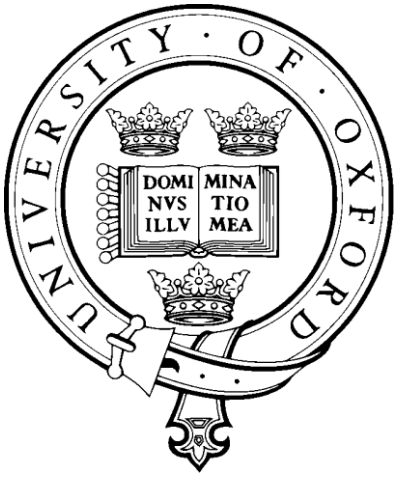
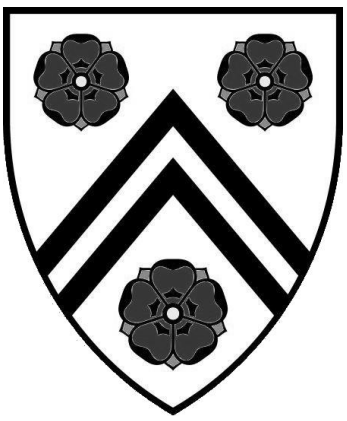
Selection Mechanisms for Working Memory

George J. Wallis

New College

DPhil Thesis, Trinity Term 2014





Selection Mechanisms for Working Memory

George J. Wallis

New College, Oxford

Trinity Term, 2014

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

ABSTRACT

The experiments in this thesis investigated the mechanisms controlling input and output gating of working memory. In chapter 3, accuracy and reaction time data from a precision/capacity working memory task with prospective and retrospective cues were analysed. The results suggest that retrocues boost performance by facilitating output gating from working memory. In chapter 4, the role of perceptual cortex in mediating the cue benefits in this task was investigated with magnetoencephalography (MEG). The pattern of alpha (8-12Hz) power in visual cortex was modulated by cue direction following both precues and retrocues, but whilst this modulation was sustained following a precue (until presentation of the memory array) it was transient following a retrocue, suggesting that a memory representation was briefly retrieved or refreshed, but that there was not a sustained biasing of top-down input to visual cortex following retrocues. This argues against the standard model of working memory as sustained attention to internal representations, and in favour of a more dynamic view in which perceptual cortex is recruited transiently, and otherwise freed up for on-going processing. In chapter 5, the role of frontal networks in precueing and retrocueing was investigated. An fMRI meta-analysis identified control networks involved in preparatory and mnemonic selection: whilst the fronto-parietal network is recruited in both cases, the cingulo-opercular network is recruited only by retrocues. This spatial pattern was replicated with a source-space ROI analysis of MEG induced-responses. These data also characterised the time-course of control network activation shedding light on their functional roles. The fronto-parietal network was activated immediately following both precues and retrocues, consistent with a direct role in top-down influence over perceptual cortex. By contrast, the cingulo-opercular network was activated following retrocues only after the perceptual refreshing event was complete, suggesting a downstream role, perhaps in selecting representations to guide action. Chapter 6 investigated the role of reward associations in controlling access to working memory, testing behavioural predictions of two theories implicating the dopamine system and basal ganglia in control of working memory. The results supported a temporal gating account in which encountering reward associated items triggers a brief (<300ms) window in which there is a boost of encoding for WM. Chapter 7 discusses the implications of the current work and suggests some future directions.

Funded by a 4-year Wellcome Trust studentship.

Approximate word count: 65,000 words

Acknowledgements

A very heartfelt thanks to my two supervisors, Kia Nobre and Mark Stokes. Both have been a fantastic source of guidance, support, and perspective. I could not have hoped to find two more scientifically rigorous and inspiring mentors. Equally, I don't think I could have found a more patient, wise and good-humoured pair to guide me through my DPhil. I hope we'll stay in touch for a long time to come.

I'm very grateful to the analysis team at OHBA who helped me with MEG – in particular, Mark Woolrich, Henry Luckhoo, Adam Baker, Hamid Mohseni and Sven Braeutigam who were all extremely generous with their time and expertise.

Nils Kolling gave much advice that contributed to this thesis, and was a constant font of wisdom on brains, but also on everything else under the sun. Franz Neubert gave me some generous and illuminating anatomy lessons without which this thesis would have looked very different, and some sage advice on what the various bits might be doing. Thanks to Nela Cicmil for the diffusion model tutorials, but mainly for the countless hours spent dissecting the life scientific. Thanks to Nick Myers for some extremely valuable discussions and suggestions. Craig Arnold was a huge help when he visited the lab and worked on the reward project with me. Laurence Hunt was an inspiring MSc supervisor, but has been generous with his assistance since then, by email, but also unwittingly, whenever some crucial piece of MEG code turned out to have 'LH' in the by-line.

Thanks also to all of the rest of my clever and friendly lab-mates in EP and OHBA with whom I've been lucky to share some time: Zita, Ian, Jess, Malcolm, Simone, Harriet, Julian, Alexandra, Gustavo, Riam, Helena, Kathryn, Rob, Josh, Theresa, Giles, Diego, Bo and Celine.

Besides running the MSc and DPhil programme, Debbie Clarke has been a very warm and welcome presence, and a great insider guide to the neuro-scientific world in Oxford.

Many thanks to the Wellcome Trust which funded this DPhil, and to New College for funding conference trips, and for looking after me as a student.

I'm grateful to my friends for making the past four and a half years in Oxford such a happy time. Thanks as ever to my loving family for supporting me and putting up with me.

Table of Contents

List of Abbreviations

AMI	accessory memory item
ANOVA	analysis of variance
BA	Brodmann area
BOLD	blood-oxygen-level-dependent
CDA	contralateral delay activity
CNV	contingent negative variation
dACC	dorsal anterior cingulate cortex
dlPFC	dorso-lateral prefrontal cortex
DPSS	discrete prolate spheroidal sequence
ECG	electrocardiogram
ECoG	electrocorticography
EEG	electroencephalography
EOG	electrooculogram
ERD	event-related desynchronization
ERF	event-related field
ERP	event-related potential
ERS	event-related synchronization
FFT	fast fourier transform
fMRI	functional magnetic resonance imaging
FPl	lateral frontal pole
FPm	medial frontal pole
HOG	horizontal occulogram
HPI	head position indicator
HRF	haemodynamic response function
Hz	hertz
ICA	independent component analysis
IFJ	inferior frontal junction
IPS	intra-parietal sulcus
LCMV	linearly-constrained minimum variance
MEG	magneto-encephalography
MEP	motor-evoked potential
MSR	magnetically shielded room
MTL	medial temporal lobe
MVPA	multi-voxel pattern analysis
NaN	not a number
OCD	obsessive compulsive disorder
PCA	principal component analysis
PET	positron emission tomography
pre-SMA	pre-supplemental motor area
ROI	region of interest
RT	reaction time
SNR	signal to noise ratio
SPL	superior parietal lobule
SSS	signal space separation
TFR	time-frequency representation
TMS	trans-cranial magnetic stimulation
TPJ	temporo-parietal junction
VOG	vertical occulogram
VSTM	visual short-term memory
WM	working memory

1 Introduction

Chapter Abstract

This thesis is concerned with the selection mechanisms that act upon working memory. This introduction first gives a broad overview of some different perspectives on selection in the brain, from both sensory and motor neuroscience, emphasising the distinction between input and output gating. The following sections consider first behavioural evidence for different modes of selection for working memory, and then neural correlates of these different selection mechanisms. The scope and structure of this thesis is then outlined in the final section.

1.1 Overview

> "We, at one glance, can perceive three glasses on a table; Funes all the leaves and tendrils of fruit that make up a grape vine. He knew by heart the forms of the southern clouds at dawn on the 30th of April, 1882, and could compare them in his memory with the mottled streaks on a book in Spanish binding he had only seen once and with the outlines of the foam raised by an oar in the Rio Negro the night before the Quebracho uprising.
>
> …
>
> With no effort, he had learned English, French, Portuguese and Latin.
> I suspect, however, that he was not very capable of thought."
>
> Jorges Luis Borges, *Funes the Memorius*

Borges' Ireneo Funes lies paralysed in his mother's backyard, tortured by a vast and pressing expanse of memories and percepts, in which every detail is as salient as every other, and nothing is abstracted, generalized, discriminated, emphasised or rejected. Funes lacks the ability to select his impressions and memories. Selection mechanisms are vital – without the ability to filter perception, or to select a coherent stream of actions, we would be highly impaired. Borges' Funes was fictional[1], but a wide range of neurological and psychiatric disorders have been associated with impaired or disordered selection, in particular schizophrenia (Kapur, 2003). Both basic scientists and clinicians therefore have a strong incentive to understand the different forms of selection that organise perception, thought and action, and to discover their neural underpinnings.

The work presented in this thesis was focussed on dissecting basic mechanisms of selection for working memory, developing prior work on the overlap and dissociations between perceptual and mnemonic selection, and additionally investigating reward

[1] Though Borges' story was likely inspired by Luria's account of his patient the Mnemonist, a real-life individual with a vast and synesthetic perceptual memory, who suffered peculiar (though less catastrophic) problems in planning his everyday life, and discriminating mental and perceptual imagery.

associations as a potential factor influencing memory selection. Section 1.2 of this introduction provides a very general overview situating working memory in the context of different perspectives on selection. Following this broad discussion, sections 1.3 and 1.4 discuss behavioural and neural evidence relevant to the questions addressed in the experimental chapters. Section 1.5 outlines the contribution of the current experiments.

1.2 Perspectives on selection mechanisms

1.2.1 Perceptual attention versus selection for action

In perception and working memory, limited neural processing resources and concomitant 'capacities' are thought to necessitate selection mechanisms: i.e. limited capacity for high level perceptual processing (O'Regan, Rensink, & Clark, 1999), and limited storage capacity for information in working memory (Fukuda, Awh, & Vogel, 2010a; Luck & Vogel, 2013). Selection determines which of many potential sensory representations are processed and retained (Desimone & Duncan, 1995).

Motor neuroscience also invokes selection. Only one action (or at most a small number) can be performed at a given time, because there is only one set of effectors, so actions must be selected from a range of action candidates.

There are differences between these two ideas of selection. Most obviously the direction of selection is reversed. In perception, the selection bottleneck is widest at the most basic level of sensory processing, and narrows as the processing hierarchy is ascended. In motor neuroscience and decision-making, the bottleneck narrows in the direction of physical action, and widens to encompass a range of action candidates (or at a higher hierarchical level, goal/task candidates – which may be a more meaningful unit

of selection). In the case of perceptual attention, the capacity limit necessitating selection is thought to be the number of representations (or volume of information) that we can process in depth. The volume of sensory stimulation arriving from the world is thought to be very large relative to the amount of information that can be processed. In the case of action the limit is physical, in that we are only capable of performing one or perhaps a small set of actions at a time. In this case, it is the set of action candidates that is thought to be very large relative to the number of actions we can perform.

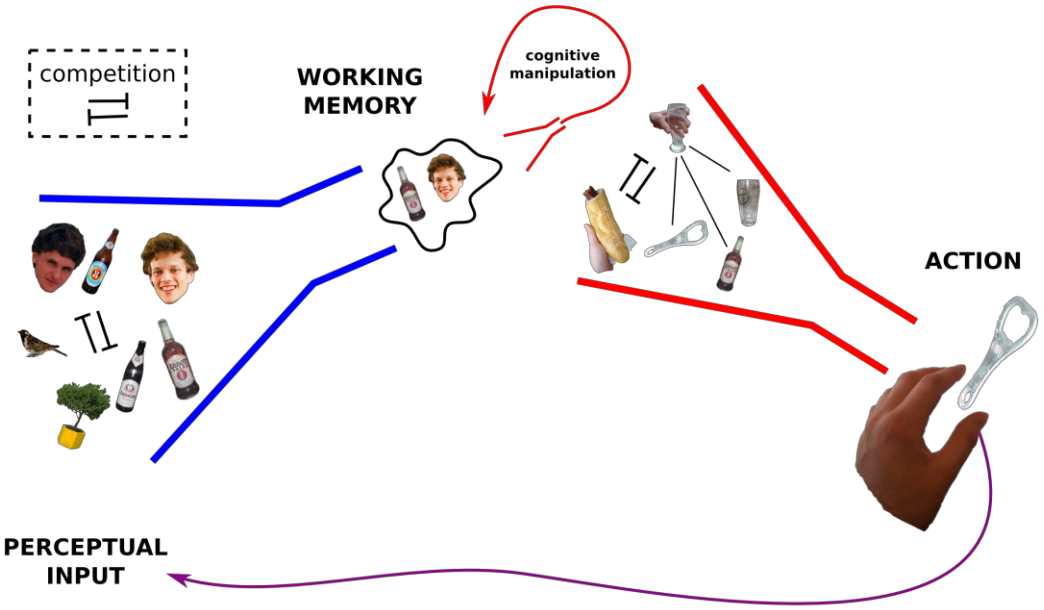


Figure 1-1 **Selection for perception, memory and action.** This figure schematizes different loci of selection in the perception-action cycle of a hypothetical human enjoying an afternoon barbeque. Cognitive manipulation is associated with an internal 'action' bottleneck, in line with the idea that selecting internal representations for manipulation has commonalities with selecting actions.

Working memory lies between perceptual selection and action control (Figure 1-1). The purpose of this broad overview is to situate working memory in relation to ideas of selection stemming from each field. A dominant perspective in cognitive neuroscience has focussed on the number and nature of the representations held in working memory (Bays & Husain, 2008; Luck & Vogel, 1997; 2013; Phillips, 1974; Sperling, 1960; W. Zhang & Luck, 2008), generalizing selection mechanisms associated with perceptual

attention in order to explain which items are represented in memory (Murray, Nobre, & Stokes, 2011; Schmidt, Vogel, Woodman, & Luck, 2002). By contrast, there has recently been an increasing interest in the functional status of items in working memory (LaRocque, Lewis-Peacock, & Postle, 2014; Oberauer & Hein, 2012; Olivers, Peters, Houtkamp, & Roelfsema, 2011), and their ability to guide action (Chatham, Frank, & Badre, 2014; LaRocque et al., 2014; Olivers et al., 2011). This speaks to classic conceptions of working memory as intimately linked with executive control (Baddeley & Hitch, 1975) and the coordination of behaviour over time (Fuster, 1990b).

1.2.2 Biased competition

In Figure 1-1, perception and action are segregated and organized serially: perceptual cortex constructs a world model, and the world model is used to inform decision-making and action selection. It has been argued that this distinction is artificial (Cisek, 2007; Rizzolatti, Riggio, & Sheliga, 1994). For example, Cisek (Cisek, 2007) takes the position that the brain is primarily a device for generating actions, and that all other cognitive functions (including perceptual and mnemonic functions) are subservient to and emerge in relation to this goal. Specifically, the brain represents a mixture of sensory input and potential associated action sequences, and these mixed representations are termed affordances (Gibson, 1977). Affordances compete, mutually inhibiting one another. By analogy with the dominant theory of attentional selection in perception (Desimone & Duncan, 1995), influences from other systems bias the competition in favour of one or other affordance. This alters the pattern of activity in both sensory and motor cortex, in line with the chosen action.

This particular position may seem extreme, but the underlying principle of competitive processing under the influence of biasing signals is important (Desimone & Duncan,

1995; Norman & Shallice, 1986). It is a potentially efficient and self-organizing system for selection, as the selection mechanism is inherent and can operate independently of top-down control. Biasing influences can be applied as and when they are adaptive. This framework naturally integrates biases arising from bottom-up differences in stimulus salience and top-down task-control signals (Miller & Buschman, 2013; Nobre & Mesulam, 2014). There is also no requirement for a homuncular controller to coordinate selection as biasing signals could simultaneously arise from a variety of sources.

Competition may act at many levels – from lateral interactions between cells tuned to different features in visual cortex (Blakemore & Tobin, 1972) to lateral antagonism trading off behavioural strategies underpinned by different brain regions (Rushworth, Kolling, Sallet, & Mars, 2012). As applied to working memory, selective attention through biased competition may filter perceptual input, determining what is encoded into memory (Murray et al., 2011; Schmidt et al., 2002). Within memory there may be competitive interactions between actively maintained items (Compte, Brunel, Goldman-Rakic, & Wang, 2000; Edin et al., 2009; X.-J. Wang, 2001) which are biased by top-down signals similar to those regulating perceptual attention, in order to control working memory content (Murray, Nobre, Clark, Cravo, & Stokes, 2013).

1.2.3 Serial organization and 'attentional episodes'

Despite the advantages discussed above, there are drawbacks to the biased competition model of selection. Hierarchically organized and sequentially executed operations are a hallmark of human thought and behaviour (Botvinick, 2008; Lashley, 1951). Constructing a sentence is a prime example. Salient first understood want just can't

you you words put most the be if to[2]. Actions, like words, will often only achieve the desired goal if strung together in the right order, and so temporal ordering of motor behaviour is crucial. The same argument has been made for cognition – in order to solve a cognitive task of any complexity, mental operations must be organised into a chain of 'attentional episodes' that have a specific temporal sequence (Duncan, 2013). Whilst the biased competition model has advantages of robustness and self-organisation, it is less clear how this system can give rise to temporally organised thought and action. Working memory may have evolved specifically to subserve behaviour organised over extended periods of time, bridging the temporal delays and sequential sub-goals required for task execution (Fuster, 1990b; 2008). If biased competition is insufficient to explain the coordination of goal directed behaviour, then it is also not likely to account fully for control of working memory.

An alternative to distributed control is a centralized system in which a controller coordinates activity across a number of slave systems. This is an efficient solution given the need to coordinate several different cognitive and motor subsystems in the service of superordinate goals (Redgrave, Prescott, & Gurney, 1999). Baddeley's classic working memory model (Baddeley & Hitch, 1975) incorporated a central executive linking a number of modality-specific mnemonic buffers. This theory has been criticised for its 'homuncular' central executive (Parkin, 1998). Baddeley's rejoinder was that a homunculus was a necessary placeholder until executive function could be explained mechanistically (Baddeley, 1998). The homunculus is not yet retired, as we still understand little about the mechanisms that might give rise to sequential organization of behaviour. However, this is a hot topic in neuroscience; for example,

[2] Or, "You can't just put the most salient words first if you want to be understood"

the interaction between basal ganglia and prefrontal cortex has inspired computational models that aim to dissect the homunculus (Hazy, Frank, & O'Reilly, 2007).

1.2.4 Selection for working memory

Two general perspectives on selection in the brain were briefly outlined above. Biased competition was loosely associated with perceptual attention (Desimone & Duncan, 1995), and the idea of a central gating/sequencing mechanism with action control (Botvinick, 2008; Redgrave et al., 1999), but both concepts are relevant for selection for working memory.

A key distinction between a biasing model of selection and a model invoking centralized coordination is that whilst the former emphasises the activity or strength of a particular representation in relation to others as the key factor determining whether it will affect behaviour (Compte et al., 2000), the latter segregates input gating (determining what is represented) from output gating (determining which representations are 'active' at a given time). This can be illustrated with the example of a complex arithmetic calculation, in which several different cognitive operations need to be performed on information retained in working memory in order to reach a final answer (Duncan, 2013). Each operation will involve a subset of the representations in memory. Memory items not used in that specific computation need to be retained in the interim, whilst extraneous sensory distractors still need to be filtered out (imagine an office mate begins relaying numbers over the phone to a colleague whilst you are doing mental maths). Even the currently irrelevant memory items are therefore in some sense attended or selected. However, they must also be prevented from interfering with the current operation – so they must at the same time be in another sense deselected or inhibited. This example illustrates that selection for working memory must involve

separable mechanisms of input and output gating. A contention of this thesis is that even in very simple working memory tasks, both input and output gating are important, and that the working memory field has tended to underplay output gating as a performance-limiting factor.

The remainder of this introduction discusses behavioural (section 1.3) and neural evidence (section 1.4) concerning different modes of selection for working memory.

1.3 Cued selection for working memory: behavioural data and functional accounts

This section discusses behavioural findings from tasks employing explicit selection cues for working memory. Different proposals about the cognitive mechanisms mediating cue use are outlined. The discussion is focussed on the retrocue benefit investigated with MEG in this thesis.

Figure 1-2 schematizes different forms of cueing that have been used to investigate selection for working memory.

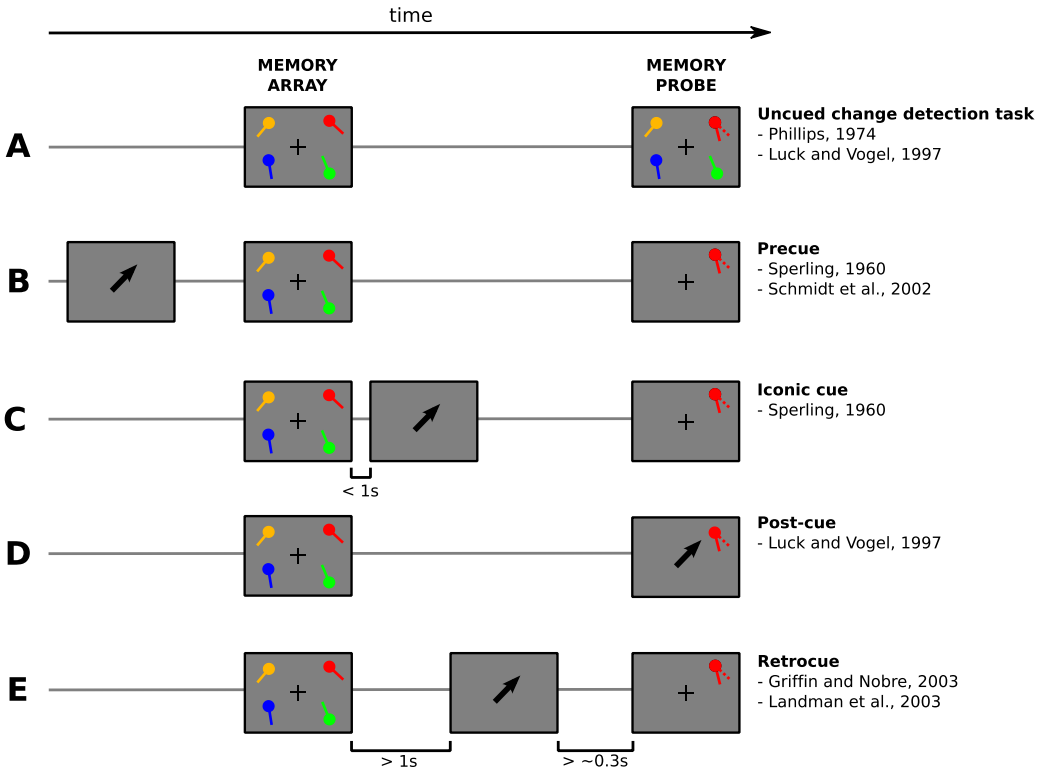


Figure 1-2 **Explicit cues for working memory selection.** The figure shows stimuli from the task presented in chapter 3 for illustration only. Some key studies are given in association with each cue type. **A – No cue, change detection task.** In the change detection task, the probe array re-displays the memory array, either unchanged or with alterations to one or more items. Subjects indicate whether the probe array is identical to or different from the memory array. **B - Precue**. Cue delivered prior to array onset. **C - Iconic cue.** Iconic memory is short lasting, and iconic cues follow the array with a short ISI (<1s). **D - Post-cue**. A central post-cue is shown here. Presenting a single probe item that unambiguously refers to one of the items in the memory array (e.g. by virtue of its spatial location) is similar to post-cueing. **E – Retrocue** Distinguished from iconic cueing by the longer [memory array : cue] ISI, typically > 1s. Also distinguished from post-cues by the cue-probe interval. Retrocues are ineffective if this interval is too short (becoming post-cues). The estimate that at least 300ms is needed between cue and probe is based on the data in Landmann et al. (2003).

A typical visual change-detection task is illustrated in the first row. Tasks of this sort have been used to investigate the capacity of visual working memory (Luck & Vogel, 1997; Phillips, 1974). A tacit assumption behind these capacity estimates has been that all the information represented in memory is available to guide a response (i.e. output gating is perfect). For example, this is made explicit in Cowan's formula for K (Cowan, 2000), a standard approach to calculating capacity based on the results of a post-cued change-detection task. As outlined below, the pattern of behavioural effects associated with retrocues call this assumption into question.

1.3.1 Iconic cues and precues

In a whole report paradigm, subjects are briefly presented with an array of items (e.g. letters) and asked to report as many as possible after a delay. Sperling (Sperling, 1960) noted that whilst whole-report paradigms tended to yield a capacity estimate for short-term retention of around four items, subjects would report that they had 'apprehended' more items than they were later able to recall. To circumvent the apparent memory limitation rendering this information inaccessible, Sperling used a part-report paradigm in which subjects were cued to some subset of the presented array very shortly after it had been displayed, the size of the cued subset lying within the four item capacity estimated from whole-report. An auditory tone, beginning at the same moment the visual array disappeared, signalled (by pitch) which of the rows in the array was to be reported. In this paradigm, given the cued row was not predictable trial by trial, it was assumed that whatever percentage of the items in the cued row was reported corresponded to the percentage of the whole array that was still represented in memory. This led to a capacity estimate of between 8 and 11 items – much higher than the ~4 items available in whole report. By varying the interval between the offset of the visual array and the presentation of the cue tone, Sperling found that this partial

report advantage lasted a very short time – around 1/6 of a second. Sperling's conclusion was that there were (at least) two qualitatively different buffers for the short-term store of information. Attention was able to selectively gate a subset of information in the short-lived store into a more robust store. Neisser later termed the short-lived high-capacity store 'iconic memory' (Neisser, 1967).

Whilst the early reports by Sperling (Sperling, 1960) and Averbach and Coriel (Averbach & Coriell, 1961) focussed on iconic cues appearing directly after the offset of the memory array, these studies also demonstrated in passing that cues appearing prior to the memory array can bias encoding. This seems a no-brainer. Trivially, if an item is not perceived it cannot be encoded, so in so far as attentional orienting determines perceptual sampling, it must also bias working memory encoding. However, it is not so clear that that the targets of covert attentional orienting will be automatically gated into memory, and this question has rarely been addressed directly. An exception is a study by Schmidt and colleagues (Schmidt et al., 2002), in which one item in a memory array more salient by virtue of 'sudden onset'[3], capturing attention, but did not predict the likelihood that the item would be probed. Nevertheless, memory was more accurate for this item[4]. Whilst this study demonstrated that incidental capture of attention can bias working memory encoding, as the authors acknowledged it did not test whether this relationship was obligatory. Additional mechanisms besides the allocation of perceptual attention might determine which stimuli are encoded into working memory. For example, dopaminergic and fronto-striatal circuitry commonly associated with action gating (Redgrave et al., 1999) but rarely linked with perceptual attention have been suggested to control working memory updating either by controlling temporal

[3] Whilst the location of all other items were foreshadowed by a placeholder stimulus, the location of the sudden onset item was not.

[4] This study also used articulatory suppression in conjunction with visual stimuli, thereby demonstrating that attention can bias encoding within purely visual working memory. The early iconic memory studies probably involved verbal recoding of the cued items.

updating windows (Braver & Cohen, 1999) or by gating specific representations (Hazy et al., 2007). This possibility is discussed in more depth in section 1.4.3 below.

1.3.2 Multiple memory buffers

Just as iconic cues were used to unmask the iconic buffer, retention-interval cues at longer latencies (or 'retrocues'; row E, Figure 1-2), have been argued to reveal a fragile, high-capacity form of working memory distinct both from the iconic store, and from robust working memory. Landman et al. (2003) performed a similar experiment to Sperling, except that they used a visual change-detection task (memory for 8 oriented bars) instead of verbal report, and they delivered cues at longer latencies than Sperling's ~200ms estimate for the persistence of iconic memory. The cues were lines extending from the fixation cross towards the item that changed between memory and probe, and were 100% valid. Without cues, or with a post-probe cue, capacity was estimated at around 4 items – commensurate with the 4-item limit typically observed in change detection tasks (Cowan, 2000; Luck, Chelazzi, Hillyard, & Desimone, 1997). However, a cue delivered in the delay interval improved performance, up to an estimated capacity of between 6 and 7 items – a 'retrocue benefit'. This was true of array-cue ISIs up to 1200ms. Performance was slightly lower when the cue appeared less than 100ms prior to the probe array, but testing at different array-probe ISIs indicated that the decrement was related to proximity of the cue to the probe array, not its latency after the memory array, suggesting that this fall-off in performance was driven by probe-interference rather than decay of the memoranda (c.f. the distinction between post-cues and retrocues). The authors concluded that there was a large-capacity visual working memory store lasting longer than the supposed time-course of iconic memory, but which was overwritten by the probe array: a 'fragile buffer'. Cueing attention within this fragile buffer could move items into a more robust store.

Sligte , Scholte and Lamme (Sligte, Scholte, & Lamme, 2008) extended these findings. They used large (up to 32-item) oriented-bar memory arrays, and delivered iconic cues, retrocues or post-cues on different trials. Retrocues resulted in capacity estimates as high as 16 items. Pattern masks (consisting in a second stimulus array with new random item orientations) presented prior to retrocues disrupted the retrocue benefit, but simply flashing the screen white did not, consistent with Landman's (2003) proposal that the fragile buffer is over-writable by new stimuli.

Although Sligte et al. (2008) proposed a tri-partite organization for working memory, with separate iconic, fragile and robust stores, these authors have since emphasised that fragile working memory may depend on a continuous range of representational buffers . The key point is not the number of buffers, but the way retention properties map to the hierarchy in visual cortex (Sligte, Vandenbroucke, Scholte, & Lamme, 2010). As the visual hierarchy is ascended, receptive fields become larger, capacity is correspondingly reduced, and robustness to interference increases as receptive fields become more specialized. The role of attention is to select information from within this representational space and render it robust. Precues, iconic cues and retrocues could all act via this common mechanism.

1.3.3 Attention and working memory as a common process

A different perspective on the retrocue benefit emphasises the potential overlap between mechanisms involved in directing perceptual attention, and in orienting attention to representations in memory. A contemporary hypothesis is that working memory maintenance consists in sustained attention to a set of internal representations, and that the capacity limit of working memory and the capacity limit of attention are one and the same (D. E. Anderson, Vogel, & Awh, 2013; Awh & Jonides, 2001; Awh, Vogel, &

Oh, 2006; Chun, 2011; Kiyonaga & Egner, 2012). In this framework, which sees working memory as a special attended state within long term memory (Cowan, 2000), directing attention within memory is analogous to orienting attention to perceptual input. This 'identity theory' naturally accounts for the retrocue benefit in terms of a redistribution towards the cued item of the attentional resource that underpins working memory. A related view is that retention of a representation in working memory is sufficient for selective attention: top-down activation of the neural representation associated with the memory item (constituting memory maintenance) also biases competitive perceptual processing (Desimone & Duncan, 1995). A key motivation for equating selective attention and memory maintenance is the finding that neural structures associated with the two constructs are highly overlapping (Nobre et al., 2004), and electrophysiological correlates of perceptual orienting and orienting within memory are also similar (Griffin & Nobre, 2003). The neural evidence is discussed in more depth below (section 1.4).

This account assumes there is a single memory buffer with a capacity identical with the capacity of selective attention (D. E. Anderson et al., 2013). Behaviourally, retrocues should therefore only work *within* the capacity of working memory[5], boosting performance for one of the memory items already selected by attention. Whilst the findings of Landman (2003) and Sligte (2008) suggest that retrocues can 'release' a capacity much higher than the ~4 item capacity of robust working memory, other investigators have not observed this high level of performance following retrocues. Astle et al. (2011) investigated both iconic cues (150ms array-cue ISI) and retrocues (1500ms array-cue ISI) for 4 or 8 item colour memory. Iconic cues boosted capacity to

[5] Though, this argument is complicated by the fact that traditional estimates of capacity (e.g. using the change detection task) may be underestimates due to interference from the probe array – an argument made elsewhere in this introduction (see also Makovski, Sussman and Jiang, 2008). However, this is predicated on either a fragile-store or output-gating account of the retrocue benefit. The specific 'attention as memory' model discussed in this section equates attentional and mnemonic capacity.

~5 items with 8 item arrays, but retrocues did not boost capacity beyond ~3 items. Matsukura and Hollingworth (2007) were also unable to replicate the very high capacity estimates of Sligte et al. (2008) and suggest that a combination of grouping strategies and extensive practice might have inflated the capacity estimate. Matsukura found that with training, a single subject's capacity estimate can rise steadily from 4 to 7 items. A second reason to doubt that retrocues retrieve items from a fragile buffer is that the retrocue benefit can survive a mask interposed between the memory array and retrocue (Makovski, 2012), suggesting that retrocues can confer a behavioural benefit by prioritizing items in a memory buffer that is already robust to perceptual interference[6].

Biasing of perceptual attention by memory maintenance, and interference with memory maintenance by interposed attentional demands, have both been put forward as behavioural evidence that a common mechanism mediates attention and working memory (Awh, Jonides, & Reuter-Lorenz, 1998; Downing, 2000; Pashler & Shiu, 1999; Soto, Heinke, Humphreys, & Blanco, 2005). Awh et al. (Awh et al., 1998) demonstrated that when a spatial location was held in memory, responses were faster on a discrimination task interposed in the retention interval when the to-be-discriminated stimulus appeared at the same spatial location as the location held in memory. They also demonstrated the converse: forcing subjects to re-orient attention during the retention interval to perform a colour discrimination impaired spatial working memory performance more than when the colour discrimination did not require a shift in attention. Using a similar paradigm, Downing (Downing, 2000) demonstrated that when an object was held in memory, its re-appearance at a different spatial location

---

[6] This seems to contradict Sligte et al.'s (2008) finding that masks interposed between memoranda and retrocue abolished the retrocue benefit. However it is possible that this was less a masking effect than a confusion effect. Sligte's masks were identical to the memory arrays, except that the item orientation was randomized. In pilot experiments conducted as part of this thesis, interposing a mask of this sort was found to impair memory performance in a task where subjects had to 'dial up' the orientation of the cued item at the end of the trial. Further analysis revealed that 50% of responses were in fact being made to the masking item, suggesting that subjects confused the two displays (i.e. the first array was not necessarily overwritten).

during the retention interval biased spatial attention towards that location, as indexed by reaction time in retention-interval discrimination task. Soto et al. (Soto et al., 2005) found that retaining an object in working memory biased a visual search task interposed in the retention interval, in which the memory item appeared as a distractor. This appeared to be obligatory, as even when the location of the memory item never coincided with the search target, it still slowed search.

1.3.4 Active and passive states in working memory

Information in working memory can bias attention – but is this truly obligatory? Several lines of evidence suggest that equating perceptual attention and memory maintenance is an over-simplification (Olivers et al., 2011; Woodman, Carlisle, & Reinhart, 2013). For one, items in working memory do not always bias visual attention. Downing (2000) suggested that this might be the case, citing pilot data in which visual search for one object held in memory was unaffected by the retention of a second object in memory. Downing and Dodds (2004) explicitly investigated this possibility, initially presenting two objects, of which one was designated the search target and the other was to be retained in memory (memory target). This second memory target was sometimes present in a subsequent search array, in which subjects had to locate the search target. Finally a memory probe appeared, which subjects compared with the memory target. Search on those trials in which the memory target appeared as a search distractor was no slower or less accurate than on the remainder of trials. Similarly, Houtkamp and Roelfsema (Houtkamp & Roelfsema, 2006) presented subjects with two search items, one on the left and one on the right side of the screen. Subjects then searched an array on the left side of the screen for the left-hand search item. The right hand search item still had to be remembered, because after completing search for the left-hand item, a search array appeared on the right, for which the

right-hand item was the target. Whilst the right-hand search item was not a target in the first search array, it was included as a distractor on 50% of trials. The authors found no slowing of the first search on target-present trials in which the second search item was present as a distractor, and neither did it capture eye movements. It did, however, slightly slow responses in target-absent trials.

These data imply that items in memory can have a different status depending on whether they are currently prioritized to guide task performance: top-down biasing by memory content is not automatic, but is recruited strategically (Woodman et al., 2013). Woodman and Luck (2007) suggested that in earlier studies like Downing (2000) memory items captured attention because there was some strategic incentive to attend to that item in the search array (e.g. in order to refresh its memory representation) even if this was not in principle necessary to perform the task. They found that during visual search, the presence of irrelevant memory items in the search array that were never coincident with the search target could even speed the search (Woodman & Luck, 2007).

Our ability to strategically modulate the effect of memory items on search suggests that the link between working memory and perceptual attention cannot be so simple as implied by the 'identity theory' (Cowan, 2000). Instead, mechanisms must exist which modulate the status of items within working memory. Similarly, working memory may have a more complex structure than single-buffer accounts (Cowan, 2000) would suggest, within which some items can be elevated to an active status (in which, for example, they can guide visual search) whilst other items remain passive (Olivers et al., 2011). Olivers et al. (2011) term the passive yet stored items 'accessory memory items' (AMIs). Oberauer has proposed that working memory may have a tripartite structure in which a single active or attended item sits at the top of the hierarchy, a small capacity-limited memory buffer (corresponding to classic '4-item' WM) lies below,

and this is part of a larger activated portion of long-term memory (Oberauer & Hein, 2012). The middle buffer is differentiated from the passive long term store by the existence of temporary bindings between each stored item and locations in a mental space corresponding to the current task; e.g. ordinal list position, or a screen location. However, only one item, and its associated bindings, can be activated at a given time. McElree (2006) argues for a simpler organisation in which the only categorical distinction to be drawn is between the currently focussed item and the remainder of LTM, within which accessibility is determined by recency and interference.

The different structural accounts introduced above all incorporate a single 'active' item. This idea of temporary prioritization of a single representation for use in action or cognition is implicit in Duncan's suggestion that organized cognition and behaviour is constructed from sequential 'attentional episodes', in which representations are sequentially activated to solve sub-goals (Duncan, 2013), as discussed above in the overview. It can also be compared to the idea of a global cognitive workspace, through which representations currently selected to control cognition or action are able to influence a wide range of brain systems (Baars, 2005).

Whatever structural account turns out to best capture the organization of memory, it is clear that if an item in memory can be temporarily prioritized for cognitive processing or action guidance, then some selection mechanism is likely to exist to control this prioritization. This mechanism could equally well be framed as perceptual (selecting a representation for processing) or 'motoric' (selection of a representation to guide cognition or action). Returning to the retrocue paradigm, it is possible that the retrocue benefit arises not because the cued representations are strengthened or protected *per se* (Pertzov, Bays, Joseph, & Husain, 2012; Sligte et al., 2008), but because retrocues facilitate the efficient selection of the relevant item to guide behaviour

('output gating'). With the aid of a retrocue, this can be accomplished in advance of the presentation of the memory probe, reducing competition for output gating at the point when the probe item appears. Behavioural observations suggest that this might be a good explanation for the retrocue benefit. Murray et al. (2013) used a precision/capacity task combined with a model-based analysis (W. Zhang & Luck, 2008) to characterise (a) the likelihood that an item in memory could be recalled at all, and (b) how precisely those items that could be recalled were represented. Retrocues had no effect on memory precision, which is not what would be expected if retrocues prevent memory decay (Pertzov et al., 2012) or reallocate maintenance resources (Bays & Husain, 2008). However, retrocues did substantially increase the probability that the cued item could be recalled at all. Crucially, this was true even comparing retrocues with a condition in which the memory probe was moved forward to the same latency as the retrocue, so this effect cannot be explained in terms of protection of the cued item from decay over the rest of the retention interval (Murray et al., 2013). Instead, retrocues render accessible items in memory that are inaccessible in uncued trials. A second phenomenon that suggests output gating may be a performance-limiting factor is 'misgating', in which the wrong item in memory appears to guide behavioural output (Bays & Husain, 2008). This has been argued to reflect inappropriate binding between different feature dimensions of the items represented in memory, but might be equally accounted for at the output stage if the wrong item was selected wholesale to guide behaviour ('misgating'). This is discussed further in chapter 3, in which retrocues are shown to almost eliminate misgating.

