & Yacoub, E. (2010). Mechanisms underlying decoding at 7 T: ocular dominance columns, broad structures, and

macroscopic blood vessels in V1 convey information on the stimulated eye. *NeuroImage, 49*(3), 1957-1964.

Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex, 45*(9), 1035-1042.

Silvanto, J., & Muggleton, N. G. (2008). New light through old windows: Moving beyond the "virtual lesion" approach to transcranial magnetic stimulation. *NeuroImage, 39*(2), 549-552.

Skinner, J. E. (1984). Central gating mechanisms that regulate event-related potentials and behavior *Self-regulation of the Brain and Behavior* (pp. 42-55): Springer.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(5408), 1657-1661.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage, 65*, 69-82.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Stuss, D., Floden, D., Alexander, M., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39*(8), 771-786.

Töpper, R., Mottaghy, F., Brügmann, M., Noth, J., & Huber, W. (1998). Facilitation of picture naming by focal transcranial magnetic stimulation of Wernicke's area. *Experimental Brain Research, 121*(4), 371-378.

Vendrell, P., Junqué, C., Pujol, J., Jurado, M. A., Molet, J., & Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia, 33*(3), 341-352.

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi:10.1152/jn.90355.2008

Walsh, V., & Pascual-Leone, A. (2003). Transcranial magnetic stimulation: a neurochronometrics of mind MIT Press. *Cambridge, Mass*.

Wardle, S., Ritchie, J. B., Seymour, K., & Carlson, T. (2015). What information is' decoded'from stimulus orientation with fMRI and MVPA? *Journal of vision, 15*(12), 993-993.

Williams, M. A., Dang, S., & Kanwisher, N. G. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature Neuroscience, 10*(6), 685-686.

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible Coding of Task Rules in Frontoparietal Cortex: An Adaptive System for Flexible Cognitive Control. *Journal of Cognitive Neuroscience, 27*(10), 1895-1911.

Woolgar, A., Jackson, J., & Duncan, J. (in press). Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis. *Journal of Cognitive Neuroscience*.

Yacoub, E., Shmuel, A., Logothetis, N., & Uğurbil, K. (2007). Robust detection of ocular dominance columns in humans using Hahn Spin Echo BOLD functional MRI at 7 Tesla. *NeuroImage, 37*(4), 1161-1177.

Yang, Z., Huang, Z., Gonzalez-Castillo, J., Dai, R., Northoff, G., & Bandettini, P. (2014). Using fMRI to decode true thoughts independent of intention to conceal. *NeuroImage, 99*, 80-92.

Yingling, C., & Skinner, J. (1977). Gating of thalamic input to cerebral cortex by nucleus reticularis thalami. *Progress in clinical neurophysiology, 1*, 70-96.

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*(5), 656-661.

Appendix A

The following paper is one that I co-authored in addition to my PhD work and is therefore not included in the main body of the thesis.

Coding of Visual, Auditory, Rule, and Response Information in the Brain: 10 Years of Multivoxel Pattern Analysis

Alexandra Woolgar[1], Jade Jackson[1], and John Duncan[2,3]

Abstract

■ How is the processing of task information organized in the brain? Many views of brain function emphasize modularity, with different regions specialized for processing different types of information. However, recent accounts also highlight flexibility, pointing especially to the highly consistent pattern of frontoparietal activation across many tasks. Although early insights from functional imaging were based on overall activation levels during different cognitive operations, in the last decade many researchers have used multivoxel pattern analyses to interrogate the representational content of activations, mapping out the brain regions that make particular stimulus, rule, or response distinctions. Here, we drew on 100 searchlight decoding analyses from 57 published papers to charac-terize the information coded in different brain networks. The outcome was highly structured. Visual, auditory, and motor networks predominantly (but not exclusively) coded visual, auditory, and motor information, respectively. By contrast, the frontoparietal multiple-demand network was characterized by domain generality, coding visual, auditory, motor, and rule information. The contribution of the default mode network and voxels elsewhere was minor. The data suggest a balanced picture of brain organization in which sensory and motor networks are relatively specialized for information in their own domain, whereas a specific frontoparietal network acts as a domain-general "core" with the capacity to code many different aspects of a task. ■

INTRODUCTION

Multivoxel pattern analysis (MVPA) of fMRI data is a powerful and increasingly popular technique used to examine information coding in the human brain. In MVPA, information coding is inferred when the pattern of activation across voxels can reliably discriminate between two or more events such as different stimuli, task rules, and participant responses (e.g., Haynes & Rees, 2006; Haxby et al., 2001). For example, if, in a certain brain region, the patterns of activation elicited in response to viewing red objects are more similar to each other than to the patterns elicited by green objects (and vice versa), we conclude that there is information in the patterns that discriminates between red and green objects and therefore codes for color. This allows infer-ence beyond traditional univariate brain mapping (e.g., this region is more active for colored objects than black and white ones) to examine the particular discrimina-tions that the region is able to make (e.g., the region carries specific information about what color was pre-sented). Information coding may be tested by compar-ing the correlation of patterns within object classes to correlations between object classes (e.g., Haxby et al.,

2001), or using a machine learning algorithm such as a pattern classifier. For example, if a classifier can be trained to discriminate between red and green objects, such that it can predict object color on an independent set of data, we conclude that the pattern of activation can be used reliably to discriminate between red and green objects. The technique has also been generalized to incor-porate multiple classes to test more complex representa-tional models (e.g., representational similarity analysis; Kriegeskorte, Mur, & Bandettini, 2008). It has been used to examine neural coding of a wide range of different task events including aspects of stimuli, task rules, participant responses, rewards, emotion, and language (e.g., McNamee, Rangel, & O'Doherty, 2013; Herrmann, Obleser, Kalberlah, Haynes, & Friederici, 2012; Woolgar, Thompson, Bor, & Duncan, 2011; Peelen & Vuilleumier, 2010; Haxby et al., 2001).

Using a "searchlight," MVPA can be used to map out the brain regions that code for each particular type of information (Kriegeskorte, Goebel, & Bandettini, 2006). For each brain voxel in turn, pattern analysis is applied to the pattern of activation across voxels in the local neighborhood (e.g., in a sphere of a fixed radius centered on the current voxel of interest), and the resulting metric, which summarizes the strength of information coding in the local neighborhood, is given to the center voxel. The resulting whole-brain map indicates where in the

[1]Macquarie University, Sydney, Australia, [2]MRC Cognition and Brain Sciences Unit, Cambridge, UK, [3]University of Oxford

brain a particular distinction is coded. This technique allows for exploratory analyses that are free from a priori hypotheses about where local patterns will be discriminative and opens the door for unpredicted findings.

After several years of searchlight MVPA, we now have an unprecedented opportunity to summarize our knowledge of information coding in the brain. This is the aim of the current paper. In the literature, most cognitive tasks comprise visual and/or auditory input, task rules, and motor output, so we focus our analysis on coding of these task features. We examine the frequency of information coding reported in five brain networks: the visual, auditory, and motor networks; the frontoparietal multiple demand (MD; Duncan, 2006, 2010) or "task-positive" (Fox et al., 2005) network; and a "task-negative" (Fox et al., 2005) or "default mode" (Raichle et al., 2001) network (DMN).

Although traditional accounts of brain organization emphasized modularity of function, several recent proposals highlight the flexibility of many brain regions (e.g., Yeo et al., 2014; Dehaene & Naccache, 2001; Duncan, 2001). For example, one of the most consistent findings in human neuroimaging is a characteristic pattern of activation in the frontoparietal MD network across a wide range of different cognitive tasks (e.g., Yeo et al., 2014; Fedorenko, Duncan, & Kanwisher, 2013; Niendam et al., 2012; Dosenbach et al., 2006; Naghavi & Nyberg, 2005; Owen, McMillan, Laird, & Bullmore, 2005; Duncan & Owen, 2000). This common activity may reflect the common need for cognitive control, one aspect of which is proposed to be the adaptive representation of task-relevant information (Duncan, 2001, 2010). Accordingly, the suggestion is that single neurons in the MD regions adjust their pattern of firing to encode the specific information currently relevant for the task, including stimuli, cues, rules, responses, etc.

The result of our review is a balanced and highly structured picture of brain organization. According to the MVPA data published in the last decade, auditory, visual, and motor networks predominantly code information from their own domain, whereas the frontoparietal MD network is characterized by domain generality, coding all four task features (visual, auditory, motor, and rule information) more frequently than other brain areas. After correcting for network area and the number of studies examining each feature, the contribution of the DMN and cortex elsewhere is minor. Although sensory and motor networks are relatively specialized for information in their own domain, the MD network appears to act as a domain-general core with the capacity to code different aspects of a task as needed for behavior.

METHODS

Paper Selection

We identified peer-reviewed papers published up until the end of December 2014 by searching PubMed, Scopus, Web of Science, HighWire, JSTOR, Oxford University Press Journals, and ScienceDirect databases with the following search terms: "MVPA searchlight," "multivariate analysis searchlight," "multivoxel analysis searchlight," and "MVPA spotlight" in any field. We additionally retrieved all the studies listed by Google scholar as citing Kriegeskorte et al. (2006) in which the procedure for searchlight MVPA was first described. This yielded 537 empirical papers (excluding reviews, comments, methods papers, or conference abstracts). Of these, we included studies that performed volumetric searchlight analysis (Kriegeskorte et al., 2006) across the whole brain of healthy adults and reported a complete list of the coordinates of peak decoding in template (MNI or TAL) space.[1] Because most tasks comprise visual or auditory input, task rules, and motor output, we focused on these task features. From each of the papers, we identified independent analyses that isolated the multivoxel representation of a single one of these task features. To achieve this, if a paper reported two or more nonindependent analyses (e.g., analyzed overlapping aspects of the same data), only one analysis was included. We excluded any analyses in which sensory and motor responses were confounded (e.g., if the same visual stimulus was associated with the same motor response). This procedure yielded a total of 100 independent analyses from 57 papers.

Characterization of Task Features

We categorized each of the 100 analyses according to what task feature they examined, namely, whether they examined the multivoxel discrimination between two or more visual stimuli, two or more auditory stimuli, two or more task rules, or two or more motor responses (Table 1). This categorization was done twice, the first time being as inclusive as possible, and the second time using stricter criteria (Table 1, second column). For the strict categorization, we excluded analyses in which the multivoxel discrimination pertained to both an aspect of the physical stimulus and a higher-level stimulus attribute such as emotion or semantic category. Analyses focusing on linguistic stimuli (e.g., written or spoken words) were not included, on the basis that representation of these stimuli would likely load on language-related processing more than visual and/or auditory information processing.