The hypothesis that the ability to retrieve (or switch attention to) the relevant memory item is a major limiting factor on performance is not prominent in the mainstream of the working memory literature, in which the emphasis on representational content and the nature of capacity limits dominates the discussion. It is however familiar from work on

task-switching (Monsell, 2003). Past work on the costs of switching focus in working memory has used reaction time as the dependent variable (Garavan, 1998; McElree, 2006). Garavan (1998) presented subjects with small groups of either triangles or rectangles on each trial, and asked them to keep a running tally of the number of each type of item they had seen. Sometimes a trial in which triangles were presented would be followed by a trial with rectangles (a switch trial), and sometimes by a trial with more triangles (a non-switch trial). Subjects pressed a key to move between trials, and there was therefore an RT measure indicating the time taken to update the running count. Non-switch trials were faster than switch trials by between 200 and 500ms (this decreased with practice), suggesting that the tallies, whilst both in memory, were not equal: one was active and one passive, and additional time was required to switch to updating the passive tally as compared to continuing to update the active tally. Garavan interpreted this as the time taken to switch attention between different elements in working memory.

Visual working memory tasks are somewhat different from than the arithmetic updating task used by Garavan, but if retrocues allow a preparatory switch within working memory we might expect them to give rise to similar RT advantages. Retrocues do speed responding: for example, Astle et al. (2011) found a robust reaction time effect of retrocueing, responses to cued items being around 150ms faster than to uncued items. However, the strength of a memory representation can also influence reaction time (Ratcliff & McKoon, 2008), so these effects do not adjudicate between a representational and an output gating account. McElree (2006) addressed this issue, investigating RT in word-list memory tasks using the response-signal speed-accuracy trade-off procedure. In an speed-accuracy tradeoff (SAT) task, subjects are forced to give their responses at certain fixed latencies after the probe item is presented. The resulting SAT curve disambiguates speed of evidence accumulation (linked to representation strength), and

the time taken to retrieve a representation (similar to switch time). In a series of experiments, McElree found that the last item to be presented in a memory list had a qualitatively different SAT curve than earlier items, consistent with an earlier start for evidence accumulation (by ~35ms). Critically, this advantage could be shifted to other items by using a retrocue. McElree proposed that this indexed the shift of a single-item focus within memory. In chapter 3, a diffusion model is applied to data from a visual working memory task with retrocues to decompose RT into retrieval and evidence accumulation time, and a similar result was obtained.

This section has outlined behavioural evidence for different forms of selection for working memory – and associated explanations for the retrocue benefit, as schematized in Figure 1-3, below. On the basis of the behavioural evidence discussed above, it seems at least two mechanisms might contribute to selection for working memory: an input gating mechanism that selects information for representation in memory, and an output gating mechanism that can select an item in memory to guide behaviour.

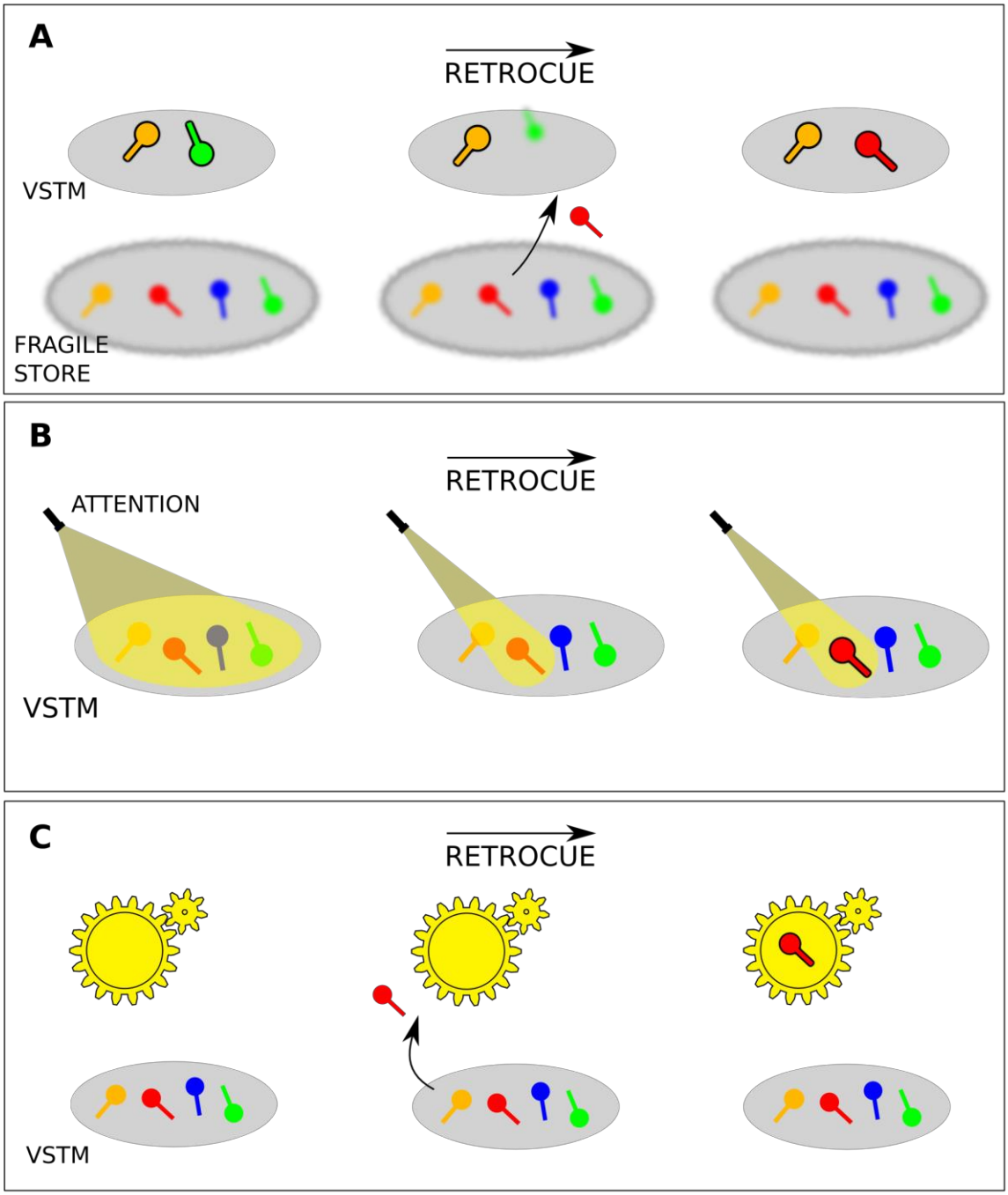


Figure 1-3 A cartoon of three proposed explanations for the retrocue benefit. **A** - the 'fragile buffer' account (Landman et al., 2003; Sligte et al. 2008). Here, the retrocue facilitates the transfer of a memory representation from a large-capacity fragile memory store to a limited-capacity, robust store. **B** - the 'memory as attention' account. Here, retrocueing re-allocates the attentional focus within memory. **C** - the output gating account. Here, retrocues allow one item in the VSTM store to become the 'active' item, able to directly guide cognition/behaviour.

How do these different modes of selection map to neural substrates? To what degree might common or distinct mechanisms control input and output gating? The following section of this introduction discusses neural mechanisms for working memory selection.

1.4 Neural mechanisms

Selective attention tasks and working memory tasks activate overlapping brain regions (Figure 1-4, below) in frontal and parietal cortex (Nee & Jonides, 2009; Nobre et al., 2004). This has been a key piece of evidence in favour of the idea that common selection mechanisms work in both the perceptual and cognitive domain, but comparison of frontal regions involved in perceptual orienting and selection from working memory suggests that additional sites might be recruited when people orient to the contents of memory (Nee & Jonides, 2009; Nobre et al., 2004). Sensory cortex is also implicated in control operations for working memory, and perhaps maintenance, but its precise role is debated (Offen, Schluppeck, & Heeger, 2009; Stokes, 2011). Finally, besides these neocortical sites, it is likely that the dopamine system and basal ganglia play a considerable (if as yet relatively unexplored) role in the control of working memory (Hazy et al., 2007).

The anatomy of frontal and parietal control regions implicated in working memory control is discussed in more depth in the context of an fMRI meta-analysis in chapter 5, and the putative role of dopaminergic and striatal mechanisms is discussed in depth in chapter 6. The discussion here aims to give a brief overview of the neural mechanisms potentially involved in control of working memory.

1.4.1 Cortical control networks for attention and working memory

In the 1930s, Jacobsen used a delayed response task to assay the effects of prefrontal lesions in monkeys, and found that whilst the animals were still capable of performing visual discriminations and dealing with familiar puzzle boxes, they had a specific deficit

in retaining information over delays (Jacobsen & Nissen, 1937). From this point on[7], the prefrontal cortex has been consistently associated with working memory, and by the 1970s, invasive electrophysiological recordings complemented lesion evidence. Sustained activity was discovered in dlPFC around the principle sulcus during the delay interval of delayed reaching and delayed alternation tasks (Fuster & Alexander, 1971; Kubota & Niki, 1971). Electrophysiologists subsequently documented similar dlPFC delay activity in visual working memory tasks (Miller, Erickson, & Desimone, 1996; Miller, Li, & Desimone, 1993). However, neural correlates of working memory are not restricted to dlPFC: parietal cortex exhibits delay-spanning activity very similar to that observed in frontal cortex (Chafee & Goldman-Rakic, 1998).

PET and fMRI have since associated similar frontal and parietal regions with working memory tasks in human subjects (for review see Curtis & D'Esposito, 2003). Do these sites mediate maintenance of active representations (Sakai, Rowe, & Passingham, 2002), manipulate or select working memory content (Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000), or code prospectively for upcoming task demands on the basis of working memory content (Stokes, 2011)? Whether or not memory maintenance even requires sustained and/or stable neural activity is a moot question (Buonomano & Maass, 2009; Mongillo, Barak, & Tsodyks, 2008; Stokes et al., 2013) that will be returned to below. The role of frontal and parietal regions in cue-driven selection for working memory is discussed here.

Working memory has been linked to frontal and parietal sites, but perceptual attention is also associated with activity in a network of regions in frontal and parietal cortex (Corbetta, Miezin, Shulman, & Petersen, 1993; Mesulam, 1981; Nobre & Mesulam, 2014;

[7] Ferrier (1876), Bianchi (1895) and others had previously found that ablating PFC impaired the organisation of behaviour, but their approach was impressionistic and they did not operationalize working memory with a behavioural task.

Nobre et al., 1997). The regions involved in controlling preparatory biasing of attention in the perceptual domain overlap with regions involved in selecting information from within working memory (Nee & Jonides, 2009; Nobre et al., 2004).

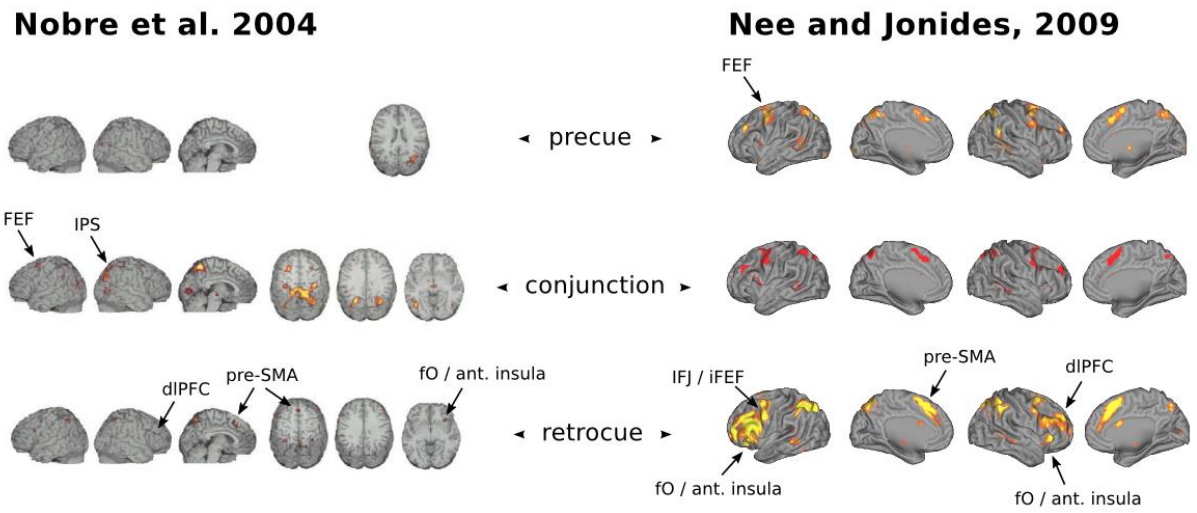


Figure 1-4 – **Common and unique activations in response to precues and retrocues**. Figures modified from Nobre et al. (2004) and Nee and Jonides (2009). Labels have been added to highlight key activations. On the left are shown fMRI activation maps from Nobre et al. (2004), which used colour as the remembered feature, and spatial cues to locations in the memory array. There are common precue/retrocue activations in FEF and IPS. Retrocues additionally recruit the dlPFC, pre-SMA and frontal operculum (fO) / anterior insula. On the right are shown fMRI activation maps from a similar study by Nee and Jonides (2009), which used words as memoranda, and cues that referred to the colour of the words as displayed on the screen. There are common activations in premotor cortex, but these are more dorsal (FEF) for precues, and more ventral (iFEF/IFJ) for retrocues. Retrocues additionally recruited the fO / anterior insula and pre-SMA. Despite using rather different memoranda and cue types, these two studies discovered remarkably similar patterns of activation.

Specifically, dlPFC (Brodmann area 46), regions in premotor cortex, and regions in the superior parietal lobule (SPL) and intra-parietal sulcus (IPS) are activated following both precues and retrocues. The overlap is consistent with the 'identity theory' of working memory and selective attention discussed above (Awh & Jonides, 2001; Kiyonaga & Egner, 2012), but is equally consistent with a more conservative account in which selection from perceptual arrays or the contents of memory recruit a common mechanism, but this mechanism is not involved in maintenance *per se*. An attractive idea is that at least some parts of the fronto-parietal network constitute a common mechanism for top-down modulation of activity in perceptual cortex. Whether this is

constitutive of memory maintenance, or is better seen as a more dynamic function that supports memory manipulation as and when this is task-relevant, is discussed below and in later chapters.

The overlap within the fronto-parietal network is not perfect (Figure 1-4), suggesting that at least parts of the system are recruited in different ways during perceptual and mnemonic orienting. In particular, a dorsal area of premotor cortex (FEF) is activated following precues, whereas a more ventral region (iFEF and/or inferior frontal junction, IFJ) is activated following retrocues (Nee & Jonides, 2009). These differences are explored in chapter 6, in which they are replicated with MEG.

Retrocues also recruit a set of brain regions not recruited by precues (Figure 1-4). These include the bilateral anterior insula or frontal operculum, and the dorsal anterior cingulate in the region of pre-SMA (Higo, Mars, Boorman, Buch, & Rushworth, 2011; Nelissen, Stokes, Nobre, & Rushworth, 2013; Nobre et al., 2004). The fO has been suggested to directly influence activity in perceptual cortex during retrocueing (Higo et al., 2011), a finding that seems *prima facie* at odds with the idea that the fronto-parietal network constitutes a common path for top-down modulation of posterior cortex. Just as the fronto-parietal sites are viewed as a functional network (Mesulam, 1981; Nobre & Mesulam, 2014), according to an influential view of frontal network structure based on clustering resting state correlations between frontal regions, the pre-SMA and fO form part of a second, cingulo-opercular control network which also incorporates lateral frontal pole (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Dosenbach et al., 2007; Power & Petersen, 2013). This network segregation is explored in chapter 6, in which the time-course of network responses to precues and retrocues are interrogated to establish which has the primary role in top-down modulation.

1.4.2 Role of sensory cortex

If working memory consists in attention to internal representations, a natural neural substrate would be for top-down influences from frontal cortex to act on sensory cortex, activating perceptual representations endogenously. Imaging and electrophysiology suggest that perceptual cortex is important for working memory (Pasternak & Greenlee, 2005). This is intuitively appealing: why duplicate representations in an extra-sensory buffer when sensory cortex already constitutes a dedicated representational resource? A stronger *a priori* argument is that a link to labelled-line representations in sensory cortex is necessary to anchor activity in flexibly allocated frontal populations (Duncan, 2001) to specific content. However, the neural data as reviewed below suggest that perceptual cortex is not the locus of active maintenance. Instead, sensory cortex may be recruited dynamically to reactivate perceptual content associated with working memory. This would free perceptual cortex up for on-going sensory processing at other times during memory retention.

Non-human primate (NHP) studies have identified retention-interval activity in perceptual cortex, with a majority of studies focussing on area IT (Chelazzi, Duncan, Miller, & Desimone, 1998; Fuster, 1990a; Fuster & Jervey, 1981; Miller et al., 1993; Miyashita & Chang, 1988), though others have found similar neural signatures in V4 (Ferrera, Rudolph, & Maunsell, 1994), MT (Bisley, Zaksas, Droll, & Pasternak, 2004) and even V1 (Super, Spekreijse, & Lamme, 2001). This has been interpreted as evidence for active storage of short-term memories in sensory cortex (Pasternak & Greenlee, 2005). Human imaging experiments have thrown up some similar results. Harrison and Tong (Harrison & Tong, 2009) applied multi-voxel pattern analysis (MVPA) to fMRI data and found that they could decode the orientation of a remembered grating from early visual cortex during the retention interval of a working memory task (Ester, Serences, & Awh,

2009; Serences, Ester, Vogel, & Awh, 2009). EEG has furnished an analogous finding in the form of contra-lateral delay activity (CDA). Vogel and Machizawa (Vogel & Machizawa, 2004) found that when subjects were cued to remember only the stimuli presented on one side of visual space, there was a sustained load-dependent negative voltage over the contralateral hemisphere relative to the ipsilateral hemisphere, in posterior electrodes. This was proposed to reflect sustained maintenance-related activity in perceptual cortex.

In section 1.3.3 above, it was argued that whilst memory tasks may recruit attentional processes, those processes might not be necessary for maintenance. Instead, as discussed above, behavioural evidence indicates that items maintained in working memory only bias attention if this is encouraged by task demands (Olivers et al., 2011; Woodman et al., 2013). An analogous argument with respect to the neural data is that apparent maintenance activity in perceptual cortex is observed only when task demands focus attention on the item in question. Most of the tasks discussed above employed memory for a single item (Chelazzi et al., 1998; Harrison & Tong, 2009; Miyashita & Chang, 1988; Serences et al., 2009; Super et al., 2001), meaning there is no reason not to either focus on that item during the retention interval (Olivers et al., 2011), or to anticipate the upcoming probe item (Stokes, 2011; Stokes, Thompson, Nobre, & Duncan, 2009b). Even for tasks with a memory load of one, the picture is mixed, with inferotemporal delay activity much more susceptible than frontal delay activity to disruption by distractors (Miller et al., 1993).

This does not mean that perceptual activity observed during the delay interval in working memory tasks is epiphenomenal. It may be related to memory updating, cognitive operations, or preparatory attention rather than memory maintenance. Cueing attention to working memory content robustly activates perceptual cortex

(Nobre et al., 2004; Yi, Turk-Browne, Chun, & Johnson, 2008). This reactivation is specific to the cued representation. For example, different areas in the ventral stream have differential sensitivity to different stimulus classes (Epstein & Kanwisher, 1998; Kanwisher, McDermott, & Chun, 2002), and this property has been exploited in studies demonstrating selective modulation in functionally specialized areas following a retrocue to stimuli from the appropriate category (Higo et al., 2011; Lepsien & Nobre, 2006; Lepsien, Thornton, & Nobre, 2011; Nelissen et al., 2013). Similarly, sensory reactivation after retrocues is retinotopic, reflecting the spatial layout of memory (Kuo, Stokes, Murray, & Nobre, 2014; Munneke, Heslenfeld, & Theeuwes, 2010; Sligte, Scholte, & Lamme, 2009). In chapter 4 of this thesis, similar retinotopic patterns of alpha power modulation in visual cortex are shown to result from both spatial precues and spatial retrocues.

The 'memory as attention' account would explain these modulations as a reallocation of the tonic top-down drive constituting maintenance. However, Lepsien et al. (2006) found that whilst only the cued item is reflected in activity in visual cortex following a retrocue, this pattern can be *reversed* by a second cue. More recently, Lewis-Peacock et al. (2012) trained pattern classifiers to decode stimulus category from fMRI data during memory retention. They used these classifiers to investigate the effects of distraction and cueing on delay-period activity, excluding from the classification any voxels in the PFC[8]. When a single item was presented for memory its category could be decoded during retention - unless distracting stimuli from another category were presented, in which case classifier evidence fell to baseline for the memory category and rose for the distractor category. Nevertheless, memory performance was unimpaired following distraction (in fact, it was slightly boosted) suggesting that the memory items must still

[8] Similar results were obtained when the classifier was trained/tested with data from inferior temporal cortex only.

have been maintained. When 2 items were presented, classifier evidence initially rose for both items, but after one of the two items was cued, only the identity of the cued item could be decoded and classifier evidence for the second item fell to baseline. Crucially, a second cue could redirect attention to the previously-cued item, and the pattern of classifier evidence then reversed in favour of the most recently cued item. Both items could be recalled at the end of the trial.

These data, which have since been replicated with EEG (LaRocque, Lewis-Peacock, Drysdale, Oberauer, & Postle, 2013), strongly suggest that sensory cortex is recruited only for the currently focussed item in memory (McElree, 2006; Oberauer & Hein, 2012; Olivers et al., 2011), and that neural measures from fMRI and EEG are not sensitive to the remainder of the items being maintained - the accessory memory items (Olivers et al., 2011). These items might either be actively maintained in a form to which the classifiers were not sensitive, or alternatively, maintenance of AMIs may not rely on sustained activity at all. Short-term synaptic potentiation resulting from calcium dynamics in the presynaptic bouton or GluR1 receptor dynamics have been proposed to mediate 'silent' synaptic working memory (M. A. Erickson, Maramara, & Lisman, 2010; Mongillo et al., 2008).

This model of working memory, in which only immediately task-relevant representations are associated with top-down activation, and other items are retained 'passively' (LaRocque et al., 2014), suggests that brain activity associated with working memory is more dynamic than the sustained top-down biasing model would imply (Cowan, 2000). Representations will be activated only as and when required, freeing up perceptual and attentional resources at other times (Hollingworth & Maxcey-Richard, 2013). Passively stored representations might even be less prone to perceptual interference. Zokaei et al. (2014) found that when perceptual cortex (MT) was

disrupted using TMS during the retention interval of a working memory task, only memory for just-cued (and therefore focussed) items were impaired, and uncued (and therefore passively stored) items were if anything recalled more successfully after TMS.

Why has the tonic biasing account been so prevalent? One explanation is that most tasks have confounded internal focus and working memory maintenance, by presenting or rendering task relevant only one item (e.g. Harrison & Tong, 2009). Another is that the dominant methods are not able to reveal dynamic temporal structure. ERPs associated with mnemonic orienting are transient (Griffin & Nobre, 2003) but this is because they are sensitive only to responses phase-locked to task events - a transient ERP difference does not imply a transient neuronal activation. fMRI is limited by the inherently slow time-course of the haemodynamic response function. By contrast, event-locked time-frequency analyses of EEG or MEG data are ideally suited to elucidating the temporal dynamics of working memory control. This is the analysis approach applied in chapters 4 and 5 of this thesis.

Returning to the general theme of selection within working memory – should cue-driven switches of focus between memory items be construed as an attentional function? To recapitulate the above discussion: retrocues produce a similar pattern of control region activation and modulations in perceptual cortex as does preparatory attention, which has led to the idea that perceptual attention and mnemonic selection are overlapping functions (Chun, 2011). The canonical model for perceptual attention is of inherently competitive processing modulated by top-down biasing (Desimone & Duncan, 1995). Working memory maintenance has also been construed as competitive (Edin et al., 2009), with top-down influences (e.g. following a retrocue) biasing competition for continued maintenance (Murray et al., 2013), via the same or similar fronto-parietal circuits as mediate perceptual biasing (Gazzaley & Nobre, 2012).

This model is only plausible if several memory representations are co-activated in a quasi-perceptual buffer. This might be an inappropriate model for working memory maintenance. Instead, re-activation in perceptual cortex following retrocues might reflect the 're-perception' or 'refreshing' of the task-relevant item. Top-down attentional biasing might be a misnomer in this context, as there is no on-going maintenance activity to bias. Rather, an activation pattern in sensory cortex is endogenously driven, akin to mental imagery (Stokes, Thompson, Cusack, & Duncan, 2009a). Despite this important functional distinction between attentional biasing and memory refreshing, the same brain regions (e.g. IPS) typically involved in top-down attentional biasing might be recruited to re-activate sensory representations associated with the cued item, constituting a common mechanism for top-down attentional biasing or cognitive recruitment of sensory cortex.

1.4.3 Fronto-striatal interactions, and the role of dopamine

As suggested early in this chapter, specific attentional mechanisms acting in sensory cortex may be enslaved to more central mechanisms of executive control (Baddeley & Hitch, 1975). Loci of executive control are thought to lie in the prefrontal cortex (Miller & Cohen, 2001). However, two related subcortical systems, the dopamine system and the basal ganglia, have also been proposed to be crucial for cognitive control (Braver & Cohen, 2000; Hazy et al., 2007; Redgrave et al., 1999), in an interaction with PFC. This theory has been developed into explicit computational models (Hazy et al., 2007; Hazy, Frank, & O'Reilly, 2006), offering a solution to the problem of the homuncular central executive in Baddeley's classic model (Baddeley & Hitch, 1975). This thesis did not directly investigate fronto-striatal interactions, although it seems likely that the gating functions studied using retrocues will recruit basal ganglia. It did investigate the relationship between reward learning and working memory gating at a behavioural

level (chapter 6). The evolving account of fronto-striatal contributions to memory control is therefore outlined briefly, focussing on the way in which it integrates reward learning with control of working memory.

The basal ganglia gate motor output (Mink, 1996). Somatosensory cortex, motor cortex, supplementary motor area and premotor cortex send projections to the striatum, and via the thalamus the striatum projects back mainly to the SMA. This 'motor loop' gates action candidates, releasing appropriate action output by disinhibiting neurons in the SMA. The motor loop is one of a number of reciprocal loops, the remaining loops connecting dlPFC, FEF, anterior cingulate and orbitofrontal cortex with the striatum (Alexander, DeLong, & Strick, 1986). By analogy with motor gating, these interactions may also gate cognitive operations (Hazy et al., 2006; 2007). Hazy et al. (2007) present a model in which two forms of gating are distinguished. Input gating determines which representations gain access to prefrontal resources; i.e. are represented in working memory. This aspect of the cortico-striatal interaction may provide the additional gating mechanism, on top of selective perceptual attention (Schmidt et al., 2002), that helps render working memory robust to attentionally demanding tasks carried out during memory maintenance (Hollingworth & Maxcey-Richard, 2013; Maxcey-Richard & Hollingworth, 2013). Without an additional gating mechanism, memory content would be quickly elbowed out by the targets of attentional orienting during on-going behaviour. In the prefrontal cortex – basal ganglia (PFC-BG) model, this is achieved by Go/NoGo gating of input to prefrontal 'stripes' (Frank, Loughry, & O'Reilly, 2001) that are responsible for maintenance. Output gating determines which items are used to guide cognition or action, a function that was discussed above in the context of sensory recruitment during working memory selection (LaRocque et al., 2014). This is similarly achieved by Go/NoGo gating, but in this case of layer 5 projection neurons in PFC, projecting either to posterior cortex or other frontal sites.

Many properties of the PFC-BG interaction make it suitable for executive control. Gating is dissociated from maintenance, resolving the trade-off between flexible updating and robust storage which is a problem for the 'attention as memory' model. It also permits sequential gating influences from cognitive towards motoric representations due to the spiralling structure of PFC-BG loops (Haber, 2003). An important property is that under the influence of a dopaminergic teaching signal, the PFC-BG system can learn which representations to gate into and out of memory (Badre & Frank, 2012; Frank & Badre, 2012). Reinforcement learning trains the PFC-BG system to gate only appropriate motor output (Hikosaka, 2005). Similarly, reinforcement learning may bias working memory gating (Hazy et al., 2007).

The evidence for a close association between the basal ganglia and working memory control is partly anatomical (Middleton & Strick, 2000). More direct evidence for the role of the PFC-BG interaction is beginning to emerge from imaging work (Callicott, 1999; Chatham et al., 2014; Lewis, Dove, Robbins, Barker, & Owen, 2004; McNab & Klingberg, 2007). For example McNab et al. (2007) demonstrated that preparing to filter a subset of representations into memory activated the striatum. Baier at al. (Baier et al., 2010) conducted lesion-behaviour brain mapping in stroke patients and found that whilst PFC lesions reduced memory capacity, lesions of the putamen were associated with impaired filtering of irrelevant information. Voytek and Knight found that basal ganglia lesions impaired accuracy on a working memory task (Voytek & Knight, 2010). Chatham and Badre (2014) found that fronto-striatal coupling was increased following a retrocue to memory content as compared to following a precue for memory encoding[9]. This might be consistent with an output gating role for the BG-PFC interaction, but the core computational model motivating their analysis predicts that the

[9] They term these cues 'context cues' – a precue is a 'context first' cue and a retrocue is a 'context last' cue

basal ganglia should also interact with premotor cortex to mediate input gating (Hazy et al., 2007). Nee and Brown (Nee & Brown, 2013), using a variant of the A-X continuous performance task (AX-CPT) that sets up hierarchical task rules (Frank et al., 2001), found that only anterior PFC changed its coupling with basal ganglia following contextual cues, and then only following high-level contextual cues, not low-level contextual cues.

Although the anatomical argument (Middleton & Strick, 2000) for a role of BG in cognitive control is compelling, direct evidence for this role is currently incomplete. There is more robust evidence for a role of dopamine in working memory control. The computational model discussed above (Hazy et al., 2007) incorporates dopamine primarily as a teaching signal to basal ganglia. However, dopaminergic afferents from the ventral tegmental area (VTA) and substantia nigra pars reticula (SNpr) terminate diffusely in prefrontal cortex (Williams & Goldman-Rakic, 1993), as well as innervating the striatum. Browsowski demonstrated that depleting prefrontal dopamine was as effective as lesioning PFC at disrupting working memory performance in monkeys (Brozoski, Brown, Rosvold, & Goldman, 1979). Braver, Cohen and colleagues (Braver & Cohen, 1999; 2000) have suggested that phasic dopamine release may act as a gating signal for working memory. This is consistent with recent imaging work by D'Ardenne et al. (2012) showing that SNpr/VTA are activated following cues to update memory.

Behaviourally, both these models link dopaminergic signalling, which is related to reward processing (Schultz, 2013), with working memory updating. This link is investigated behaviourally in chapter 6.

1.5 Scope and structure of this thesis

This introduction has emphasised the role of potentially dissociable input and output gating mechanisms in control of working memory, yet the bulk of neural evidence suggests that common mechanisms mediate perceptual gating and mnemonic selection. The first experiment in this thesis was a large-scale magnetoencephalography (MEG) study exploring the cortical dynamics of preparatory and mnemonic selection, that aimed to clarify the role of different frontal control networks and of perceptual cortex in mediating selection for working memory, by exploiting the high spatial and temporal resolution of MEG.

Chapter 2 introduces magnetoencephalography (MEG). The first part of this chapter gives a theoretical introduction to the technique and associated analysis approaches. The second part documents specific analysis steps used in the work presented in chapters 4 and 5.

Chapter 3 presents a detailed analysis of behavioural data from the MEG study. Reaction times are decomposed using a diffusion model, demonstrating that both precues and retrocues excise a decision component from the RT. When considered alongside precue and retrocue effects on accuracy and misgating, these data strongly suggest that the retrocue benefit results mainly from a facilitation of output gating.

Chapter 4 investigates the effects of precues and retrocues on patterns of activity in sensory cortex. Spatiotopic modulation of alpha power in sensory cortex is a well-established correlate of preparatory attention orienting (van Ede, de Lange, Jensen, & Maris, 2011; Worden, Foxe, Wang, & Simpson, 2000). In chapter 4, retrocues and precues are shown to give rise to a very similar pattern of alpha lateralization. This predicts behavioural performance trial-by-trial. Decoding approaches reveal that this

attentional alpha power modulation is retinotopic and resembles the event-related desynchronization observed following a physical stimulus. Critically, the time-course of alpha power modulation differs between precues and retrocues: whilst precues give rise to a sustained lateralization of alpha power, retrocues give rise to a punctate effect. This constrains the functional interpretation of alpha lateralization following a retrocue, arguing against a modulation of tonic top-down modulation or a preparatory attentional effect, and suggesting instead that this effect reflects transient recruitment of sensory cortex associated with refreshing information in working memory.

Chapter 5 investigates networks mediating control over working memory. A meta-analysis of fMRI studies investigating predictive and retrodictive cueing is conducted to obtain spatial ROIs constraining the MEG source analysis. Frequency-domain responses at virtual electrodes corresponding to these spatial ROIs are extracted using a beamforming approach. The spatial patterns of brain activity over this net of virtual electrodes replicated the patterns of activation predicted by the fMRI meta-analysis, dissociating the fronto-parietal and cingulo-opercular networks. The MEG induced responses allowed a fine-grained characterization of the activation time-courses in each control site. The timecourse of activity in fronto-parietal sites, when compared with the time-course of alpha lateralization established in chapter 4, suggested a direct role for the fronto-parietal network in top-down biasing. By contrast, the frontal operculum and pre-SMA had dissociable time-courses, activating later in the epoch following retrocues only. This segregation suggests that these two networks might be more associated, respectively, with input and output stages in memory gating.

Chapter 6 presents three behavioural experiments that aimed to characterise the effects of reward associations on selection for working memory. As discussed in section 1.4.3,

above, both dopaminergic mechanisms and cortico-striatal interactions have been proposed to contribute to control over working memory. However, the behavioural implications of these models have been little investigated behaviourally – paradigms employing explicit selection cues have dominated the study of memory control (cf. chapters 3, 4 and 5).

The first two experiments presented in chapter 6 revealed that stimulus-reward associations had a robust effect on working memory performance, but surprisingly the encoding biases observed were not item-specific. Instead, they were more consistent with a temporal gating mechanism for working memory that is driven by reward associations. The third experiment employed rapid sequential stimulus presentation to investigate this possibility. The results of this experiment were consistent with the idea that perceiving an item with a high reward value opens a brief gating window for working memory. This mechanism may be coupled with sequential sampling of the environment (e.g. through eye movements) to mediate an item-specific encoding bias.

Chapter 7 is a general discussion of the implications of the work presented in this thesis, and of possible directions for future work. It discusses three key themes: (1) the role of perceptual cortex in working memory operations; (2) the roles and interactions of different frontal networks in coordinating memory operations, and (3) the need to integrate work on striatal and dopaminergic control over cognitive processes with cortico-centric views of working memory control.

2 Magnetoencephalography (MEG)

Chapter Abstract

This chapter introduces MEG, used in much of the work presented in this thesis. The first part discusses recording and interpreting the MEG signal, with a focus on the beamformer method for source reconstruction. The second part details pre-processing steps and gives and overview of the sensor-space data.

2.1 Recording the MEG signal

Electrical activity in neurons generates measurable changes in the electromagnetic fields in and around the brain. There are currently two major non-invasive techniques in human cognitive neuroscience that can provide data on neural processes with millisecond temporal resolution: electroencephalography (EEG) and magnetoencephalography (MEG). EEG measures electrical potentials at the scalp, whereas MEG measures changes in the magnetic field around the head. This section discusses the acquisition and analysis of the MEG signal.

The magnetic signal generated by the brain is very small relative to magnetic noise in the environment (Figure 2-1, panel A). In order to record this signal, extremely sensitive superconducting quantum interference device (SQUID) detectors are coupled with an environment physically shielded from magnetic noise (the magnetically shielded room, MSR). The MEG system used to record the data presented in this thesis was an Elekta Neuromag system, with 306 SQUIDs. The typical setup is photographed in Figure 2-1, panel D.

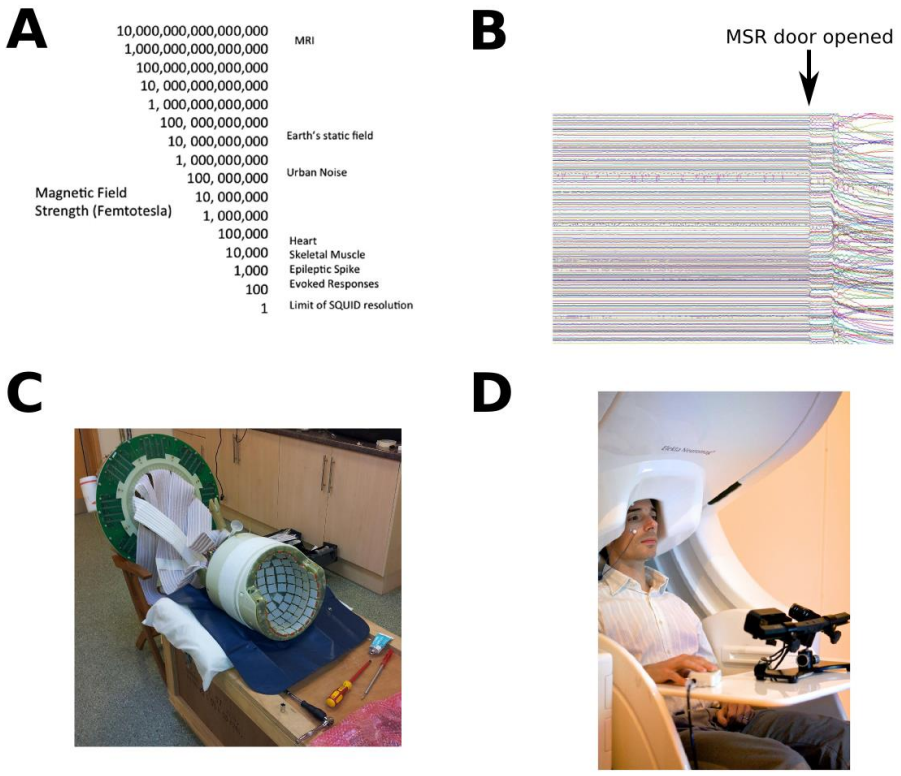


Figure 2-1 **The MEG setup.** **A** – Magnetic signals from the brain are very weak relative to ambient fields: around 8 orders of magnitude weaker than the earth's magnetic field. **B** – The importance of physical shielding. Opening the door to the magnetically shielded room (MSR) during recording disrupts the signal. **C** – The MEG sensors from OHBA's Elekta system removed for maintenance. Compare with Figure 2-2. The sensors are normally submerged in liquid helium at ~4 Kelvin, to keep the SQUID components superconductive. **D** – Dr. Malcolm Proudfoot of OHBA models a typical MEG setup. The subject is seated with their head resting in the 'cap' moulded into the bottom of the liquid helium dewar. An eyetracker placed on a table in front of the subject monitors eye movements. The vertical EOG electrodes can be seen attached above and below the right eye. *Panels A, B and D adapted with permission from Proudfoot et al., 2014.*

Each SQUID in the system is coupled with the magnetic field near the subject's head via a flux transformer (Figure 2-2) – a superconducting loop of wire. A small coil apposed to the SQUID couples the SQUID to the flux transformer; this is termed the signal coil. Further coils (termed pickup coils) are responsible for sampling the magnetic field. Current is generated when there is a magnetic field component 'threading' these pickup coils (Figure 2-2). Because the coils super-conduct, even a static magnetic field will generate current. The geometry of the pickup coils determines the features of the magnetic field to which the SQUID is sensitive. The Elekta system houses two sensor types: 102 magnetometers and 204 planar gradiometers. The magnetometer pickup

coil is a single flat turn, measuring the field component orthogonal to the scalp. The gradiometer pickup coils are wound as a flat figure of eight, measuring the difference between the orthogonal component threading the two halves: i.e. the spatial gradient in the field component orthogonal to the scalp, along a spatial axis tangential to the scalp. Each sensor location has two planar gradiometers, the figure-eight loops oriented at right angles to one another, so together they fully sample the spatial gradient in the plane tangential to the scalp.

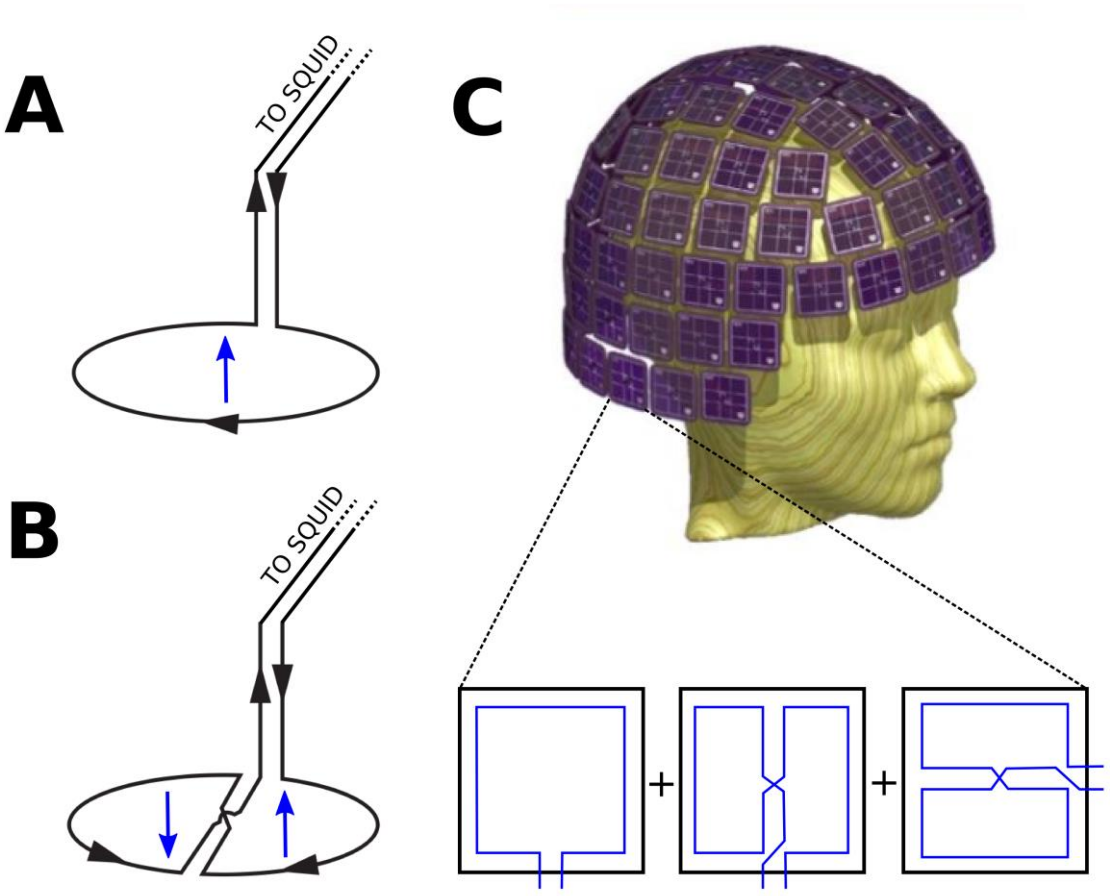


Figure 2-2 (**A**) flux transfer geometry for a magnetometer. Blue arrow indicates threading field. (**B**) flux transfer geometry for a planar gradiometer (**C**) arrangement of sensor chips in the Elekta Neuromag system.

Modelling a hypothetical brain source as a current dipole, its field projection can be visualised. A source dipole oriented tangentially to the scalp (Figure 2-3, panel **A**) will

produce a ring field in the plane orthogonal to the scalp (Figure 2-3, panel **B**). The locations with the strongest orthogonal components to the field (activating magnetometers) are indicated with magenta arrows, and the location with the strongest gradient of the orthogonal component of the field (activating planar gradiometers) with a green arrow. Planar gradiometers are activated by dipoles underneath the sensor, and this activation will depend on the orientation of the dipole relative to the orientation of the gradiometer (green arrow, Figure 2-3). A dipole activates magnetometers in adjacent locations, with positive responses to one side of the dipole, and negative responses to the other (magenta arrows, Figure 2-3). In contrast, a dipole oriented orthogonally to the scalp (Figure 2-3, panel **C**) produces a ring field in a plane tangential to the scalp that activates neither sensor type (Figure 2-3, panel **D**).

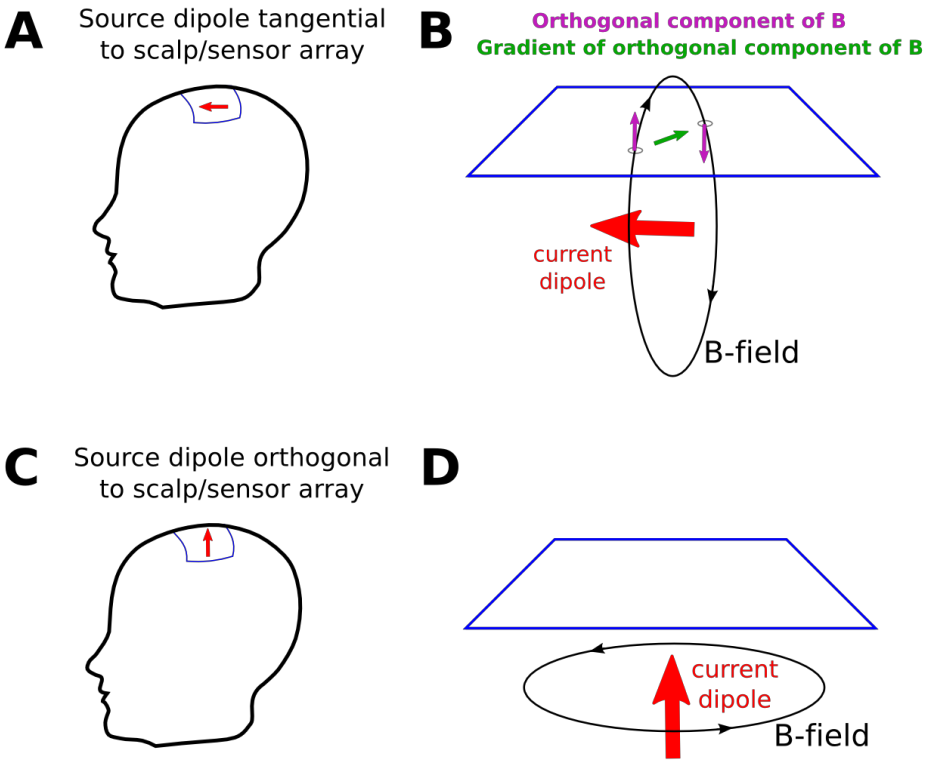


Figure 2-3 **A simple schematic showing the relation between a source dipole and activity in the MEG sensor array.** See Figure 2-4 for a more detailed depiction of the lead fields for each sensor type. **A, C** Current dipoles oriented tangentially and orthogonally to the scalp/sensor array. **B, D** Magnetic fields associated with tangential and orthogonally oriented current dipoles, intersecting the plane of the sensor array.

Figure 2-4, below, is adapted from Malmivuo et al. (1997) and shows a cut through the lead field of a magnetometer (panel **A**) and a planar gradiometer (panel **B**) in an idealized spherical head model. Due to Helmholtz's reciprocity theorem, the magnetic field produced by passing a changing electrical current through the sensor coil is identical with the sensitivity of that sensor to dipolar sources[10]. The plots in Figure 2-4 were generated by modelling the magnetic field produced by current flow through the MEG pick-up coils. The region of maximum sensitivity is shaded in green. The greater sensitivity for more superficial sources is apparent for both sensor types.

[10] The sensitivity patterns of the pick-up coils are therefore very similar to the magnetic fields produced by equivalently-wound TMS coils (other than the difference in scale).

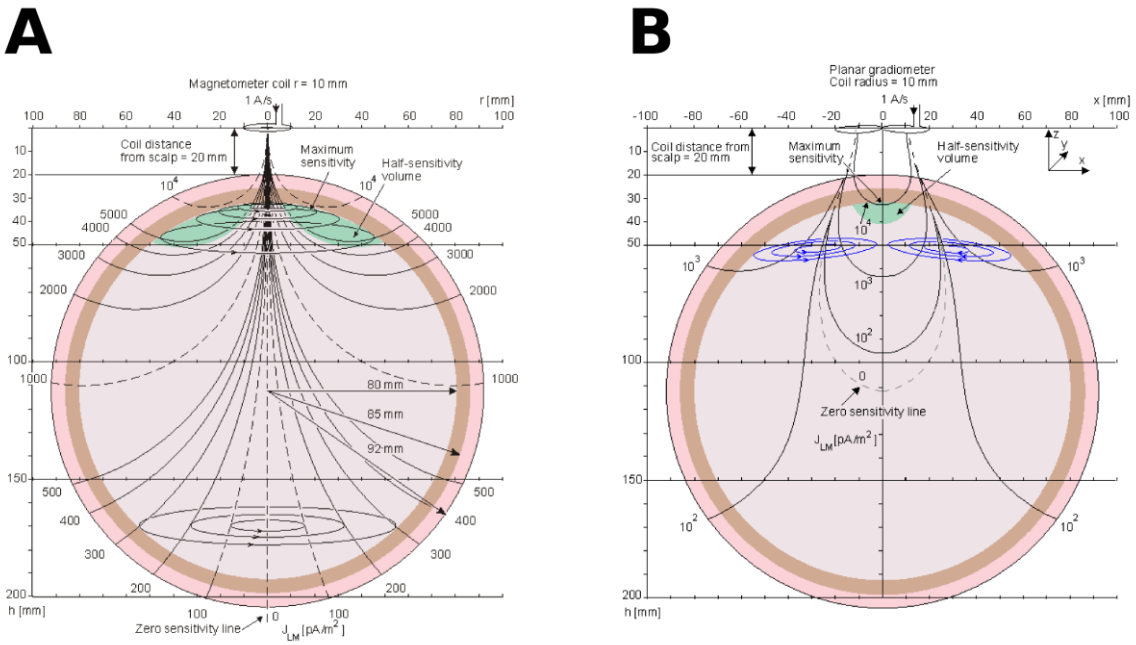


Figure 2-4 **Lead fields for a magnetometer (A) and a planar gradiometer (B).** Fields are modelled in an idealised 'spherical head'. *Figure adapted from Malmivuo et al., 1997*

Individual planar gradiometer responses depend on the orientation of the source dipole. A synthetic sensor that is insensitive to the source dipole orientation can be generated by taking the Cartesian sum of the signal from the two gradiometers mounted on the same chip. This gives a field map sampled at 102 locations, in which source dipoles are roughly underneath the synthetic sensors. Dipole orientation for a given cortical source may well differ across subjects (due to differences in cortical anatomy, or head position in the scanner), but the insensitivity of the combined gradiometer to source orientation means that the combined gradiometer map can be averaged across subjects. Differences in single-subject head position relative to the sensor array will still 'smear' these group maps, but they give a sense of the pattern of activity without having to apply source reconstruction algorithms. For example, the group-level sensor space topography of a visual ERF is shown in Figure 2-5, spatially interpolated and smoothed across the sensors. The activation is strongest in combined gradiometers over the parietal/occipital cortex.

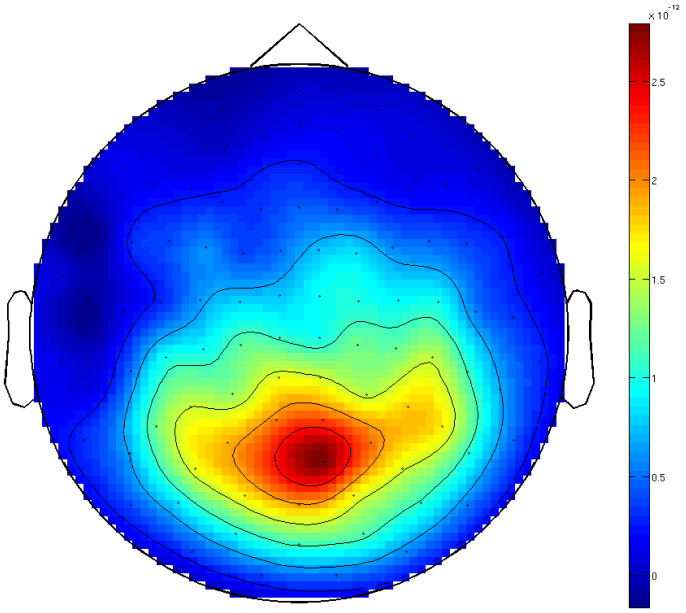


Figure 2-5 A group-averaged visual ERF, plotted over the combined gradiometers.

2.2 Neural generators of the M/EEG signal

The two main forms of neuronal electrical activity are the fast axonal action potentials that convey signals between neurons, and the dendritic potentials that arise from the summation of post-synaptic potentials. For two reasons, the dendritic potentials are thought to be the main generators of the magnetic field measurable outside of the head. Firstly, action potentials are short-lived, whereas post-synaptic potentials last an order of magnitude longer (from ~10ms to ~100ms). Post-synaptic potentials are therefore more likely to overlap in time, allowing their magnetic field to summate. Secondly, action potentials (APs) involve spatially adjacent dipoles of opposite orientation: ahead of the propagating AP, current flows in a retrograde direction inside the axon, and in a forward direction in the extracellular medium; behind the AP the directions of current flow are reversed. These dipoles with opposite orientation produce magnetic fields that will cancel when they are summated. In contrast, the current in the dendrites resulting from post-synaptic activity tends to flow in one direction, so gives rise to a single dipole.

The direction of the current dipole within a pyramidal cell varies depending on the location of stimulation of the dendrite, and whether the post-synaptic potential is excitatory (an EPSP) or inhibitory (an IPSP). An active current sink occurs at an excitatory synapse, and an active source at an inhibitory synapse. Neurons are embedded in a conductive medium (the extracellular fluid), and the synaptic source/sink will generate passive currents in this medium. The magnetic field generated is attributable to the flow of current inside of the dendritic axis, because the return currents in the extracellular medium tend to cancel.

The magnetic field produced by any given neuron is too small to be detected outside of the head: signals from multiple neurons must summate in order to generate a measurable signal. The detectability of neuromagnetic signals is therefore determined by whether or not fields from different neurons summate additively, which depends upon the geometry of the neuronal sources. The post-synaptic potentials that form the dominant contribution to the measured field outside of the head form a dipole along the dendritic axis. In order to summate, the dipoles from multiple neurons must be co-aligned in space (dendritic axes must be co-aligned). This is true for the pyramidal neurons of the cortex, the apical dendrites of which are aligned radially to the cortical surface. This organization of dipole generators, leading to linear summation of the B-field, is termed 'open-field'. Disorganized neuronal structures, or structures that are organized but with a more complex structure such that they have a 'closed' field, do not contribute to the magnetic field outside of the scalp as the fields due to individual neurons will tend to cancel.

Two factors affect the detectability of cortical sources: distance from the sensors, and dipole orientation. Since the magnetic field falls off rapidly with distance from the generator, neuronal activity in parts of cortex close to the scalp is more detectable than activity in deep brain regions. Post-synaptic currents in the apical dendrites of cortical pyramidal neurons are therefore likely to dominate the MEG signal, as these neurons are superficial, and have an open field organization. Each neuron contributes a source dipole aligned with its apical dendrite. As shown in Figure 2-3, neurons oriented orthogonally to the scalp surface will produce fields that are parallel to the scalp and therefore have no effect on the MEG sensors. Neurons oriented parallel to the scalp surface produce ring fields that are orthogonal to the scalp, and will therefore cause flux changes in the MEG sensors (Figure 2-3). Neurons lying in the walls of the cortical sulci are therefore more detectable to MEG than neurons lying on the apex of a gyrus or the

bottom of a sulcus. The insensitivity of MEG to gyral sources may be minor relative to the effect of source depth. A modelling study by Hillebrand and Barnes (2002) using simulations to estimate detectability of sources at different cortical locations found that there is only a narrow gyral strip of ~2mm width that is poorly detectable, and that gyral sources are otherwise readily detectable due to their proximity to the scalp. Groups of neurons whose activity will contribute less, on account of their disorganized structures, include the interneurons of the cortex, and neurons in subcortical nuclei - though, there have been reports attributing MEG activity to subcortical sources; e.g. thalamus (Roux, Wibral, Singer, Aru, & Uhlhaas, 2013)

An excitatory post-synaptic potential delivered to the dendritic arbour of a pyramidal cell would generate a dendritic current flowing from the upper to the lower layers of the cortex, but an IPSP at the soma would generate the same direction of current flow. Current flow from the deep to shallow layers (i.e. retrograde in the apical dendrite) is also possible. The MEG signal could therefore potentially represent either excitatory or inhibitory afferent input to the apical dendrites of the pyramidal cells. Either way, the current source underlying the MEG signal must be synchronized between neurons. The current dipole of a cortical pyramidal cell is of the order 0.2pAm and the weakest current dipoles we can measure with MEG are of order 10nAm, so around 50,000 cells must be active to generate a detectable signal. These neurons must co-activate within a small enough time window (determined by the timescale of the post-synaptic potentials) that their magnetic fields summate.

2.3 Source space and the inverse problem

2.3.1 Beamforming

If we knew the configuration of currents in the brain, we would be able to straightforwardly predict the activations in the MEG sensors. This *forward model* depends on a model of the head conductance, and the sensitivity of the sensors to the magnetic field. This is a well-posed problem, in that the solution is unambiguous: if we improved our predictive model we would be able to improve our prediction to an arbitrary precision. The forward model corresponds to the projection patterns pictured in Figure 2-3, and can be captured mathematically in a set of 'lead fields' for each sensor:

$$\boldsymbol{B} = \boldsymbol{LQ}$$

Equation 1

Here, **B** is an [M x 1] vector of activations in the M MEG sensors, and the [Nx1] matrix **Q** is the strength of activity at each of N locations in the brain[11]. **L** is the lead field matrix, an [M x N] array in which each column constitutes the lead field for a given sensor, for the brain locations in **Q**.

In contrast to the forward model, taking a set of activations in the MEG sensors and reconstructing the brain activations that generated the signal is an ill-posed problem, meaning that there is no unique solution. There are an infinite number of current configurations in the three-dimensional brain space that could give rise to any given

[11] For example, at each location we might model an equivalent source current dipole. The matrix **Q,** with a single entry for each location, assumes the orientation of the source is known (e.g. constrained to be orthogonal to the cortical surface). In practice there are various ways of dealing with source dipole orientation, which are discussed elsewhere in the text.

activation pattern in the two-dimensional space of the MEG sensors. This ill-posed problem is the one we need to solve in order to map MEG signals to their cortical generators. The problem is analogous to the one we face when viewing a film on screen, when a 3D scene is compressed into a 2D representation. Our visual system easily solves the inverse problem by making assumptions that constrain the reconstruction of the 3D scene: e.g. the typical size of known objects in the scene constrains scale and depth. The same approach is used to solve the MEG inverse problem: reasonable assumptions are chosen in order to constrain the inverse solution. Our brain likely uses a hierarchical generative model to reconstruct relevant information about the outside world from sensory input (Friston, 2009; R. P. Rao & Ballard, 1999). The assumptions constraining the reconstruction in this case consist in the structure of the world-model. Although generative model approaches are being developed for MEG analysis (David et al., 2006; Kiebel, Garrido, Moran, Chen, & Friston, 2009), and may in the future become the go-to method, the technique used in the work presented here was more basic. Beamforming, a method borrowed from radar (Van Veen & Buckley, 1988), constructs spatial filters that aim to pass activity from a target location in brain space, and block activity from elsewhere (Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1998). This is done independently for each point on a grid spanning the brain space.

The key assumption in the beamformer approach is that sources in the brain produced uncorrelated time-series, i.e. that the covariance matrix for the set of brain sources is zero off-diagonal. Mosher (2003) gives the general form of a linear inverse solution (compare with Equation 1, the forward model) for source currents **Q** with lead field **L,** sensor measurements **B,** and covariance matrices **C:**

$$\mathbf{Q} = \mathbf{C_j L^T C_b^{-1} B}$$

Equation 2

There are two covariance matrices: $\mathbf{C}_j$, the source current covariance, and $\mathbf{C}_b$, the sensor data covariance. As stated above, the beamformer assumes that sources are uncorrelated: i.e, off-diagonal elements in the source current covariance matrix are zero. Given this assumption, we only need to compute the diagonal entries for $\mathbf{C}_j$ – i.e., the variance at each source. Denoting a given source θ, the corresponding diagonal element can be computed from the sensor covariance (Mosher et al., 2003):

$$\sigma_\theta^2 = (\mathbf{L}_\theta^T \mathbf{C}_b^{-1} \mathbf{L}_\theta)^{-1}$$

Equation 3

Substituting this expression into Equation 2 gives the beamformer solution for activity at location θ:

$$Q_\theta = (\mathbf{L}_\theta^T \mathbf{C}_b^{-1} \mathbf{L}_\theta)^{-1} \mathbf{L}_\theta^T \mathbf{C}_b^{-1} \mathbf{B} = \mathbf{W}_\theta^T \mathbf{B}$$

Equation 4

An [M x 1] (where M is the number of MEG sensors) set of weights **W** is derived for each point in source space. The weights depend only on the forward model, and the covariance of the sensor data. The dot product of $\mathbf{W}^T$ for a given source with the sensor data **B** gives an estimate of the activity at that source at each point in time.

A key theoretical limitation of the beamformer is the assumption that sources are uncorrelated. Practical challenges include estimating the sensor covariance matrix,

specifying the forward model **L**, deciding how to spatially sample the brain space, and combining data over sessions and subjects.

2.3.2 The correlated sources assumption

The beamformer relies on the assumption that the brain sources giving rise to the MEG signal are uncorrelated. To what degree is the assumption valid, and what will happen when it is violated?

Van Veen et al. (1998) modelled situations in which the beamformer algorithm was used to detect activity from two sources, which in different simulations were nearby or distant, and correlated or uncorrelated. When sources were uncorrelated, the beamformer was able to discriminate two source peaks even when they were fairly close (4cm distant)[12]. For fully correlated, spatially separated sources, there is a cancellation effect: neither source is detected. As the correlated sources are moved closer together in space, this cancellation switches to merging: a single source is detected midway between the two true sources. Partly correlated sources lead to partial cancellation/partial merging. The 'merging' behaviour for nearby sources is beneficial: within a given brain region (i.e. within a cortical area of a few millimetres square) neurons probably do generate a correlated signal (the LFP), but for these millimetre scale local correlations the sources will be merged, not cancelled (otherwise, the beamformer would cancel all brain activity).

Given some signal remains (partial cancellation) when sources are partially correlated, the beamformer can be considered robust to modest violations of the uncorrelated

[12] This simulation study was geared to beamforming for EEG. Spatial resolution for uncorrelated sources in MEG is potentially rather better than this: Hillebrand and Barnes (Hillebrand & Barnes, 2005) suggest that in a realistic setup, the beamformer can potentially discriminate between uncorrelated sources separated by less than half a millimetre.

sources assumption (for distant sources). Hillebrand and Barnes (Hillebrand & Barnes, 2005) argue that fully correlated activity in separated brain regions is implausible, given that local connectivity dominates in the brain. However, this argument ignores the possibility that a thalamic pacemaker (for example) might coordinate activity across distant brain regions (Womelsdorf et al., 2007). It is unclear whether this inherent coherence would be strong enough to cause source cancellation. As an empirical reference point, there certainly are *stimulus-driven* brain-states in which activity is sufficiently correlated in distant brain areas that the beamformer fails. Cyclic amplitude-modulated auditory stimuli give rise to a correlated steady state evoked response in left and right auditory cortex, which will cancel using a standard beamformer (Brookes et al., 2007). It is possible that evoked responses, which briefly synchronize activity in perceptual cortex, have a similar effect. Whilst cancellation is likely for distant correlated sources (which reduces detectability of true sources, but will not give rise to spurious activations), nearby sources may merge. Evoked responses appear to generate particularly strong activity on the midline, which would be consistent with merging of lateralized activations that are highly correlated.

To summarize, for induced-response and evoked-response analyses, the uncorrelated sources assumption may occasionally reduce sensitivity or merge separated sources where distant sources are correlated, but the consensus is that these limitations are outweighed by the potentially high spatial resolution of the beamformer approach (Hillebrand & Barnes, 2005). It has also been argued that the beamformer is an appropriate source reconstruction methodology for connectivity analyses that rely on phase synchronization between spatially separated brain regions (Hadjipapas, Hillebrand, Holliday, Singh, & Barnes, 2005), an application for which the ability to detect correlated sources is particularly important. However, this only holds when the duration of these interactions is short relative to the length of data over which the

beamformer solution is estimated (i.e. the sites are correlated over <30% of the recorded time-points).

2.3.3 Estimating data covariance

As outlined above, the beamformer relies on an estimate of the data covariance at the sensor level, $\mathbf{C}_b$. In estimating the covariance matrix, there is a trade-off between having sufficient data to obtain a robust estimate, and the fact the covariance estimate will differ for different frequency bands and different states of brain activity. In order to best meet the inherent assumption that the covariance is stationary, it is best to estimate the covariance over as targeted a temporal and spectral window as possible (Gross et al., 2013).

The problem of estimating a large covariance matrix from limited data can be ameliorated by regularizing the covariance matrix. This involves amplifying the magnitude of the diagonal entries of the covariance matrix relative to the off-diagonal entries by some regularization coefficient. However, because the off-diagonal entries confer the high spatial resolution of the beamformer, this reduces spatial resolution. In the limit that only the diagonal elements of the covariance are used, the beamformer solution becomes a dipole fit (Hillebrand & Barnes, 2005). The strategy used in the analyses presented elsewhere in this thesis was to allow the epoching to dictate the temporal window over which the covariance was estimated – the trials were divided into epochs of ~2s length around trial events. Spectral windowing was achieved by breaking the analysis down into two broad bands: a low frequency band between 3 and 30Hz, and a high frequency band between 30 and 120Hz. The covariance matrix was then regularized using the top 59 PCA components of the data (Woolrich, Hunt, Groves, & Barnes, 2011).

2.3.4 The forward model

An important part of the beamformer algorithm is the forward model of source dipole projections onto the sensor space, captured in the lead field matrix **L.** This projection depends on the geometry and conductivity of the brain, skull and scalp. The most basic forward model assumes a spherical head with a single conductive shell (Salmelin, Kringelbach, & Hansen, 2010). This single sphere model can be improved by adding spherical harmonics to make the head-shape more realistic (Nolte, 2003). This single shell model can be tailored to the head-shape of individual subjects, based on their structural MRI scan.

At each grid point, a source dipole could have any given orientation. Source orientation can be accounted for by modelling three orthogonally oriented dipoles at each grid point (so the lead field for a given sensor has 3*N entries, where N is the number of grid points used in the source space analysis). However, the single shell approach used here constrained the dipole orientations to the plane tangential to the MEG sensors (as radially oriented sources will not be detectable). PCA was then used to extract a scalar time-course for each grid-point.

2.3.5 Spatial sampling

The dimensionality of the raw MEG data (~64, due to redundancy between the 306 sensors) is low relative to the number of points in even a coarse grid spanning the brain space (~3500 grid-points for an 8mm grid), which might suggest that the source space map should be fairly smooth. This is not necessarily the case as the spatial resolution is not homogeneous in source space, and depends on source amplitude (Barnes, Hillebrand, Fawcett, & Singh, 2004). Barnes et al. (2004) characterized the smoothness

of the beamformer solution in source space for real MEG data, based on the correlation between the beamformer weights for adjacent locations on the source space grid. They reported this smoothness as the full-width half-maximum (FWHM) of an equivalent Gaussian point spread function. They discovered that the FWHM, for realistic data, could easily be as low as 1-2mm.

This is substantially finer than the spatial sampling of the source space grid in a typical analysis. For computational reasons (the number of source dipoles increasing as the inverse cube of the sampling resolution), in whole-brain analyses, this is typically not made much finer than 6mm. The source-space grid-step used in the analyses presented here was 8mm. This could in principle under-sample the source space, and fail to pick up on some activity. However, regularizing the beamformer covariance estimate will increase the smoothness of the source space solution. Barnes et al. (2004) deliberately used no regularization in order to estimate the maximum realistic spatial resolution. For the analyses presented in the current thesis, the covariance was regularized based on the top 59 PCA components of the data, which will smooth the source space solution.

2.3.6 Subject-level and group-level statistics

For the work presented in the current thesis, each subject completed three recording sessions. The beamformer weights are computed separately for each session. This places some constraints on the analysis pipeline. An important limitation is that each source-space time-course has a valence that depends on the orientation of the source dipole. For a given dipole, two exactly opposite orientations are always equally plausible, and the extracted signal for the two orientations will be of opposite sign. This means that the time-domain signal may well be phase-reversed across adjacent

voxels in one beamformer solution, or at the same voxel under different beamformer solutions. This precludes direct averaging of the source-space time-domain signals across sessions, as the arbitrary phase reversals will cancel. This makes the computation of source space ERFs troublesome. A solution is to rectify the data. In the GLM framework, this can be done either before the computation of contrasts ('COAPE' – contrast of absolute parameter estimates) or after ('ACOPE' – absolute contrast of parameter estimates). This facilitates averaging, but the resulting all-positive ERFs obscure the characteristic ERP positivities/negativities familiar from EEG.

For time-frequency analyses, the solution is to perform the time-frequency transformation separately for data from each beamformer solution. The resulting power time-course is inherently rectified and can be averaged across subjects. Phase-based spectral analyses, such as coherence or phase-locking value, must still be computed within session, and then averaged at the subject and group level, as these are sensitive to phase reversals.

Parametric statistics are computed at the group level, e.g. by averaging and generating t-statistics, or by modelling the data using a GLM. However, the averaged signal can often be non-gaussian (e.g. the heavy-tailed, all positive time-frequency data). More problematically, time-frequency data are very high dimensional, and the multiple-comparisons problem is pervasive in MEG analysis (Gross et al., 2013). To deal with these two problems, statistical inference was based on comparing true data to permutation distributions, usually defined over a clustering measure in some combination of time/frequency/brain space (Maris & Oostenveld, 2007).

2.4 MEG analysis

2.4.1 Preprocessing pipeline

The MEG system has to be highly sensitive to record a brain signal many orders of magnitude weaker than the earth's magnetic field (Figure 2-1, panel A). This also renders it sensitive to artefactual noise, defined as that part of the MEG signal not generated by brain activity. In particular, muscle activity associated with eye-blinks, eye movements, and neck movements can contaminate the signal. The system is also prone to sensor glitches (for example, an air bubble under a SQUID can corrupt the signal from a channel). The aim of the preprocessing pipeline is to isolate the brain signal from the artefactual signals. The pre-processing pipeline is schematized in Figure 2-6.

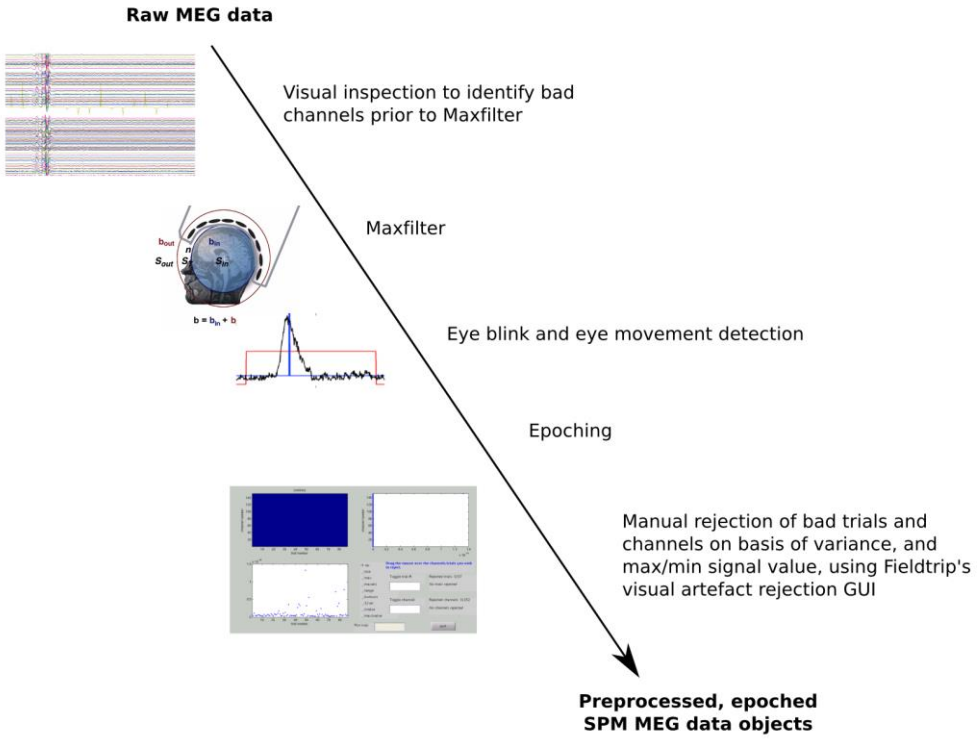


Figure 2-6 **Preprocessing pipeline**

2.4.1.1 Artefact rejection: Signal Space Separation

Neuromag's Maxfilter software implements an algorithm (Maxwell filtering) for spatial denoising by signal space separation (Taulu, Kajola, & Simola, 2004). The signal from the sensors is mathematically decomposed into a set of basis functions which have different coefficients for signals arising from within a sphere centred on the sensor array (i.e. brain signal), from an intermediate hollow sphere encompassing the sensors, and from the space outside an outer sphere (i.e. environmental noise). These spaces are illustrated in Figure 2-7. By retaining only the inner-sphere signal components, the MEG signal can be de-noised. This de-noised signal is then projected back onto the sensors to produce de-noised sensor space data for further analysis. Head movements can be compensated for by transforming the basis set with respect to the sensor array, before re-projecting the signal. The data presented in this thesis were movement corrected within-session on the basis of the continuous head position signal recorded from the head-coils (HPI coils), but were not movement compensated between sessions or subjects. Note that this will lead to 'smearing' of the sensor space signal when averaging across sessions/subjects, but that source space analyses take the head position into account as part of the forward model calculation, and are therefore able to correct for differences in head position between sessions and subjects.

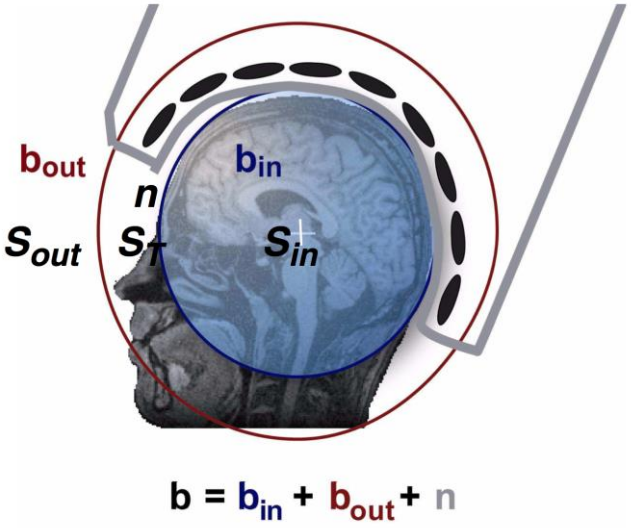


Figure 2-7 **Spatial filtering using the Maxfilter.** The signal is decomposed into $\mathbf{b}_{in}$, the signal arising from an inner sphere encompassing the brain, and $\mathbf{b}_{out}$, the signal arising from the space outside a larger sphere surrounding the sensors. Only $\mathbf{b}_{in}$ is retained in the filtered signal. A third intermediate space lies between the two spheres. If a channel is associated with a substantial signal projection from this intermediate space, this suggests a glitch in the sensor. This allows Maxfilter to automatically detect bad channels. *Figure modified from the Eleta Neuromag Maxfilter 2.1 user's manual.*

The Maxfilter algorithm is not robust to severe sensor artefacts, particularly those introduced by the signal processing hardware in the MEG system. These characteristic square-wave or saw-tooth artefacts often contaminate a few channels, but will propagate to the rest of the channels when the Maxfilter is applied. The first step in preprocessing (Figure 2-6) was therefore to visually inspect the raw MEG data and tag bad channels before Maxfiltering.

2.4.1.2 Artefact rejection: eye blinks

Subjects were asked to blink between but not during trials whilst performing the cognitive task in the MEG, but most subjects failed to completely suppress blinks during trials. Blinks were more likely to occur in the retention interval between cues than during cue presentation. The probability of a blink within each task epoch, broken down by task condition, is shown averaged across all subjects in Figure 2-8.

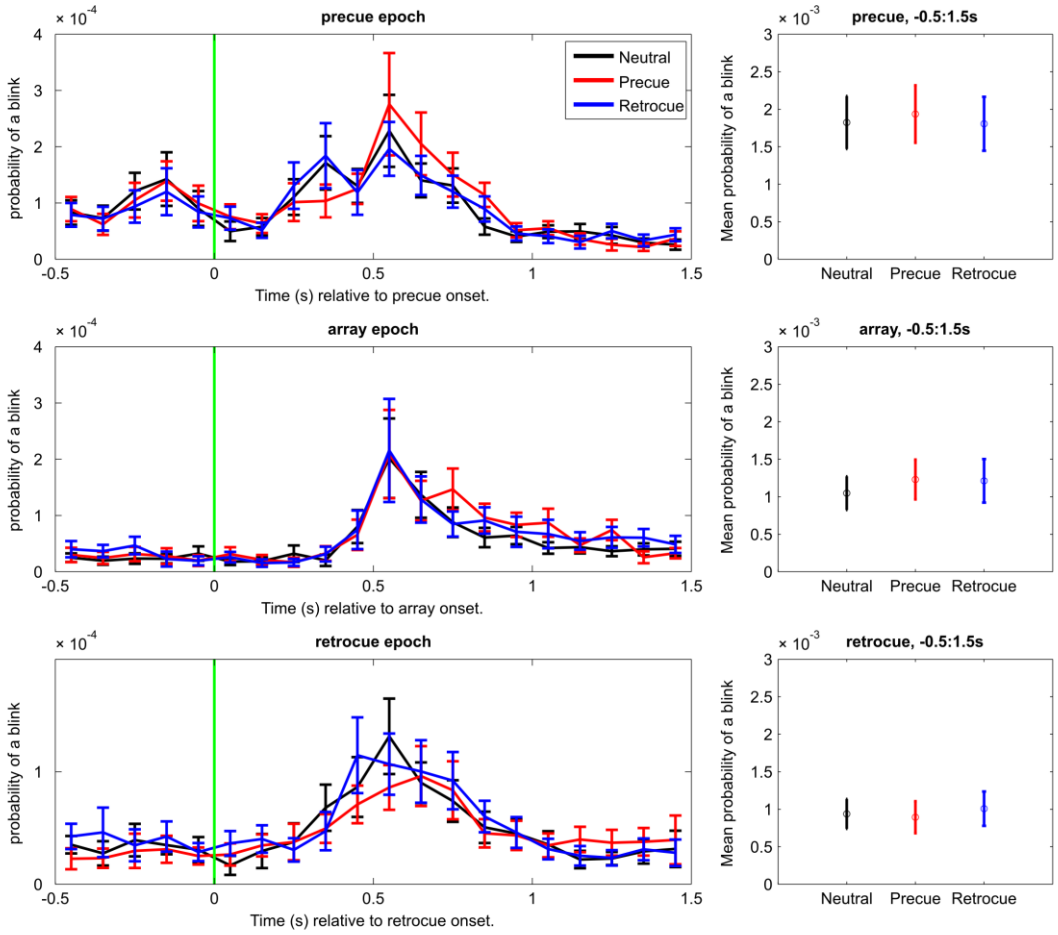


Figure 2-8 **The probability of a blink in each of the three task epochs.** Subjects were more likely to blink in periods when no cues or stimuli were presented. The probability of blinking did not differ depending on the cue condition for that trial.

Eyeblinks introduce artefact, as the muscles involved produce a magnetic signal large relative to the signals produced by the brain, and they also induce a burst of alpha activity in the visual cortex. One approach to the problem of signal contamination by eyeblinks would be to reject any trial containing an eyeblink, but this may be impractical if it results in discarding large quantities of data, as is the case with the MEG dataset discussed in this thesis. An alternative approach is to find and unmix eyeblink components using ICA. However, eyeblink components can split across ICA components (for example if the subject moves during a scan, the change in the pattern of activity at the sensors before and after the movement can force ICA components to split

to accommodate the change in topography). This method is therefore not guaranteed to remove eyeblink artefact. However, the sensor topography of the eyeblink component (as well as the time-course) is stereotyped, and eyeblink IC time-courses were extracted for each subject as part of preprocessing by manually identifying the component – this is a useful backup if the EOG and/or eyetracker data are noisy.