Analyses pertaining to the discrimination of visual stimuli included discrimination of stimulus orientation, position, color, and form. Additional analyses pertaining to the semantic category of the visual stimulus (e.g., animals vs. tools; Simanova, Hagoort, Oostenveld, & van Gerven, 2014) and stimuli that were consistently associated with different rewards (e.g., face vs. currency, where a picture of currency indicated a monetary reward; Clithero, Smith, Carter, & Huettel, 2011) were included in our lenient categorization but excluded from the strict categorization. In our strict categorization, we also excluded two further studies in which there was a possibility that the

301

Table 1. Multivoxel Decoding Analyses Included in This Study

Category	Included in Strict Categorization?	Study	Decoding Analysis	Searchlight Size	Threshold at Which Results Were Reported	Number of Participants
Visual	Yes	Bode and Haynes (2009)	Target stimuli (dynamic color patterns)	4 voxel radius	$p < .001$ uncorr	12
Visual	Yes	Mayhew et al. (2010)	Radial vs. concentric glass pattern stimuli (young adults)	9 mm radius (av. 98 voxels)	$p < .05$, $k = 5$ mm^2	10
Visual	Yes	Mayhew et al. (2010)	Radial vs. concentric glass pattern stimuli (older adults)	9 mm radius (av. 98 voxels)	$p < .05$, $k = 5$ mm^2	10
Visual	Yes	Bogler, Bode, and Haynes (2011)	Most salient visual quadrant of grayscale pictures of natural scenes	6 voxel radius	$p < .05$, FWE	21
Visual	Yes	Carlin, Calder, Kriegeskorte, Nili, and Rowe (2011)	View-invariant gaze direction	5 mm radius	$p < .05$ FWE	18
Visual	Yes	Kahnt et al. (2011)	Stimulus orientation (low contrast Gabor in upper right visual field)	4 voxel radius	$p < .0001$, $k = 20$, cluster level corr $p < .001$	20
Visual	Yes	Kalberlah et al. (2011)	Which of 4 spatial locations participant is attending and responding to	12 mm radius	$p < 10-5$ at vertex level, $p < .005$ at cluster level	18
Visual	Yes	Vickery et al. (2011)	Visual stimulus (coin showing heads vs. tails side)	27 voxel cube	$p < .001$ uncorr, $k = 10$	17
Visual	Yes	Woolgar, Thompson, et al. (2011)	Stimulus position	5 mm radius	$p < .001$ uncorr	17
Visual	Yes	Woolgar, Thompson, et al. (2011)	Background color of screen (within rule)	5 mm radius	$p < .001$ uncorr	17
Visual	Yes	Guo, Preston, Das, Giesbrecht, and Eckstein (2012)	Target present vs. absent in full color natural scenes	9 mm radius (153 voxels volume)	$p < .005$, uncorr	12
Visual	Yes	Hebart, Donner, and Haynes (2012)	Direction of motion (dynamic random dot patterns)	10 mm radius	$p < .00001$, uncorr, $k = 30$	22
Visual	Yes	Peelen and Caramazza (2012)	Perceptual similarity of 12 familiar objects	8 mm radius	$p < .05$ Bonferroni, $k = 5$	15
Visual	Yes	Peelen and Caramazza (2012)	Pixelwise similarity of 12 familiar objects	8 mm radius	$p < .05$ Bonferroni, $k = 5$	15

Table 1. (*continued*)

Category	Included in Strict Categorization?	Study	Decoding Analysis	Searchlight Size	Threshold at Which Results Were Reported	Number of Participants
Visual	Yes	Reverberi et al. (2012a)	Visual cue (two visually unrelated abstract shapes coding for the same rule)	4 voxel radius (3 × 3 × 3.75 mm voxels)	$p < .05$ cluster corr	13
Visual	Yes	Billington, Furlan, and Wann (2013)	Congruency in depth information (congruent looming and binocular disparity cues vs. incongruent looming and binocular disparity cues)	6 mm radius	$p < .01$, Bonferroni corr	16
Visual	Yes	Bode, Bogler, and Haynes (2013)	Black and white photograph (piano vs. chair)	4 voxel radius	$p < .05$, FWE cluster corr	15
Visual	Yes	Mayhew and Kourtzi (2013)	Radial vs. concentric Glass pattern stimuli (young adults)	9 mm radius, av. 98 voxels volume	$p < .05$, cluster threshold 5 mm^2	10
Visual	Yes	Mayhew and Kourtzi (2013)	Radial vs. concentric Glass pattern stimuli (older adults)	9 mm radius, av. 98 voxels volume	$p < .05$, cluster threshold 5 mm^2	10
Visual	Yes	Clarke and Tyler (2014)	Low-level visual features (early visual cortex model)	7 mm radius	$p < .05$, FDR, $k = 20$	16
Visual	Yes	Pollmann et al. (2014)	Gabor patches differing in both color (red/green) and spatial frequency	3 voxels (10.5 mm) radius (123 voxels volume)	$p < .001$	15
Visual	No	Clithero et al. (2011)	Image of face vs. currency (currency influenced participant payment), within participants analysis	12 mm radius (max 123 voxels)	$p < .05$ Bonferroni correction	16
Visual	No	Vickery et al. (2011)	Visual stimulus (photograph of hand making rock/paper/scissor action)	27 voxel cube	$p < .001$ uncorr	22
Visual	No	Christophel, Hebart, and Haynes (2012)	Complex artificial stimuli consisting of multicolored random fields (STM storage of visual stimuli during delay phase)	4 voxel radius	$p < .05$ FWE, $k = 20$	17
Visual	No	Bode, Bogler, Soon, and Haynes (2012)	Category of visual stimulus (piano/chair/noise), high-visibility condition	4 voxel radius	$p < .0001$ uncorr	14

Uncorrected Proof

303

Visual	No	Carlin et al. (2012)	Left vs. right head turn (silent video clips)	5 mm radius	$p < .05$ FWE	17
Visual	No	Gilbert, Swencionis, and Amodio (2012)	Black vs. white faces (color photographs, collapsed over trait and friendship judgment tasks)	3 voxel radius	$p < .05$ FWE	16
Visual	No	Kaplan and Meyer (2012)	Discrimination between five 5-sec video clips showing manipulation of different objects (plant, tennis ball, skein of yarn, light bulb, set of keys) within-subject analysis	8 mm radius (average 75 voxels)	>maximum value given by a permutation test in a occipital spherical ROI for each subject	8
Visual	No	Linden, Oosterhof, Klein, and Downing (2012)	Visual category (face/body/scene/flower) (encoding phase)	100 voxels volume	$p < .05$, voxelwise uncorr, cluster correction using Monte Carlo simulation	18
Visual	No	Linden et al. (2012)	Visual category (face/body/scene/flower) (delay 1 phase)	100 voxels volume	$p < .05$, voxelwise uncorr, cluster correction using Monte Carlo simulation	18
Visual	No	Murawski, Harris, Bode, Dominguez, and Egan (2012)	Subliminal presentation of apple logo vs. neutral cup	3 voxel radius	$p < .05$ FWE cluster corr	13
Visual	No	Peelen and Caramazza (2012)	Conceptual level information about 12 familiar objects: kitchen vs. garage, and rotate vs. squeeze	8 mm radius	$p < .05$ Bonferroni, $k = 5$	15
Visual	No	Weygandt, Schaefer, Schienle, and Haynes (2012)	Food vs. neutral images (normal weight control group)	3 voxel radius	$p < .001$ uncorr $k = 10$ and $p < .05$ FWE	19
Visual	No	McNamee et al. (2013)	Visual category (food/money/trinkets)	20 mm radius	$p < .05$ FDR $k = 10$	13
Visual	No	Clarke and Tyler (2014)	Semantic features of visual stimuli	7 mm radius	$p < .05$, FDR, $k = 20$	16
Visual	No	Clarke and Tyler (2014)	Category of visual stimuli (animals vs. fruits vs. vegetables vs. vehicles vs. tools vs. musical instruments)	7 mm radius	$p < .05$, FDR, $k = 20$	16

Table 1. (*continued*)

Category	Study	Included in Strict Categorization?	Decoding Analysis	Searchlight Size	Threshold at Which Results Were Reported	Number of Participants
Visual	Clarke and Tyler (2014)	No	Animal and plants vs. nonbiological visual stimuli	7 mm radius	$p < .05$, FDR, $k = 20$	16
Visual	Clarke and Tyler (2014)	No	Animal visual stimuli (a model in which patterns for animal stimuli are similar to one another and all other stimuli are dissimilar from animals and from each other)	7 mm radius	$p < .05$, FDR, $k = 20$	16
Visual	Simanova et al. (2014)	No	Category of visual stimulus (animal vs. tool) – photographs	2.5 voxels, 8.75 mm radius (33 voxel volume)	$p < .001$, FDR	14
Visual	FitzGerald, Schwartenbeck, and Dolan (2014)	No	Discriminate between two abstract visual stimuli when attending to visual stimuli (stimuli differentially associated with financial reward)	6 mm radius (31 voxels volume)	$p < .05$ FWE-corrected	25
Visual	FitzGerald et al. (2014)	No	Discriminate between two abstract visual stimuli when attending to concurrently presented auditory stimuli (visual stimuli differentially associated with financial reward)	6 mm radius (31 voxels volume)	$p < .05$ FWE-corrected	25
Auditory	Alink, Euler, Kriegeskorte, Singer, and Kohler (2012)	Yes	Direction of auditory motion (rightwards vs. leftwards)	1.05 cm radius (72 voxel volume)	$T > 4.0$, $k > 4$	19
Auditory	Lee, Janata, Frost, Hanke, and Granger (2011)	Yes	Ascending vs. descending melodic sequences	2 neighboring voxels (max 33 voxels)	$p < .05$ cluster corr	12
Auditory	Linke, Vicente-Grabovetsky, and Cusack (2011)	Yes	Frequency-specific coding (discriminate pure tones in four frequency ranges)	10 mm radius	$p < .005$ FDR	16 (9 with two sessions, 7 with one session)
Auditory	Giordano et al. (2013)	Yes	Pitch (median)	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	Giordano et al. (2013)	Yes	Loudness (median)	6.25 mm radius	$p < .0001$, $k = 20$	20