A simpler approach is simply to identify periods of time in which there was an eyeblink, and not include that data in the analysis. Blinks were detected by taking the first principal component of the vertical EOG, eyetracker dropout, and ICA eyeblink time-course, to make a single eyeblink vector robust to noise in any one signal. This was then convolved with a 50ms standard deviation Gaussian kernel (approximating a matched filter for a blink), and the second derivative of the smoothed signal was thresholded by hand to identify blink times. A window was then defined around each blink, from 200ms before to 300ms after the peak. If two blinks were closer than 1000ms apart, the window was extended to include them both. An example segment of raw (unsmoothed) blink trace and the corresponding time windows from which data were rejected is shown in Figure 2-9.

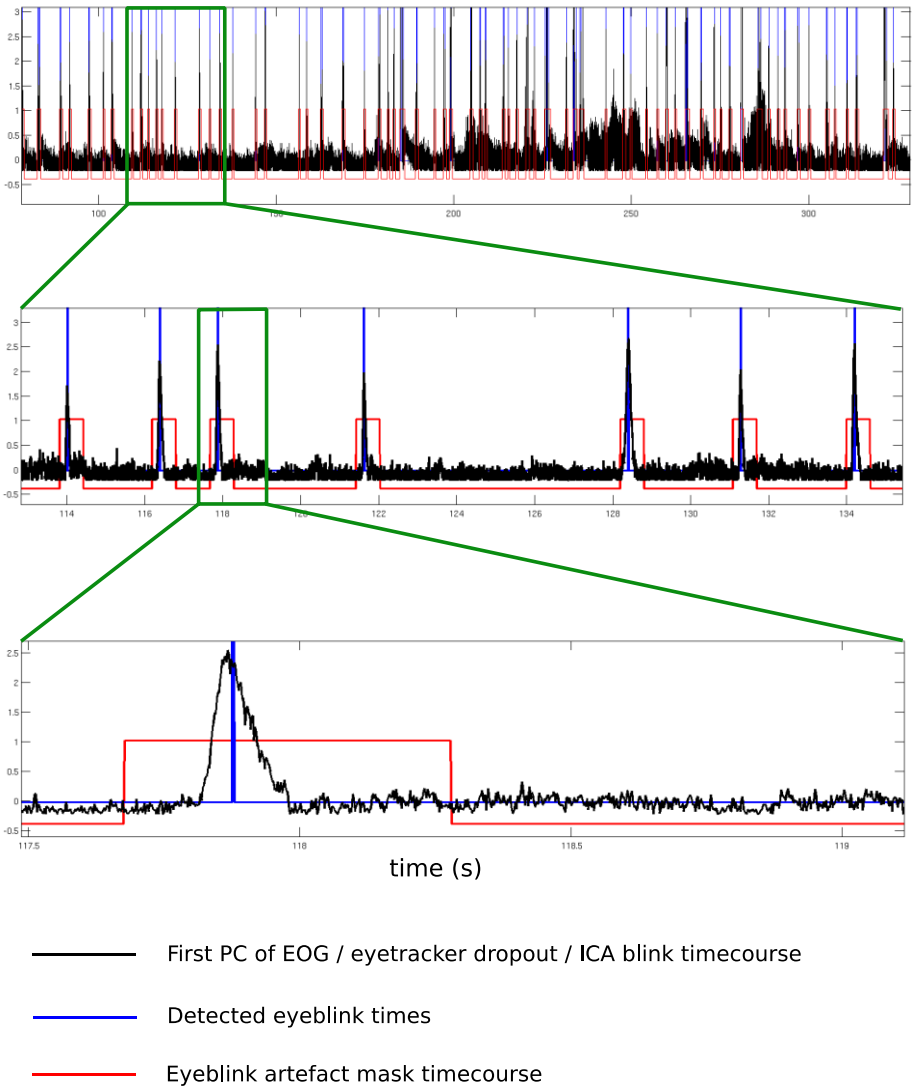


Figure 2-9 **Detecting and removing eyeblinks**. A semi-automated algorithm was used to detect times at which a blink occurred. A masking time-course was created based on this information, and data in a window 200ms prior to until 300ms following each blink was excluded from the analysis (including calculation of beamformer weights)

2.4.1.3 Artefact rejection: eye movements

Muscle activations associated with eye movements contribute artefact to the MEG signal. The eyetracker traces were inspected for eye movements. Most subjects successfully maintained fixation during trials, but a subset of subjects made saccades during trials. Within-subject, the pattern was consistent, a given subject either maintaining fixation on

all trials, or making regular saccades. Four subjects making regular saccades were excluded from the analysis.

2.4.2 Overview of task data

This section provides a general overview of the task data, as a reference for chapters 4 and 5. Data were averaged over the combined planar gradiometer array (see section 2.3.6, above). The sensors were divided into three pools: a posterior pool (corresponding roughly to occipital/parietal cortex), a central pool, and a frontal pool.

2.4.2.1 Evoked responses

Group-averaged evoked responses for each cue condition are shown in Figure 2-10, for each of the three pools of combined planar gradiometers. Note that the planar gradiometer trace is inherently rectified, so these traces do not resemble ERPs as recorded using EEG. Informative cues are associated with a substantial boost to the ERF, as compared to their neutral-cue counterparts (neutral cues were white squares). The signal is much stronger in the posterior sensor group than the frontal group. There are three likely reasons for this:

1. Occipital/parietal cortex is physically closer to the sensor array than is frontal cortex, using the typical MEG configuration, so the signal from cortex is stronger relative to the ambient noise
2. Differences in head position between subjects (especially rotational shifts) will cause a greater alteration of the signal pattern in frontal than in posterior sensors. Subjects rest the back of the scalp against the dewar cap, so the spatial variability between subjects is least for the parieto-occipital sensors.
3. Responses in occipital/parietal cortex are likely to be more tightly phase-locked to the stimulus than those in frontal cortex which will have a more variable

latency. The ERF depends on phase-locked responses and out-of-phase signals will cancel.

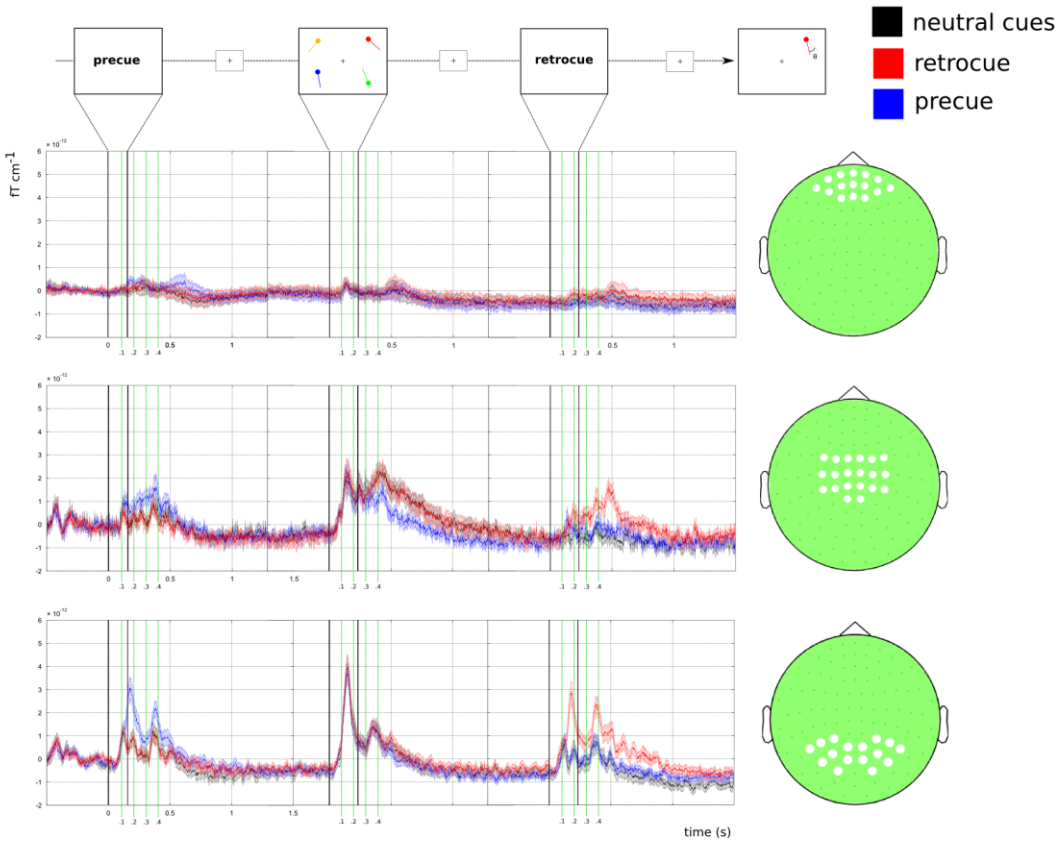


Figure 2-10 **Overview of the event-related field (ERF) to task events.** Informative cues are associated with a substantial boost to the ERF, as compared to their neutral-cue counterparts (neutral cues were white squares).

2.4.2.2 Induced responses

The sensor-space data were transformed to the time-frequency domain using a Hanning taper analysis spanning 4 cycles of the underlying frequency. The resulting time-frequency maps are shown below for each sensor group separately.

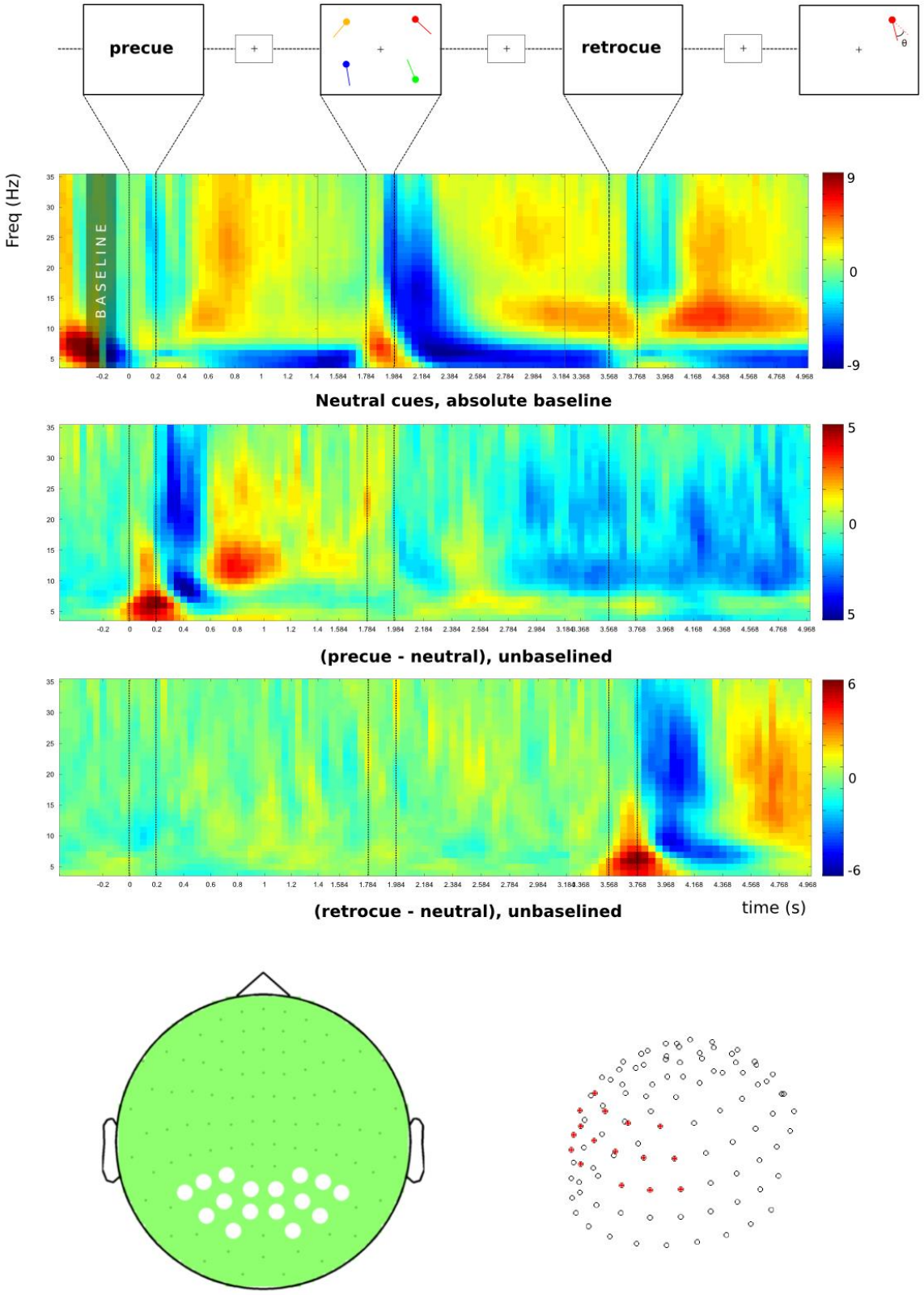


Figure 2-11 **Induced responses, posterior sensors.** The theta (<8Hz), alpha (8-14Hz) and beta (18-30Hz) bands are clearly evident. Theta and alpha tend to vary independently, but the alpha and beta band co-vary. Note the burst of theta activity associated with ERFs (compare with Figure 2-10). ERFs also reliably induce an alpha desynchronization: the event related desynchronization (ERD).

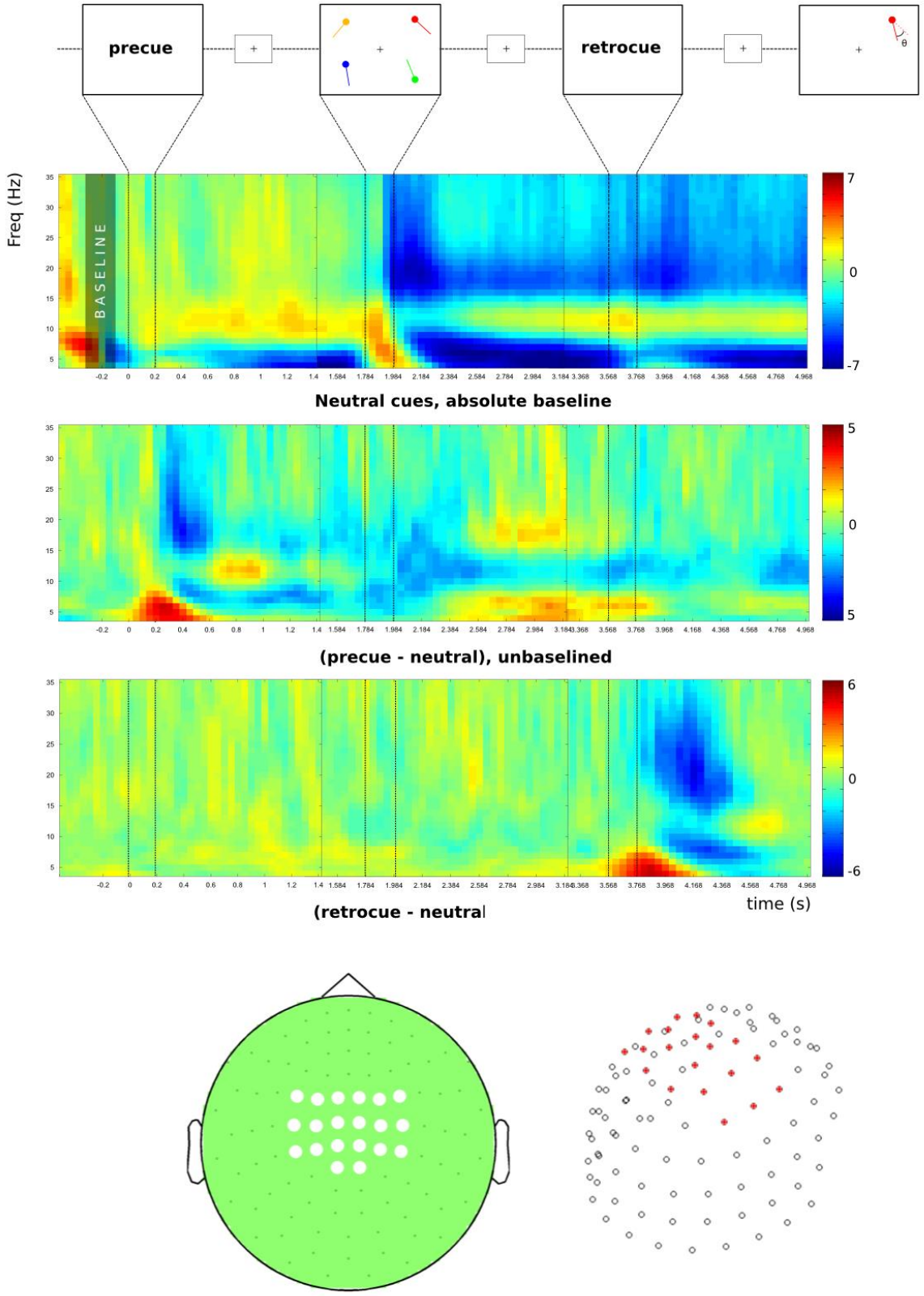


Figure 2-12 **Induced responses, central sensors.** Note the greater desynchronization in beta-band power as compared with posterior sensors.

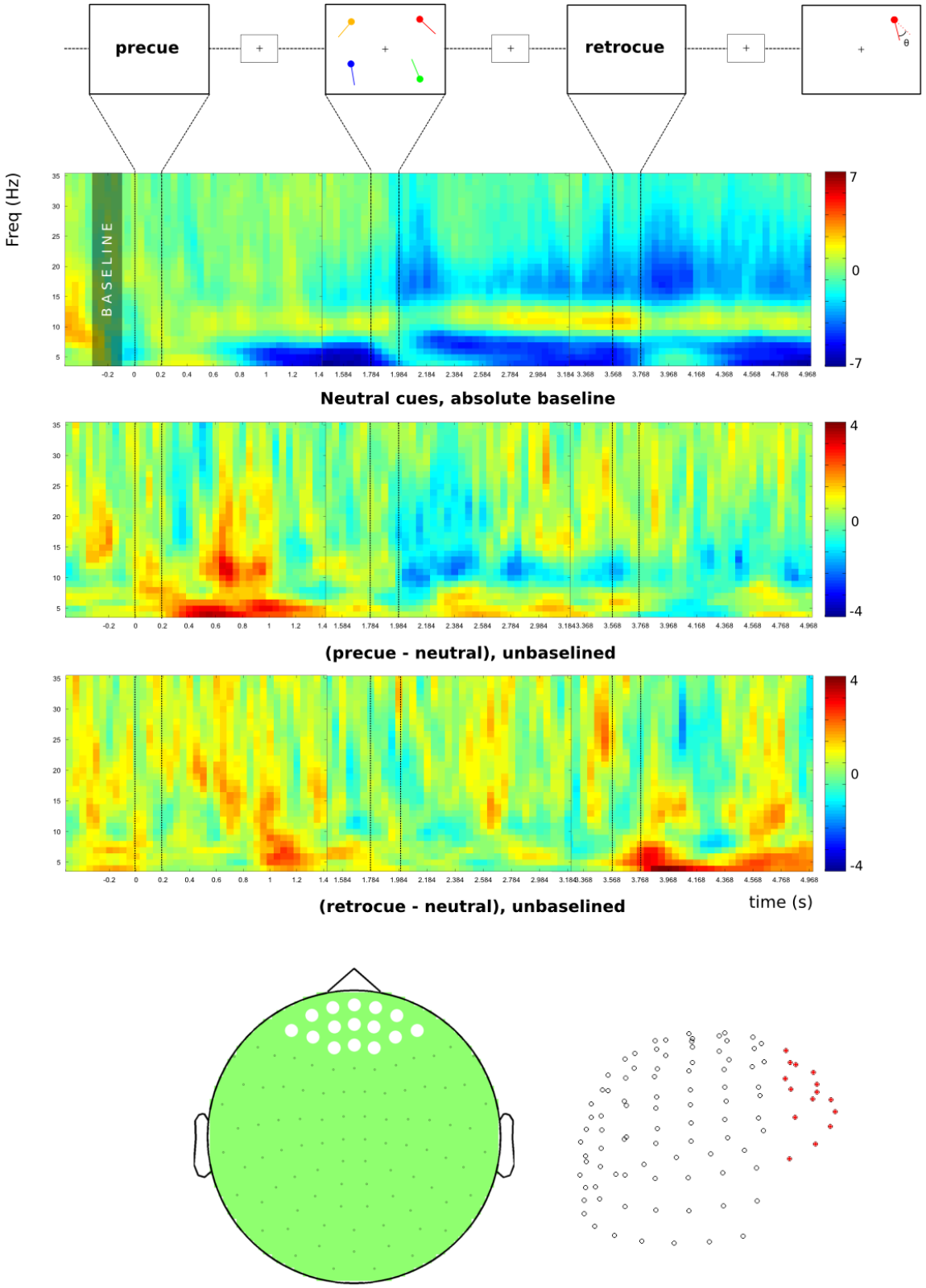


Figure 2-13 **Induced responses, frontal sensors.** Note the visibly higher level of noise in the signal (compare with Figure 2-10). Frontal sensors are further from the cortex than posterior sensors. Theta synchronization can nevertheless be observed following informative cues.

2.5 Summary

For the purposes of the work presented in this thesis, three key advantages of MEG were important. Firstly, MEG has very high temporal resolution. Crucially, the signal is generated directly by neuronal electrical activity, and not filtered through a slow haemodynamic response function, as is the case with fMRI. Secondly, MEG has reasonable spatial resolution when combined with source reconstruction techniques such as beamforming[13]. This allows activity to be mapped to the brain with some anatomical specificity. However, all approaches to solving the inverse problem (including the beamforming approach) are addressing a fundamentally ill-posed problem, and as such depend on assumptions to constrain the solution. These assumptions are inevitably imperfect, and the source reconstruction technique suffers from associated limitations, as discussed in sections 2.3.1 through 2.3.6, above. Thirdly, MEG data in the sensor space is rich in information and amenable to multivariate approaches, which are used in chapters 4 and 5 to characterize information in the signal without needing to spatially select data (either at the sensor level or in source space). The multivariate approach is complementary to the source reconstruction approach: whilst the latter depends critically on prior assumptions and tends to expand the dimensionality of the (already high-dimensional) sensor-space data (introducing the potential for a multiple comparisons problem, or even for biased selection of spatial features), the former collapses the dimensionality of the data and is relatively assumption-free. Key results in the current thesis are approached both using multivariate approaches in sensor space, and source-space data.

---

[13] The relative merits of EEG and MEG, in terms of their ability to spatially localize activity, has been the subject of a long and on-going debate - in which an important theme is the much higher cost of MEG as compared to EEG. As EEG was not collected as part of this thesis this question was not relevant for this chapter.

3 Retrocues boost performance by facilitating output gating

Chapter Abstract

Retrocues (Griffin & Nobre, 2003; Landman et al., 2003), like precues (Murray et al., 2011; Schmidt et al., 2002), confer a behavioural advantage in working-memory tasks. Precues are usually assumed to boost performance by allowing people to selectively encode only task-relevant items from a memory array, reducing effective memory load (Schmidt et al., 2002). The retrocue benefit is puzzling from this perspective: the items in memory have already been encoded and are no longer perceptually present when the cue is presented, so how can retrocues produce behavioural benefits almost as robust as precues do? The implication is that working-memory performance (and corresponding capacity estimates) may strongly depend on processes besides the transformation from a perceptual representation to a mnemonic buffer: for example, output gating to allow the appropriate item(s) in memory to drive behaviour. This chapter discusses behavioural data collected as part of the MEG project discussed in later chapters. A precision/capacity working-memory task was used to replicate previous work on cued selection in short-term memory (Murray et al., 2013), in a forced choice variant in order to obtain both reaction time (a proxy for retrieval time) and recall accuracy as dependent variables. Both precues and retrocues have been previously observed to affect RT (Astle et al., 2011; Griffin & Nobre, 2003; Rerko, Souza, & Oberauer, 2014) but the way in which retrocues can speed subsequent recall has not been modelled. Here the standard mixture model is fitted to the accuracy data (W. Zhang & Luck, 2008), replicating the findings of previous work, but the accuracy and RT data are also modelled using a well established model of decision making in forced-choice tasks: the drift-diffusion model (Ratcliff & McKoon, 2008). Precues and retrocues are both found to reduce the non-decision time by the same fixed amount, supporting the idea that cues give rise to a qualitative state change of the cued representation that facilitates decision making, rather than changing the quality or robustness of the cued representation.

3.1 Introduction

3.1.1 Task design

In order to investigate the effects of retrocues on memory performance, Murray et al. (Murray et al., 2013) employed a forced-choice variant of the precision/capacity task (Bays & Husain, 2008; W. Zhang & Luck, 2008) that has recently come to the fore in VSTM research. In a typical version of this task, a number of items are presented onscreen, each item randomly assigned some value on a given feature dimension. The current task uses oriented 'keyhole' stimuli. The probe item re-displays one of the items in the memory array, but the probe is rotated either clockwise or anticlockwise relative to the original memory item. Subjects judge, on each trial, whether the rotation was clockwise or anticlockwise. By manipulating the magnitude of the angular displacement, we can gauge both the chance that a subject is guessing on any given trial (guess rate), and the precision with which the probed item is stored in memory (on non-guess trials). Over the range of different angular displacements of the probe item, the probability of responding 'clockwise' on a given trial follows a sigmoidal psychometric function. This is illustrated in Figure 3-1. The function asymptotes before reaching 0 for large anticlockwise changes and 1 for large clockwise changes. Intuitively, the magnitude of this asymptote, the point beyond which any decrease in the difficulty of the comparison with the probe item would not boost performance any further, reflects the proportion of trials in which the subject is simply guessing with respect to the probed memory item. In contrast, the slope of the psychometric curve, for which the smaller changes in orientation are most diagnostic, gives an estimate of the precision of the memory representation.

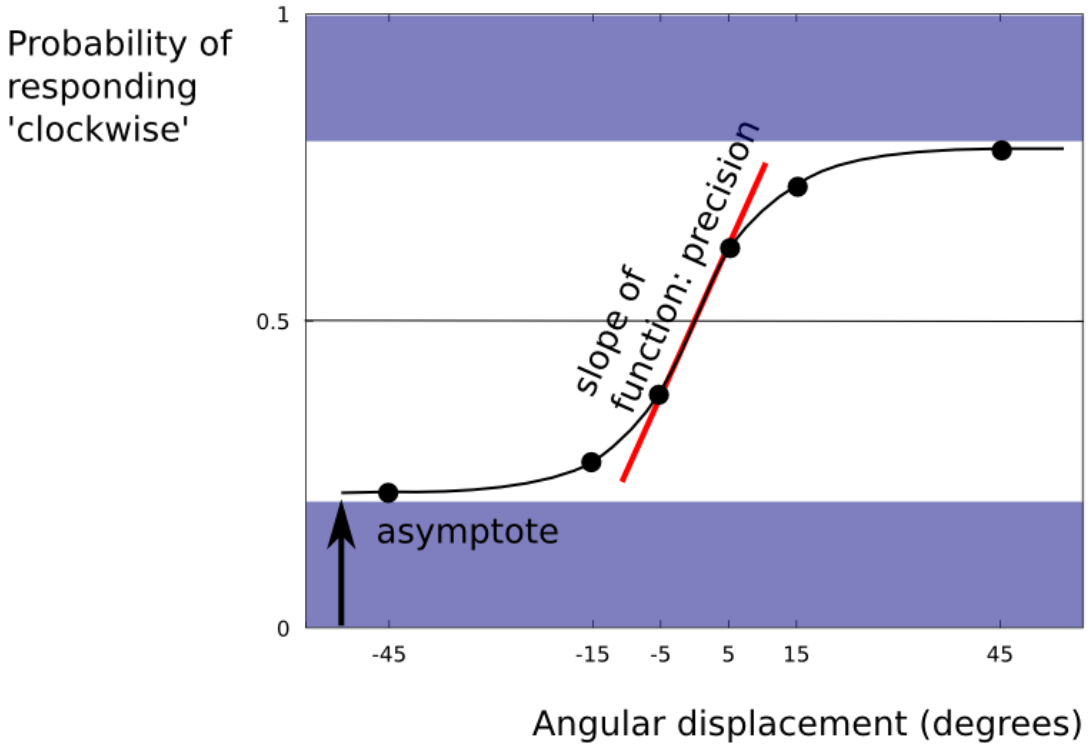


Figure 3-1 **Hypothetical data from a precision/capacity forced-choice task, illustrating the mixture-model fit.** The model incorporates a proportion of trials in which subjects are guessing with respect to the probed item, which determines the asymptote of the function. The slope of the function indexes the precision with which items are represented in memory.

This sigmoid response function can be fitted using a mixture model (W. Zhang & Luck, 2008) (described in more detail in section 3.2 below, and in appendix 9.1), yielding a parameter capturing guess rate and a parameter capturing precision. Gororaptis and colleagues (2011) tested the effect of *precues* in a precision/capacity task, and found that they affected both precision (increased) and guess rate (decreased). This has been taken as evidence for preferential encoding of the cued item. Combining the precision/capacity approach with *retrocueing*, Murray et al. (2013) found that guess rate was robustly reduced by retrocues, but precision was not affected (Murray et al., 2013). This suggests that retrocues affect the accessibility of items in memory, but not the quality with which they are represented. The efficacy of retrocues reveals that output gating may be as much a performance limiting factor in working memory tasks as is

input gating – though, this factor is ignored by many studies that have attempted to characterise memory capacity.

The current study aimed to investigate both the brain mechanisms that determine biased encoding, but also the mechanisms responsible for changing the accessibility of items in memory. A precision/capacity task was therefore developed in which one third of trials were uncued, one third of trials had a valid spatial precue, and one third of trials had a valid spatial retrocue. This allowed a replication of the results of Murray et al. (2013) and Gorgoraptis et al. (2011) in a single experiment, and a quantitative comparison of the effects of each cue type on performance. By using a forced-choice response, reaction time was available as a dependent measure, in addition to accuracy. This extra information allowed the effects of the two cue types to be dissected by applying a model that accounts for retrieval speed as well as response accuracy.

3.1.2 Modelling reaction time data: motivation and rationale

A common assumption in the working memory literature is that if a representation is present in memory, then it should be able to guide action. Therefore, a failure to respond appropriately to a memory probe is taken to indicate that an item was not represented in memory (W. Zhang & Luck, 2008). However, the retrocue benefit reveals a potential disconnect between whether an item is represented, and whether it can guide action. It has been proposed that items in working memory can inhabit passive or active states, determining whether they can be immediately used to guide responses (McElree, 2006; Oberauer & Hein, 2012; Olivers et al., 2011). A key mechanism underlying the retrocue benefit might be a state change to the cued representation, rendering it the active item. Precues should have a similar effect, but with additional benefits of optimising encoding of the cued item. Reaction time to the

probed item gives an indication of the readiness with which the probed memory item can guide action. We therefore analysed the data from our working memory task using both the standard mixture model, but also using a long-standing model that captures both choice and RT data, the drift-diffusion model.

Garavan measured RT costs associated with switching between internal representations (Garavan, 1998). Subjects were presented with small groups of either triangles or rectangles, trial by trial, and asked to keep a running tally of the number of each type of item they had seen. Sometimes a trial in which triangles were presented would be followed by a trial with rectangles (a switch trial), and sometimes by a trial with more triangles (a non-switch trial). Subjects pressed a key to move between trials, and there was therefore a measure of the time taken to update the running count on each trial. Non-switch trials were faster than switch trials by between 200 and 500ms (a value that decreased with practice). Garavan's conclusion was that the tallies, whilst both in short-term memory, were generally not of equal status: one was active and one passive, and additional time was required to switch to updating the passive tally as compared to continuing to update the active tally.

Updating a memory trace, as in Garavan's task, is not necessarily the same process as retrieving a memory item. However, similar RT effects have been found for retrieval. McElree investigated the interaction between internal attentional focus and retrieval speed using serially presented word lists (McElree, 2006). A confound when using RT to infer differences in retrieval time between WM representations is that if the representations differ in strength, this can also influence RT by changing both the speed with which evidence for the response is accumulated and the maximum attainable strength of evidence for the response. An RT difference could therefore reflect either retrieval time or memory strength. McElree addressed this problem by using the

speed-accuracy trade-off procedure (SAT) in which subjects are forced to respond at various points following the memory probe stimulus by the presentation of a subsequent tone indicating "respond now!". Manipulating the interval between probe item and response tone allows the experimenter to build up an SAT curve, measuring accuracy at different probe-response latencies. For very short probe-tone latencies, performance is at chance – the interval over which performance remains at chance is assumed to correspond to the initial retrieval time. Performance then begins to rise exponentially from this intercept point up to some asymptote, at which point increasing the probe-tone interval no longer increases accuracy[14]. By fitting an exponential curve to these data, the intercept parameter (initial retrieval time), rate parameter (rate of evidence accumulation) and asymptote parameter (performance ceiling) can be estimated. McElree and Dosher (McElree & Dosher, 1989) presented sequential lists of 3 or 6 words. They used the SAT approach to quantify the time course of retrieval following a probe word – subjects had to indicate whether the probe item had been in the preceding list. They found that all their fitted parameters were identical across serial positions except for the very last item in the list, for which the retrieval time was reduced. Their interpretation was that the last item in the list was still in the 'focus of attention' and was therefore retrieved more quickly. McElree (McElree, 2006) has since shown that this difference in retrieval time for word lists can also be induced by retrospectively cueing subjects to certain items in memory. Two groups of three semantically related words (e.g. "CAT, MOOSE, WOLF, DOCTOR, COP, LAWYER") were presented. Subjects were then given a semantic cue to one of the groups (e.g. "ANIMALS" or "PROFESSIONS"). Subsequent retrieval time was 45ms faster for words from the cued set. This demonstrated that the last-item effect did not depend purely

[14] The retrieval time is comparable to the initial component of non-decision time in a diffusion model, and the rising segment to diffusion model's evidence accumulation period.

on a lingering perceptual representation, but rather that the availability of a set of items in memory can be dissociated from the strength with which they are represented, and is under executive control.

McElree's experiment involved cueing a *category* of items from among multiple categories held in memory (see also Lepsien et al., 2011), and in Garavan's task subjects switched between counters that were associated with different tasks/stimulus categories. In tasks in which all items are of a homogeneous category, scholars have tended to overlook the role of accessibility in determining performance. In the task presented in this chapter, items were all of the same type and single items were cued. The working hypothesis was that both precues and retrocues would allow the cued item to achieve the active status, and that we would therefore see a comparable RT advantage in cued versus uncued trials of either cue type. Previous retrocueing experiments have found robust effects of retrocueing on RT – for example, Astle et al. found that retrocues reliably speeded responses by 150-200ms. This finding is replicated here. In order to dissect effects of memory strength from retrieval time, we modelled the RT distributions using Ratcliff's Drift Diffusion Model (DDM; Ratcliff & McKoon, 2008). The details of this model are discussed in greater depth in the methods section below. Essentially, the model decomposes reaction time as the sum of a non-decision period (encompassing perceptual processing, memory item retrieval, and motor response time) and a decision period, during which evidence accumulates for one or other response option until it hits one or other bound, triggering a response. Non-decision time, rate of evidence accumulation, and bound separation differentially contribute to the shape of the reaction-time distribution. By modelling the whole RT distribution, the drift diffusion model can therefore dissociate the effects of task manipulations on retrieval time, rate of evidence accumulation, and response strategy.

3.2 Methods

3.2.1 Experimental procedure

In order to investigate the neural correlates of prospective and retrospective cueing (chapters 4 and 5) we implemented a forced-choice precision/capacity WM task (Murray et al., 2013; similar to Murray, Nobre, Astle, & Stokes, 2012) with both prospective and retrospective cues, to be used in conjunction with MEG.

Four 'keyhole' memory items were presented for encoding on every trial. Following a delay period of 3 seconds, a single probe item appeared. This probe item re-displayed one of the items in the memory array, except that it was rotated either clockwise or anticlockwise relative to the memory item. On one third of trials, a spatial precue indicted the spatial location of the relevant memory item prior to the onset of the array, and on a separate third of trials, a spatial retrocue indicated the location of relevant item during the retention interval. On the remaining trials ('neutral cue' trials) no information was given about which item would be probed. Cues were always valid. A task schematic with trial timings is given in Figure 3-2.

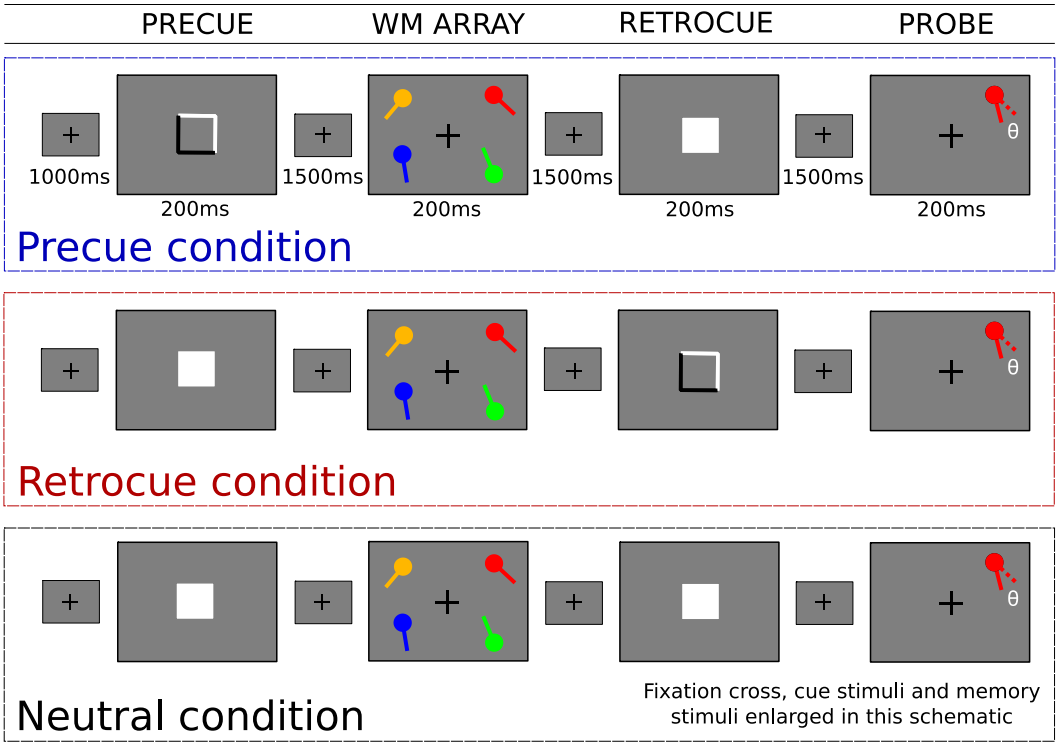


Figure 3-2 **Task schematic, precision/capacity task with precues/retrocues.** Four 'keyhole' memory items were presented in the memory array. A probe item appeared at the end of the trial. The probe item was matched to the corresponding memory item in both location and colour, but was rotated either 5, 15 or 45 degrees clockwise/anticlockwise. Subjects made a forced-choice response, indicating whether the rotation was clockwise or anticlockwise. On one third of trials, no information was given about which item would be probed (bottom row); on one third of trials a precue appearing 1.5s prior to the memory array indicated which item would be probed (top row); on one third of trials, a retrocue appearing 1.5s into the retention interval indicated which item would be probed (middle row).

The probe colour matched the colour of the probed item (thus providing an additional piece of information as to which memory item was being probed, besides spatial location). The probe stimulus was rotated about the circular end body of the original memory item, either through 5 degrees (small rotation), 15 degrees (medium rotation) or 45 degrees (large rotation) clockwise or anticlockwise (with equal probability). Subjects pressed one of two buttons with either their index or middle finger of their right hand; the left-hand button (index finger) indicated an anti-clockwise angular change, and the right-hand button (middle finger) a clockwise rotation. In this forced-choice paradigm, subjects always registered a response, irrespective of any confidence criterion. Subjects were also encouraged always to make an effort to

respond based on memory even if they even if felt they had retained very little or no information about the probed item.

The behavioural task was run in two separate sessions: one training session prior to the MEG session, and a further session in the MEG scanner. The task was split into blocks of 36 trials between which there were self-timed rest breaks. The training session contained 6 blocks (216 trials in total) and the MEG session contained 9 blocks (324 trials total).

In both the training session and the MEG session, subjects were instructed to maintain fixation during the trials, and eye movements were monitored using an Eyelink 1000 infrared video eyetracker (binocular). During the training session, stimuli were presented on an LCD screen at a viewing distance of 80cm, and a chinrest was used to stabilize the head. During the MEG session, stimuli were back-projected onto a screen at a viewing distance of 85cm. In both training and MEG sessions, the WM stimuli were presented at an eccentricity of 6 degrees visual angle from the fixation cross, and each stimulus subtended 1.2 degrees visual angle.

Of the 50 subjects who participated in the MEG experiment, one subject's MEG session was abandoned mid-way through the behavioural task as the subject was very drowsy, and a further subject's responses on the VSTM task were at chance in both training and MEG sessions across all conditions. Behavioural data were therefore analysed for 48 subjects. To reduce noise from trials in which the subjects were settling into the task, it was decided to discard the first 50 trials from the training block, as well as the first 10 trials from the MEG session (though including these trials in the analysis does not substantively change any aspect of the reported results).

Any trials in which the RT was shorter than 0.1s or longer than 10s (the median reaction time was 1.1s) were dropped from the analysis, as they likely represented anticipation responses or a lapse in task engagement respectively. For most subjects (40/48), five or fewer trials were excluded on the basis of RT. The maximum number of trials rejected for any one subject was 20.

3.2.2 Mixture model of accuracy data

Clockwise/anticlockwise changes were collapsed together for modelling, factoring out any clockwise/anticlockwise response bias. First a mixture model was fitted for the accuracy data (W. Zhang & Luck, 2008) in which responses across trials are assumed to come from one of two distributions: a uniform 'guess' distribution, representing trials in which subjects had no information about the probed item and chose a response at random, and a Von Mises distribution (the circular variant of a Gaussian distribution) that represents the fidelity of the WM representation on trials when subjects were not guessing. The probability that a subject made the correct response on a given trial depends on the precision of the Von Mises distribution **k** (higher values indicate a tighter distribution), the probability that subjects are guessing on any particular trial **pGuess**, and the size of the orientation change of the probe stimulus relative to the corresponding memory item on that trial denoted $\boldsymbol{\theta}$. This is captured in Equation 5.

$$p(correct) = (pGuess * 0.5) + (1 - pGuess) * \ vonmisescdf\,(k, \theta)$$

Equation 5

'vonmisescdf' is the cumulative density function of the Von Mises distribution, which can be evaluated numerically. Full details of the model fitting are given in Appendix 9.1. The model was fitted for each subject and each condition separately using a maximum

likelihood approach, constrained by loose prior distributions on both parameters (the priors were the same across all subjects and all conditions).

3.2.3 Misgating analysis

In the mixture model analysis, it is assumed that the *unprobed* items have no influence on responding. However, several studies have shown that unprobed items do indeed influence responding. For example Bays and colleagues (Bays, Catalao, & Husain, 2009; Gorgoraptis et al., 2011) include responses to unprobed items in their mixture model, as an additional distribution, and find up to 30% of responses are made to non-target items (Bays et al., 2011 - orientation memory; Bays et al., 2009 - colour memory). In the current task, the orientation of each memory item was randomly assigned independently of the orientation of the other items. The effect of unprobed items could therefore be evaluated simply by running an analogous analysis to that for the probed item, examining the probability of responding 'clockwise' or 'anticlockwise' depending on whether the probe item was rotated clockwise or anti-clockwise of the non-target item. Trials were binned by the difference between the probe orientation and the orientation of each of the three un-probed items. The orientation difference between the probe item and a given unprobed item spanned the full circle between -180 and 180 degrees. This range was divided into eight bins: clockwise/anti-clockwise 0-45°, 45-90°, 90-135° and 135-180°, and for each bin the probability of responding clockwise was calculated. The analysis was performed for each of the three unprobed items in a given trial separately, sorting them with respect to the location of the probed item: i.e. same hemifield, horizontally opposed and diagonally opposed.

3.2.4 Diffusion model of accuracy and RT data

To dissect the RT effects, both the accuracy and RT data were modelled with a drift diffusion model (Ratcliff & McKoon, 2008). The DDM accounts for both accuracy and RT by assuming that each time a decision is made, an internal decision variable accumulates evidence for one or other decision in a biased random walk. The variable accumulates between two absorbing bounds. Each bound is associated with a response. Once the diffusion variable hits either bound, a response is made. The key fitted parameters are **a**, the separation of the bounds, **v**, the mean drift rate of the decision variable, and **Ter**, the non-decision time (which captures any part of the RT that does not include evidence accumulation– e.g. time for sensory processing, any other internal processes occurring prior to or after the main decision process, and time for motor aspects of response production). A fourth variable, bias towards one or other bound, is not included in the current model, as left/right responses are collapsed together and coded as correct/incorrect.

Error responses are usually slower than correct responses in tasks demanding accurate responses. The DDM with parameters as stated so far will always predict correct and error RT distributions of equal latency, but the DDM can account for slower error responses if mean drift rate and the DV start point vary across trials. For example, on those trials where drift rate happens to be higher, both correct and error RTs are shorter, but correct responses become more likely, whereas trials with lower drift rate have longer correct and error RTs, but errors become more likely. Across all possible drift rates, there is a larger proportion of (slower) errors for low drift rates, and a larger proportion of (faster) correct responses for high drift rates. In aggregate, the error distribution therefore becomes slower than the correct response distribution. By

including cross-trial variability in drift rate, the diffusion model can therefore capture this aspect of the behavioural data.

DDM modelling typically requires a large number of within-subject trials per condition bin in order to obtain good parameter estimates. In the current experiment, we had a modest number of trials per subject (540 per subject, split between three cue conditions and three probe difficulties), but a relatively large number of subjects (48 for behavioural analysis). Hierarchical model fitting is a technique in which the large amount of data available over the whole group of subjects is used to constrain the fit for each subject, compensating for limited trial numbers. We used the freely available Python package HDDM (Hierarchical Drift Diffusion Model; Wiecki, Sofer, & Frank, 2013) to fit the DDM for each condition separately. Group-level posterior distributions on each model parameter are estimated simultaneously with the posteriors for subject-wise parameters in a Bayesian hierarchical model (using Markov Chain Monte Carlo to estimate the posteriors). Data from each of the three cue conditions were fitted using separate models. We allowed the model to estimate a different drift rate, bound separation, and non-decision time for each orientation shift magnitude, within cue condition. Group and subject-wise parameters were estimated for drift rate **v**, bound separation **a**, and non-decision time **Ter**. Cross-trial distributions **sv** (drift rate variability across trials), **sz** (diffusion start point variability across trials) and **st** (non-decision time variability across trials) were estimated at the group level only, for computational reasons. The group posteriors can be used to obtain confidence estimates for each fitted parameter.

In order to draw conclusions from a model fit, the fit needs to be reasonable. One way to 'sanity check' the model fit is to generate predicted data from the fitted parameters (Rubin, 1984). If the model fit is sensible, the observed data should resemble the

predicted data. For the diffusion model, 500 parameter values sampling the posterior distribution for each subject were used to generate 500 predicted datasets for each subject. The RT distributions were binned into a histogram within subjects for each predicted dataset, and the value for each bin averaged across datasets, yielding a single histogram per subject. These subject-wise predicted RT histograms were averaged across subjects, yielding a group-level predicted RT histogram. The observed RT data were binned in the same way for each subject, and averaged across subjects. The predicted and observed data can then be compared directly (see 3.3.3 below, Figure 3-9). A similar procedure was followed for the predicted and observed accuracy data.

3.3 Results

Trials were grouped according to the orientation change of the probe item relative to the memory item. There were six bins: small, medium and large orientation change, for anticlockwise and clockwise changes. Figure 3-3 summarizes the group-level accuracy and reaction time data.

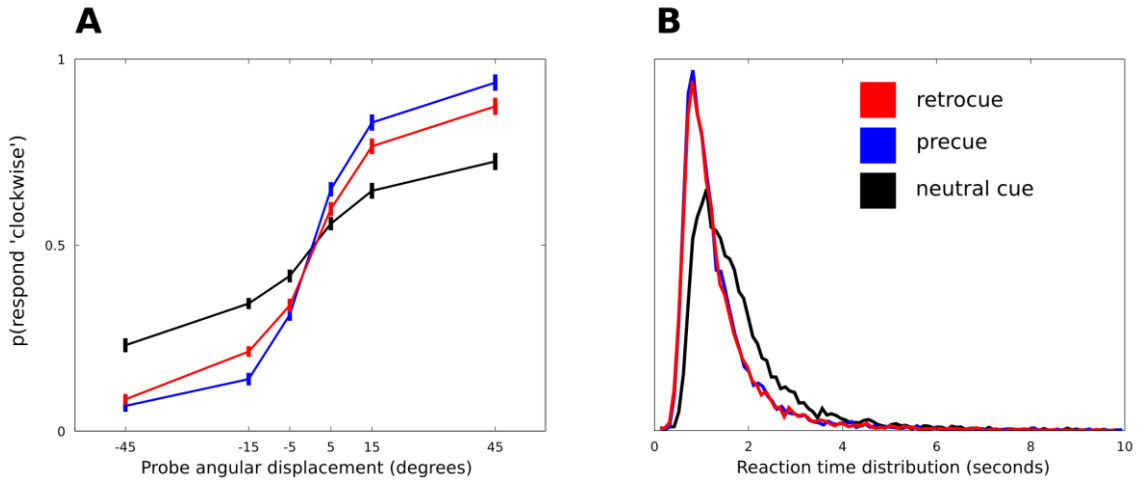


Figure 3-3 **Summary plot of data from the precision/capacity task. A** - accuracy data: the proportion of clockwise responses made for each combination of angular displacement and cue condition. Error bars are ±1SEM, at the group level. **B** - RT distributions, correct and incorrect trials collapsed together

The proportion of clockwise responses across bins had a characteristic sigmoid shape, in which performance is better for the large orientation changes than for the small changes. Both precues and retrocues improved performance compared to the neutral cue condition. The RT data also showed the typical cueing effect: RT is reduced for both precues and retrocues, and the RT distributions for the two cue types are strikingly similar. By inspection there are two effects of cueing on reaction time: a 'leftwards shift' of the distribution to shorter RTs, and a reduction in variance. A diffusion model can account for both of these changes as discussed in 3.3.3 below.

3.3.1 Mixture model

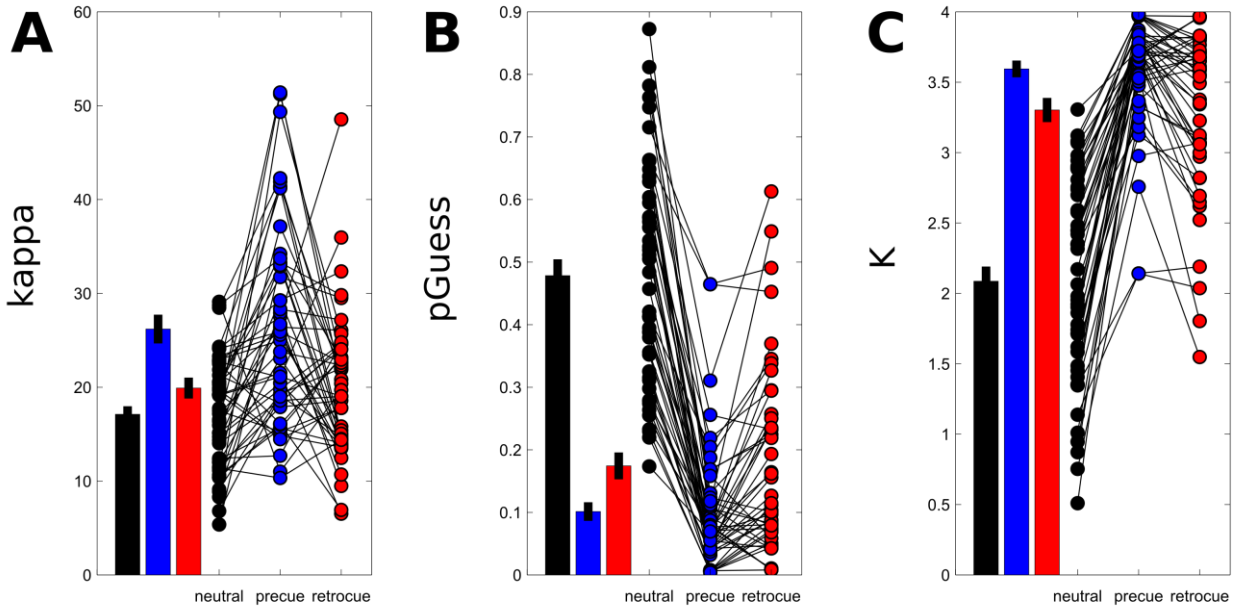


Figure 3-4 **Mixture-model fit parameters.** **A** – memory precision. **B** – guess rate **C** – capacity, K (4*(1-pGuess)). The bar plots on the LHS of each plot give group-averaged data. Subject-wise data are shown as separate points on the RHS of each plot. Bar plot error bars are ±1SEM over subjects. Precues increase precision and decrease guess rate, whereas retrocues decrease guess rate but do not substantially affect precision.

Fit parameters for the standard mixture model are shown in Figure 3-4. Also shown is an estimate of capacity, K, which is a transformation of the guess-rate parameter (given there were four items in the memory array, K = (1-pGuess)*4). By inspection, precues increase precision and decrease guess rate, whereas retrocues similarly reduce guess rate, rate but have a much less substantial effect on precision. Both precues and retrocues increase the 'effective capacity' (as indexed by K).

A repeated-measures ANOVA with factor CUE (3 levels: neutral cues, precue, retrocue) was run separately for the precision and guess-rate parameter estimates. In both cases

there was a significant main effect of cue condition: pGuess, $F(1.74,82.3) = 203.1$, $p < 0.0001$, Greenhouse-Geiser corrected; kappa, $F(2,94) = 17.1$, $p < 0.0001$.

Paired-sample post-hoc t-tests were used to test for pairwise differences between the cue conditions. The guess rate was significantly lower in the precue than neutral condition ($t(47) = 16.95$, $p < 0.0005$), and also in the retrocue condition than neutral condition ($t(47) = 14.55$, $p < 0.0005$). The guess rate in the precue condition was significantly lower than guess rate in the neutral condition ($t(47) = -4.62$, $p < 0.0005$). Kappa was significantly higher in the precue condition than the neutral condition ($t(47) = 5.63$, $p < 0.0005$). The precision was also marginally higher in the retrocue condition than the neutral condition ($t(47) = 2.04$, $p = 0.047$). Notably, precision was significantly higher in the precue condition than the retrocue condition ($t(47) = 3.56$, $p = 0.001$). This pattern is broadly consistent with the prior finding that retrocues improve performance primarily by modulating guess rate, rather than memory precision (Murray et al., 2013). In the current data, precision was only marginally higher after retrocueing than in the neutral condition, whereas precues had a much more substantial impact on precision. In contrast, the change in guess rate due to precues and retrocues was of comparable magnitude.

3.3.2 Misgating analysis

Subjects' responses were analysed relative to the unprobed items. Figure 3-5 shows the effect of unprobed items on responding. Responses are influenced by the unprobed items ('misgating'[15]) in the neutral condition, but less so in the precue or retrocue conditions. These data were modelled using a repeated-measures ANOVA, with factors

[15] 'Misbinding' is the term introduced by Bays and colleagues to describe this effect. This term implies that features are bound with the wrong object representations in memory. This is not the interpretation proposed here, where 'misgating' might be more appropriate – the wrong item is suggested to have been selected wholesale to drive memory. 'Misgating' is therefore used in this thesis.

Quadrant Location (relative to probed item, 3 levels), Orientation Change (8 levels), and Cue Condition (3 levels). The three-way interaction between Quadrant Location, Orientation Change and Cue Condition was not significant ($F(28,1316)=0.768$, $p = 0.802$). There was a main effect of Orientation Change ($F(7,1316)=13.65$, $p<0.0005$), confirming that the unprobed items affected responding, and there was an interaction between Orientation Change and Cue Condition ($F(14,1316) = 4.1$, $p < 0.0005$), indicating that the misgating effect differed between the cue conditions. There was a weak interaction between Orientation Change and Quadrant Location ($F(14,1316) = 1.73$, $p = 0.046$), suggesting that the relative location of the unprobed item to the probed item affected the degree to which it influenced behaviour. By inspection, the diagonally opposed unprobed item had less effect on responses than the horizontally opposed (closer) items.

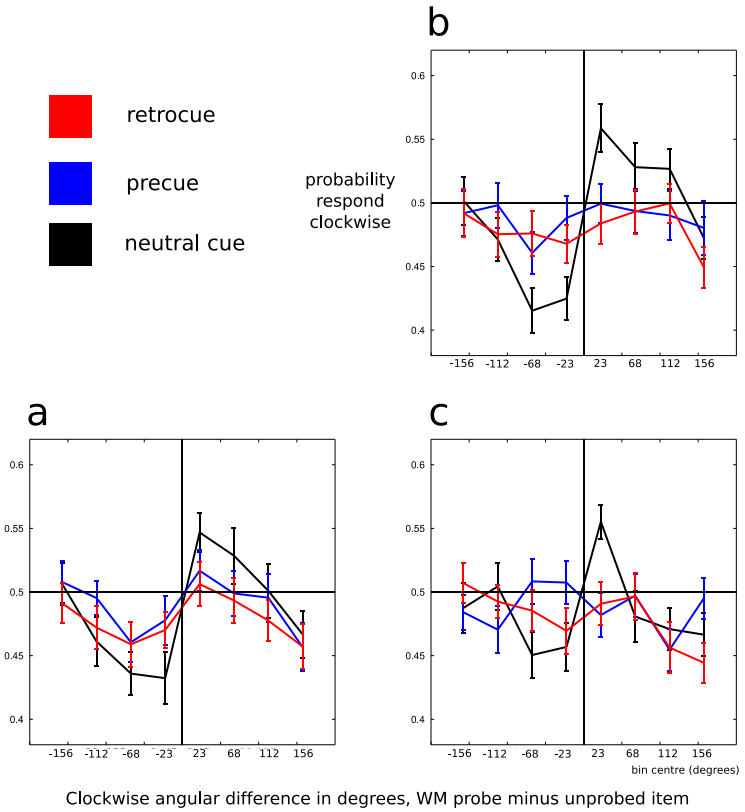


Figure 3-5 **Influence of unprobed items on responding.** Separate plots: effect of items (a) vertically opposed (same hemifield), (b) horizontally opposed, and (c) diagonally opposed to probed quadrant. Unprobed items influence responding ('misgating'). This effect is slightly stronger for the directly apposed quadrant locations to the probed location than for than the diagonally opposite quadrant (Orientation Bin * Quadrant interaction, p = 0.046)

A two-way ANOVA was performed for each cue condition separately to investigate the evidence for unprobed item effects within each condition. For the precue condition alone, there was no evidence that the orientation of unprobed items influenced responses ($F(7,658)=0.837$, $p = 0.557$). In contrast, for the retrocue condition, there was evidence for a misgating effect ($F(7,658)=2.95$, $p = 0.005$), as there also was for the neutral condition ($F(7,658)= 17.4$, $p < 0.0005$).

There is statistical evidence for misgating in neutral and retrocue trials, but not precue trials. However, plotting the misgating effect collapsed across quadrant and clockwise/anticlockwise rotation (Figure 3-6) makes it apparent that the degree of misgating was substantially elevated on neutral trials, compared to either precue or

retrocue trials, and was not substantially different between the two cue conditions. The Cue*Orient interaction (capturing differences in misgating between conditions) was broken down using separate ANOVAs for each pair of conditions. There was no Cue*Orient interaction for between precue and retrocue conditions ($F(7,329) = 0.598$; $p = 0.757$) but there was a significant Cue*Orient interaction between neutral and precue conditions ($F(7,329) = 5.87$, $p < 0.0005$) and between neutral and retrocue conditions ($F(7,329) = 5.711$, $p < 0.0005$).

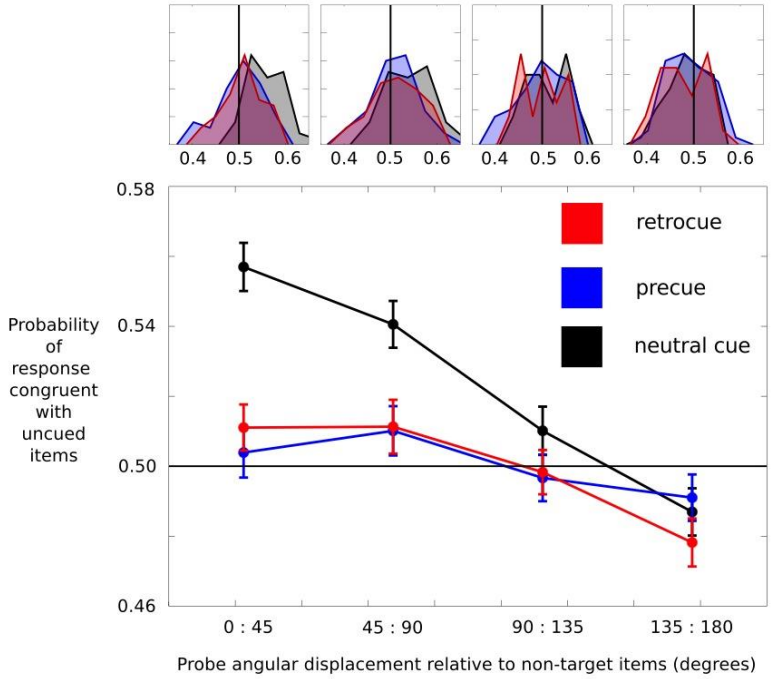


Figure 3-6 **Influence of unprobed items on responding, collapsed over quadrants, and collapsing clockwise and anticlockwise responses.** Error bars are ±1SEM. Misgating strength did not differ between precue and neutral conditions (Cue*Orient interaction, precue/retrocue, n.s.) but misgating in both cue conditions was significantly weaker than for the neutral condition (Cue*Orient interaction neutral/precue $p<0.0005$, neutral/retrocue, $p<0.0005$). Both precues and retrocues significantly reduce the effect of the unprobed items on responding. For each bin, cross-subject histograms are shown, demonstrating that the majority of subjects showed a misgating effect.

3.3.3 Drift diffusion model

The forced-choice responses in the precision/capacity WM task yield a dependent variable of reaction time as well as accuracy. The previously observed reduction in

reaction time with retrocueing (Astle et al., 2011; Griffin & Nobre, 2003; Rerko et al., 2014) is also evident in the current data (Figure 3-3). The reaction times for precue and retrocue trials are slower than for neutral trials. Below, these distributions are broken down for correct and incorrect responses (Figure 3-7). Both precues and retrocues speed RT for correct responses by around 200ms compared to uncued trials (lower left panel, Figure 3-7), demonstrating that the difference in RT between the cue conditions is not driven by a different admixture of slower incorrect response trials. The reaction time for correct responses was also dependent on the shift magnitude of the probe item (i.e. the difficulty on that trial) – responses were faster for larger orientation shifts.

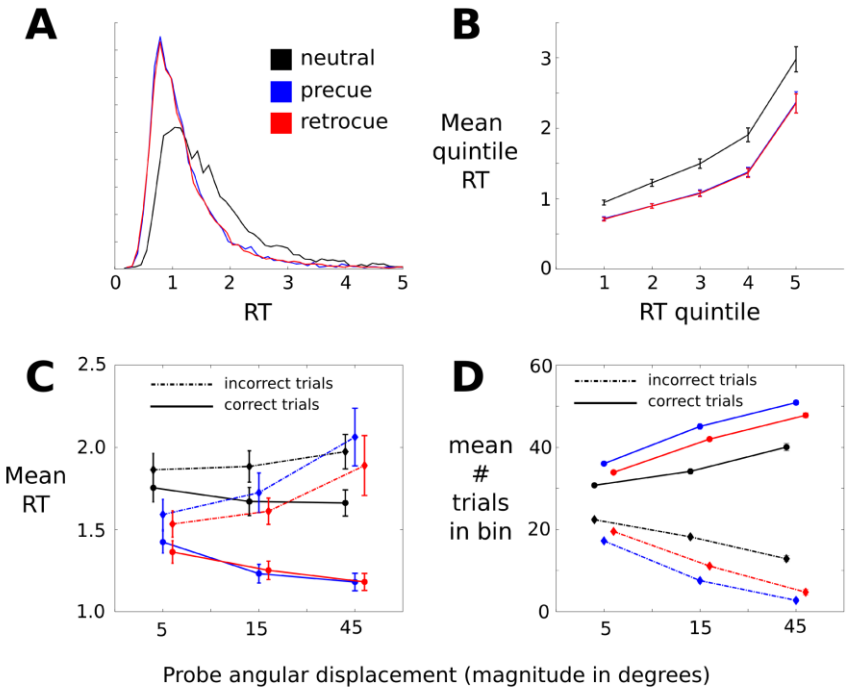


Figure 3-7 **Summary plot of RT data**. **A**: Precue and retrocue RT distributions have a lower central tendency, and lower variance than neutral cue RT distributions. **B**: group RT distributions expressed as the median within each quintile, for correct and incorrect trials together. This representation emphasises the similarity between RT distributions for precues and retrocues, and the consistent RT speeding across the full distribution as compared to neutral trials. **C**: mean RT by cue-condition and probe angular displacement for correct (circles, solid lines) and incorrect (diamonds, dotted lines) separately. Incorrect trials are slower than correct trials. For correct trials, RT decreases with increasing orientation change magnitude. For incorrect trials, this trend is reversed. **D:** trial numbers in each bin of the bottom-left plot. There are few error trials for cued trials with an easily discriminable probe orientation change.

We fitted the RT/accuracy data using a hierarchical drift diffusion model. The estimated parameter values and their confidence intervals are given in Figure 3-8.

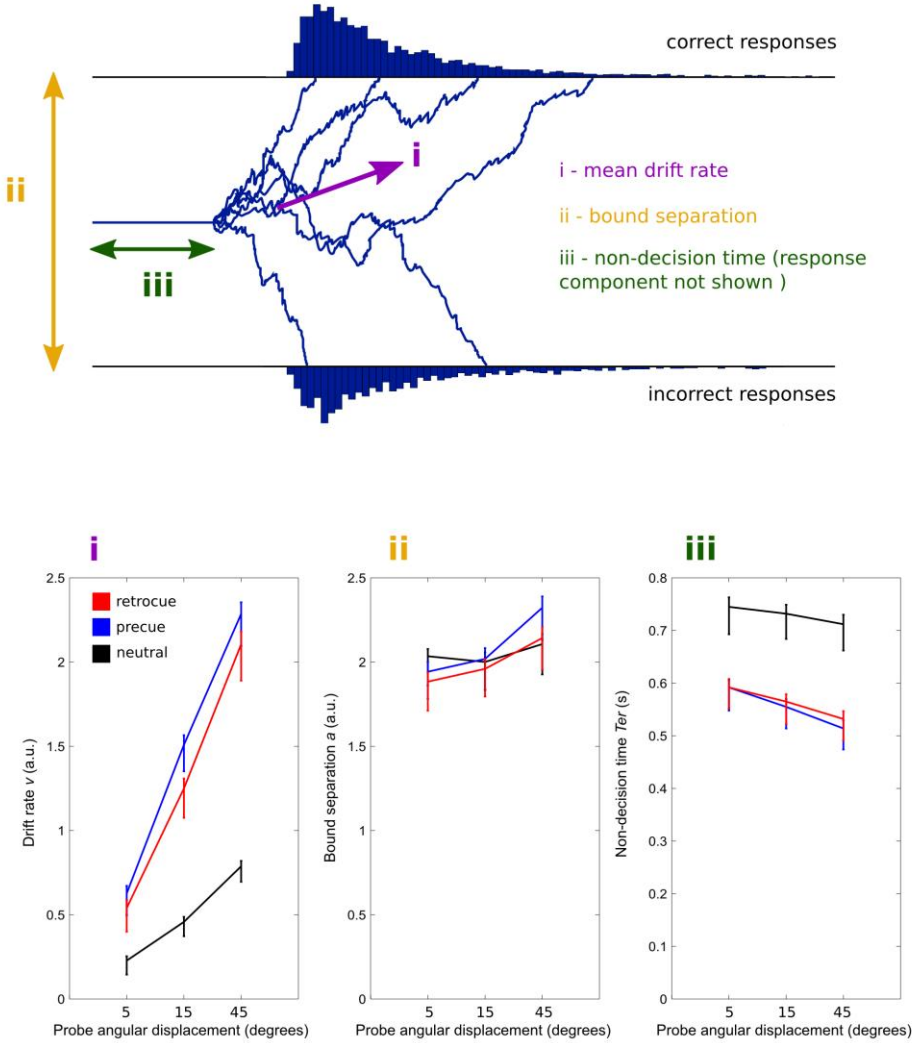


Figure 3-8 **Key parameters from the diffusion model**. **The top panel** illustrates the essential parameters of the diffusion model. Six instances of a decision variable diffusing towards to one or other bound and triggering a response are shown. The mean drift rate (i) determines the ratio of correct/incorrect trials, along with the bound separation. The bound separation (ii) also plays a role in determining the ratio of correct/incorrect trials. The non-decision time (iii) has both a 'sensory' component (occurring before the diffusion process begins) and a 'motor' component (occurring once the diffusion process is complete), though this is simplified to show just the sensory part above. All three parameters influence the reaction-time distributions. **The bottom panels** show the parameters fitted to the current data by the diffusion model. Key effects of precues and retrocues are to increase mean drift rate (i) and decrease non-decision time (iii). Error bars are 2.5% upper and lower confidence bounds based on the inferred group posterior distribution for each parameter.

Drift rate co-varies with the magnitude of angular displacement of the probe, and is much larger for precue and retrocue trials than for neutral trials. Bound separation is

not different between the cue conditions. There is a clear difference in non-decision time between the neutral and cued trials, suggesting that an additional retrieval process is interposed between probe presentation and response when the identity of the probed item is not known in advance. The mean difference in non-decision time was 176ms between the precue and neutral trials, and 167ms between retrocue and neutral.

To check how well the diffusion model fitted the data, data were simulated on the basis of the fitted parameters (as described in section 3.2.4 above). If the diffusion model fit is good, this simulated data should resemble the observed data. The predicted and observed reaction time distributions are shown in Figure 3-9.

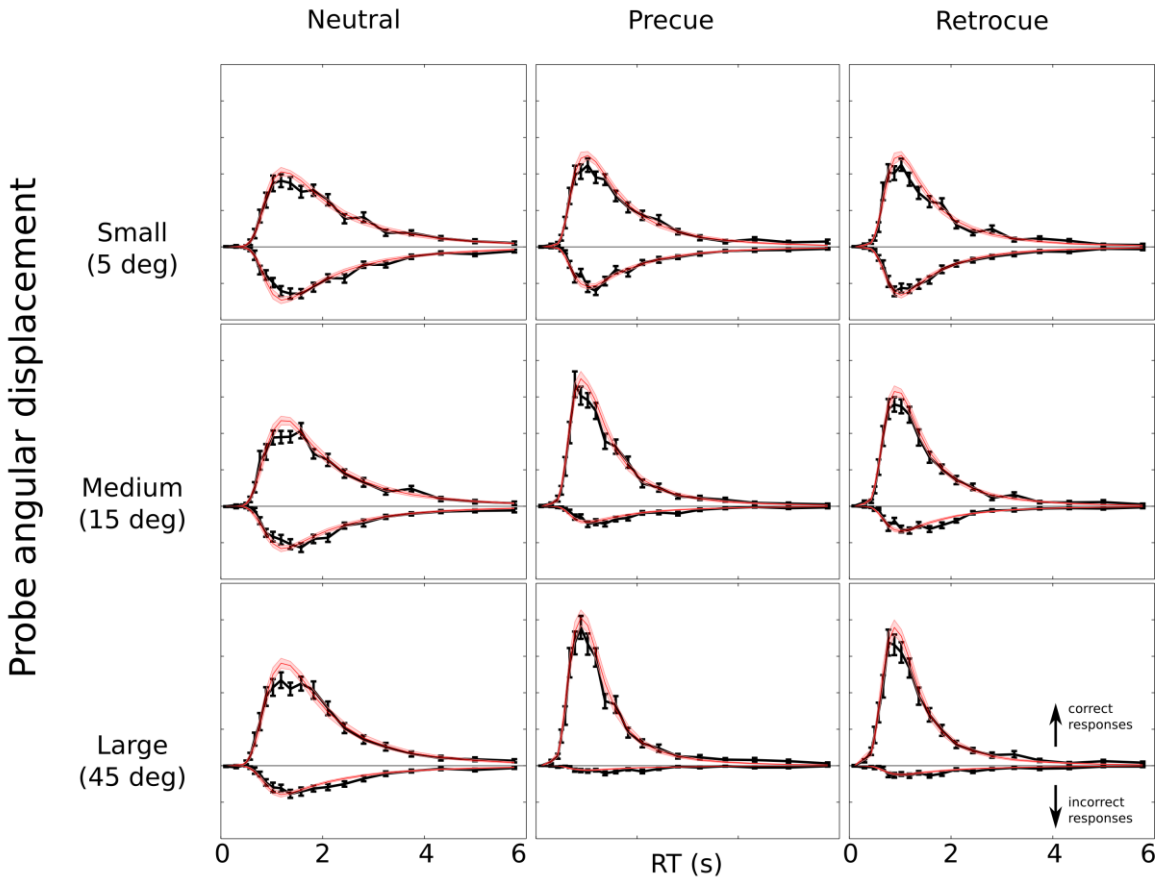


Figure 3-9 **Group-averaged RT histograms (black lines), and group-averaged predicted RT distributions from the diffusion model (red shaded).** Incorrect response RT histograms are flipped negative. The diffusion model adequately predicts the true RT distributions, though it slightly overestimates the proportion of early responses, particularly for incorrect responses in the neutral condition.

The mixture model, described above, is fitted to the accuracy data only. A posterior predictive check for the mixture model can be run in order to compare the predictions the two different approaches make for the accuracy data. The predicted accuracy data from each model, for each condition/angular displacement bin, is shown together with the observed accuracy data in Figure 3-10. The group-level means of the subject-wise median RTs (i.e. a summary metric for the distributions shown in Figure 3-9) are also shown.

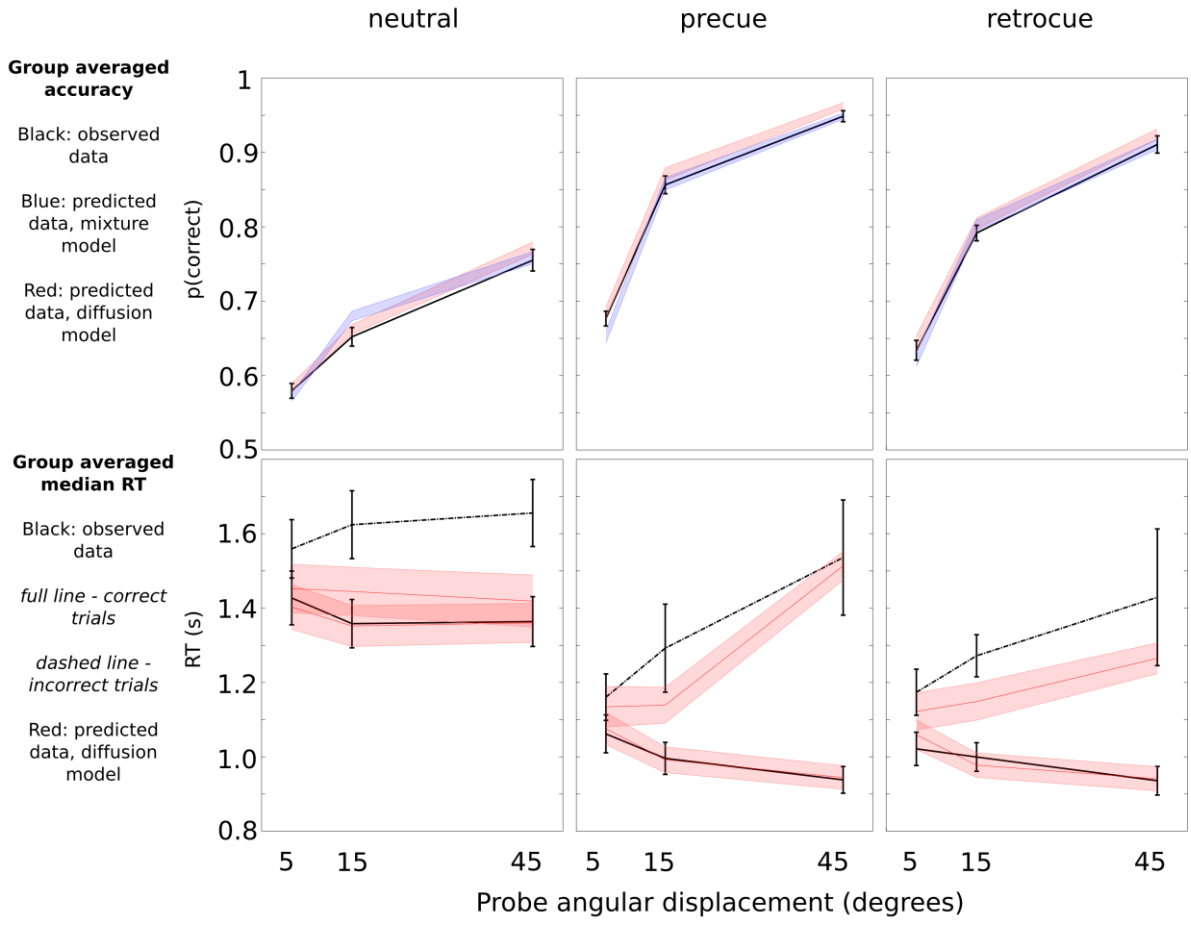


Figure 3-10 **Group level accuracy and mean RT, alongside model predictions.** **Top panels**: black lines, accuracy data. Red shading, diffusion model predictions for the accuracy data. Blue shading, mixture model predictions for the accuracy data. The two rather different models make similar predictions for the shape of the accuracy curves, and neither model describes appreciably more variance. **Bottom panels**: full black line, correct trial RT means; dotted black line, incorrect trial RT means. Red shading, RT predictions from the diffusion model. The diffusion model predicts median RT in correct trials successfully. It is less successful in predicting the median RT for the error trials, although it does predict the increase in RT with increasing shift magnitude for the two cued conditions.

The diffusion model predicts reaction time distributions and median RT reasonably well, though it tends to underestimate median RTs for the incorrect trials (in the neutral cue condition particularly). Interestingly, for the precue and retrocue conditions the diffusion model predicts the observed slowing of RT with increasing shift magnitude in the incorrect trials. The diffusion model predicts this slowing by the same logic as described above, by which cross-trial variability in drift rate gives rise to slower error trials than correct trials. When the magnitude of the angular displacement of the probe item is larger, mean drift rate is higher. Error responses become less probable, and sample the more extreme end of the across-trial distribution of drift-rates. This has the net effect of slowing error RTs.

3.4 Discussion

The mixture model (W. Zhang & Luck, 2008) is a well established model for precision/capacity working memory data. In the current study, the mixture model was fitted to data from a task in which a single item was cued either in advance of or following the memory array. This replicated the prior finding that precues affect memory precision as well as guess rate (Gorgoraptis et al., 2011) whereas retrocues primarily affect guess rate (Murray et al., 2013). Reaction times were also faster for both precues and retrocues, confirming previous reports (e.g. Astle et al., 2011; Griffin & Nobre, 2003; Rerko et al., 2014). However, this reaction-time effect has not been modelled previously. Comparing the cued and uncued RT distributions, cueing appears both to decrease the variance in RT but also to shift the entire curve towards shorter RTs. This suggests that both cue types eliminate a fixed component of RT otherwise present on uncued trials. In order to test this quantitatively, a drift diffusion model was fitted to the data. The change in RT was found to result in part from a reduction in non-decision time, of similar magnitude for precues and retrocues (corresponding to the leftward shift in the distribution). Additionally, evidence accumulation was faster in cued trials, resulting in higher accuracy and an additional reduction in RT.