305

Modality	Includes MD	Study	Decoding	Searchlight size	Threshold	N
Auditory	Yes	Giordano et al. (2013)	Spectral centroid (interquartile range)	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	Yes	Giordano et al. (2013)	Harmonicity (median)	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	Yes	Giordano et al. (2013)	Loudness (cross-correlation)	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	Yes	Lee, Turkeltaub, Granger, and Raizada (2012)	/ba/ vs. /da/ speech category (3 voxel searchlight)	3 voxel radius	$p < .001$ voxelwise uncorr and $p < .05$ clusterwise-corrected	13
Auditory	Yes	Lee et al. (2012)	/ba/ vs. /da/ speech category (reanalysis of a previous data set)	3 voxel radius	$p < .001$ voxelwise uncorr and $p < .05$ clusterwise-corrected	12
Auditory	Yes	Merrill et al. (2012)	Hummed speech prosody (rhythm + pitch) vs. speech rhythm	6 mm radius	$p < .05$ cluster size-corrected	21
Auditory	Yes	Merrill et al. (2012)	Hummed song melody (pitch + rhythm) vs. musical rhythm	6 mm radius	$p < .05$ cluster size-corrected	21
Auditory	Yes	Jiang, Stecker, and Fine (2014)	Apparent direction of unambiguous auditory motion (left vs. right, 50% coherence, sighted subjects)	2 mm radius (33 voxels volume)	$p < .001$ corrected	7
Auditory	Yes	Klein and Zatorre (2014)	Musical category (minor third vs. major third)	3 voxels radius (max 123 voxels volume)	$p < .001$ (uncorrected)	10
Auditory	No	Giordano et al. (2013)	Human similar	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	No	Giordano et al. (2013)	Living similar	6.25 mm radius	$p < .0001$, $k = 20$	20
Auditory	No	Kotz et al. (2013)	Emotion of vocal expression (angry, happy, sad, surprised or neutral)	9 mm radius (3 × 3 × 3 mm voxels)	$p < .0001$ uncorr	20
Auditory	No	Merrill et al. (2012)	Spoken sentences (words + rhythm + pitch) vs. hummed speech prosody (rhythm + pitch)	6 mm radius	$p < .05$ cluster size-corrected	21
Auditory	No	Merrill et al. (2012)	Sung sentences (words + pitch + rhythm) vs. hummed song melody (pitch + rhythm)	6 mm radius	$p < .05$ cluster size-corrected	21
Auditory	No	Simanova et al. (2014)	Category of sound	2.5 voxels, 8.75 mm	$p < .001$, FDR	14

Woolgar, Jackson, and Duncan 7

306

Table 1. (*continued*)

Category	Included in Strict Categorization?	Study	Decoding Analysis	Searchlight Size	Threshold at Which Results Were Reported	Number of Participants
			(animal vs. tool)	radius (33 voxel volume)		
Auditory	No	Boets et al. (2013)	Speech sounds (sounds differing on both consonant and vowel vs. neither), typical readers	9 mm radius (123 voxels volume)	$p < .001$ voxelwise uncorr and $p < .05$ FWE clusterwise	22
Auditory	No	Zheng et al. (2013)	Speech stimuli vs. noise during passive listening	4 mm radius	$p < .05$, FWE	20
Auditory	No	Jiang et al. (2014)	Reported apparent direction of ambiguous auditory motion (left vs. right), 0% coherence, sighted subjects	2 mm radius (33 voxels volume)	$p < .001$ corrected	7
Rule	Yes	Haynes et al. (2007)	Intended task addition vs. subtraction	3 voxel radius	$p < .005$ uncorr	8
Rule	Yes	Bode and Haynes (2009)	Task rules (stimulus–response mapping)	4 voxel radius	$p < .001$ uncorr	12
Rule	Yes	Greenberg, Esterman, Wilson, Serences, and Yantis (2010)	Type of attentional shift to make (shift attention to alternate location vs. shift attention to alternate color)	27 voxel cube	$p < .05$ cluster correction	8
Rule	Yes	Vickery et al. (2011)	Participants upcoming choice (switch or stay relative to last choice)	27 voxel cube	$p < .001$ uncorr	22
Rule	Yes	Woolgar, Thompson, et al. (2011)	Stimulus–response mapping rule	5 mm radius	$p < .001$ uncorr	17
Rule	Yes	Hebart et al. (2012)	Stimulus–response mapping rule	10 mm radius	$p < .00001$, uncorr, $k = 30$	2
Rule	Yes	Momennejad and Haynes (2012)	Task participants are intending to perform (parity or magnitude task) during maintenance phase	4 voxel radius	$p < .0001$ uncorr, $k = 0$	12
Rule	Yes	Momennejad and Haynes (2012)	Task participants are performing (parity or magnitude task) during retrieval phase	4 voxel radius	$p < .0001$ uncorr, $k = 0$	12

307

Rule	Yes	Momennejad and Haynes (2012)	Time delay after which participants should self-initiate a task switch (15, 20 or 25 sec), during maintenance phase	4 voxel radius	$p < .0001$ uncorr, $k = 0$	12
Rule	Yes	Nee and Brown (2012)	Higher level task context (first delay period)	10 mm radius	$p < .05$ cluster-corrected ($p < .005$ with $k = 66$ – extent thresh given using simulations in AlphaSim)	21
Rule	Yes	Nee and Brown (2012)	Higher and lower level task context (second delay period)	10 mm radius	$p < .05$ cluster-corrected ($p < .005$ with $k = 66$ – extent thresh given using simulations in AlphaSim)	21
Rule	Yes	Reverberi et al. (2012a)	Rule representation (e.g., if there is a house press left)	4 voxel radius ($3 \times 3 \times 3.75$ mm voxels)	$p < .05$ cluster corr	13
Rule	Yes	Reverberi et al. (2012b)	Stimulus response mapping "rule identity" (if furniture then A, if transport then B vs. if furniture then B, if transport then A)	4 voxel radius	$p < .05$ corrected	14
Rule	Yes	Reverberi et al. (2012b)	Order in which rules are to be applied	4 voxel radius	$p < .05$ corrected	14
Rule	Yes	Soon et al. (2013)	Task (search for house/face/car/bird) during preparatory period	20 mm radius	$p < .05$ Bonferroni-corrected against 55%, 25 mm cluster threshold	15
Rule	Yes	Zhang et al. (2013)	Current rule: attend to direction of motion, color or size of dots in a random dot kinematogram	2 voxel radius	$p < .001$ uncorr $k = 35$	20
Rule	Yes	Helfinstein et al. (2014)	Safe vs. risky choice to be taken by the participant on the following trial	Not reported	>60%, whole-brain cluster-corrected $p < .05$ via comparison with 1,000 random permutations	108

Uncorrected Proof

Table 1. (continued)

Category	Included in Strict Categorization?	Study	Decoding Analysis	Searchlight Size	Threshold at Which Results Were Reported	Number of Participants
Rule	Yes	Jiang and Egner (2014)	Congruency of face/word compound stimulus (congruent vs. incongruent) in a face gender categorization task	3 voxel radius (3 × 3 × 3 mm voxels)	$p < .05$ corrected	21
Rule	Yes	Jiang and Egner (2014)	Congruency of decision with side of response (congruent vs. incongruent)	3 voxel radius (3 × 3 × 3 mm voxels)	$p < .05$ corrected	21
Rule	Yes	Wisniewski, Reverberi, Tusche, and Haynes (2014)	Stimulus–response mapping rule (selected by participant)	4 voxel radius	$p < .001$ FWE corr	14
Rule	No	Gilbert (2011)	Dual task (2-back + prospective memory task) vs. single task (2-back)	3 voxel radius	$p < .05$ FWE, $k = 10$	32
Rule	No	Ekman, Derrfuss, Tittgemeyer, and Fiebach (2012)	Prepare for upcoming color or motion task vs. neutral condition in which upcoming task was not known (no preparation possible)	8 mm radius	nonparametric permutation test, FDR threshold at $q = 0.05$, peaks more than 24 mm apart	9
Rule	No	Li and Yang (2012)	Task set: categorize glass patterns as radial or concentric when stimuli vary on angle vs. categorize glass patterns as radial or concentric when they vary on signal level	9 mm radius (voxels 3 × 3 × 3 mm)	$p < .01$, cluster correction using Monte Carlo simulation	20
Motor	Yes	Soon et al. (2008)	Left vs. right button press intention (before conscious decision indicated)	3 voxel radius	$p < .05$ FWE	14
Motor	Yes	Soon et al. (2008)	Left vs. right button press (execution)	3 voxel radius	$p < .05$ FWE	14
Motor	Yes	Soon et al. (2008)	Left vs. right response preparation (cued when to make choice)	3 voxel radius	$p < .001$ uncorr	14

309

Motor	Bode and Haynes (2009)	Motor response (leftward vs. rightward joystick movement)	4 voxel radius	$p < .05$ FWE	12
Motor	Woolgar, Thompson, et al. (2011)	Button press response (index vs. middle finger)	5 mm radius	$p < .001$ uncorr	17
Motor	Bode et al. (2012)	Button press response (index/middle/ring finger of right hand)	4 voxel radius	$p < .0001$ uncorr	14
Motor	Bode et al. (2013)	Button press (right index vs. middle finger)	4 voxel radius	$p < .05$, FWE cluster corr	15
Motor	Carp et al. (2011)	Left vs. right index finger tapping	12 mm radius	$p < 1e-7$, $k = 50$	37
Motor	Colas and Hsieh (2014)	Motor bias (whether participant will later press left or right button, operated with index finger of each hand, prestimulus display)	5 voxel sided cube	$p < .025$	14
Motor	Colas and Hsieh (2014)	Left vs. right button press (index finger of each hand)	5 voxel sided cube	$p < .01$, $k = 9$	14
Motor	Huang et al. (2014)	Left vs. right key press (left vs. right hand)	5 voxels (15 mm) cube	$p < .01$ cluster-corrected	14

corr = corrected; uncorr = uncorrected; FDR = false discovery rate correction; FWE = family-wise error correction; k = cluster extent threshold.

visual stimulus could evoke representation of motor actions. These were videos of head turns (Carlin, Rowe, Kriegeskorte, Thompson, & Calder, 2012) and photos of hands in rock/paper/scissor pose (Vickery, Chun, & Lee, 2011).

Analyses pertaining to the coding of auditory information included discrimination of the direction of auditory motion, pitch, loudness, and melody. Analyses pertaining to the semantic category of sound (e.g., animals vs. tools; Simanova et al., 2014) or emotion of vocal expression (Kotz, Kalberlah, Bahlmann, Friederici, & Haynes, 2013) were also included in our lenient categorization and excluded from the strict categorization.

Analyses pertaining to the discrimination of task rules included discrimination of different stimulus–response mappings (e.g., Bode & Haynes, 2009), intended tasks (e.g., addition vs. subtraction; Haynes et al., 2007) and task set (e.g., attend to motion vs. color vs. size; Zhang, Kriegeskorte, Carlin, & Rowe, 2013). Two analyses were included in our lenient categorization and excluded from the strict categorization. One was an analysis that discriminated a dual from single task (Gilbert, 2011), which was excluded from the strict categorization because of the obvious confound with difficulty (for discussion, see Woolgar, Golland, & Bode, 2014; Todd, Nystrom, & Cohen, 2013), and the other pertained to discrimination of task set where the stimuli were very similar but not identical between the two tasks (Li & Yang, 2012).

Analyses pertaining to the discrimination of motor responses included discrimination of different button presses and the direction of joystick movement during response preparation and execution. One analysis that discriminated between left and right finger tapping (Carp, Park, Hebrank, Park, & Polk, 2011) was also excluded from the strict categorization, because it was not clear whether the side to tap was cofounded with a visual cue. Two further studies were excluded from our stricter analysis, because it was unclear which of two possible motor responses was modeled (Colas & Hsieh, 2014; Huang, Soon, Mullette-Gillman, & Hsieh, 2014).

Analyses

Our first analysis quantified the prevalence of visual, auditory, rule, and motor information coding in different brain networks. We focused on Visual, Auditory, and Motor networks (capitalized to distinguish from visual, auditory, and motor task features), the frontoparietal MD network (Fedorenko et al., 2013; Fox et al., 2005; Duncan & Owen, 2000), and the DMN (Fox et al., 2005; Raichle et al., 2001). Our definition of the MD network was taken from the average activation map of Fedorenko et al. (2013), which is freely available online at imaging. mrc-cbu.cam.ac.uk/imaging/MDsystem. This map indicates the average activation for high relative to low demand versions of seven tasks including arithmetic, spatial

and verbal working memory, flanker, and Stroop tasks. Thus, the MD network definition is activation based: It indexes regions that show a demand-related univariate increase in activity across tasks. The map is symmetrical about the midline because data from the two hemispheres were averaged together in the original paper (Fedorenko et al., 2013). We used the parcellated map provided online in which the original average activation map was thresholded at $t > 1.5$ and then split into anatomical subregions (imaging.mrc-cbu.cam.ac.uk/imaging/MDsystem). This map includes restricted regions of frontal, parietal, and occipitotemporal cortices as well as a number of small subcortical regions. We only included frontal and parietal regions. The resulting 13 MD regions were located in and around the left and right anterior inferior frontal sulcus (aIFS; center of mass [COM] +/−35 47 19, 5.0 cm^3), left and right posterior inferior frontal sulcus (pIFS; COM +/−40 32 27, 5.7 cm^3), left and right anterior insula/ frontal operculum (AI/FO; COM +/−34 19 2, 7.9 cm^3), left and right inferior frontal junction (IFJ; COM +/−44 4 32, 10.1 cm^3), left and right premotor cortex (PM; COM +/−28 −2 56, 9.0 cm^3), bilateral ACC/pre-SMA (COM 0 15 46, 18.6 cm^3), and left and right intraparietal sulcus (IPS; COM +/−29 −56 46, 34.0 cm^3). Visual, Auditory, Motor, and DMN networks were taken from the whole-brain map provided by Power et al. (2011), which partitions the brain into networks based on resting state connectivity. The Visual network consisted of a large cluster of 182.6 cm^3 mm covering the inferior, middle, and superior occipital, calcarine, lingual and fusiform gyri, and the cuneus (BA 17, 18, 19, 37), with COM at MNI coordinates 1 −79 6, plus small clusters in left BA 37 (0.22 cm^3, COM −54 −65 −21) and right inferior parietal lobe (0.17 cm^3, COM 26 −55 55, BA 7). The Auditory network comprised two large clusters in left and right superior temporal gyrus and rolandic operculum (23.4 cm^3 in each hemisphere, with COM at −51 −22 12 and 52 −19 10, BA 22, 42). The Motor network comprised a large cluster over the precentral and postcentral gyri, paracentral lobule and SMA (107.7 cm^3, COM 1 −25 60, BA 4, 5, 6), and small clusters in the SMA at the midline (0.04 cm^3, COM 3 7 72) and left and right middle temporal gyrus (0.07 cm^3 with COM −48 −64 11 and 0.02 cm^3 with COM 55 −60 6). The DMN comprised six main clusters around the precuneus (extending to mid cingulate cortex, 43.9 cm^3, COM −1 −51 31, BA 7, 23), ventral ACC, and orbital frontal cortex extending dorsally along the medial part of the superior frontal gyrus (107.2 cm^3, COM −2 42 24, BA 9, 10, 11, 32), left and right angular gyrus (12.2 cm^3, COM −43 −66 34; 10.6 cm^3, COM 47 −62 32; BA 39), and left and right middle temporal lobe (18.7 cm^3, COM −58 −17 −13; 15.0 cm^3, COM 58 −11 −17, BA 21, 20). To ensure that the networks did not overlap, the MD network was masked with each of the other networks. Therefore, our definition of the MD network pertained to voxels that were not part of the Visual, Auditory, Motor, or DMN networks. To serve as a comparison

311

with our five principal networks, all other voxels in the voxelwise map of Power et al. (2011), which corresponds to the anatomical labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) and excludes the cerebellum, ventricles, and large white matter tracts, were considered as a residual, Other network. Definitions of the five principal networks are depicted in Figure 1.

For each of our task features, we counted the number of decoding peaks that were reported in each of our six networks, including Other (any decoding peaks reported using TAL coordinates were converted to MNI152 space using tal2mni; imaging.mrc-cbu.cam.ac.uk/downloads/MNI2tal/tal2mni.m). To visualize these data, for each task feature and network, we divided the relevant tally by the number of reported analyses for that task feature and the volume of the network and plotted them on a stacked bar chart. We visualized the data from the lenient and strict categorization separately. Using data from the strict categorization, we then carried out a series of chi-square analyses to test for statistical differences in the distribution of information coding across the networks. First, we carried out a one-way chi-square analysis on the total number of decoding peaks in each network. For this, the observed values were the raw numbers of decoding peaks (across all task features) reported in each network, and the expected values were set proportional to the volume of each network. This analysis tests whether the distribution of information coding between the networks is predicted by network volume. Second, we carried out a chi-square test of independence to assess whether the distribution of information about each task feature (visual, auditory, motor, and rule decoding

points) was independent of network (MD, Visual, Auditory, Motor, DMN, and Other). Finally, where significant effects were found in these first two analyses, we carried out a series of post hoc analyses considering each task feature and region separately to clarify the basis for the effect. For each task feature separately, we compared the distribution of observed coding (tally of decoding points in each network) to that predicted by the relative volumes of the six networks. This was done using chi-square (visual and rule information) or the equivalent exact goodness of fit multinomial test for situations where >20% of expected values were <5 (motor and auditory information; implemented in R version 3.2.2 (Team, 2015) using the XNomial package (Engels, 2014)). Finally we asked whether the tally of observed coding in each of the five principal networks separately was greater than that in Other, using a one-way chi-square test or a one-tailed exact binomial test where any expected value was <5.

Our second analysis concerned subdivisions within the MD network. Although the observation of the MD activation pattern in response to many kinds of demand emphasizes the similarity of their response, we do expect that there will be some functional differences between the different regions (e.g., Fedorenko et al., 2013). To explore this, we first carried out a one-way chi-square comparing the total number of decoding peaks reported in the seven different MD regions (aIFS, pIFS, AI/FO, IFJ, PM, ACC/pre-SMA, IPS; data pooled over hemispheres). Next, we divided the MD regions into two subnetworks: a frontoparietal (FP) subnetwork, comprising the IPS, IFJ, and pIFS MD regions, and a cingulo-opercular (CO)
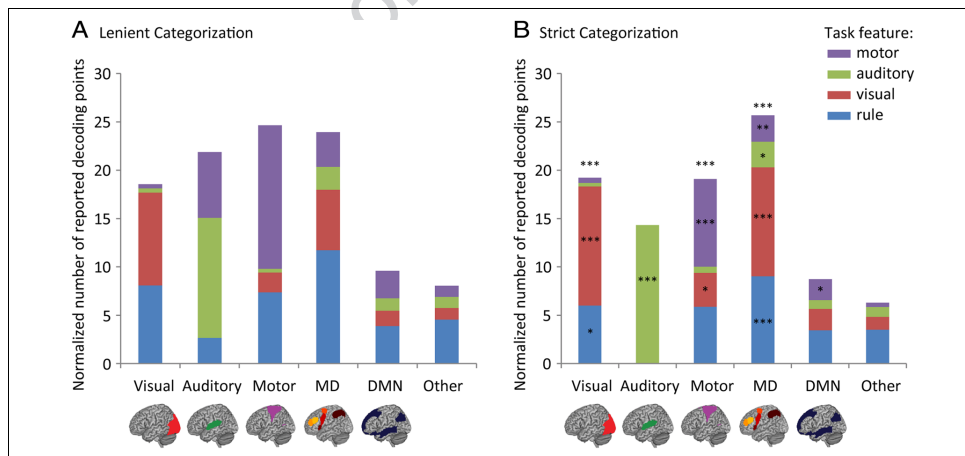

Figure 1. Number of significant decoding points reported in each network, after correcting for the number of analyses examining coding of each task feature and network volume. Asterisks indicate significance of chi-square or exact binomial goodness of fit tests examining whether there was more coding in each principal network compared with Other for all peaks (above bars) or for each task feature separately (asterisks on colored bar segments). Statistical testing was carried out for the strict categorization data only. *$p < .05$, **$p < .01$, ***$p < .00001$.

subnetwork comprising ACC/pre-SMA, AI/FO, and aIFS MD regions (Power & Petersen, 2013; Power et al., 2011; Dosenbach et al., 2007). We carried out one-way chi-square test comparing the total number of decoding peaks reported in the two subnetworks to each other and to coding in Other. We again used chi-square or the equivalent exact test (Freeman & Halton, 1951) to test for independence between subnetwork and task feature and to compare coding of each feature between the two subnetworks. Statistical testing was again carried out for the "strict" categorization data only.

RESULTS

We summarized 100 independent decoding analyses, reported in 57 published papers, that isolated the multivoxel representation of a single one of the following task features: visual and auditory stimuli, task rules, or motor output. First, we compared information coding in each of our five a priori networks of interest, with Other included as a baseline. The data, shown in Figure 1, suggest a highly structured distribution. For data from the strict categorization (Figure 1B), we used a series of chi-square analyses and exact tests to examine the statistical differences between networks. First we asked whether there was more decoding in some networks compared with others, over and above the differences expected due to variation in network volume (see Methods). Indeed, the total number of decoding peaks varied significantly between the six networks even after network volume was accounted for (χ^2 (5, $n = 365$) = 157.16, $p < .00001$). Second, we asked whether there was a relationship between the distribution of coding of the different task features and the different brain networks. This chi-square test of independence was also highly significant (χ^2 (15, $n = 365$) = 172.34, $p < .00001$), indicating a significant relationship between task feature and brain network. We carried out a series of post hoc analyses to clarify the basis for these effects. For this, we considered each task feature separately and compared the number of reported points to the number that would be expected based on the relative volumes of the six networks. For all four task features separately, coding differed significantly between networks (visual information: χ^2 (5, $n = 153$) = 188.37, $p < .00001$; auditory information: exact test $p < .00001$; rule information: χ^2 (5, $n = 151$) = 29.47, $p = .00002$; motor information: exact test $p < .00001$). For visual information, compared with expectations based on network volume, coding in the Visual (χ^2 (1, $n = 84$) = 140.71, $p < .00001$), Motor (exact test, $p = .015$), and MD (χ^2 (1, $n = 77$) = 119.65, $p < .00001$) networks was significantly more frequent than coding in Other. No such difference was seen for visual information coding in the DMN and Auditory networks (ps > .13). Auditory information coding was reported more frequently in the Auditory (exact test, $p < .00001$) and MD (exact test, $p = .043$) networks compared with

Other (for DMN, Motor, and Visual networks compared with Other, ps > .68). Rule information coding was reported more frequently in the MD (χ^2 (1, $n = 99$) = 21.06, $p < .00001$) and Visual (χ^2 (1, $n = 89$) = 5.02, $p = .03$) networks compared with Other (equivalent tests for DMN, Auditory and Motor networks, ps > .09). Motor information was coded more frequently in the Motor (exact test, $p < .00001$), MD (exact test, $p = .008$), and DMN (exact test, $p = .019$) networks compared with Other (equivalent tests for Visual and Auditory networks, ps > .61). Therefore, relative to Other, the MD network showed more coding of all four task features (visual, auditory, rule, and motor), the DMN showed more coding of motor information, the Motor network showed more coding of motor and visual information, the Visual network showed more coding of visual and rule information, and the Auditory network showed more coding of auditory information.