The means by which retrocues boost performance is debated. Some have suggested that retrocues improve the quality of representation of the cued item in working memory, perhaps due to the consolidation of information stored in a fragile format (Sligte et al., 2008) that would otherwise be overwritten by the probe stimulus. An alternative explanation is that rather than changing the representation *per se*, retrocues allow people to resolve competition between items in memory for action control, in advance of the presentation of the probe stimulus. A recent proposal is that rather

than being equally attended and equally available to guide action/cognition (Cowan, 2000), at any one time, a single item in working memory is prioritized and rendered 'active' whereas the remainder of the items in memory are in a 'passive' state (LaRocque et al., 2014; Oberauer & Hein, 2012; Olivers et al., 2011). Retrocues may trigger the promotion of the cued item to the active state. The results presented here are consistent with this proposal. When a single item is cued, either prospectively or retrospectively, reaction time to the probe item is reduced by a fixed amount. This suggests that the probed item is not immediately accessible on neutral trials, requiring an additional retrieval or orienting process following the presentation of the probe item that can be circumvented when the probed item is cued in advance. This finding is conceptually close to prior findings from task switching (Garavan, 1998) and memory for word lists (McElree, 2006), in which a reaction-time cost is seen when subjects need to switch between items in memory to make a response. However, the visual WM literature has not typically considered re-orienting between maintained representations to be a limiting factor on performance, instead focussing on the representational content of short-term memory. A contemporary view holds that short-term memory maintenance consists in sustained attention to a small set of internal representations (Awh & Jonides, 2001; Cowan, 2000; Gazzaley & Nobre, 2012; Kiyonaga & Egner, 2012) with the capacity limit considered identical with the scope of attention (D. E. Anderson et al., 2013), a view that is inconsistent with a cost to switching between representations in the short-term store. Similarly, changes in representational content, either through retrieval of information from a fragile buffer (Sligte et al., 2008) or by biasing competition between representations in memory (Edin et al., 2009; Murray et al., 2013) have been proposed to explain the retrocue benefit.

Why does cueing affect accuracy, as well as RT? We might expect that if retrocues do not alter the content of memory, only changing the functional status of the cued item,

then people would take longer to retrieve an item when they had not received a cue, but they would be able to retrieve it eventually. However, this intuitive view may underestimate the degree to which retrieval and response-selection processes limit performance. One possibility is that even if the probed item is represented in memory when the probe stimulus is presented, there is more likely to be a failure of output gating (Hazy et al., 2007) if the probe item must be processed and the memory item retrieved all at once. One way in which output gating could fail is if the wrong item is selected to guide action. The misgating (or 'mis-selection') results presented above support this idea. Unprobed items were much more likely to influence responding in neutral cue trials than in precue or retrocue trials. There was also some evidence that the effect of unprobed items was greater for items that had been closer to the probed item (horizontally adjacent or vertically adjacent, as compared to diagonally opposite; Figure 3-5). This is in line with neural evidence suggesting that items which were presented in a spatial array are represented in memory on a corresponding spatial map (Kuo et al., 2014; Munneke, Belopolsky, & Theeuwes, 2012; Sligte et al., 2009). Memory items may be more confusable at retrieval if they are closer in 'mnemonic space'.

Two models were used to analyse these data that make different assumptions about the processes underlying performance. The mixture model splits trials into guess trials and non-guess trials, whereas the diffusion model explains accuracy differences between cue conditions in terms of difference in strength of evidence. However, alternatives to the mixture model have been proposed on the basis of accuracy data alone. The first challenge to a discrete-resource model proposed that a continuous short-term memory resource is shared between however many items are in memory (Bays & Husain, 2008). This did not account well for the observed data with large set sizes (W. Zhang & Luck, 2008), being outperformed by a slots-plus-averaging model.

More sophisticated models have recently been proposed which can account for apparent guessing with large set sizes. Van den Berg et al. (van den Berg, Shin, Chou, George, & Ma, 2012) proposed a model in which the precision with which a given item is encoded varies across items and trials, but on average decreases with larger set sizes. The authors systematically compared a number of proposed models. Their variable precision model outperformed the slots-plus-averaging model in a formal model comparison. A key point from the paper by van den Berg and colleagues (2012) is that the slots-plus-averaging and variable-precision models in fact make rather similar predictions about the shape of the response distributions. If drift rate in the diffusion model is taken to index evidence accumulation (Ratcliff & McKoon, 2008), and the rate of evidence accumulation corresponds to the fidelity with which an item is stored in memory, then the diffusion model is roughly analogous to this variable-precision model.

Realistically, neither model is likely to fully capture the complex processes that mediate the working memory task we used. Parameter estimates from each model help capture different aspects of the behavioural data, but should not be taken to correspond directly to underlying cognitive components. The conceptual model advocated above, allowing for different states in memory and failures of output gating, goes beyond the slot-model assumptions that motivated the mixture model. The diffusion model allowed us to dissect the RT distributions and discover an additional retrieval time in uncued trials, but cannot account for phenomena like choosing the incorrect item to guide responding. A more complex accumulator model, incorporating a biased race between memory items for control of action, might be able to capture both reaction time components and the effect of unprobed items on behaviour. Such a model would have the attractive property of making explicit the role of output gating in determining performance. In race models, the number of competing items correlates positively with reaction time, if units representing items inhibit one another (Usher & McClelland, 2001). Future

research, gathering large number of trials per subject, could help construct a more realistic accumulator model that explains performance in terms of both of output gating (which item in memory should guide behaviour?) and the decision process itself (comparison of memory content with the probe item). In general, combining the DDM with a real-time brain recording (e.g. MEG) could also help develop a mechanistic understanding of how retro-cueing modulates these behavioural parameters. Analysing the interval between probe presentation and response might reveal sequential processes associated with retrieval and evidence accumulation. Their timing and topography could be compared across cue condition. A similar analysis has been reported for value-based decision making, in which a biophysical winner-takes-all decision network predicted mass-action dynamics similar to those observed in ventro-medial PFC (Hunt et al., 2012).

4 Alpha-power modulation reveals transient reactivation of perceptual cortex during mnemonic selection

Chapter abstract

Both precues and retrocues can boost working memory performance (chapter 3). Electrophysiological recordings give us a window onto the brain mechanisms of cued selection that mediate this change in performance. Preparatory attention is known to modulate the pattern of alpha power in occipital and parietal cortex. In this chapter the effects of precues and retrocues on alpha oscillations in posterior cortex are compared. Retrocueing subjects to one of a number of items already in memory is shown to evoke similar spatiotopic patterns of alpha power modulation as have previously been associated with preparatory attention. The good spatial resolution afforded by MEG was exploited to dissociate changes in alpha power in parietal and occipital cortex, and to demonstrate that the pattern of alpha activity in response to cue stimuli is specific to the retinotopic location of the cued item resembling the induced response to a physical stimulus. The time-course of alpha modulation was characterized, showing that whilst precues induce a sustained alteration in alpha power that persists until the presentation of the memory array, retrocues give rise to a transient modulation of the pattern of alpha power in perceptual cortex, consistent with a brief reactivation of the cued stimulus in memory. This has implications for the role of attention in short-term memory, constituting evidence against the hypothesis that working memory consists in sustained attention to internal representations, and supporting the hypothesis that perceptual cortex is instead transiently recruited to furnish perceptual content during re-prioritization of a working memory representation.

4.1 Introduction

4.1.1 Working memory as sustained attention to internal representations

A standard hypothesis is that working memory maintenance consists in an internally-directed focus of the same attentional mechanisms that are responsible for biasing perceptual attention (Awh et al., 2006; Awh & Jonides, 2001; Chun, 2011; Cowan, 2000; Gazzaley & Nobre, 2012). This view is party motivated by the observation that preparatory attention and working memory appear to share neural substrates. One line of evidence has come from the retrocue paradigm: retrocues give rise to similar changes in neural measures as do preparatory attentional cues. Several fMRI studies (Lepsien & Nobre, 2006; Nobre et al., 2004) have shown that retrocues activate a fronto-parietal network that has been associated with prospective attentional control (Corbetta, 1998; Nobre et al., 1997), suggesting that similar top-down control mechanisms may be engaged in internal and perceptual selection. EEG measures also suggest neural overlap: Griffin and colleagues (Griffin & Nobre, 2003) showed that the evoked response to both precues and retrocues were similarly lateralized in parieto-occipital electrodes, with a larger N1 negativity contralateral to the cued side. On the basis of these similarities, selection within working memory has been suggested to result from biasing of competition between maintained representations analogous to the biasing signal proposed to mediate perceptual attention (Desimone & Duncan, 1995; Gazzaley & Nobre, 2012). A close cousin of this idea is the hypothesis that perceptual cortex is involved in the maintenance of working memory content. For example, multi-voxel pattern analysis (MVPA) of fMRI data has been presented as evidence in favour of a role for perceptual cortex in WM maintenance. Harrison and colleagues

(Harrison & Tong, 2009) found that during the retention interval of a working memory task, MVPA of visual cortex activity could decode the orientation of a grating stimulus held in memory. This was argued to reflect active maintenance of the memory representation in visual cortex.

Perceptual attention and shifts of focus in working memory share neural correlates, and perceptual cortex seems to be in some recruited in some capacity during WM tasks. However, this does not necessarily imply that mechanisms of perceptual attention are identical with mechanisms of memory maintenance, or that delay activity in perceptual cortex is the basis for memory maintenance. Likewise, the apparent neural overlap between perceptual attention and orienting within working memory does not necessarily imply that selection within working memory is mechanistically identical to biasing of perceptual input. An alternative view is that perceptual cortex is an important resource for working memory, but one that is only recruited transiently, when perceptual content associated with a given working memory item needs to be interrogated/retrieved. This dynamic account sees top-down recruitment of visual cortex occurring as and when required, with perceptual and attentional mechanisms otherwise freed up for on-going processing during the maintenance interval (Hollingworth & Maxcey-Richard, 2013).

The temporal resolution of fMRI is limited by the slow (~4s) time-course of the hemodynamic response function (HRF), which makes it a poor tool to characterize the rapid temporal dynamics of top-down control predicted in this framework. However, M/EEG recordings have a much higher temporal resolution, sampling at the millisecond scale. Conveniently, top-down modulation associated with preparatory attention also has a well-characterized and reliably observed correlate in these modalities in the form of modulation of the alpha band (10Hz) oscillation. In the current study, MEG was used

to investigate the spatio-temporal effects of retrocues upon the distribution of alpha power in the perceptual and parietal cortex.

4.1.2 Alpha oscillations and selective attention

The alpha or Berger rhythm at around 10Hz is a major oscillatory signal in the human brain, constituting the most prominent component in the M/EEG power spectrum, and the primary resonant frequency of visual cortex as indexed using flickering stimuli (Herrmann, 2001). Parieto-occipital alpha has long been known to change with arousal and stimulation, synchronizing when subjects close their eyes, and desynchronizing during visual processing (Adrian & Matthews, 1934). This suppression of parieto-occipital alpha power with stimulation is termed event-related desynchronization (ERD), and contrasts with the event-related synchronization (ERS) typically seen in the theta and gamma bands (Klimesch, 2012). Alpha has therefore been thought of as the idling rhythm of the cortex, alpha power decreasing when the cortex is involved in active processing (Pfurtscheller, Stancák, & Neuper, 1996).

Alpha power is modulated by attention, decreasing in the parts of cortex that correspond to the focus of attention. For example, if a task requires attention to the auditory modality but not visual, alpha power will be increased in visual cortex in comparison to a task requiring attention to the visual modality but not auditory (Foxe, Simpson, & Ahlfors, 1998). Within modality, alpha power lateralizes if preparatory attention is directed towards one side of space, with a relative decrease in power in the contralateral (relevant) cortex, and an increase in ipsilateral (irrelevant) cortex. In visual tasks (Rihs, Michel, & Thut, 2007; Worden et al., 2000) the pattern of alpha modulation in the M/EEG sensor space reflects the quadrant to which attention was directed, suggesting that alpha power modulations in visual cortex are retinotopically

specific to the attended location in the visual field. Lateralization effects are not specific to visual cortex: directed attention also lateralizes alpha power (Haegens, Händel, & Jensen, 2011a; Haegens, Luther, & Jensen, 2012) in somatosensory cortex (in which context it is known as the mu rhythm). Alpha lateralization in the visual cortex has been shown to index the degree of certainty with which spatial attention is directed (Gould, Rushworth, & Nobre, 2011), indicating that alpha lateralization is under flexible top-down control, closely tracking the strength of top-down bias. Alpha modulation is also closely linked with behavioural effects of attention: a number of studies have demonstrated a correlation between the magnitude of alpha lateralization in directed-attention tasks and behavioural performance (Bonnefond & Jensen, 2012a; Gould et al., 2011; Haegens et al., 2012; Haegens, Händel, & Jensen, 2011a; Haegens, Osipova, Oostenveld, & Jensen, 2010; Händel, Haarmeier, & Jensen, 2011; Sauseng et al., 2009b; Thut, Nietzel, Brandt, & Pascual-Leone, 2006; van Ede et al., 2011). Spontaneous (uncued) alpha lateralization also modulates performance in detection tasks, and van Ede and colleagues (van Ede et al., 2011) have used this 'baseline' relationship to quantify the proportion of variance in behavioural performance explained by modulation of contralateral alpha power by cues, finding that around 30% of the behavioural benefit from cueing can be predicted from the associated changes in alpha lateralization.

Despite the extensive body of evidence on alpha modulation by preparatory attention, it is not clear exactly what cortical function alpha oscillations index. Pre-stimulus alpha desynchronization in the task-relevant parts of cortex is *prime facie* consistent with the original idling hypothesis of the alpha rhythm, as preparatory attention can increase the excitability of task-relevant cortical areas (Fries, Womelsdorf, Oostenveld, & Desimone, 2008; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Luck et al., 1997). The link between alpha power and cortical excitability has been investigated directly

with trans-cranial magnetic stimulation (TMS), in two key ways. The size of the motor evoked potential (MEP) resulting from a TMS pulse to the motor cortex is a direct index of excitability in the motor system. Sauseng and colleagues correlated the size of MEPs with pre-pulse EEG power in a range of frequency bands, and found that increased alpha power in the stimulated area of motor cortex predicted a smaller MEP (Sauseng, Klimesch, Gerloff, & Hummel, 2009a). TMS can also be used to entrain oscillations in cortex, giving a means to directly probe the causality in the otherwise correlative link between alpha power and excitability. Romei and colleagues found that unilateral TMS stimulation at 10Hz impaired detection of a low-contrast stimulus on the contralateral side, and facilitated detection on the ipsilateral side, but the same effect was not seen for stimulation at 5Hz or 20Hz (Romei, Gross, & Thut, 2010). This finding suggests that the alpha oscillation is not merely an epiphenomenal correlate of cortical idling, but directly contributes to, or necessarily induces, cortical inhibition. This is borne out in electrophysiological work in non-human primates (NHPs). Haegens et al. (2011) have shown that in macaque premotor cortex, alpha power is negatively correlated with spike rate, and alpha phase also predicts spike rate with more spiking in the troughs of the alpha oscillation (Haegens, Nácher, Luna, Romo, & Jensen, 2011b). However, this was not true of every brain region – in primary somatosensory cortex a U-shaped curve of firing rate with alpha power was observed. As discussed at the end of this chapter, the 10Hz oscillation may mediate inhibition in a functional as opposed to a literal sense, as an 'offline' state for cortical regions not currently involved in processing (Hanslmayr, Staudigl, & Fellner, 2012).

On the basis of the widely posited link between alpha and inhibition, several investigators have proposed that attentional modulation of the alpha rhythm may not merely index top-down excitation, but may be part of an active top-down inhibitory mechanism (Bonnefond & Jensen, 2012b; Payne, Guillory, & Sekuler, 2013). Jensen has

suggested that the brain controls information transmission from sensory cortex by changing the pattern of inhibition (Ole Jensen, 2010) to block irrelevant input, and this inhibition is mediated by alpha synchronization. A number of studies have shown that cueing attention to one side of space results in an ipsilateral increase in alpha power, as compared to a baseline period before the cue was presented (S. P. Kelly, Lalor, Reilly, & Foxe, 2006; Rihs et al., 2007). This ipsilateral synchronization has been argued to reflect active inhibition of distracting input. However, a purely facilitatory top-down attentional influence could give rise to a boost in inhibition at unattended sites by low-level mechanisms within perceptual cortex. This 'see-saw' effect might extend across the hemispheres: enhanced target detection has been observed in the ipsilateral visual field to TMS virtual lesions of the parietal cortex (Hilgetag, Théoret, & Pascual-Leone, 2001), an effect replicated by Romei and colleagues who found that stimulus detection was facilitated on the ipsilateral side to alpha-frequency TMS (Romei et al., 2010). The fact that ipsilateral synchronization is observed even when there are no distractors to inhibit (S. P. Kelly et al., 2006) is consistent with such an automatic 'see-saw' mechanism. It is also difficult to say what the appropriate baseline for alpha power should be, even when visual stimulation is equated before and after an attentional cue. If alpha power in the contralateral cortex increases following the attentional cue in a directed attention task, this could either reflect active inhibition of the irrelevant neural ensembles, or a return towards a default idling state from a state of enhanced excitability. The default state during task performance may be for visual cortex to be in a higher state of excitability than during the inter-trial interval, with processing resources evenly allocated over the whole visual field. Synchronization of alpha activity in the ipsilateral hemisphere after an attentional cue may therefore represent a return towards the more synchronized resting baseline as attention is

withdrawn from the task-irrelevant part of the visual field, not an active inhibitory signal *per se.*

4.1.3 Alpha oscillations and retrocueing

The majority of studies investigating the link between alpha oscillations and attention have manipulated preparatory attention for upcoming sensory stimuli. By contrast, in the current task there are two types of selection cue. Precues were expected to replicate the previously observed pattern of alpha lateralization with preparatory attention. Given prior evidence that retrocues recruit similar mechanisms of top-down control as are associated with preparatory attention, it was predicted that alpha in parieto-occipital cortex would lateralize in response to a retrocue in a similar manner. To date, no study has investigated alpha lateralization following retrocues during WM maintenance. Parieto-occipital alpha lateralization has been investigated in the context of retrieving items from long-term memory (Waldhauser, Johansson, & Hanslmayr, 2012), but this study used lateralized retrieval cues which could themselves have driven alpha lateralization. We predicted that alpha lateralization would be observed following a central symbolic retrocue, and in the absence of lateralized visual stimulation.

This was not a foregone conclusion; it has also been proposed that when attention is directed to internal representations, alpha power should increase in perceptual cortex, 'blocking out' perceptual input in favour of internal processing (Cooper, Croft, Dominey, Burgess, & Gruzelier, 2003). From that perspective, retrocues might be expected to boost parieto-occipital alpha in order to block out visual distraction whilst the content of memory is manipulated, or even give rise to a reversed pattern of lateralization in visual cortex, to protect memory representations in downstream regions such as parietal

cortex from interfering input. However, to preview the results alpha lateralized as predicted following a retrocue.

Preparatory alpha lateralization/desynchronization tracks temporal expectations about the time of appearance of the cued item (Rohenkohl & Nobre, 2011; van Ede et al., 2011), implying that the time-course of lateralization closely tracks the time-course of top-down attentional allocation. The time-course of alpha modulation in response to retrocues was therefore taken to index the time-course of top-down modulation of sensory cortex in response to a retrocue. Cowan and others have suggested that working memory maintenance consists in sustained attention to internal representations (D. E. Anderson et al., 2013; Awh & Jonides, 2001; Kiyonaga & Egner, 2012). If this were the case then we might expect the alpha lateralization in response to a retrocue to persist throughout the remainder of the delay period, reflecting a sustained shift in attention (c.f. contralateral delay activity, CDA; (Vogel & Machizawa, 2004)). A second possibility is that alpha lateralization following a retrocue is associated with preparatory attention for the upcoming probe item, and should ramp up towards the moment at which the probe appears (Stokes, 2011). A final possibility is that if short term memories are maintained independently of perceptual cortex, then changes in alpha power in perceptual cortex might be more punctate, reflecting a short-lived reactivation of the perceptual trace associated with the cued item, in order to 'reconfigure' short-term memory.

4.2 Methods

4.2.1 MEG recordings

Fifty subjects participated in the experiment (26 female, 24 male; mean age 24 years, range 19 to 34). All subjects were healthy, had normal or corrected-to-normal vision, and were right-handed. Ethical approval was obtained from the NHS South Central Berkshire ethics committee (11/SC/0053). Of these subjects, one performed at chance in the behavioural task, four made eye movements during the task, and one was too drowsy to complete the MEG session. A further six subjects were rejected because their MEG data was of poor quality (one due to failed co-registration, five due to severe contamination by artefacts). Data from 38 subjects were therefore retained for the task analysis.

The precue/retrocue task used here is described in chapter 3. Subjects performed this task whilst MEG data were recorded, having already completed a training session in a previous visit to the lab. The MEG system was a Neuromag Elekta system with 204 planar gradiometers and 102 magnetometers (further details are given in the methods chapter).

Eye and heart artefacts were recorded during MEG acquisition. A vertical pair of EOG electrodes were placed above and below the left eye to detect blinks, and a horizontal pair was applied with one electrode to the left of the left orbit, and the second electrode in the corresponding position on the right, in order to detect lateral eye movements. An electrode was applied to both the left and right wrist in order to record the ECG. The subject's head position was monitored during the scan using an active magnetic signal. Four magnetic emitting coils (head position indicator, HPI coils) were taped to

the subject's head, one behind each ear, and one on each side of the forehead. The position of these coils relative to the nasion and two peri-auricular landmarks was digitised using a Polhemus 3D tracking system (Polhemus, EastTrach 3D). Finally the Polhemus probe was used to obtain a set of ~100 points recording the shape of the subject's scalp surface.

Once subjects were seated in the MEG system, the chair was raised until the top of the subject's scalp made contact with the top of the MEG cap. Subjects were provided with a fibre-optic button box with which to make responses. They held this in their right hands, making responses by lifting their index or middle finger. Eye movements were tracked at 1000Hz using a binocular infra-red video eyetracker (Eyelink 1000). Stimuli were projected onto a screen around 85cm in front of the subject. The task is described in detail in a previous chapter. WM stimuli were presented at 6 degrees visual angle eccentricity from the fixation cross, and each stimulus subtended 1.2 degrees visual angle.

MEG data were recorded in three blocks. The subject's head position was acquired at the beginning of each recording. During the recording, the HPI coils continuously transmit their locations. This signal was used to compensate for head movements during the scan. Subjects completed three blocks in each recording block, for a total of nine blocks of the task (36 trials per block, 324 trials in total).

4.2.2 Computing forward models from structural MRIs

A structural MRI was acquired for each subject using a Siemens 3T scanner (OCMR, John Radcliffe Hospital). SPM's *spm_eeg_inv_mesh* was used to compute the transformation that mapped a set of canonical meshes for the cortical surface, skull and scalp to each subject's individual anatomical MRI, and this transformation was used to define a

forward model tailored to each subject's head shape. A single shell forward model was used in beamformer source reconstruction (Nolte, 2003).

Co-registration computes the spatial relationship between the head model and the HPI points, which are in turn localized in the scanner space by the scanner software. This links the head model to the MEG space. Co-registration between the head model and the HPI was initialized using SPM's canonical fiducial points, which were matched to the fiducial positions digitized using the Polhemus system. This first approximation was optimised by an iterative robust fit of the scalp mesh surface (from the structural scan) to the scalp surface points recorded with the Polhemus whilst setting up the subject for the MEG.

4.2.3 MEG preprocessing

The MEG preprocessing pipeline is described in detail in the methods chapter. In brief, the data were first inspected to remove channels corrupted by noise associated with the MEG system (e.g. digital artefacts such as square wave jumps), and then denoised using Elekta's Maxfilter Signal Space Separation (SSS) algorithm. This decomposes the data into a set of spherical harmonic basis functions, and rejects any components coming from outside of a sphere defined around the subject's head (i.e. environmental noise), before re-projecting the data onto the MEG sensors. This last step can be used to compensate for head movements, by transforming the position of the spherical basic function representation relative to the sensors before re-projecting the data. Head movements were compensated for within-session using the continuous HPI information.

Data were then epoched and visually inspected for artefacts using Fieldtrip's visual artefact rejection tool. Bad trials and channels were tagged on the basis of abnormal variance, kurtosis and maxima/minima in the time-domain data. Eyeblinks were

detected from the EOG and eye-tracker data using a semi-automatic algorithm, and data from 200ms prior to and until 300ms following each blink was not used in any part of the subsequent analysis, including estimation of the beamformer weights.

4.2.4 Task epochs

The task is described in detail in chapter 3. The MEG analyses are RAM and time intensive, so to make them more manageable, the trials were divided into a series of shorter epochs (Figure 4-1) and analyses were run separately on these epochs. Where timings are given in the text unless otherwise stated they are relative to the precue, array or retrocue (depending on the epoch in question).

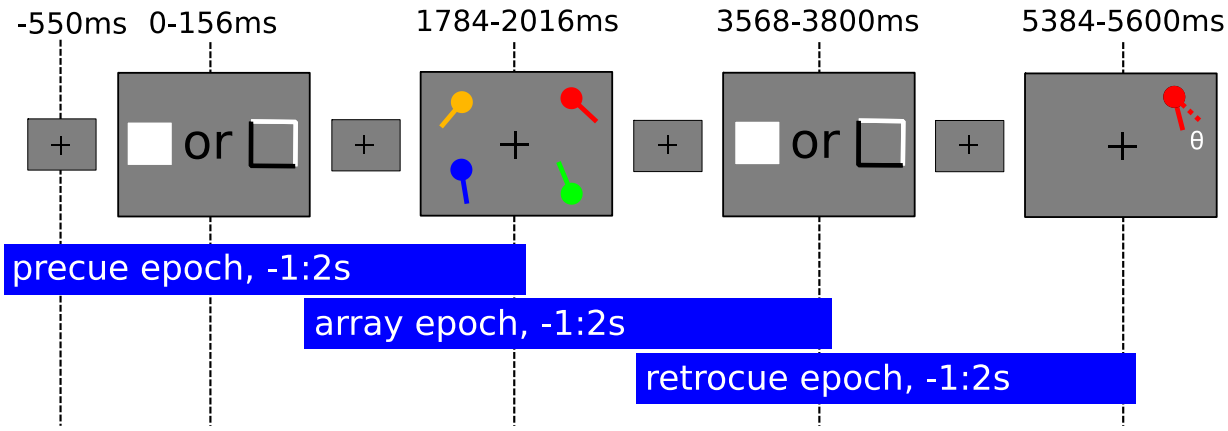


Figure 4-1 **Trial timings, and analysis epochs**. Onset and offset times are given for the cues, WM array and probe. In the text, timings are within-epoch; i.e. relative to the event defining the epoch.

4.2.5 Sensor space analyses

4.2.5.1 Group-averaged topographies

The time-domain sensor-space signal from both the interval between the precue and the working memory array ('precue interval'), and the interval following the retrocue

('retrocue interval') was transformed to the frequency domain in 50ms steps, using a Hanning taper/FFT algorithm with a taper spanning 4 cycles of the filtered frequency, for frequencies between 3 and 30Hz in 1Hz steps. The resulting power spectra were averaged over trials within each cue condition. The power time-series in the planar gradiometer pairs were then combined and the magnetometers discarded, giving a 102 channel combined planar gradiometer map of sensor space power. The between condition subtractions [precue left minus precue right] and [retrocue left minus retrocue right] were computed within-subject for the precue and retrocue epochs respectively.

Sensor-space cluster permutation statistics (2000 permutations) were computed for these topographies by permuting cue left/cue right condition labels (using Fieldtrip's *ft_freqstatistics*). Clusters were formed in space (by sensor proximity) and time. The data from each trial were subdivided into three epochs (precue epoch, array epoch, retrocue epoch) as described in the methods chapter; permutation tests were run over each of these epochs separately, over the time range 0-1.7s for each epoch.

4.2.5.2 Classifying alpha topographies to decode cue direction

A pattern classification approach was used to characterise the cue-direction information carried in the alpha-band topography. A single trial was set aside, and for the remaining trials, the planar gradiometer topography of power in a given time-frequency bin was averaged separately for trials in which attention was cued to each of the four quadrants. The topography in the left-out trial was correlated with each of these four averages and the correlation values Fisher transformed, yielding four numbers representing that trial's similarity to each of the quadrant-averaged patterns. This leave-one-out analysis was repeated for all trials, and for all time-frequency bins within an epoch for frequencies between 3 and 30Hz in 1Hz steps. These were converted into

a classification score by classifying each trial into one of the quadrant-cueing groups, on the basis of whichever correlation was highest for that trial (for each time-frequency bin independently). In contrast to the sensor-space topographies presented above, the orthogonal gradiometers were not combined for this analysis, as the technique makes use of patterns within-subject, and is therefore robust to changes in source orientation between subjects.

Data were collapsed over trials by reorganizing the classification values relative to the quadrant that was cued on a given trial. Classification probabilities were sorted into four categories: match (proportion of times a left-out trial was classified correctly), misclassification within the same hemifield (the proportion of times trials were misclassified as having been cued to the quadrant in the same visual hemifield as the actually cued quadrant), misclassification with the horizontally apposed location, and misclassification with the diagonally apposed location. This summarizes classification performance as a confusion matrix between the quadrants. The classification values were normalized by subtracting chance classification (25% correct). Classification maps were converted into a t-statistic at the group level.

A similar analysis was performed in order to test whether or not the pattern of alpha responses to an attentional cue resembled the pattern of alpha desynchronization in response to a physical stimulus. For this analysis, it was first established that the alpha ERD to the probe stimulus permitted quadrant decoding by running a similar analysis to that described above for the cue epochs. Having established robust decoding of probe stimulus quadrant, the topographies were averaged between 0.3s and 0.5s post-cue, a time window for which classification was high in the alpha band, but which largely excludes the additional ERF decoding seen in the theta band for the probe analysis, between 0 and 0.3s post cue. These topographies were then correlated against the

trialwise topographies from the precue and retrocue epochs. The leave-one-out approach was used to avoid the possibility of autocorrelation in the signal on any given trial driving classification (e.g. due to slow drift in background noise across trials). The results were then converted into a confusion matrix as described above.

4.2.5.3 Cross-temporal correlation analysis

In order to establish the temporal stability of cue-induced brain states in an unbiased manner, a cross-temporal correlation analysis was performed (King & Dehaene, 2014). The full set of trials were randomly subdivided into two groups of equal size within-subjects (randomly discarding trials to equalize trial numbers if more trials had survived preprocessing in one condition than another). Within each of the two halves, data were averaged within-condition and contrasts were computed. This yielded two independent estimates of the topographies for each contrast (contrasts: [informative cue minus neutral cue], indexing stability of cue induced effects; [cue left minus cue right], indexing stability of cue laterality effects).

The analyses presented here focuses on the alpha band (8-12Hz) as this carried the most information about cue laterality. As in previous analyses (Stokes et al., 2013), one half of the data was designated the training data, and the other half the test data. The topographies were extracted from the training data for each time-point in the epoch. This topography was correlated with the topography at every time-point in the test data, building up a cross-temporal correlation matrix. If a state is transient, then it gives rise to high correlation values mainly on the diagonal of this matrix. By contrast, stable states will give rise to off-diagonal correlations (King and Dehaene, 2014). The data were split randomly, and each time this analysis is run a slightly different result will be obtained. The analysis was therefore performed 20 times and the results averaged. This bootstrapping procedure stabilizes the estimate of the correlation structure.

4.2.6 Source space analyses

Data were mapped to the source space using an LCMV beamformer (Van Veen et al., 1998) with data covariance regularized using the top 59 PCA components of the data (Woolrich et al., 2011). The beamformer algorithm (described in more detail in the methods chapter) constructs a spatial filter for each point in brain space based on the covariance structure of the data. Sensor-space data was band-pass filtered between 3 and 30Hz prior to covariance estimation. Sections of data containing eye-blinks were excluded from the beamformer weights estimation, as described in the methods chapter. We constructed source space maps by beamforming activity from points arranged on a square 8mm grid in the brain space.

In order to characterise the time-course of alpha lateralization, virtual electrodes were defined for each of the occipito-parietal ROIs, and time-series for each virtual electrode were transformed into a time-frequency representation using Fieldtrip's Hanning taper algorithm, choosing a Hanning window of 4 cycles of the underlying frequency, for frequencies between 3 and 30Hz in 1Hz steps. The TFRs were averaged across task conditions within-subject, masking out any times associated with eyeblinks (by setting data from these periods to NaN, and using MATLAB's *nanmean*). The condition-averages were subtracted within-subject for the [precue left minus precue right] and [retrocue left minus retrocue right] conditions. These data were averaged at the group-level to create time-frequency maps of cue-related activity for each ROI. Significance testing was performed by forming clusters in the time and frequency dimension for each ROI and comparing cluster area to one-sample permutation distributions of cluster area (derived by sign-flipping the subtractions for 50% of subjects).

4.2.7 Investigating correlations between alpha lateralization and behaviour

In order to investigate the correlation between alpha lateralization and behavioural performance, a lateralization index was defined that had a higher value when the pattern of lateralization on a given trial was congruent with the hemifield in which the probe stimulus subsequently appeared, and lower when the pattern of lateralization was incongruent. The sensor space classification analysis was used to extract the lateralization measure, as this analysis inherently adapts to the structure of each subject's data, in contrast to the sensor and source space group analyses which are susceptible to spatial variability across subjects. The lateralization measure for each trial was derived by subtracting that trial's correlation with the mean cue right topography from its correlation with the cue left topography, creating a measure that is positive for trials in which the data resembled the average data from cue left trials, and negative for trials in which the data resembled the average data from cue right trials. The sign of this measure was inverted for trials in which the right hemisphere was subsequently probed, to yield a measure that would be positive when the pattern of activity was 'congruent' with the subsequently cued side.

The relationship between the trialwise lateralization measure and trial-by-trial performance (in terms of accuracy and reaction time) was investigated for each condition separately (NB for the neutral condition, this amounts to testing whether or not spontaneous fluctuations in alpha lateralization influence performance). We tested the relationship between the trialwise measure and accuracy (correct/incorrect) using a logistic regression, in which the orientation shift magnitude of the probe stimulus at the end of each trial (relative to the memory stimulus) was also included as a regressor of no interest, to factor out the effect of this parameter on accuracy. Similarly, a linear

regression was used to test the relationship between trialwise measures and reaction time, again including orientation shift magnitude as a regressor of no interest. In both cases, both the trialwise and shift magnitude regressors were normalized to zero mean and unit standard deviation. This analysis was repeated for each time-frequency point. One-sample cluster permutation tests (sign-flipping 50% of subjectwise beta maps, 5000 permutations forming clusters in frequency and time) were used to test for significant deviations of the beta-weights from zero at the group level.

4.3 Results

4.3.1 Alpha lateralization in response to prospective and retrospective cues

4.3.1.1 Mass-univariate analysis

Previous work on preparatory attention using M/EEG has revealed a characteristic pattern of activity in the alpha band when subjects are cued to attend to one side of visual space. Here, the induced response to precues in the current task is shown to replicate this pattern. This is compared to the induced response to a retrocue to establish whether or not a similar pattern of activity is observed when subjects were cued to attend to an internal representation.

A sensor space analysis was performed first. For the contrast precue left minus precue right, the expected pattern of right hemisphere desynchronization and left hemisphere synchronization over parieto-occipital sensors emerged. The evolving topography is shown in Figure 4-2, in which sensors belonging to a significant cluster are marked with a cross. There was a significant positive cluster over the LHS parieto-occipital sensors ($p = 0.0005$). The right hemisphere desynchronization did not reach significance ($p = 0.10$).

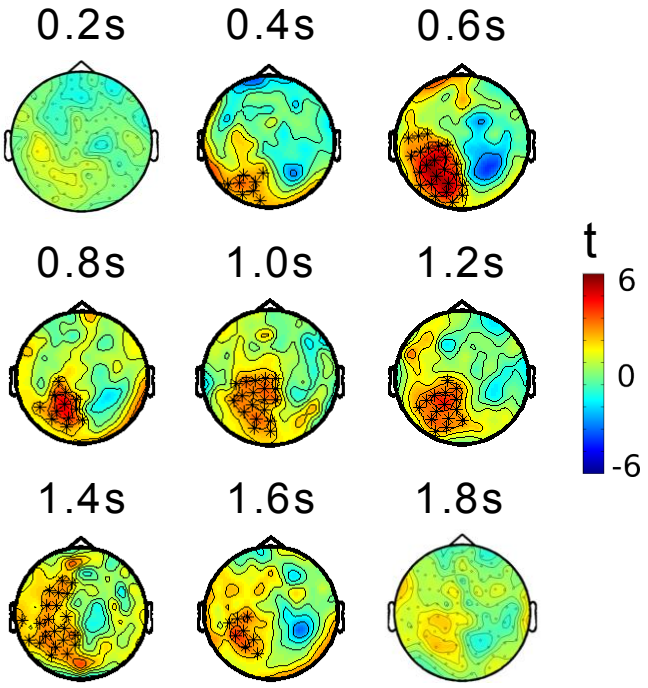


Figure 4-2 **Precue left minus precue right, precue epoch**. There is one significant positive cluster, p = 0.0005. Times spanned by the significant cluster are indicated with an emboldened ring around the topoplots. Sensors belonging to the significant cluster are marked with a cross. Timings are relative to precue onset. This pattern replicates the pattern or alpha lateralization previously observed with precues: alpha is increased on the ipsilateral side to the cued hemifield, relative to alpha on the contralateral side. NB, the [left-right] contrast does not permit ipsilateral synchronization and contralateral desynchronization to be distinguished. See Figure 4-11 for time-courses of alpha power separated left and right occipital/parietal ROIs.

Data from the array and retrocue epochs were analysed with respect to precue direction to test whether this lateralization pattern persists during and following encoding of the WM items. No significant clusters were discovered in the array epoch, but the lateralization of alpha power returned in the retrocue epoch, in the run-up to the presentation of the probe stimulus. The topographies from this epoch are shown in Figure 4-3.

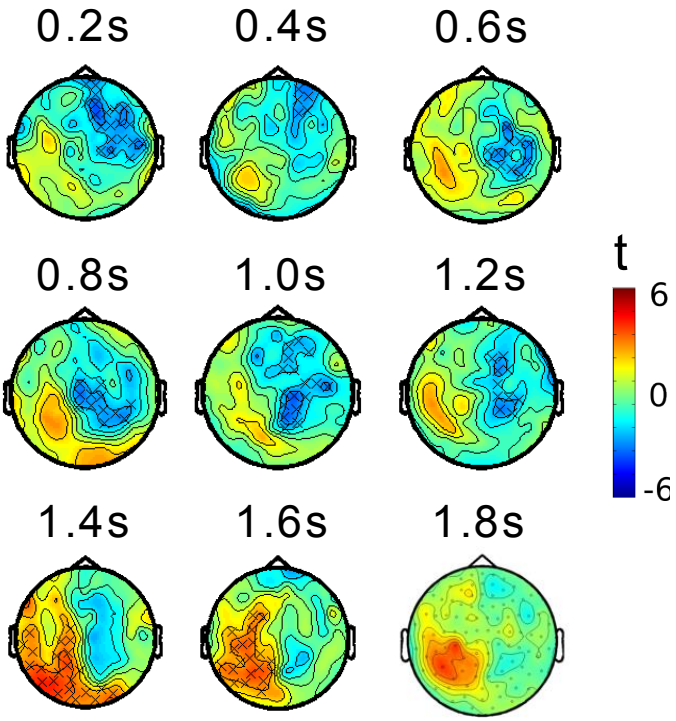


Figure 4-3 **Precue left minus precue right, retrocue epoch.** One significant positive cluster, p = 0.011. Two negative clusters, first p = 0.011, second p = 0.016. Timings are relative to retrocue onset. Times spanned by the significant clusters are indicated with a bold ring around the topoplots. Sensors belonging to a significant cluster are marked with a cross. This figure shows lateralization following a precue returning towards the end of the trial and the appearance of the probe stimulus.

The sensor-space alpha topography for the contrast [retrocue left minus retrocue right] is plotted in Figure 4-4. A similar pattern of lateralization is observed as with precues, with desynchronization of alpha power contralateral to the cued side, and ipsilateral synchronization.