Our second series of analyses concerned subdivisions within the MD network, again using data from the strict categorization. First, we looked at total number of decoding peaks in each region, combining across task feature (visual, auditory, motor, rule). There was no evidence for a difference between the seven MD regions, again compared with expectations based on region volume (data collapsed over hemisphere, χ^2 (6, $n = 93$) = 5.77, $p = .45$). Second, we asked whether there were differences in the reported representational content of two putative subnetworks, an FP subnetwork (IPS, IFJ, and pIFS), proposed to support transient control processes, and a CO network (ACC/pre-SMA, AI/FO, and aIFS), proposed to support sustained control processes (Dosenbach et al., 2007). The data are shown in Figure 2. There was no evidence for a difference in the frequency of information coding in these two subnetworks (χ^2 (1, $n = 84$) = 2.62, $p = .11$), with encoding in both subnetworks more frequent than encoding in Other (FP: χ^2 (1, $n = 178$) = 124.28, $p < .00001$; CO: χ^2 (1, $n = 132$) = 23.99, $p < .00001$). Interestingly, however, there was a significant relationship between subnetwork and information type (Freeman–Halton extension of Fisher's exact test, $p = .002$), suggesting that the two networks had different representational profiles. The dissociation was driven by more coding of visual information in FP than CO (χ^2 (1, $n = 41$) = 6.65, $p = .010$) and more coding of motor information in CO than in FP (two-tailed binomial exact test, 0% of motor points in FP was less than the 69.2% predicted based on the two subnetwork volumes, $p = .009$). Visual points were reported in all FP regions as well as in ACC–pre-SMA and AI/FO, and motor points were reported in ACC/pre-SMA and aIFS. There was no difference in coding between the subnetworks for rule or auditory information, ps > .48. The pattern of results did not change if ROIs were restricted to gray matter or if coordinates reported in TAL were converted to MNI using the tal2icbm_spm routine provided with GingerALE (www.brainmap.org/icbm2tal/) instead of tal2mni.

Figure 2. Number of significant decoding points reported in each MD subnetwork after correcting for the number of analyses examining coding of each task feature and subnetwork volume. Asterisks indicate significance of chi-square or exact binomial goodness of fit tests examining whether there was more coding in each subnetwork compared with Other for all peaks (above bars) or for each task feature separately (asterisks on colored bar segments) and comparing coding of each task feature between the two subnetworks (asterisks above colored horizontal lines). Statistical testing was carried out for the strict categorization data only. $*p < .05$, $**p < .01$, $***p < .00001$.
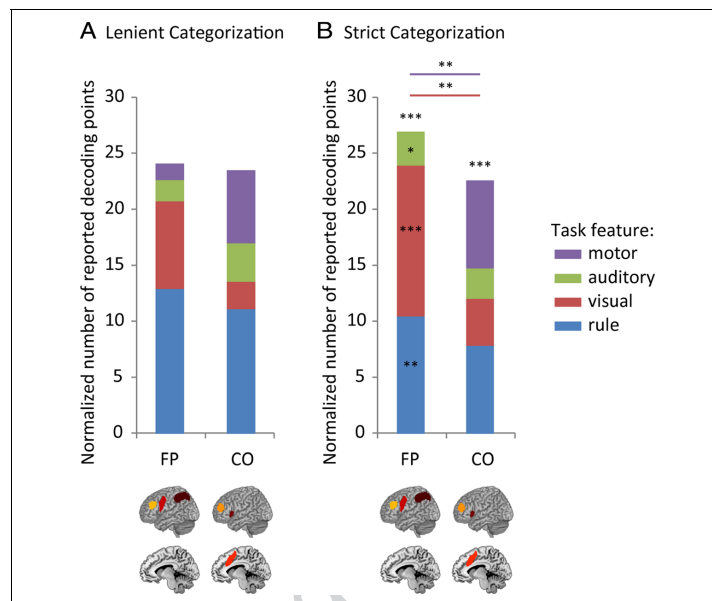

To aid the reader in visualizing the data, we generated a whole-brain decoding map from the lenient categorization. For this, the peak decoding coordinates reported in each analysis were projected onto a single template brain, smoothed (15 FWHM Guassian kernel) and thresholded (≥3 times the height of a single peak). The resulting map indicates regions most commonly identified as making task-relevant distinctions in the literature. As can be seen in Figure 3, regions of maximum reported decoding corresponded well with our a priori networks. Information coding was frequently reported in the MD network (bilateral ACC/pre-SMA, right AI/FO, left IFJ, left and right aIFS, right pIFS, left PM, and left and right IPS), Visual network (BA 18/19) extending to inferior temporal cortex, Auditory network (left and right superior temporal gyrus), and the Motor network (left and right precentral and postcentral gyri). Additional small regions of frequent decoding were found in the dorsal part of the right middle frontal gyrus (BA 9/8), the ventral part of the right inferior frontal gyrus (BA 45/47), a ventral part of the left precuneus (BA 30), and the right temporal parietal junction (BA 21). We similarly generated whole-brain decoding

Figure 3. Brain regions where significant decoding of visual, auditory, rule, and motor information was most frequently reported in the literature. Areas of maximal decoding are shown rendered on left and right hemisphere and on the medial surface ($x = -4$). To create this visualization, all the decoding peaks were projected onto a single template brain, smoothed, and summed, and the resulting image was thresholded at 3 times the maximum height of a single smoothed peak.

Figure 4. Brain regions where significant decoding of (A) visual, (B) auditory, (C) rule, and (D) motor information was most frequently reported in the literature. To create this visualization, the decoding peaks for each task feature (lenient categorization) were projected onto a single template brain, smoothed, and summed, and the resulting image was thresholded at 1.2 times the maximum height of a single smoothed peak. (E) Maps from A to D flattened and overlaid at 50% transparency.
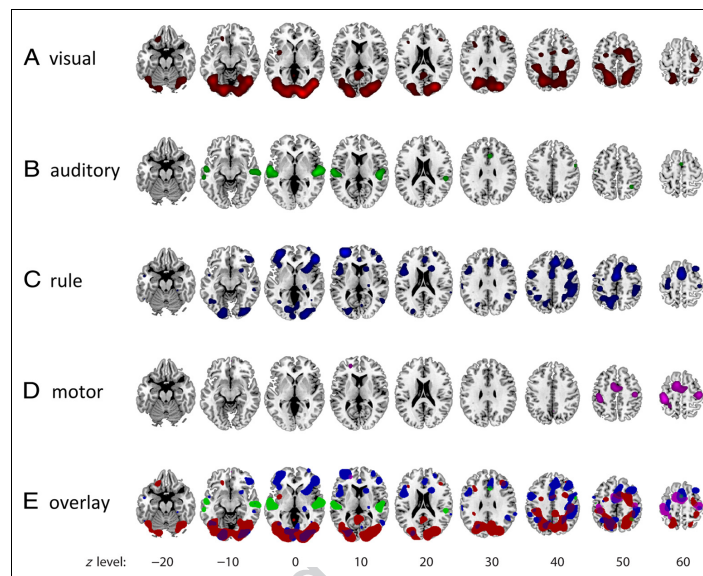
maps for each task feature separately (using a lower threshold of 1.2 * single peak height to account for the smaller number of data points in this visualization). As can be seen in Figure 4, the result was a reassuring picture in which visual information was predominantly found to be encoded in the visual cortex, with some additional contribution from frontal and parietal lobes, auditory information was predominantly reported in the auditory cortex, and motor information was primarily coded in motor cortices. Rule was the most diffusely coded task feature, represented in frontal, parietal, and occipitotemporal cortices. These maps did not change markedly if the strict categorization data were used instead.

DISCUSSION

The human brain is a massively parallel complex system. In the past three decades, PET and fMRI technologies have allowed us to probe the function of different parts of this system by assessing what regions are active in different tasks. In the last decade, MVPA has taken this endeavor to a new level, enabling us to study what aspects of stimuli, rules, and responses are discriminated in the local pattern of multivoxel activation in different brain regions. In this paper, we summarized the current state of the literature, drawing on 100 independent analyses, reported in 57 published papers, to describe the distribution of visual, auditory, rule, and motor information pro-

cessing in the brain. The result is a balanced view of brain modularity and flexibility. Sensory and motor networks predominantly coded information from their own domain, whereas the frontoparietal MD network coded all the different task features we examined. The contribution of the DMN and voxels elsewhere was minor.

The observation that the MD network codes information from multiple domains fits well with an adaptive view of this system. Consistent with the observation of similar frontoparietal activity across many tasks (e.g., Yeo et al., 2014; Fedorenko et al., 2013; Duncan & Owen, 2000; Dosenbach et al., 2006), the proposal is that these regions adapt their function as needed for the task in hand (Duncan, 2001, 2010). To support goal-directed behavior in different circumstances, they are proposed to be capable of encoding a range of different types of information, including the details of auditory and visual stimuli that are relevant to the current cognitive operation (Duncan, 2010). Support comes from single unit recordings, in which the firing rates of prefrontal and parietal cells have been shown to code task rules (e.g., Sigala, Kusunoki, Nimmo-Smith, Gaffan, & Duncan, 2008; Wallis, Anderson, & Miller, 2001; White & Wise, 1999), behavioral responses (e.g., Asaad, Rainer, & Miller, 1998; Niki & Watanabe, 1976), auditory stimuli (e.g., Romanski, 2007; Azumo & Suzuki, 1984), and visual stimuli (e.g., Freedman & Assad, 2006; Freedman, Riesenhuber, Poggio, & Miller, 2001; Hoshi, Shima, & Tanji, 1998; Rao, Rainer, & Miller, 1997). Further support for an adaptive view of this system comes

from the observation that the responses of single units in prefrontal and parietal regions adjust to code different information over the course of single trials (Kadohisa et al., 2013; Stokes et al., 2013; Rao et al., 1997) and make different stimulus distinctions in different task contexts (Freedman & Assad, 2006; Freedman et al., 2001). Accordingly, in human functional imaging, the strength of multivoxel codes in the MD system has been found to adjust according to task requirements, with perceptual discrimination increasing under conditions of high perceptual demand (Woolgar, Williams, & Rich, 2015; Woolgar, Hampshire, Thompson, & Duncan, 2011), rule discrimination increasing when rules are more complex (Woolgar, Afshar, Williams, & Rich, 2015), and a greater representation of visual objects that are at the focus of attention (Woolgar, Williams, et al., 2015). These regions are also thought to make qualitatively different distinctions between visual stimuli in different task contexts (Harel, Kravitz, & Baker, 2014). The data presented here emphasize the extent of flexibility in these regions, suggesting they are capable of representing task relevant information from visual, auditory, rule, and motor domains.

Although each of the individual MD regions are known to respond to a wide range of cognitive demands (e.g., Fedorenko et al., 2013), it nonetheless seems likely that the different regions will support somewhat different cognitive functions. Several organizational schemes have been proposed for the pFC, including a rostrocaudal axis along which different regions support progressively more abstract control processes (Badre & D'Esposito, 2007; Koechlin & Summerfield, 2007), ventral and dorsal segregation based on the modality of the information being processed (Goldman-Rakic, 1998), different types of attentional orienting (Corbetta & Shulman, 2002) or what the information will be used for (O'Reilly, 2010), and a medial/lateral segregation based on conflict monitoring and task set implementation (Botvinick, 2008), although some of these accounts have been challenged experimentally (Crittenden & Duncan, 2014; Grinband et al., 2011). One prominent subdivision of the MD system draws a distinction between an FP subnetwork comprising the MD regions on the dorsal lateral prefrontal surface and the IPS and a CO subnetwork comprising cortex around ACC/pre-SMA, AI/FO, and aIFS. This distinction is born out in analysis of resting state connectivity (Power & Petersen, 2013; Power et al., 2011), and the two subnetworks have been ascribed various different functions, for example, supporting transient versus sustained control processes (Power & Petersen, 2013; Dosenbach et al., 2007), "executive" versus "salience" systems (Seeley et al., 2007), and transformation versus maintenance of information (Hampshire, Highfield, Parkin, & Owen, 2012). In our data, there was no evidence for differences in the frequency with which information coding was reported in the seven (bilateral) MD regions separately. Subdividing the MD system into FP and CO subnetworks also resulted in comparable levels of coding overall in

each subnetwork. However, there was a significant difference in the profile of task features coded by these two subnetworks, with more coding of visual information in FP than in CO and more coding of motor information in CO than in FP. In CO, motor points were reported both in ACC/pre-SMA region known to support motor function and also in the aIFS. Clarification of the basis of the subnetwork coding difference and how we should interpret it will require further work.

Visual, auditory, and motor regions principally coded information from their own domain. However, the visual and motor networks also showed some domain generality, with coding of other task features. Particularly salient was the overlap between the maps for visual and rule information in the visual cortex (Figure 4E). In our review, it was difficult to completely rule out confounds between domains. For example, task rules were usually cued visually, meaning that the visual properties of the cues, as much as representation of the abstract rules per se, could drive discrimination between rules. However, there are some cases of rule coding in the visual cortex where this explanation is not sufficient. For example, we previously reported that discrimination between two stimulus–response mapping rules in the visual cortex generalizes over the two visual stimuli used to cue each rule (Woolgar, Thompson, et al., 2011). Similarly, Zhang et al. (2013) found that rule discrimination in the calcarine sulcus generalized over externally cued and internally chosen rules, and Soon, Namburi, and Chee (2013) reported rule discrimination in the visual cortex when rules were cued with an auditory cue. In some cases, rule discrimination in the visual cortex may reflect different preparatory signals, for example, if the two rules direct attention to different visual features (e.g., Zhang et al., 2013) or object categories (e.g., Soon et al., 2013), but this is not always the case: the two rules of Woolgar, Thompson, et al. (2011) required attention to the same features of identical visual stimuli. Intriguingly, both rule and response coding has previously been reported in the firing rates of single units in V4 of the macaque visual cortex (Mirabella et al., 2007).

In the motor cortex, the majority of reported coding was for discrimination between motor movements, but this region also showed appreciable coding of visual stimuli. Interestingly, population level responses in the primary motor cortex of the macaque have been reported to encode visual stimuli and stimulus–response mapping rules (e.g., Riehle, Kornblum, & Requin, 1994, 1997; Zhang, Riehle, Requin, & Kornblum, 1997). In the MVPA papers we studied, it was often difficult to say precisely what aspects of a stimulus underpinned a given multivoxel discrimination. For example, visual presentation of a familiar object might evoke representation of its associated properties in other sensory domains (e.g., implied somatosensory properties when watching manual exploration of objects; Kaplan & Meyer, 2012). We excluded any papers in which there were obvious associations between

our task features, and in our stricter analysis, we also excluded any studies in which higher-level features such as semantic category differed between decoded items or cases where items might evoke representations of associated motor actions. The remaining points of visual discrimination in the motor cortex were for discrimination between Gabor patches differing in color and spatial frequency (Pollmann, Zinke, Baumgartner, Geringswald, & Hanke, 2014), the spatial location of a target (Kalberlah, Chen, Heinzle, & Haynes, 2011), radial versus concentric glass patterns (Mayhew & Kourtzi, 2013; Mayhew, Li, Storrar, Tsvetanov, & Kourtzi, 2010), and between two abstract shapes cuing the same rule (Reverberi, Gorgen, & Haynes, 2012a). In one study, radial and concentric patterns had been associated with differential button presses during training, whereas during scanning, participants performed an unrelated task (Mayhew et al., 2010), and in all other cases, any button press responses given by participants were orthogonal (Mayhew & Kourtzi, 2013) or unrelated (Pollmann et al., 2014; Reverberi et al., 2012a; Kalberlah et al., 2011; Mayhew et al., 2010) to the visual discrimination.

A few of the studies we included reported multivoxel coding in the DMN. In some cases, the reported discrimination in the DMN reflected participant intentions, such as coding of internally selected task choices (Momennejad & Haynes, 2012; Vickery et al., 2011; Haynes et al., 2007) or externally instructed task rules (Soon et al., 2013; Nee & Brown, 2012) during preparatory periods, the time delay after which participants will self-initiate a switch (Momennejad & Haynes, 2012), and the button which the participant intends to press (Soon, Brass, Heinze, & Haynes, 2008). In other cases, it reflected aspects of active tasks including current rule (Zhang et al., 2013; Reverberi et al., 2012a; Reverberi, Gorgen, & Haynes, 2012b) and stimulus (e.g., orientation of a Gabor (Kahnt, Grueschow, Speck, & Haynes, 2011), concentric versus radial glass patterns (Mayhew & Kourtzi, 2013), and harmonicity of a sound (Giordano, McAdams, Zatorre, Kriegeskorte, & Belin, 2013). Interestingly, this network has recently been reported to show activation during task switching as well as multivoxel discrimination between the tasks being switched to (Crittenden, Mitchell, & Duncan, 2015). Additionally, we recently reported multivoxel discrimination between stimulus–response mapping rules in the precuneus, overlapping a major node of the DMN, during an active stimulus–response task (Woolgar, Afshar, et al., 2015). Those data suggest a role for DMN that is qualitatively different from the internally driven activities such as mind wandering and introspection with which this network is more typically associated (e.g., Buckner, Andrews-Hanna, & Schacter, 2008).

There was more coding of motor information in the DMN than in Other, but all five DMN motor coding points came from a single study (Soon et al., 2008). Four of these points corresponded to discriminatory activation in preparation of a left versus right button press at a time point before the participant had indicated their conscious intention to press a button, and the remaining point was for response preparation when participants were cued to make a choice. There were no motor coding points in the DMN during button press execution.

An important challenge for MVPA is to account for variables that differ between conditions on an individual participant basis, such as differences in RT (Woolgar et al., 2014; Todd et al., 2013). Because MVPA is usually carried out at the level of individual participants, with a directionless summary statistic (e.g., classification accuracy) taken to the second level, any effect of difficulty, effort, attention, time on task, trial order (etc.) will not average out at the group level. This may be a particular concern in regions such as the MD and DMN networks, which are known to show different overall activity levels according to task demand. It is difficult to estimate the extent to which these factors have contributed to the data analyzed here. Some of the included studies matched their conditions for difficulty (e.g., Zhang et al., 2013), explicitly accounted for differences in RT in their analysis (e.g., Woolgar, Thompson, et al., 2011), or used designs in which difficulty was unlikely to artifactually drive coding (e.g., passive viewing, Kaplan & Meyer, 2012), but many did not. Other studies sought to account for univariate effects of difficulty that could drive multivariate results, for example, by normalizing the multivoxel patterns to remove overall activation differences between conditions at the level of individual participants (e.g., Gilbert, 2011). However, because the effect of difficulty would not necessarily manifest as an overall activation difference, this could still fail to remove the effect of difficulty on decoding. In our stricter analysis, we excluded analyses in which there was an obvious difference in difficulty between discriminated conditions, but most studies did not report whether there were any differences between conditions on an individual participant basis. Note, though, that we have previously examined the extent to which trial by trial differences in RT contribute to decoding in empirical data and found the contribution to be minor (Crittenden et al., 2015; Erez & Duncan, 2015; Woolgar et al., 2014).

We summarized 100 independent analyses, reported in 57 published papers, that isolated the multivoxel representation of visual and auditory sensory input, task rules, or motor output. The results confirm the power of the MVPA method, with predominant coding of visual, auditory, and response distinctions in the expected sensory and motor regions. Outside sensory and motor areas, the results are also structured, with a specific network of frontal and parietal regions involved in coding several different types of information. Consistent with the observation of similar frontoparietal activity across many tasks and the suggestion that neurons of these regions adapt their function as needed for current behavior (Duncan 2001), frontoparietal cortex codes information from across sensory and task domains.

Reprint requests should be sent to Alexandra Woolgar, ARC Centre of Excellence in Cognition and Its Disorders, Department of Cognitive Science, Macquarie University, Sydney, New South Wales 2109, Australia, or via e-mail: alexandra.woolgar@mq.edu.au.

Note

1. In two cases, coordinates were not reported, but a list of peaks was sent by email to AW.

REFERENCES

Alink, A., Euler, F., Kriegeskorte, N., Singer, W., & Kohler, A. (2012). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Human Brain Mapping, 33,* 969–978.

Asaad, W. F., Rainer, G., & Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron, 21,* 1399–1407.

Azumo, M., & Suzuki, H. (1984). Properties and distribution of auditory neurons in the dorsolateral prefrontal cortex of the alert monkey. *Brain Research, 298,* 343–346.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19,* 2082–2099.