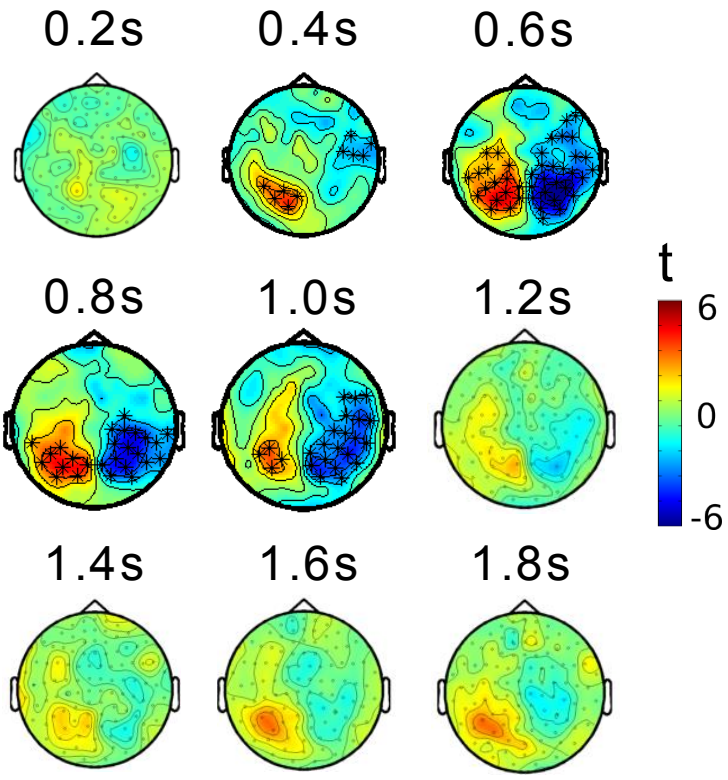


Figure 4-4 **Retrocue left minus precue right, retrocue epoch**. One significant positive cluster, p = 0.0065. One significant negative cluster, p = 0.001. Times spanned by the significant clusters are indicated with an emboldened ring around the topoplots. Retrocues produce a similar spatial pattern of alpha lateralization to precues, with a relative increase in alpha power ipsilateral to the cued side, and a decrease contralateral to the cued side.

These changes in alpha power were mapped into the brain space using an LCMV beamformer. Source space projections (section 4.2.6) of the data plotted in sensor space in Figure 4-2 and Figure 4-4 are shown in Figure 4-5.

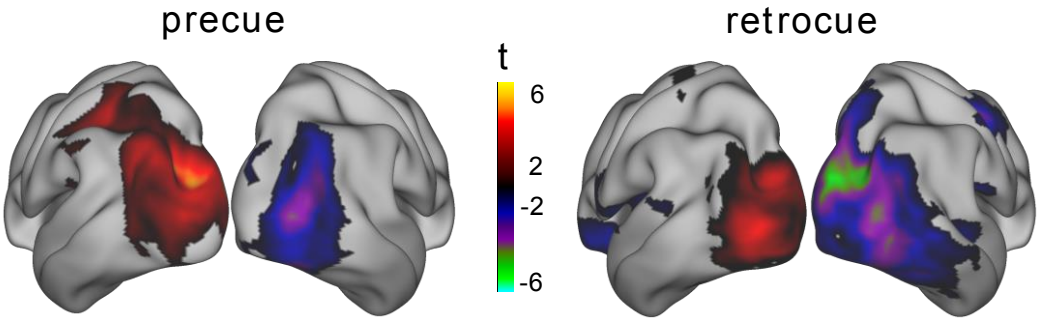


Figure 4-5 **Source space projection, [cue left – cue right] contrast, 0.6s post cue**. T-statistics of power differences in the 8-12Hz band are shown projected onto the CARET brain. The power differences are thresholded at t = ±2.

Alpha lateralization is strongest in the occipital pole, but extends dorsally towards IPS. Spatial ROIs (derived from the meta-analysis presented in chapter 5) were used to extract time frequency representations and time-courses for the laterality effects (section 4.3.1.3 below).

4.3.1.2 Decoding cued quadrant from alpha power

Having established a basic lateralization effect for both precues and retrocues, a simple multivariate approach was used to test (a) whether the brain activity carrying information about direction of cueing was band-limited (i.e. to the alpha band) and (b) whether changes in the topography of alpha power were specific to the cued quadrant, and (c) whether these patterns were similar to patterns of activity evoked by presentation of physical stimuli limited to one quadrant of the screen (specifically, the probe stimuli).

Classification confusion matrices are plotted in Figure 4-6 and Figure 4-7, for the precue epoch and retrocue epoch respectively.

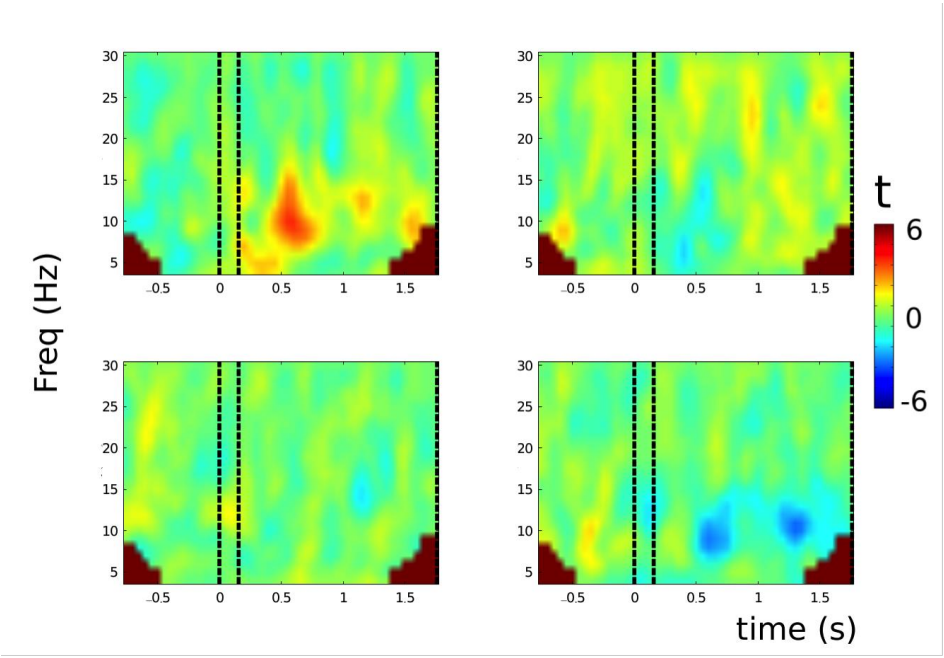


Figure 4-6 **Precue quadrant classification in the precue epoch. Top-left**, correctly classified; **bottom-left**, classified as same hemifield; **top-right**, classified as horizontally opposed quadrant; **bottom-right**, classified as diagonally opposed quadrant.

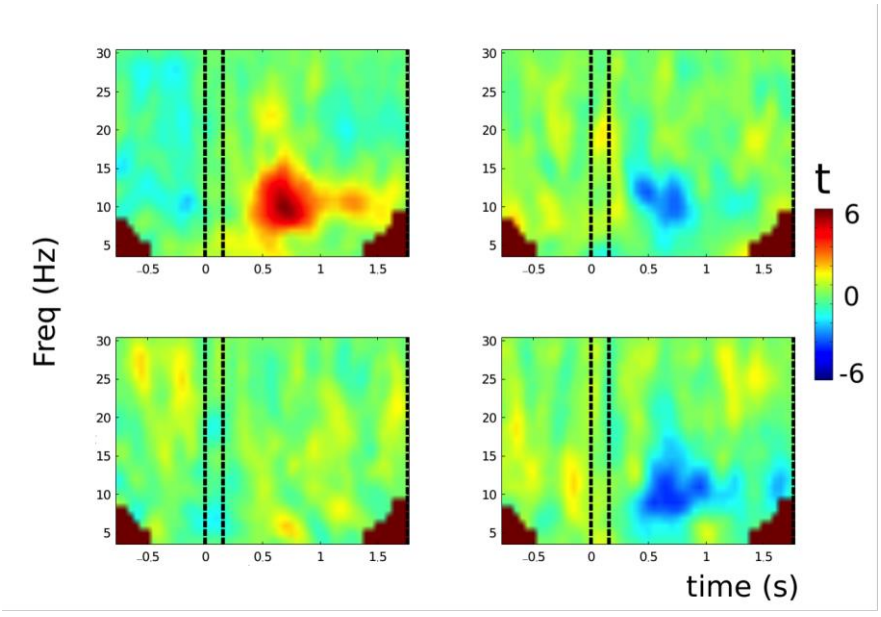


Figure 4-7 **Retrocue quadrant classification, retrocue epoch. Top-left**, correctly classified; **bottom-left**, classified as same hemifield; **top-right**, classified as horizontally opposed quadrant; **bottom-right**, classified as diagonally opposed quadrant.

Classification performance was highest in the alpha band. By inspection, the topography can discriminate cued quadrant, not just cued side. This was tested statistically by averaging classification values over the lower two and upper two quadrants to create a top versus bottom classification map, and by averaging data over the left hand and right hand quadrants to create a left versus right classification map. A one-sample cluster permutation distribution was calculated for each of these maps by randomly flipping 50% of the signs of the subjectwise maps (5000 permutations) and forming clusters with a t-statistic threshold of 2.5.

Classification performance was significantly above chance for both pairwise tests. This demonstrates that the pattern of alpha power changes is specific to the cued quadrant, but does not reveal whether the pattern is similar between conditions, or whether it is similar to the pattern of alpha desynchronization seen when a physical stimulus is presented. In order to test this, the pattern of activity in response to the probe stimuli was compared with the pattern of activity observed with precues and retrocues. Probe stimuli desynchronize alpha in the visual cortex (Klimesch, 2012), an effect that persists beyond the initial ERF. First, the same leave-one-out analysis as described above was run for the probe stimuli. This was able to robustly classify probed quadrant, and additionally give rise to a lateralized ERF which permits quadrant decoding in the theta range immediately following the appearance of the probe stimulus (Figure 4-8).

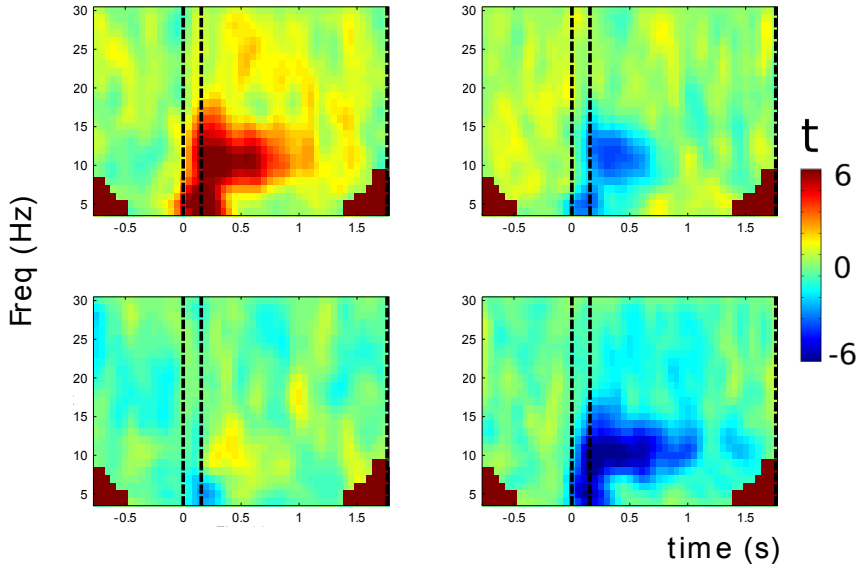


Figure 4-8 **Probe quadrant classification.** **Top-left**, correctly classified; **bottom-left**, classified as same hemifield; **top-right**, classified as horizontally opposed quadrant; **bottom-right**, classified as diagonally opposed quadrant. Overall, the response to probe presentation is a bilateral event-related desynchronization (ERD) in the alpha band (raw induced responses to the probe stimulus not shown here)

In order to test whether the pattern of activity in response to the probe stimulus was able to classify the direction of attentional cueing, the averaged topographies in response to probe stimuli from the within-epoch cross-validation analysis presented in Figure 4-8, were averaged between 0.3 and 0.5s post-cue and correlated with the topography of each trial in the precue and retrocue epoch (retaining the leave-one-out approach to avoid signal auto-correlation from driving classification). The results are presented in Figure 4-9 and Figure 4-10, divided into up/down and left/right classification in the same way as described above for the within-epoch cross-validation.

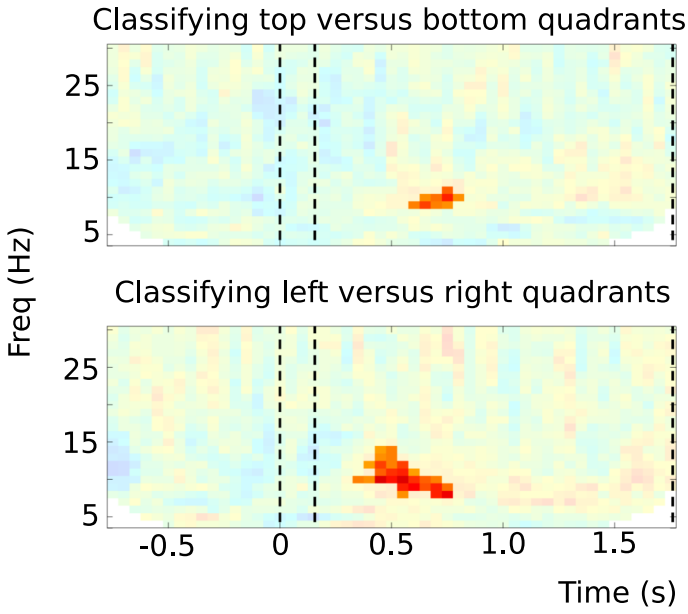


Figure 4-9 **Decoding precue direction, using the topography of induced responses to the probe stimuli, in the precue epoch.** Only time-frequency points belonging to a significant cluster have full colour saturation. Top vs. bottom, $p = 0.029$. Left vs. right, $p < 0.0002$. Up-down classification is weaker than left-right, but the presence of a significant cluster in both cases indicates that the alpha pattern is specific to the cued quadrant.

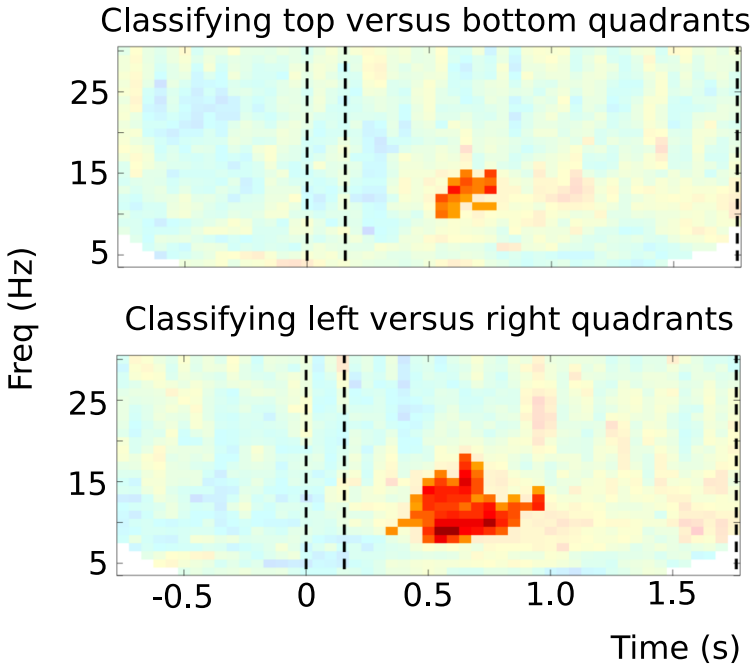


Figure 4-10 **Decoding retrocue direction, using the topography of induced responses to the probe stimuli, in the retrocue epoch.** Only time-frequency points belonging to a significant cluster have full colour saturation. Top vs. bottom, $p < 0.0002$. Left vs. right, $p < 0.0002$.

The alpha-power topographies following probe presentation are able to classify the attended quadrant in the precue and retrocue epochs, demonstrating that attention to a quadrant in space gives rise to retinotopically organized changes in alpha power similar to those induced by a physical stimulus.

4.3.1.3 Time-frequency analysis for source-space virtual sensors

Alpha power lateralizes in response to both prospective and retrodictive cues, and the pattern is retinotopic and quadrant specific. The time-course of alpha power modulation in different anatomical ROIs was characterised by reconstructing timeseries of alpha power at virtual electrodes in source space using an LCMV beamformer. ROIs were derived from the meta-analysis of fMRI studies presented in the following chapter. Section 4.2.6 (above) describes the procedure followed to obtain time-frequency representations for each ROI. The whole brain maps of alpha power presented in Figure 4-5 localize the alpha power changes to visual and parietal cortex, and in agreement with this pattern, across all the 26 ROIs used for the MEG analysis, significant modulations of alpha power by cue laterality were found in occipital and parietal ROIs only (occipital, IPS0, mid-IPS, anterior IPS). In order to visualize the time-course of alpha power modulation, the TFRs were averaged between 8 and 12 Hz. Figure 4-11 shows the time-course of 8-12Hz activity following a precue. Data from all of the occipito-parietal ROIs are shown: occipital cortex, IPS0, mid-IPS, SPL and anterior IPS. Baselined averages are shown in the left hand panels of each subplot, and the contrasts [cue left minus cue right] and [informative cue minus neutral cue] are shown in the right hand panels.

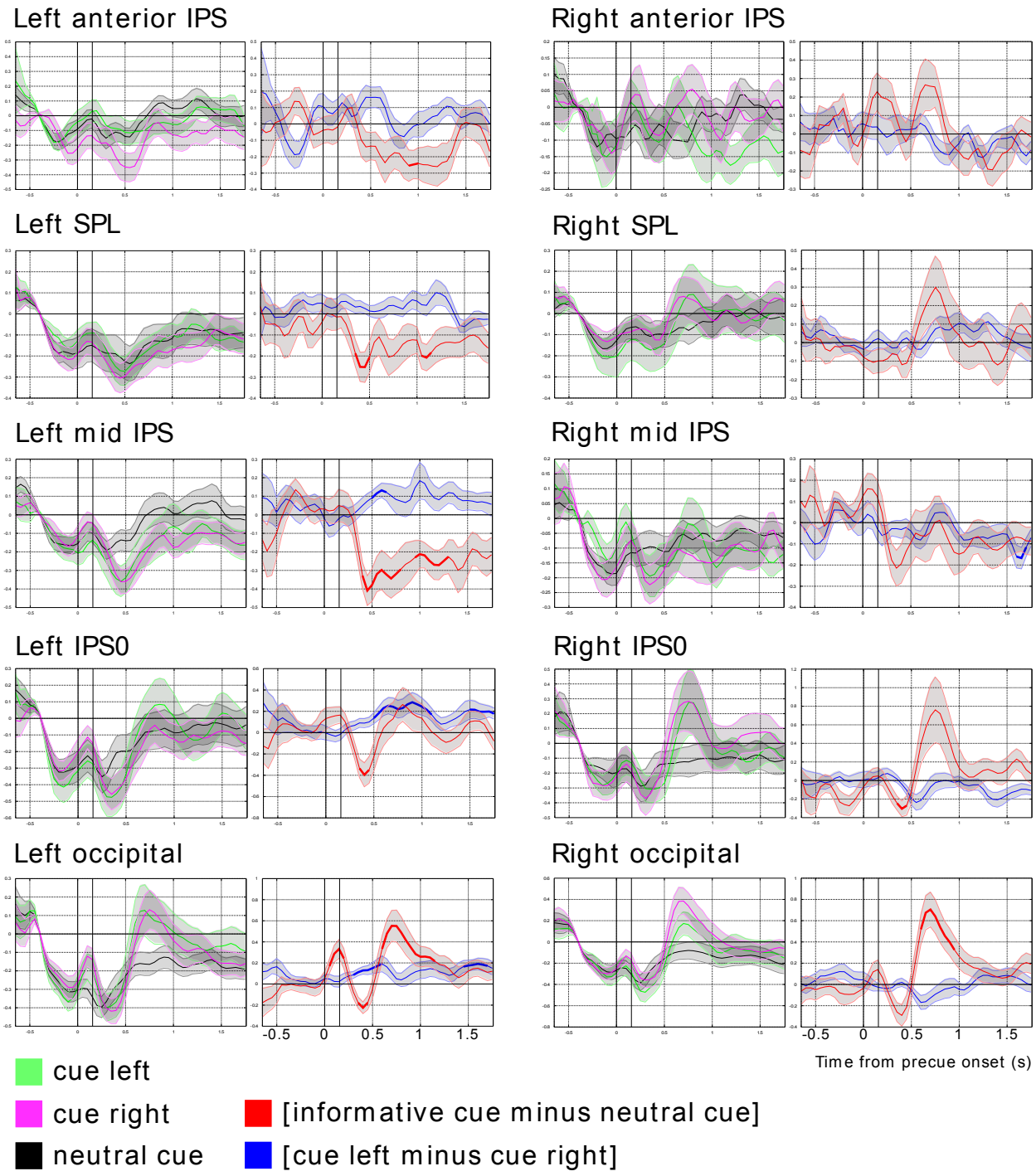


Figure 4-11 **Pattern of alpha (8-12Hz) activity in parietal and occipital ROIs, following a precue.** Data are shown for each hemisphere separately. The black/green/magenta plots show averaged time-course baselined between -0.5 and -0.3s. The blue/red plots show the two main task contrasts. Significant deviations from zero (1D cluster permutation statistic; $p < 0.05$) are shown by a thickening of the time-course (for the contrasts only). Error bars are ±1SEM.

The dynamics of alpha activity in occipital and parietal cortex are dissociable (Figure 4-11). Lateralization of alpha is strongest in occipital cortex and IPS0, and much less evident in anterior parietal cortex. Lateralization is generally stronger in left

hemisphere ROIs (compare with Figure 4-2). In addition to the lateralization contrast [cue left minus cue right] shown in blue, the contrast between [informative minus uninformative cue] is shown in red. This latter contrast reveals that precues desynchronize alpha in mid IPS from 1s onwards (the role of left mid-IPS as a node in the fronto-parietal attention network is discussed in chapter 5). At the same time, alpha is synchronized in occipital cortex. Alpha in occipital and parietal cortex is therefore dissociated: lateralization effects are restricted to the visual cortex, whereas the overall pattern of cue-induced activity is of opposite sign in occipital and parietal cortex. This contrasts with an MEG study of precue-induced alpha lateralization in which effects extended from occipital pole to the central sulcus (Siegel, Donner, Oostenveld, Fries, & Engel, 2008).

This pattern can be compared to the pattern of alpha lateralization that follows a retrocue (Figure 4-12).

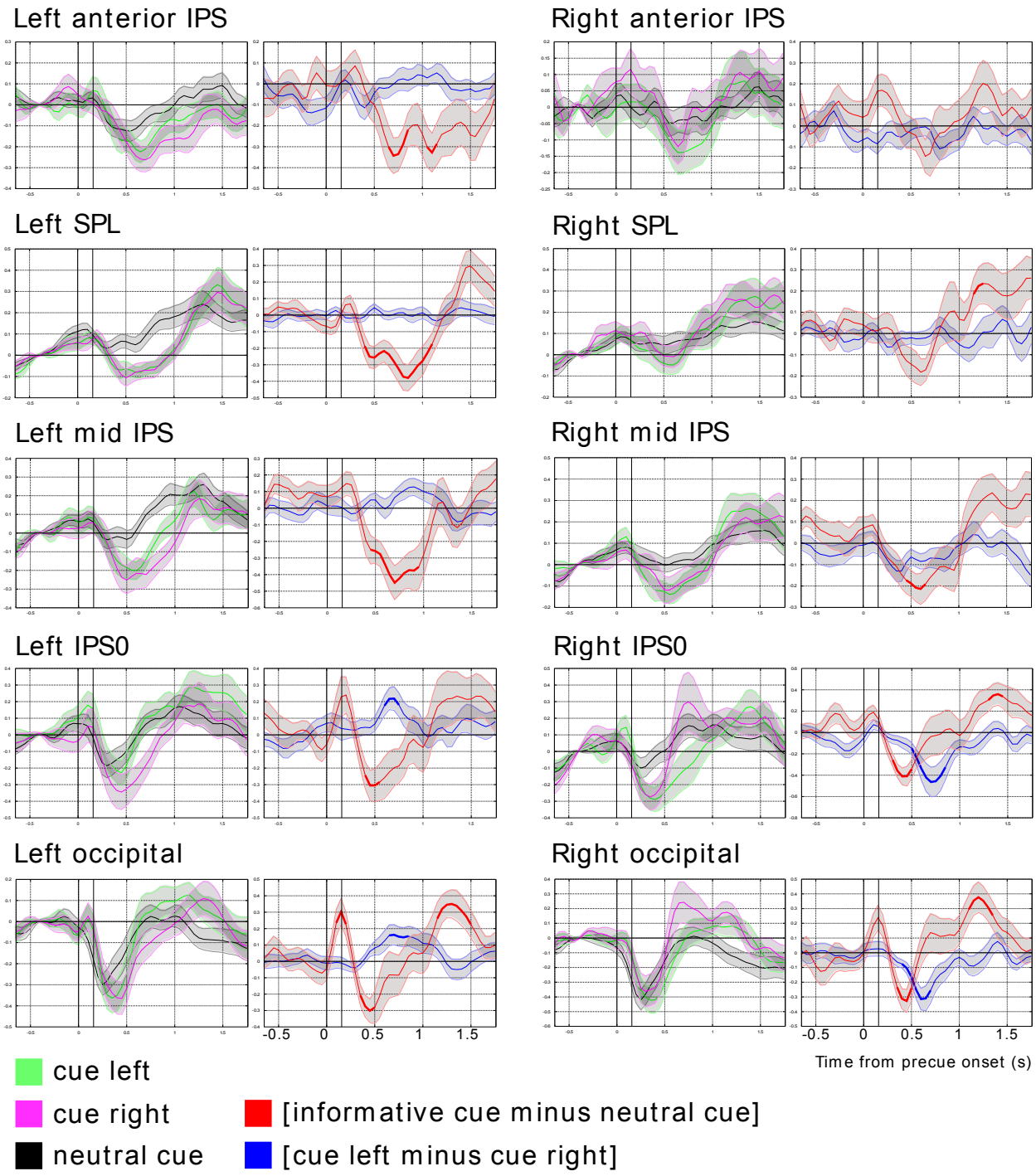


Figure 4-12 **Pattern of alpha (8-12Hz) activity in parietal and occipital ROIs, following a retrocue.** Data are shown for each hemisphere separately. The black/green/magenta plots show averaged time-course baselined between -0.5 and -0.3s. The blue/red plots show the two main task contrasts. Significant deviations from zero (1D cluster permutation statistic; $p < 0.05$) are shown by a thickening of the time-course (for the contrasts only). Error bars are ±1SEM.

The spatial pattern is similar to that for a precue: alpha lateralizes predominantly in occipital cortex and IPS0, though for retrocues the pattern is stronger and more bilateral

than for precues (compare with Figure 4-4). There is a similar focal desynchronization in left mid IPS/SPL (discussed further in chapter 5).

4.3.1.4 Stability of cue-induced brain states

To characterize the time-course of control operations following precues and retrocues, the trials within each condition were randomly split into two groups, and the induced response topographies of contrasts performed on each half separately were correlated across time (i.e. correlating the topography at time A - 'train time' – with the topography at time B – 'test time' – for all combinations of A and B). This analysis was performed on the alpha topographies for the cue-effect [informative cue minus neutral cue] and cue-laterality [cue left minus cue right] contrasts.

The cross-temporal correlation matrix for the cue-effect contrast (capturing the temporal stability of changes in topography due to cue use) showed that after precues, a stable state emerged from ~0.6s post-cue until the presentation of the memory array (Figure 4-13, panel A; 'square' in upper-right, from 0.6s). Retrocues gave rise to a different pattern (Figure 4-13, panel B). A brief period of stability between ~0.6 and ~1.2s following the cue subsequently evolved into a second state with an uncorrelated topography, lasting until probe presentation. The control network activations corresponding to this temporal pattern are explored in chapter 5.

The laterality contrast produced a similar pattern, though the signal-to–noise ratio was lower than for the cue-contrast as trial numbers for each estimate are halved. Precues gave rise to a sustained lateralized alpha topography from 0.6s until memory array presentation (Figure 4-13, panel C; rough square in upper right), whereas retrocues produced a more transient pattern of alpha lateralization between ~0.5 and ~1s post cue, which then subsided between ~1s post-cue and probe onset (Figure 4-13, panel D).

In order to compare the time-course of alpha lateralization for precues and retrocues a lateralization index was defined. For each pair of bilateral ROIs the sign for the [cue left minus cue right] contrast in the RHS ROI was flipped, and this inverted signal was added to the raw contrast for the left hand side; i.e. {[cue left minus cue right]$_{\text{LHS ROI}}$ – [cue left minus cue right]$_{\text{RHS ROI}}$}. This index showed the strongest effects in IPS0, for which it is plotted in Figure 4-13, panels E and F. Following precues, alpha lateralization begins with the same latency as following retrocues, but persists until array onset. In contrast, the alpha lateralization associated with retrocues peaks around 600ms post-cue, and has returned to baseline by the time the probe stimulus is presented at the end of the epoch.

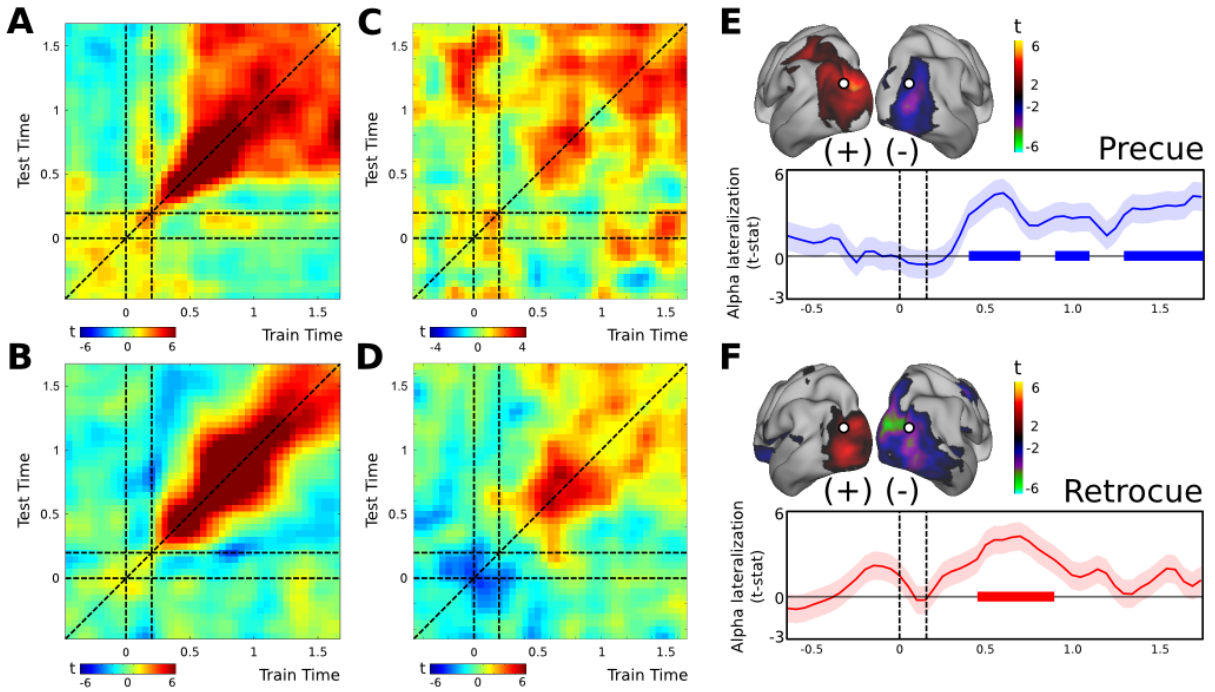


Figure 4-13 **Time-course of control.** Panels **A** and **B** show the split-half cross-temporal alpha (8-12Hz) topography correlation structure for the [precue minus neutral] contrast and [retrocue minus neutral contrast] respectively. For precues, a stable topography emerges ~0.6s post-cue that lasts until the onset of the memory array. For retrocues, there is a brief period of stability between 0.5 and 1.2s post-cue, but this state is then replaced from 1.2s post-cue, until probe onset. Panels **C** and **D** show the analogous correlation structure for the [precue left minus precue right] and [retrocue left minus retrocue right] alpha topographies. The precue alpha lateralization is stable between ~0.6s post-cue and the memory array, with a slight 'dip' at ~1s post-cue. In contrast, the retrocue alpha lateralization is stable between ~0.5 and ~1s post-cue. Panels **E** and **F** show an alpha lateralization index calculated for the IPS0 virtual electrode. Alpha lateralization is persistent following precues (with a slight dip ~1s post-cue). Following retrocues, alpha lateralization is transient, returning to near- baseline level by 1s post-cue.

4.3.1.5 Alpha lateralization predicts behaviour

The relationship between alpha lateralization and behaviour was investigated as described in section 4.2.7: a trialwise index for alpha lateralization relative to the cued hemisphere was derived from the sensor-space classification analysis, for both the precue and retrocue epochs. The relationships between this measure and RT and accuracy were assessed using a linear regression and a logistic regression, respectively. The shift magnitude of the probe stimulus at the end of each trial was included as a second predictor variable in both cases (this parameter affects both accuracy and RT, so its inclusion as a nuisance regressor should render the GLM more sensitive to correlations with the MEG signal).

The analysis was repeated for each of the three conditions, and for each time-frequency bin. For the precue epoch, only differences in the precue condition could reflect a direct effect of cueing, and for the retrocue epoch only differences in the precue and retrocue conditions could directly reflect cueing. However, a link to behaviour in the remaining condition/epoch bins was also conceivable: previous studies have demonstrated that spontaneous pre-stimulus fluctuations in alpha lateralization can predict behavioural performance in sensory discrimination tasks (van Ede et al., 2011). Results for the precue and retrocue epochs are plotted in Figure 4-14 and Figure 4-15 respectively.

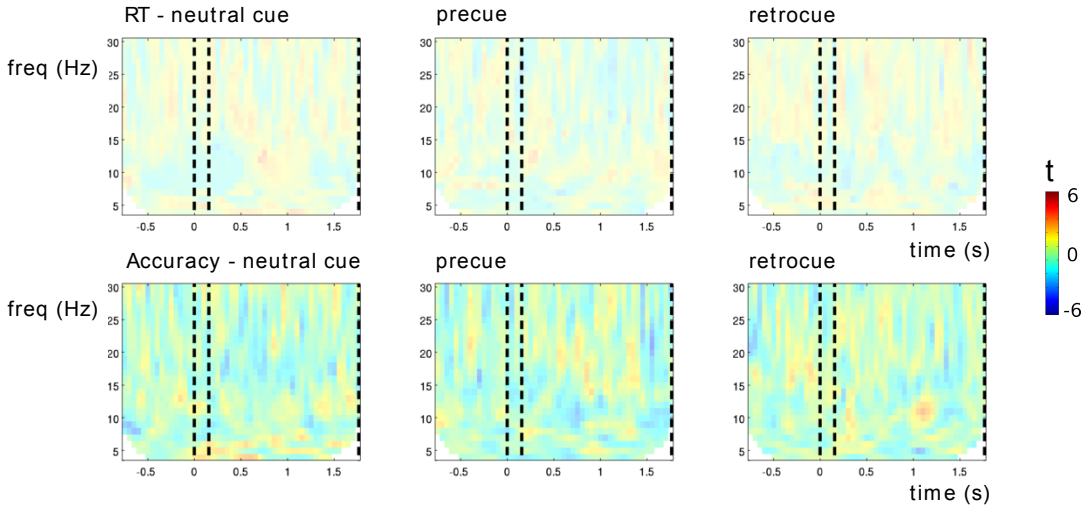


Figure 4-14 **Group-level t-statistic of subjectwise beta weights for the linear regression of trialwise lateralization in the precue epoch (for frequencies between 3 and 30Hz) with RT (top panels) and the logistic regression of lateralization with accuracy (bottom panels).** Significant deviations from zero were tested using a one-sample cluster permutation test (clustering in frequency and time). No significant clusters were found.

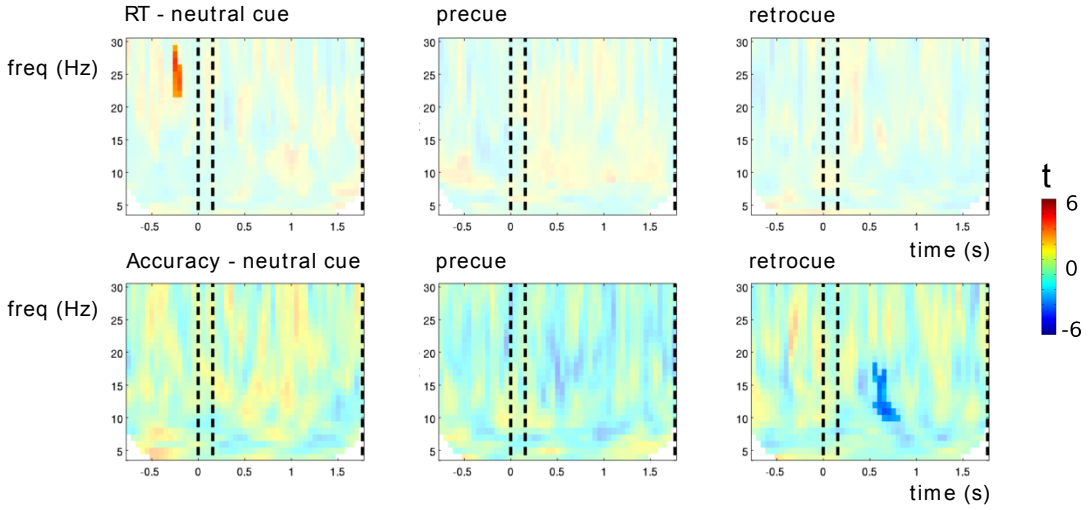


Figure 4-15 **Group-level t-statistic of subjectwise beta weights for the linear regression of trialwise lateralization in the retrocue epoch (for frequencies between 3 and 30Hz) with RT (top panels) and the logistic regression of lateralization with accuracy (bottom panels).** Significant deviations from zero were tested using a one-sample cluster permutation test (clustering in frequency and time). A single significant negative-going cluster was found for the logistic regression with accuracy, in the retrocue condition (in the alpha/beta band; p = 0.026). This is at the expected time and frequency for a correlation with cue induced lateralization, as it overlaps with the main effect (Figure 4-10). A significant positive-going cluster was also found for the RT regression, prior to retrocue onset in the neutral condition (beta-band, p = 0.01). There was no prior reason to expect a cluster in this condition/time/frequency bin.

Whilst there were no significant correlations between lateralization index and accuracy for the precue analysis, there was a significant negative relationship between accuracy

and alpha lateralization following the retrocue (more lateralization was associated with worse performance). Whilst this effect was statistically modest ($p = 0.026$) it was in the expected time and frequency bins (i.e., it overlaps with the main effect of alpha lateralization). This was not the case for an unexpected second significant cluster, which appeared prior to neutral cue onset in the neutral condition.

The negative relationship in the retrocue epoch was characterized by extracting the trialwise lateralization index for the peak time-frequency bin for this effect, median splitting the behavioural data within-subject on the basis of this index, and running mixture model analyses separately for each half of the data (Figure 4-16, Figure 4-17).

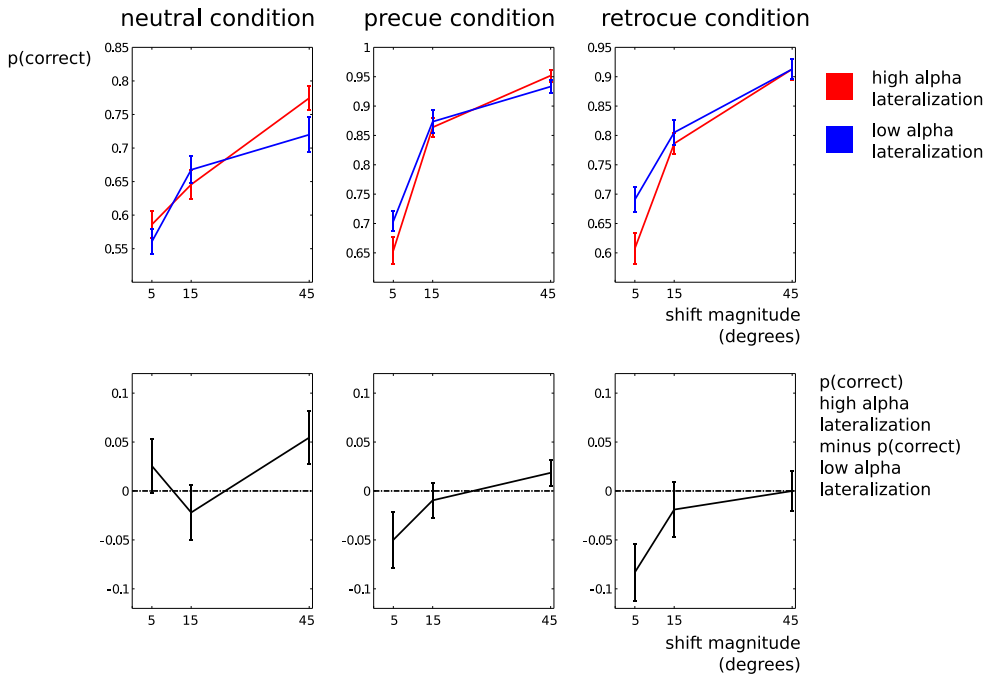


Figure 4-16 **Accuracy data median split by retrocue lateralization index, 0.6s post-cue at 11Hz (the time/frequency point at which the correlation with behaviour was strongest).** The right hand panels show data from the retrocue epoch. The accuracy effect is mainly for the smallest orientation shift of the probe (5 degrees), with no change in performance for the largest shift (45 degrees). In terms of the precision/capacity account, this implies an effect on the precision of the stored representations, but not a categorical effect on whether an item is represented or not. This was tested by running the mixture model for each of the median-split halves of each subject's dataset (Figure 4-17).