Billington, J., Furlan, M., & Wann, J. (2013). *Cortical responses to congruent and incongruent stereo cues for objects on a collision path with the observer.* Elsevier.

Bode, S., Bogler, C., & Haynes, J. D. (2013). Similar neural mechanisms for perceptual guesses and free decisions. *Neuroimage, 65,* 456–465.

Bode, S., Bogler, C., Soon, C. S., & Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *Neuroimage, 59,* 1924–1931.

Bode, S., & Haynes, J. D. (2009). Decoding sequential stages of task preparation in the human brain. *Neuroimage, 45,* 606–613.

Boets, B., Op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., et al. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science, 342,* 1251–1254.

Bogler, C., Bode, S., & Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Current Biology, 21,* 1667–1671.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences, 12,* 201–208.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124,* 1–38.

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology, 21,* 1817–1821.

Carlin, J. D., Rowe, J. B., Kriegeskorte, N., Thompson, R., & Calder, A. J. (2012). Direction-sensitive codes for observed head turns in human superior temporal sulcus. *Cerebral Cortex, 22,* 735–744.

Carp, J., Park, J., Hebrank, A., Park, D. C., & Polk, T. A. (2011). Age-related neural differentiation in the motor system. *PLoS One, 6,* 1–6.

Christophel, T. B., Hebart, M. N., & Haynes, J. D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *Journal of Neuroscience, 32,* 12983–12989.

Clarke, A., & Tyler, L. K. (2014). Object-specific semantic coding in human perirhinal cortex. *Journal of Neuroscience, 34,* 4766–4775.

Clithero, J. A., Smith, D. V., Carter, R. M., & Huettel, S. A. (2011). Within- and cross-participant classifiers reveal different neural coding of information. *Neuroimage, 56,* 699–708.

Colas, J. T., & Hsieh, P. J. (2014). Pre-existing brain states predict aesthetic judgments. *Human Brain Mapping, 35,* 2924–2934.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Crittenden, B. M., & Duncan, J. (2014). Task difficulty manipulation reveals multiple demand activity but no frontal lobe hierarchy. *Cerebral Cortex, 24,* 532–540.

Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2015). Recruitment of the default mode network during a demanding act of executive control. *Elife, 4.* doi:10.7554/eLife.06481.

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition, 79,* 1–37.

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 11073–11078.

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron, 50,* 799–812.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience, 2,* 820–829.

Duncan, J. (2006). EPS Mid-Career Award 2004: Brain mechanisms of attention. *Quarterly Journal of Experimental Psychology (Hove), 59,* 2–27.

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences, 14,* 172–179.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475–483.

Ekman, M., Derrfuss, J., Tittgemeyer, M., & Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences, U.S.A., 109,* 16714–16719.

Engels, B. (2014). *XNomial: Exact goodness-of-fit test for multinomial data with fixed probabilities.* R package version 1.0.1. CRAN.R-project.org/package=XNomial.

Erez, Y., & Duncan, J. (2015). Discrimination of visual categories based on behavioral relevance in widespread regions of frontoparietal cortex. *Journal of Neuroscience, 35.* doi:10.1523/JNEUROSCI.1134-15.2015.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal

cortex. *Proceedings of the National Academy of Sciences, U.S.A., 110,* 16616–16621.

FitzGerald, T. H., Schwartenbeck, P., & Dolan, R. J. (2014). Reward-related activity in ventral striatum is action contingent and modulated by behavioral relevance. *Journal of Neuroscience, 34,* 1271–1279.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 9673–9678.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature, 443,* 85–88.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291,* 312–316.

Freeman, G. H., & Halton, J. H. (1951). Note on exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika, 38,* 141–149.

Gilbert, S. J. (2011). Decoding the content of delayed intentions. *Journal of Neuroscience, 31,* 2888–2894.

Gilbert, S. J., Swencionis, J. K., & Amodio, D. M. (2012). Evaluative vs. trait representation in intergroup social judgments: Distinct roles of anterior temporal lobe and prefrontal cortex. *Neuropsychologia, 50,* 3600–3611.

Giordano, B. L., McAdams, S., Zatorre, R. J., Kriegeskorte, N., & Belin, P. (2013). Abstract encoding of auditory objects in cortical activity patterns. *Cerebral Cortex, 23,* 2025–2037.

Goldman-Rakic, P. S. (1998). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive.

Greenberg, A. S., Esterman, M., Wilson, D., Serences, J. T., & Yantis, S. (2010). Control of spatial and feature-based attention in frontoparietal cortex. *Journal of Neuroscience, 30,* 14330–14339.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage, 57,* 303–311.

Guo, F., Preston, T. J., Das, K., Giesbrecht, B., & Eckstein, M. P. (2012). Feature-independent neural coding of target detection during search of natural scenes. *Journal of Neuroscience, 32,* 9499–9510.

Hampshire, A., Highfield, R. R., Parkin, B. L., & Owen, A. M. (2012). Fractionating human intelligence. *Neuron, 76,* 1225–1237.

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences, U.S.A., 111,* E962–E971.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293,* 2425–2430.

Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7,* 523–534.

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology, 17,* 323–328.

Hebart, M. N., Donner, T. H., & Haynes, J.-D. (2012). Human visual and parietal cortex encode visual choices independent of motor plans. *Neuroimage, 63,* 1393–1403.

Helfinstein, S. M., Schonberg, T., Congdon, E., Karlsgodt, K. H., Mumford, J. A., Sabb, F. W., et al. (2014). Predicting risky choices from brain activity patterns. *Proceedings of the National Academy of Sciences, U.S.A., 111,* 2470–2475.

Herrmann, B., Obleser, J., Kalberlah, C., Haynes, J. D., & Friederici, A. D. (2012). Dissociable neural imprints of perception and grammar in auditory functional imaging. *Human Brain Mapping, 33,* 584–595.

Hoshi, E., Shima, K., & Tanji, J. (1998). Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *Journal of Neurophysiology, 80,* 3392–3397.

Huang, Y. F., Soon, C. S., Mullette-Gillman, O. A., & Hsieh, P. J. (2014). Pre-existing brain states predict risky choices. *Neuroimage, 101,* 466–472.

Jiang, F., Stecker, G. C., & Fine, I. (2014). Auditory motion processing after early blindness. *Journal of Vision, 14,* 4.

Jiang, J., & Egner, T. (2014). Using neural pattern classifiers to quantify the modularity of conflict-control mechanisms in the human brain. *Cerebral Cortex, 24,* 1793–1805.

Kadohisa, M., Petrov, P., Stokes, M., Sigala, N., Buckley, M., Gaffan, D., et al. (2013). Dynamic construction of a coherent attentional state in a prefrontal cell population. *Neuron, 80,* 235–246.

Kahnt, T., Grueschow, M., Speck, O., & Haynes, J. D. (2011). Perceptual learning and decision-making in human medial frontal cortex. *Neuron, 70,* 549–559.

Kalberlah, C., Chen, Y., Heinzle, J., & Haynes, J. D. (2011). Beyond topographic representation: Decoding visuospatial attention from local activity patterns in the human frontal cortex. *Wiley Periodicals, 21,* 201–210.

Kaplan, J. T., & Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *Neuroimage, 60,* 204–212.

Klein, M. E., & Zatorre, R. J. (2014). Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. *Cerebral Cortex,* bhu003.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences, 11,* 229–235.

Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. D. (2013). Predicting vocal emotion expressions from the human brain. *Human Brain Mapping, 34,* 1971–1981.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 3863–3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.

Lee, Y., Janata, P., Frost, C., Hanke, M., & Granger, R. (2011). Investigation of melodic contour processing in the brain using multivariate pattern-based fMRI. *Neuroimage, 57,* 292–300.

Lee, Y. S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience, 32,* 3942–3948.

Li, S., & Yang, F. (2012). Task-dependent uncertainty modulation of perceptual decisions in the human brain. *European Journal of Neuroscience, 36,* 3732–3739.

Linden, D. E., Oosterhof, N. N., Klein, C., & Downing, P. E. (2012). Mapping brain activation and information during category-specific visual working memory. *Journal of Neurophysiology, 107,* 628–639.

Linke, A. C., Vicente-Grabovetsky, A., & Cusack, R. (2011). Stimulus-specific suppression preserves information in auditory short-term memory. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 12961–12966.

319

Mayhew, S. D., & Kourtzi, Z. (2013). Dissociable circuits for visual shape learning in the young and aging human brain. *Frontiers in Human Neuroscience, 7,* 75.

Mayhew, S. D., Li, S., Storrar, J. K., Tsvetanov, K. A., & Kourtzi, Z. (2010). Learning shapes the representation of visual categories in the aging human brain. *Journal of Cognitive Neuroscience, 22,* 2899–2912.

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience, 16,* 479–485.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., et al. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology, 3,* 76.

Mirabella, G., Bertini, G., Samengo, I., Kilavik, B. E., Frilli, D., Della Libera, C., et al. (2007). Neurons in area V4 of the macaque translate attended visual features into behaviorally relevant categories. *Neuron, 54,* 303–318.

Momennejad, I., & Haynes, J.-D. (2012). Human anterior prefrontal cortex encodes the "what" and "when" of future intentions. *Neuroimage, 61,* 139–148.

Murawski, C., Harris, P. G., Bode, S., Dominguez, D. J., & Egan, G. F. (2012). Led into temptation? Rewarding brand logos bias the neural encoding of incidental economic decisions. *PLoS One, 7,* e34155.

Naghavi, H. R., & Nyberg, L. (2005). Common fronto-parietal activity in attention, memory, and consciousness: Shared demands on integration? *Consciousness and Cognition, 14,* 390–425.

Nee, D. E., & Brown, J. W. (2012). Rostral-caudal gradients of abstraction revealed by multi-variate pattern analysis of working memory. *Neuroimage, 63,* 1285–1294.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience, 12,* 241–268.

Niki, H., & Watanabe, M. (1976). Prefrontal unit activity and delayed response: Relation to cue location versus direction of response. *Brain Research, 105,* 79–88.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in Neurosciences, 33,* 355–361.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). *n*-Back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25,* 46–59.

Peelen, A., & Vuilleumier (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30,* 10127–10134.

Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience, 32,* 15728–15736.

Pollmann, S., Zinke, W., Baumgartner, F., Geringswald, F., & Hanke, M. (2014). The right temporo-parietal junction contributes to visual feature binding. *Neuroimage, 101,* 289–297.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron, 72,* 665–678.

Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology, 23,* 223–228.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 676–682.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science, 276,* 821–824.

Reverberi, C., Gorgen, K., & Haynes, J. D. (2012a). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex, 22,* 1237–1246.

Reverberi, C., Gorgen, K., & Haynes, J. D. (2012b). Distributed representations of rule identity and rule order in human frontal cortex and striatum. *Journal of Neuroscience, 32,* 17420–17430.

Riehle, A., Kornblum, S., & Requin, J. (1994). Neuronal coding of stimulus–response association rules in the motor cortex. *NeuroReport, 5,* 2462–2464.

Riehle, A., Kornblum, S., & Requin, J. (1997). Neuronal correlates of sensorimotor association in stimulus–response compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1708–1726.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebral Cortex, 17(Suppl. 1),* i61–i69.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience, 27,* 2349–2356.

Sigala, N., Kusunoki, M., Nimmo-Smith, I., Gaffan, D., & Duncan, J. (2008). Hierarchical coding for sequential task events in the monkey prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 11969–11974.

Simanova, I., Hagoort, P., Oostenveld, R., & van Gerven, M. A. (2014). Modality-independent decoding of semantic information from the human brain. *Cerebral Cortex, 24,* 426–434.

Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11,* 543–545.

Soon, C. S., Namburi, P., & Chee, M. W. L. (2013). Preparatory patterns of neural activity predict visual category search speed. *Neuroimage, 66,* 215–222.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron, 78,* 364–375.

Team, R. C. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Todd, M. T., Nystrom, L. E., & Cohen, J. D. (2013). Confounds in multivariate pattern analysis: Theory and rule representation case study. *Neuroimage, 77,* 157–165.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.

Vickery, T. J., Chun, M. M., & Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron, 72,* 166–177.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411,* 953–956.

Weygandt, M., Schaefer, A., Schienle, A., & Haynes, J. D. (2012). Diagnosing different binge-eating disorders based on reward-related brain activation patterns. *Human Brain Mapping, 33,* 2135–2146.

White, I. M., & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research, 126,* 315–335.

Wisniewski, D., Reverberi, C., Tusche, A., & Haynes, J.-D. (2014). The neural representation of voluntary task-set selection in dynamic environments. *Cerebral Cortex,* bhu155.

Woolgar, A., Afshar, S., Williams, M. A., & Rich, A. N. (2015). Flexible coding of task rules in frontoparietal cortex: An adaptive system for flexible cognitive control. *Journal of Cognitive Neuroscience, 27,* 1895–1911.

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. *Journal of Neuroscience, 31,* 14592–14599.

Woolgar, A., Thompson, R., Bor, D., & Duncan, J. (2011). Multivoxel coding of stimuli, rules, and responses in human frontoparietal cortex. *Neuroimage, 56,* 744–752.

Woolgar, A., Williams, M. A., & Rich, A. N. (2015). Attention enhances multivoxel representation of novel objects in frontal, parietal and visual cortices. *Neuroimage, 109,* 429–437.

Yeo, B. T., Krienen, F. M., Eickhoff, S. B., Yaakub, S. N., Fox, P. T., Buckner, R. L., et al. (2014). Functional specialization and flexibility in human association cortex. *Cerebral Cortex.* doi:10.1093/cercor/bhu217.

Zhang, J., Kriegeskorte, N., Carlin, J. D., & Rowe, J. B. (2013). Choosing the rules: Distinct and overlapping frontoparietal representations of task rules for perceptual decisions. *Journal of Neuroscience, 33,* 11852–11862.

Zhang, J., Riehle, A., Requin, J., & Kornblum, S. (1997). Dynamics of single neuron activity in monkey primary motor cortex related to sensorimotor transformation. *Journal of Neuroscience, 17,* 2227–2246.

Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *Journal of Neuroscience, 33,* 4339–4348.

Ethics Approvals

Office of the Deputy Vice-Chancellor
(Research)

Research Office
Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia
T: +61 (2) 9850 4459
http://www.research.mq.edu.au/
ABN 90 952 801 237

MACQUARIE
University
SYDNEY·AUSTRALIA

31 May 2016

Miss Jade Jackson
Department of Cognitive Science
Faculty of Human Sciences
Macquarie University
NSW 2077

Dear Miss Jackson

Reference No: HE27NOV2009-R00180

Title: *How does the brain adapt to visual input deficits*

This letter is to confirm that the ethics application cited above met the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007 – Updated May 2015) (the *National Statement*).

The application received approval from the Macquarie University Human Research Ethics Committee on 11 March 2010.

The above project was conducted by Miss Jade Jackson, PhD candidate, under the supervision of Dr. Alexandra Woolgar (Primary Supervisor), Associate Professor Anina Rich (Co-Supervisor) and Professor Mark Williams (Associate Supervisor).

Please do not hesitate to contact me if you have any questions.

Yours sincerely

Dr Karolyn White
Director, Research Ethics & Integrity
Chair, Macquarie University Human Research Ethics Committee

4 February 2014

Associate Professor Mark Williams
Department of Cognitive Science
Faculty of Human Sciences

Dear Associate Professor Williams

RE: *Investigating the nature of visual, auditory and tactile stimuli*

Thank you for submitting the above application for ethical and scientific review. Your application was first considered by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences)) at its meeting on 24 October 2013 at which it was requested that the application be resubmitted to be reviewed by the HREC (Medical Sciences) out of session.

The HREC (Medical Sciences) requested that you provide further clarifications delegating authority to approve the application to the Ethics Secretariat upon receipt of a satisfactory response.

The requested information was received with correspondence on 4 February 2014 and I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted at:

- Macquarie University

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007) (the *National Statement*).

This letter constitutes ethical and scientific approval only.

Details of this approval are as follows:

Reference No: 5201300541

Approval Date: 4 February 2014

The following documentation has been reviewed and approved by the HREC (Medical Sciences):

Documents reviewed	Version	Date
MQ Human Research Ethics Committee Application Form	2.2	May 2013
Correspondence from Kimberley Weldon responding to the HREC's feedback.		Received 04/02/2014
MQ Participant Information and Consent Form– '*fMRI using visual stimuli*'	2	31/10/2013
MQ Participant Information and Consent Form – '*fMRI using auditory stimuli*'	2	31/10/2013

Office of the Deputy Vice-Chancellor (Research)

Research Office
C5C Research HUB East, Level 3, Room 324
MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

Phone   **+61 (0)2 9850 4194**
Fax   +61 (0)2 9850 4465
Email   ethics.secretariat@mq.edu.au

8 July 2014

Dr Paul Sowman
Department of Cognitive Science
Faculty of Human Sciences
MACQUARIE UNIVERSITY NSW 2109

Dear Dr Sowman

RE: *Stimulating and Recording the Brain in Studies of Cognitive Control*

Thank you for submitting the above application for ethical and scientific review. Your application was considered by the Macquarie University Human Research Ethics Committee (HREC (Medical Sciences)) at its meeting on 29 May 2014 at which further information was requested to be reviewed by the Ethics Secretariat.

The requested information was received with correspondence on 25 June and 7 July 2014.

I am pleased to advise that ethical and scientific approval has been granted for this project to be conducted at:

- Macquarie University

This research meets the requirements set out in the *National Statement on Ethical Conduct in Human Research* (2007 – Updated March 2014) (the *National Statement*).

Details of this approval are as follows:

Reference No: 5201400585

Approval Date: 8 July 2014

The following documentation has been reviewed and approved by the HREC (Medical Sciences):

Documents reviewed	Version no.	Date
Macquarie University Ethics Application Form	2.3	July 2013
Short Protocol		
Correspondence from Dr Paul Sowman responding to the issues raised by the HREC (Medical Sciences)		Received 25/6/2014 & 7/07/2014
Flyer entitled *Stimulating and recording the brain in studies of cognitive control*	1	
Newspaper Advertisement	1	
Study advertisement in SONA-system	1	

1

University of Reading

Academic and Governance Services

Whiteknights House
Whiteknights, PO Box 217
Reading RG6 6AH

phone +44 (0)118 378 7119
fax +44 (0)118 378 8979
email m.j.proven@reading.ac.uk

Dr Eva Feredoes
School of Psychology and Clinical
Language Sciences
University of Reading
RG6 6AL

6 November 2015

Dear Eva

UREC 15/45: Filtering irrelevant information: A concurrent fMRI-TMS study investigating the nature of top down signals from the pre-frontal cortex. *Provisional opinion*

Thank you for the response (email dated 27 October 2015, including attachments, refers) addressing the issues raised by the UREC Sub-committee at its September 2015 meeting. On the basis of the revised documentation, I can confirm that the Chair is pleased to confirm a favourable ethical opinion.

Please note that the Committee will monitor the progress of projects to which it has given favourable ethical opinion approximately one year after such agreement, and then on a regular basis until its completion.

Please also find attached Safety Note 59: Incident Reporting in Human Interventional Studies at the University of Reading, to be followed should there be an incident arising from the conduct of this research.

The University Board for Research and Innovation has also asked that recipients of favourable ethical opinions from UREC be reminded of the provisions of the University Code of Good Practice in Research. A copy is attached and further information may be obtained here:

http://www.reading.ac.uk/internal/res/QualityAssuranceInResearch/reas-RSqar.aspx.

Yours sincerely

Dr M J Proven
Coordinator for Quality Assurance in Research (UREC Secretary)
cc: Dr John Wright (Chair); Professor Laurie Butler (Head of School)

This letter and all accompanying documents are confidential and intended solely for the use of the addressee

School of Psychology and Clinical Language Sciences
Research Ethics Committee
Application Form
Academic Year 2011/2012

SECTION 1: APPLICATION DETAILS

1.1

Project Title: Filtering irrelevant information: A concurrent fMRI-TMS study investigating the nature of top down signals from prefrontal cortex

Date of Submission: Proposed start date: 24.09.15   Proposed End Date: 24.09.17

1.2

Principal Investigator: Eva Feredoes
Office room number: 1S15   Internal telephone: x5011
Email address: e.a.feredoes@reading.ac.uk   Alternative contact telephone: 07968737685
Other applicants
Name: Jade Jackson Student   Institution/Department: Macquarie University, Dept. of Cognitive Science
Email: jade.jackson@mq.edu.au

....

1.3

Project Submission Declaration

I confirm that to the best of my knowledge I have made known all information relevant to the Research Ethics Committee and I undertake to inform the Committee of any such information which subsequently becomes available whether before or after the research has begun.

I understand that it is a legal requirement that both staff and students undergo Criminal Records Checks when in a position of trust (i.e. when working with children or vulnerable adults).

I confirm that if this project is an interventional study, a list of names and contact details of the subjects in this project will be compiled and that this, together with a copy of the Consent Form, will be retained within the School for a minimum of five years after the date that the project is completed.

Signed   *Eva Feredoes.*   (Principal Investigator)   Date: 20/08/15
                Jade Jackson   (Student)   Date: 20/08/15

1.4

University Research Ethics Committee Applications
Projects expected to require review by the University Research Ethics Committee (research involving NHS patients, Social Services clients, research involving potential for distress to participants) must be reviewed by the Chair of the School Ethics Committee or the Head of School before submission.

Signed…………………………… (Chair of School Committee) Date:…27-Aug-2015…
Signed…………………………… (Head of School)   Date:……………………