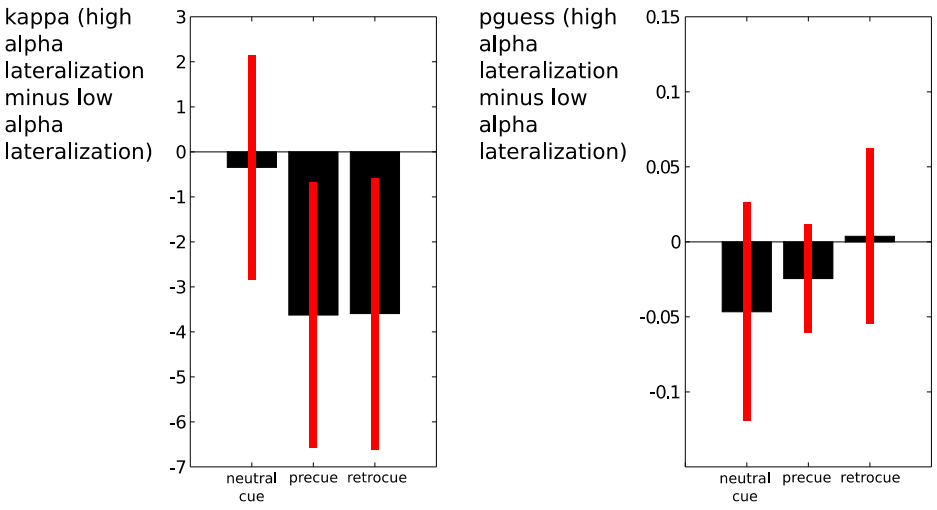


Figure 4-17 **Differences in mixture-model parameter estimates between high and low alpha lateralization trials (retrocue epoch, 0.6s post-cue at 11Hz).** Red error bars are 95% CI. For each subject, the mixture model was run separately for the high alpha lateralization and low alpha lateralization trials (median split), for each condition independently. There was no effect on the guess rate parameters, but kappa was lower in both the retrocue and also the precue conditions for trials in which alpha lateralization had been higher.

Accuracy in trials in which the alpha lateralization index was low was subtracted from accuracy for trials in which the alpha lateralization index was high (Figure 4-16, lower panels). The difference in accuracy was largest for the smallest change in probe orientation relative to the memory item (5 degrees) with no change in accuracy for the largest orientation shift (45 degrees). Surprisingly, a similar effect was observed for the precue condition (retrocue epoch) in which there was no main effect in the time-frequency resolved analysis (Figure 4-15, centre panel in the lower row). The mixture model (chapter 3) was run separately for high and low lateralization trials (Figure 4-17). Kappa (memory precision) was lower in the high lateralization trials for both precue and retrocue conditions. Guess rate was unaffected. The magnitude of alpha lateralization following a retrocue appears to selectively co-vary with the precision of the items in memory – the lower the precision, the more alpha is lateralized.

4.4 Discussion

What role does perceptual cortex play in working memory maintenance? Recently fMRI work demonstrating that the contents of working memory can be decoded from visual cortex (Harrison & Tong, 2009) has encouraged the belief that working memory maintenance (at least for simple visual features) depends on sustained activity in visual cortex (Gazzaley & Nobre, 2012; Stokes, 2011). BOLD activity in perceptual cortex has been shown to be modulated in response to retrocues triggering selection from within working memory (Higo et al., 2011; Kuo et al., 2014; Lepsien et al., 2011; Lepsien, Griffin, Devlin, & Nobre, 2005; Munneke et al., 2012; Sligte et al., 2009). The MEG data presented here characterize the time-course with which retrocues modulate activity in perceptual cortex. Retrocues robustly modulate alpha in perceptual cortex in a retinotopic manner – the modulation observed had a similar spatial pattern to the ERD induced by a physical stimulus. However, this change in the alpha pattern was punctate, lasting until around 1s post-cue. There was no evidence for a sustained biasing of activity lasting until the probe stimulus was presented. This activity was therefore unlikely to simply reflect preparation for the probe stimulus, as this would be expected to ramp up in preparation for the probe item. Nor did it represent a shift in a sustained top-down biasing signal to perceptual cortex responsible for active maintenance of the memory items, as whilst this might manifest itself immediately after the retrocue, it would also be expected to last for the reminder of the retention interval. These data suggest that perceptual cortex is recruited transiently when an item in memory is prioritized, in order to 're-perceive' the cued item. A emerging perspective sees working memory maintenance and utilization as a dynamic interplay between active and passive states (Mongillo et al., 2008; Olivers et al., 2011; Stokes et al., 2013). In chapter 3, reaction time data from this task were shown to support the hypothesis

that retrocues cause the cued item to move to a qualitatively distinct prioritized state. Whilst sustained activity in perceptual cortex seems not to be necessary for maintenance, its recruitment following a retrocue may be a necessary step in prioritising the cued item in memory.

Trial-by-trial or subjectwise co-variation between an electrophysiological effect and behavioural performance is often taken to imply that the effect of interest reflects an important causal mechanism, and not an epiphenomenal process[16]. Previous studies investigating alpha lateralization in response to predictive attentional cues have reported that alpha lateralization predicts behavioural performance – both trial by trial (Haegens et al., 2010; 2012; Haegens, Händel, & Jensen, 2011a; Thut et al., 2006; van Ede et al., 2011) and across subjects (Bonnefond & Jensen, 2012a; Gould et al., 2011; Händel et al., 2011; Sauseng et al., 2009b). A positive relationship between alpha lateralization and behaviour in validly cued trials is generally reported: the more alpha is lateralized, the better performance on the task. Investigating the co-variation between alpha lateralization and behavioural performance across trials in the current data revealed an unexpected *negative* correlation between retrocue-induced alpha lateralization and behavioural performance: the more alpha lateralization on a given trial, the worse the accuracy on that trial. A *post hoc* explanation for this finding might be that more 'work' is performed to retrieve the cued item when it is weakly represented in memory than when it is represented with high fidelity (remaining agnostic over whether this is an explicit strategic effect driven by effortful deployment of top-down resources on trials in which the representation was weak, or a more implicit property arising from lower-level mechanisms).

---

[16] Though, this correlation, whilst seeming to link brain and behavior more 'tightly', really has the same status as the main effects of experimental manipulations, which can also be seen as inducing a correlation between behavior and brain states (except in this case across conditions).

Decomposing the behavioural correlation using the mixture model, precision of recall was found to co-vary with the degree of alpha lateralization, but guess rate did not. This is consistent with the explanation in terms of representation strength: i.e., that the degree of alpha lateralization reflects the precision of the representation of the cued item on a given trial, which affects the ease of retrieval. For most subjects, the guess rate was low in retrocue trials (the mean guess rate for retrocue trials included in the MEG analysis was 0.16, which compares with 0.115 in precue trials and 0.455 in neutral cue trials). As such there is a ceiling effect for guess rate – there are few trials in which the cued item is not successfully retrieved following a retrocue. However, the precision-capacity design unmasks differences in the precision of these prioritised representations. Precision is a property retrocues do not alter (Murray et al., 2013), and which seems likely to be fixed following encoding.

The current results did not replicate the oft-reported positive correlation between the degree of alpha lateralization in the precue interval and behavioural performance (van Ede et al., 2011). A possible explanation for this discrepancy is that previous studies have generally employed tasks involving sensory discrimination or detection. In these tasks, where perceptual limits determine task difficulty, the strength of prospective attentional bias may crucially determine the efficiency of sensory sampling and therefore task performance. The role of the precue in the current task was to permit selective gating of the cued item into working memory. As argued above, prioritizing a single item may involve an all-or-nothing selection process. As such, it is unsurprising that alpha lateralization in precue trials did not correlate with guess rate: at least for the special case of cueing a single item (as in the current task), if the cue is used at all, it is likely to be used successfully (Murray et al., 2011). A correlation of preparatory alpha lateralization with memory precision might have been more expected: precues influence precision as well as guess rate. However, this change in precision may not reflect an

encoding limit, instead reflecting the reduction in memory load from four to one item with a precue: memory load determines memory precision, at least for loads below memory capacity (Bays & Husain, 2008; W. Zhang & Luck, 2008).

Was the pattern of alpha modulation observed here attributable to ipsilateral synchronization of alpha, or contralateral desynchronization? Alpha has been suggested to reflect top-down active inhibition, blocking transmission from irrelevant parts of cortex to control information transmission (Ole Jensen, 2010). This has been proposed to explain ipsilateral increases in alpha power in attentional cueing tasks. However, inspecting the raw time-courses of alpha power in left and right hemisphere ROIs (left hand plots from Figure 4-11 and Figure 4-12) reveals that there is no simple correspondence between synchronization/desynchronization in one or other hemisphere, and the alpha lateralization effect. In the raw signal effects of cue processing are superimposed on lateralization effects. In response to both precues and retrocues (either to the left or right side, and as compared to neutral cues), alpha was initially briefly synchronized (probably reflecting the ERF), followed by a desynchronization (perhaps reflecting cue processing), which then rebounds to synchronization. The alpha lateralization effect for both cue types spanned these cue-associated global power changes. It therefore seems likely that multiple alpha generators were contributing to the signal recorded from each ROI, including but not limited to the generators for which activity lateralized. Moreover, inspecting the raw alpha power beginning of the precue epoch reveals no obvious 'resting' baseline for alpha power before cue onset: the mere presentation of the fixation cross 550ms before precue onset strongly reduces bilateral alpha power (presumably because vigilance is increased in preparation for the trial). Absolute levels of alpha power following the cues are therefore difficult to interpret. The most parsimonious explanation of the alpha patterns observed in the current dataset is that they reflect a focal top-down

signal which modulates cortical excitability, down-regulating alpha power – an endogenously-driven analogue of an event-related desynchronization (ERD). Indeed, the quadrant-directed cues gave rise to topographies that resemble the topography of the ERD induced by a physical stimulus appearing in the corresponding quadrant (Figure 4-8).

A strong version of this hypothesis would imply that alpha desynchronization and spike rate are in a tight inverse relationship, but whilst this is observed in many cases it is not an absolute rule, as discovered by Haegens et al. (2011b): in primary somatosensory cortex of monkeys, a U-shaped relationship was seen, in which the lowest and highest alpha power bins were associated with the highest spike rate. A more general hypothesis would allow alpha power and spike rate to decouple. Alpha desynchronization might be a signature of readiness or ability to code information (Hanslmayr et al., 2012): a gating, as opposed to a driving signal. The key idea here is that highly synchronized alpha might correspond to highly correlated firing across the neural population. This correlated activity is redundant (low in entropy). Speculatively, it is also predictable by downstream brain areas - especially if they receive a copy of the alpha drive from a thalamic site, e.g. the pulvinar (Saalmann, Pinsk, Wang, Li, & Kastner, 2012). By analogy with the role of efference copy in suppressing self-generated and therefore predictable sensory input (Bell, 1989), this might serve as an 'ignore me' signal for cortex not involved in active processing. Spikes propagating to connected cortical areas can be ignored when observed in the context of this default signal, and the system is therefore robust to noise perturbations, or firing driven by irrelevant input. By contrast, when alpha desynchronizes, the coding potential of the population is increased because spiking becomes less redundant (Hanslmayr et al., 2012), and downstream brain areas might also be rendered more sensitive to incoming spikes, as they are not redundant on an alpha "null signal". Taken together, this 'alpha

gating' hypothesis resembles the inhibition account (Ole Jensen, 2010) in that alpha desynchronization is tightly coupled to attentional routing of information, but it differs in that spike rate and alpha can be decoupled (expanding the coding potential of the desynchronized state, which might for example involve little or no spiking if there is no stimulus: a valid coding state).

In general, whether or not *inhibitory* top-down drive to perceptual cortex contributes to cognitive control remains an open question (Nelissen et al., 2013). Some have argued that inhibition in sensory cortex as a result of top-down mechanisms chiefly results from local competitive interactions, as opposed to direct top-down suppression (Miller & Cohen, 2001; Munakata et al., 2011). In so far as our data can address this question, at no point in the task did alpha power in perceptual cortex exceed the power observed in the inter-trial interval – even in the epoch following a precue, when there were three stimuli soon to appear, which it would have been adaptive to inhibit. The peak cue-induced change in alpha power in this epoch was midway between cue and memory array. An inhibitory signal to suppress the three irrelevant stimuli in the WM array would have been expected to ramp-up towards the presentation of the memory array.

5 Sequential activation of frontal networks during selection from working memory

Abstract

A number of frontal and parietal sites have been implicated in working memory control on the basis of fMRI experiments (Nee & Jonides, 2009). Resting state connectivity analyses partition these sites into two major frontal networks (Dosenbach et al., 2007). It is not clear how this network grouping maps to functional components of top-down control. Whilst some have suggested that fronto-parietal regions are the primary source of top-down regulation for working memory (Nobre et al., 2004), linking this to their role in selective attention (Corbetta et al., 1993), others have argued that the fO, a node of the cingulo-opercular network, plays the primary role (Higo et al., 2011). The poor temporal resolution of fMRI may be responsible for this lack of clarity, as the BOLD signal is unable to distinguish between control operations that may follow one another in rapid succession. In this chapter, time-resolved MEG induced responses in frontal and parietal control sites are used to clarify the role of different frontal regions in control of working memory, during a precue/retrocue task. A meta-analysis of fMRI experiments was performed to evaluate regional involvement in precueing and retrocueing, and to derive a set of ROIs for MEG analysis. Source space induced responses were analysed at these ROIs. The spatial patterns replicate the activity patterns observed using fMRI, and the time-courses dissociate the frontal networks and clarify their roles in control of working memory.

5.1 Introduction

Working memory is regulated by control mechanisms that can gate encoding (Murray et al., 2011; Schmidt et al., 2002) and also modulate on-going maintenance and access to stored information (Griffin & Nobre, 2003; Landman et al., 2003; Olivers et al., 2011). A number of regions in prefrontal cortex and parietal cortex have been implicated in control of working memory, including intra-parietal sulcus (IPS), dorsolateral prefrontal cortex (dlPFC), frontal eye field (FEF), iFEF/inferior frontal junction (IFJ), anterior insula/frontal operculum (aI/fO), and pre-supplementary motor area (pre-SMA) (Nee & Jonides, 2009; Nobre et al., 2004). A challenge for cognitive neuroscience is to link the roles of these frontal regions to the structure of cognitive control, on the assumption that functional anatomy and the structure of cognition are isomorphic (Fuster, 2003). There are two sides to this problem. One (theoretical) is the construction of a cognitive model that accurately partitions the cognitive processes involved in control of working memory. The other (technical) is mapping the spatiotemporal patterns of brain activity that correspond to control over working memory.

A first step in understanding the role of different frontal regions is to characterize anatomical subdivisions and connexions, independent of a specific task. Classic maps of frontal anatomy were based on cytoarchitecture (Brodmann, 1909; Economo & Koskinas, 1925; Sanides, 1964), but neuroimaging technology has furnished new approaches, and recently DTI and resting state fMRI have been used to parcellate PFC by clustering regions with similar patterns of anatomical (M. Beckmann, Johansen-Berg, & Rushworth, 2009; Tomassini et al., 2007) and functional connectivity (Neubert, Mars, Thomas, Sallet, & Rushworth, 2014; Sallet et al., 2013; S. Zhang, Ide, & Li, 2012). Functional connectivity measures have also been used to infer the network structure

linking distinct prefrontal regions (Dosenbach et al., 2007; Petersen & Posner, 2012; Seeley et al., 2007). Dosenbach (2007) and Seeley (2007) grouped prefrontal sub-regions into two networks on the basis of resting state connectivity. Adopting Dosenbach's terminology, a fronto-parietal network incorporating dlPFC, precentral gyrus, IPS and IPL was dissociated from a cingulo-opercular network incorporating the pre-SMA, the anterior insula / frontal operculum, and the lateral frontal pole. These regions overlap with fronto-parietal sites implicated in control of working memory. The networks may play dissociable roles in task control, a possibility that is explored in this chapter.

In order to establish mappings between frontal sites and cognitive processes mediating working memory control, brain activity can be recorded whilst people are performing WM tasks. In this chapter, a precue/retrocue working memory control task is used, in which subjects can either select items at encoding, or from within working memory. fMRI is the mainstay of structure-function mapping and has been applied to precue and retrocue tasks, demonstrating that retrospective cueing to working memory content appears to share neural mechanisms with prospective cueing for working memory encoding (Griffin & Nobre, 2003; Nobre et al., 2004) - which itself has a neural signature similar to that associated with preparatory spatial attention. fMRI data from prospective attention tasks, selective encoding tasks, and the retrocue task reveal activation in a common fronto-parietal network (Corbetta et al., 1998; Nobre et al., 1997), suggesting that this set of brain regions is invoked whether top-down control is exerted to filter sensory input, or to select between items already in memory (Nobre et al., 2004). This anchor to spatial attention is a useful clue to the neural/cognitive processes involved in the more complex task of selecting an item from within working memory. The shared neural signature has informed functional hypotheses, contributing to the view that shared mechanisms of spatial attention act upon

perceptual and mnemonic representations to prioritise task-relevant content (Gazzaley & Nobre, 2012).

However, in fMRI additional regions besides the fronto-parietal areas are seen to activate following a retrocue. These include the anterior insula (Lepsien et al., 2005) and/or frontal operculum (Higo et al., 2011; Nelissen et al., 2013), as well as the dorsal anterior cingulate / pre-supplementary motor area (Nee & Jonides, 2009; Nobre et al., 2004). Opinions differ on which cortical regions are most important for prioritizing a representation within working memory. Studies that have focussed on fronto-parietal network involvement emphasise the link with preparatory attention and suggest that fronto-parietal regions are the critical control sites that permit retrocues to boost performance (Nobre et al., 2004). In contrast, Higo and colleagues (Higo et al., 2011), and more recently Nelissen and colleagues (Nelissen et al., 2013) have suggested that the frontal operculum is the key region involved in selection from within working memory, modulating memory representations in sensory cortex.

Cognition is highly dynamic in time (Duncan, 2013) compared with the inherent haemodynamic limit on the temporal resolution of fMRI (the mainstay of structure-function mapping). fMRI may be too blunt a tool to dissociate activations associated with different cognitive operations that last for only a few tenths of a second, and follow one another in close succession. This inability to dissociate rapid sequential operations may be responsible for the lack of clarity over the key regions involved in preparatory and internal selection. In this chapter, MEG, a method that measures neural activity directly with millisecond temporal resolution, is used to derive cue-induced activation time-courses for different frontal regions and clarify their roles in control of working memory. MEG has acceptable spatial resolution but is inferior in this regard to fMRI, so a meta-analysis of pre-existing fMRI studies of precueing and

retrocueing was first performed to distil spatial information about the brain regions recruited by each cue type. ROIs for MEG analysis were derived from the meta-analysis results, and induced responses in key regions of interest were extracted from the source-space MEG data. The spatial pattern of activations replicated the fMRI data. The fine-grained time-course of these activations informs competing theories about the role of different frontal sites in cognitive control. Whole-brain activity mapping was also performed to verify the spatial specificity of the MEG responses. Finally in order to test whether the source space ROIs are really measuring activity in the intended network regions, the task ROIs were used in a functional connectivity analysis of resting and task data, replicating networks found using fMRI and characterizing their spectral signatures.

5.2 Meta-analysis of fMRI studies of prospective and retrospective cueing

A number of fMRI studies have investigated the brain activations associated with prospective and/or retrospective cueing. Several have noted overlap between the regions recruited by each cue type, but differences are also observed (Nee & Jonides, 2009; Nobre et al., 2004). The meta-analysis presented here has the twin aims of deriving a set of spatial ROIs to guide MEG analyses of the brain responses to precues and retrocues, and of reviewing the similarities and differences between the pattern of brain regions that have been associated with prospective and retrospective selection.

5.2.1 Meta-analysis methods

Imaging studies were searched using Pubmed and Google Scholar. We included only fMRI studies. Further selection criteria and lists of the included studies for each cue type are given below. For each contrast for which activation foci were extracted, we recorded the number of subjects in that experiment. The Activation Likelihood Estimation (ALE) approach (Eickhoff et al., 2009; Turkeltaub, Eden, Jones, & Zeffiro, 2002) inversely weights each data point by the number of subjects in the experiment (see 5.2.1.3 below). Some studies were reported in the MNI coordinate system, others in Talairach coordinates. The meta-analysis was performed in MNI coordinates, and coordinates reported in Talairach space were converted to MNI space using GingerALE's **tal2icbm** function (Turkeltaub et al., 2002).

5.2.1.1 Selecting retrocue studies

There are a number of published studies using some variant of the retrocue paradigm. Experiments were included in our meta-analysis if they involved a late cue (>1s

post-array; i.e., not iconic) to focus on an item already in memory, and the key contrast(s) chosen from each study isolated the brain response to this cue. The studies we reviewed used a range of different stimulus classes (words, face/scene images, visual patterns, spatial locations). Studies also differ in terms of memory load, with some studies presenting a single item and then 'refreshing' that item, and others presenting up to four items. Despite these differences in task parameters, we reasoned that the core control regions involved in internal orienting would be common across tasks, even if the class of the memory stimuli or the memory load differed. Some studies subtracted activity in a control condition in which instead of a cue to attend to an item in memory, the memory item was presented again. Others did not include a sensory control. For this reason activations in sensory cortex will be inconsistent across studies. Some studies used a spatially distributed memory array and spatial cue, others used non-spatial cues (e.g. cueing by referring back to a given position in a sequential presentation, or a simple 'refresh' cue where there had been only one memory item). We chose 17 studies comprising 22 experiments (different experiments having independent subject groups; 301 subjects / 255 activation foci in aggregate) for inclusion in the meta-analysis. These studies are listed in Table 1.

5.2.1.2 Selecting precue studies

Prospective cueing (or Posner cueing) is a ubiquitous manipulation in cognitive psychology. To restrict our meta-analysis to a manageable number of studies, we included only studies that cued spatial attention with an endogenous (symbolic) foveal cue. We included only those studies that could differentiate responses to the cue from subsequent responses to the cued targets. 8 studies (101 subjects / 128 foci) were included in the meta-analysis of precueing. These studies are listed in Table 2.

RETROCUE META-ANALYSIS							
Study	**Year**	**Study #**	**Experiment #**	**N subjects**	**Spatial cue?**	**Space /** software	**Notes**
Rowe et al. (2000), Science, 288(5471), 1656–1660.	2000	1	1	6	1	**Talairach,** SPM99	Memory for location of three circles. Spatial retrocue to one circle.
Rowe & Passingham (2001), NeuroImage, 14(1), 77–86	2001	2	2	6	1	**Talairach,** SPM99	Memory for location of three sequentially presented dots. Numerical retrocue (by sequence location) to one dot.
Raye et al. (2002), NeuroImage, 15(2), 447-453	2002	3	3	12	0	**Talairach**, NIS/AFNI/Talairach Daemon	Memory for single words. Refresh cue contrasted with re-presentation of word.
‘’	‘’	3	4	8	1	‘’	Memory for three words in a column. Retrocue to one word contrasted with reading word again.
Johnson et al. (2003), Cerebral Cortex, 13(3) 265-273	2003	4	5	14	0	**Talairach,** NIS, AFNI	Memory for single words. Refresh cue contrasted with re-presentation of word.
Nobre et al. (2004), Journal of Cognitive Neuroscience, 16(3), 363–373	2004	5	6	10	1	**Talairach**, SPM99	Memory for colour of 2 or 4 items bound to location. Spatial retrocue to one item.
Johnson et al. (2004), Psychological Science, 15(2), 127–132.	2004	6	7	14	0	**Talairach,** AFNI	Memory for single words. Refresh cue contrasted with re-presentation of word.
Lepsien et al. (2005), NeuroImage, 26(3), 733–743	2005	7	8	10	1	**MNI,** FSL	Memory for colour of 2 or 4 items bound to location. Spatial retrocue to one item.
Johnson et al. (2005), Cognitive, Affective, & Behavioral Neuroscience, 5(3), 339–361	2005	8	9	14	0	**Talairach,** NIS, AFNI	Memory for single words. Refresh cue.
‘’	‘’	8	10	12	0	‘’	Memory for people, outdoor scenes, or words. Refresh cue.
‘’	‘’	8	11	17	1	‘’	Memory for spatial location. Refresh cue.
‘’	‘’	8	12	15	0	‘’	Memory for word. Refresh cue.
‘’	‘’	8	13	15	1	‘’	Memory for words. Refresh cue.
Lepsien & Nobre (2006), Cerebral Cortex 17(9), 2072–2083	2007	9	14	14	0	**MNI,** SPM2	Memory for face and house stimuli. Retrocue by category.

Yeh et al. (2007)Brain Research.	2007	10	15	10	1	**Talairach,** SPM99	Memory for colour of four items (tagged by location). Spatial retrocue to one item.
Johnson et al. (2007), NeuroImage, 37(1), 290–299	2007	11	16	15	1	**Talairach,** SPM2	Memory for face and scene stimuli. Retrocue to one item.
Yi et al. (2008), Journal of cognitive neuroscience	2008	12	17	8	0	**Talairach,** SPM2	Memory for scene stimuli. Refresh cue.
Raye et al. (2008), Journal of Cognitive Neuroscience, 20(5), 852–862.	2008	13	18	29	0	**Talairach,** AFNI, Talairach Daemon	Memory for either 1 or 3 words. Refresh cue.
Johnson et al. (2009), Journal of Cognitive Neuroscience, 21(12), 2320–2327	2009	14	19	14	1	**MNI,** SPM5	Memory for face and scene. Retrocue to face or scene.
Nee & Jonides (2009), NeuroImage, 45(3), 963–975	2009	15	20	18	0	**MNI,** SPM2	Memory for either three or six words. Retrocue to sets of three words.
Roth et al. (2009), NeuroImage, 48(3), 601–608	2009	16	21	22	0	**MNI,** SPM5, BioimageSuite, AFNI	Memory for single words. Refresh cue.
Higo et al. (2011), PNAS, 108(10), 4230–4235	2011	17	22	21	0	**MNI,** FSL	Two of house, face, or torso presented sequentially (pair always from different categories). Numerical cue to select one item.

Table 1 **Studies included in the retrocue meta-analysis**

PRECUE META-ANALYSIS							
Study	**Year**	**Study #**	**Experiment #**	**N subjects**	**Spatial cue?**	**Space /** software	**Notes**
Mangun et al. (2000), Nature Neuroscience, 3(3), 284–291	2000	1	1	6	1	**MNI,** SPM97	Discrimination task, left/right checkerboard stimuli. Left/right precue.
Corbetta et al. (2000), Nature Neuroscience, 3(3), 292–297	2000	2	1	13	1	**Talairach**	Detection task (flashed asterisk). Left/right precue.
Giesbrecht, B et al. (2003), NeuroImage, 19(3), 496–512	2003	3	1	10	1	**MNI,** SPM99	Discrimination task. Location precue.
Nobre et al. (2004), Journal of Cognitive Neuroscience, 16(3), 363–373.	2004	4	1	10	1	**Talairach,** SPM99	Memory encoding for 2 or 4 coloured items. Quadrant-wise spatial precue.
Woldorff et al. (2004), Journal of Cognitive Neuroscience, 16(1), 149–165.	2004	5	1	20	1	**Talairach,** SPM99	Detection task, faint target. Left/right precue.
Wilson et al. (2005), NeuroImage, 25(3), 668–683	2005	6	1	16	1	**MNI,** SPM99	Discrimination task. Spatial precue.
de Haan et al. (2008), Brain Research, 1204, 102–111.	2008	7	1	12	1	**MNI,** SPM2	Cue to one of several peripheral stimuli.
Egner et al. (2008), Journal of Neuroscience, 28(24), 6141–6151	2008	8	1	14	1	**MNI,** SPM5	Search for oddball target. Spatial precue.

Table 2 **Studies included in the precue meta-analysis**

5.2.1.3 ALE meta-analysis

ALE (Activation Likelihood Estimation) is a method for performing meta-analysis of functional imaging studies (Laird et al., 2009; Turkeltaub et al., 2002). The freely available software tool GingerAle (Eickhoff et al., 2009) was used to run the ALE analysis. Activation foci from each experiment were converted into Gaussian 'activation likelihood' distributions around the foci, where the variance of the distribution is inversely related to the number of subjects in the experiment. The activation likelihood distributions are combined within each experiment by taking the maximum of the foci distributions (thereby avoiding an overweighting of locations with two nearby local maxima), and between experiments by taking the union of the foci distributions across experiments. This yields a single ALE map across studies. The ALE map is converted to a p-value map by finding the probability of each ALE value across the whole brain space. This p-value map is thresholded at p=0.001 to form clusters. P-values are assigned for each cluster on the basis of their volume, by comparison with a null distribution of cluster volumes (null hypothesis of no common spatial patterns across studies). This distribution is obtained by running the ALE analysis on 5000 simulated datasets with the same properties as the true data (i.e. number of experiments, subjects per experiment, number of foci per experiment), other than that the foci locations are randomised. Clusters with a p-value less than 0.05 were considered significant.

5.2.2 Meta-analysis results

The ALE analysis yields a set of significant activation clusters. The spatial extent of the significant clusters from each analysis is shown in the FSL MNI standard brain in Figure 5-1, and in a surface projection onto the CARET cortical mesh in Figure 5-2. A table of significant clusters with local maxima is given for each analysis separately (precue, Table 4; retrocue, Table 3).

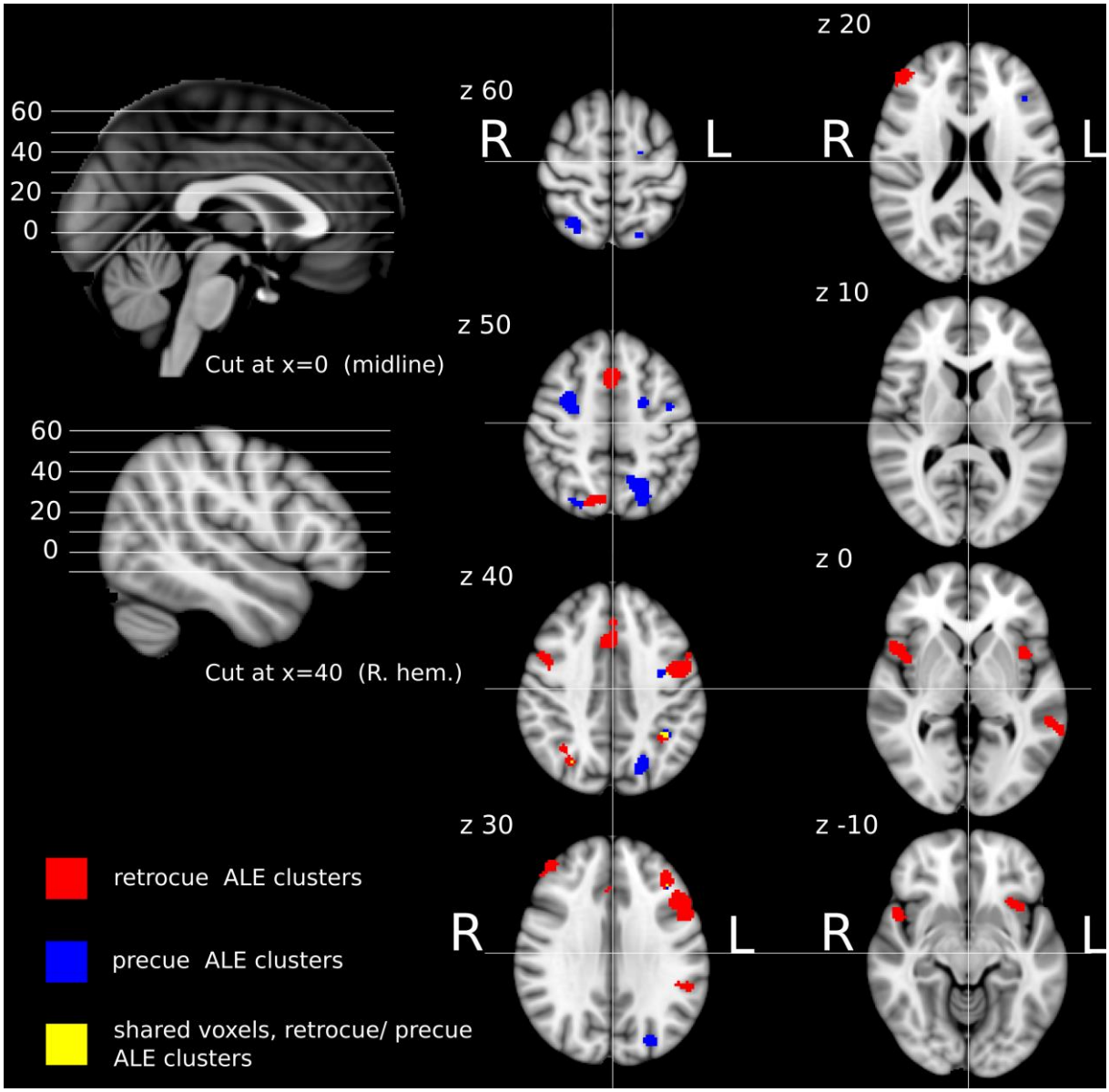


Figure 5-1 **Significant activations from the precue and retrocue meta-analysis** displayed on the FSL's MNI152 brain (shown in radiological convention).

RETROCUE META-ANALYSIS					
Label	**Cluster #**	**Local Maxima (MNI coordinates)**			**Cluster volume (mm^3)**
		x	**y**	**z**	
Right anterior MFG	1	44 34	46 50	24 16	1696
Left anterior MFG	2	-40	36	28	944
Left precentral Left anterior MFG "	3	-50 -50 -54	-2 22 12	40 28 34	5040
Right anterior insula	4	48	12	-4	2056
Left anterior insula	5	-40	14	-4	1304
dACC / pre-SMA "	6	0 -2	18 30	48 36	3272
Right precentral	7	48	6	42	440
Left MTG "	8	-58 -66	-38 -42	0 0	824
Left TPJ "	9	-56 -48	-38 -38	32 32	480
Left IPS	10	-38	-48	44	488
Right SPL "	11	12 18	-70 -62	52 54	1400
Right IPS " "	12	34 38 30	-58 -44 -64	42 44 38	1048

Table 3 **Retrocue meta-analysis**: significant clusters and their local maxima

PRECUE META-ANALYSIS					
Label	**Cluster #**	**Local maxima (MNI coordinates)**			**Cluster volume (mm^3)**
		x	**y**	**z**	
Left anterior MFG	1	-40	30	24	472
Right FEF	2	32	0	50	1560
Left precentral " / left FEF	3	-36 -42	-4 -4	42 52	464
Left FEF "	4	-22 -22	-2 -8	50 58	512
Left IPS	5	-40	-48	36	544
Right SPL	6	26	-56	58	1112
Left IPS " "	7	-22 -18 -22	-66 -58 -68	54 54 40	3960
Right SPL "	8	22 28	-72 -68	50 44	768
Right IPS0/V7 "	9	34 34	-78 -74	26 26	584
Left IPS0 / V7	10	-28	-78	30	400
Left occipital	11	-46	-70	-10	800
Right occipital "	12	34 34	-80 -84	14 14	800

Table 4 **Precue meta-analysis**: significant clusters and their local maxima

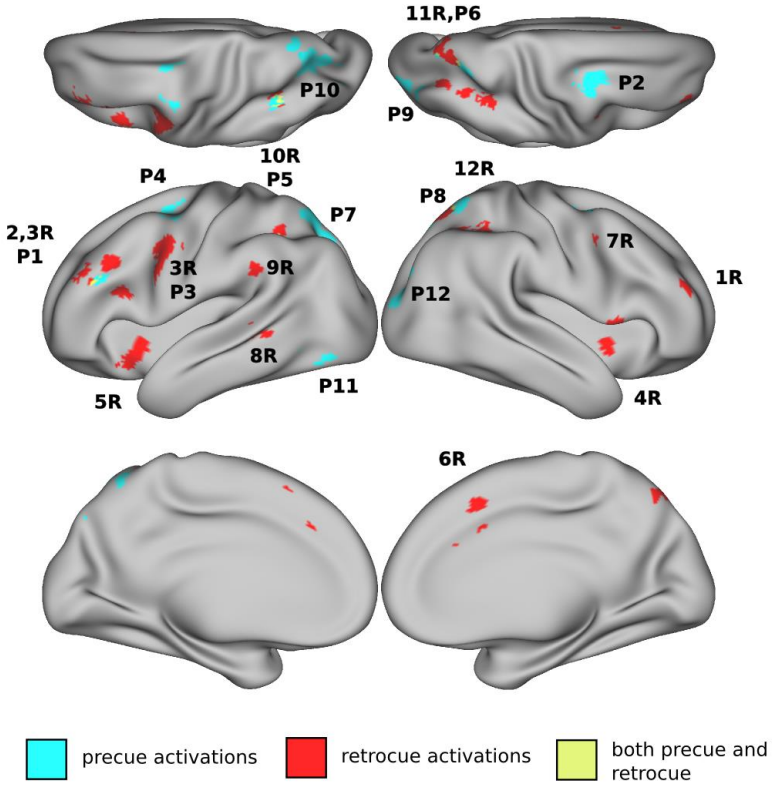


Figure 5-2 **Meta-analysis results displayed on the CARET cortical surface.** , '**P#**' labels indicate the location of a cluster from the precue analysis, '**#R**' labels indicate the location of a cluster from the retrocue analysis (as listed in Table 4 and Table 3, above). NB. **P11** is not visible on the cortical surface.

5.2.3 Meta-analysis discussion

5.2.3.1 Dual networks for cognitive control

Spatial attention is an example of a cognitive function that has been described at the network level (Mesulam, 1990; Petersen & Posner, 2012; Posner & Petersen, 1990). Following the introduction of brain imaging tools, a distributed network of frontal and parietal regions termed the dorsal or fronto-parietal attention network was identified using PET (Corbetta et al., 1993; Nobre et al., 1997) and further characterized with fMRI (Corbetta et al., 1998; Gitelman et al., 1999). By contrast, attempts to map the anatomical substrates of cognitive control and working memory have tended to focus on single regions in the prefrontal cortex – in particular, the dorsolateral prefrontal cortex (Koechlin, Ody, & Kouneiher, 2003; Rowe et al., 2000) and anterior cingulate cortex

(MacDonald, 2000). More recently, the emphasis has shifted to defining control networks (Dosenbach et al., 2008; Duncan, 2010; Petersen & Posner, 2012). Dosenbach and colleagues (Dosenbach et al., 2008) and Seeley and colleagues (Seeley et al., 2007) have independently formulated dual-network accounts of cognitive control, incorporating the fronto-parietal network previously associated with selective attention, and distinguishing it from a second cingulo-opercular network.

Resting-state fMRI allows us to characterize functional networks in the brain by observing which regions tend to covary in BOLD activation when people are not performing any task (C. F. Beckmann, DeLuca, Devlin, & Smith, 2005). Dosenbach and colleagues first identified task control regions by looking for cortical areas that activated across a wide range of cognitive tasks (Dosenbach et al., 2006). They then performed a graph analysis of the pairwise functional coupling between these regions, as indexed by the correlation between their BOLD signals in a resting state scan (Dosenbach et al., 2007). They found that the co-activation pattern clustered task control regions into two major networks: a fronto-parietal network, and a cingulo-opercular network. Seeley et al. (Seeley et al., 2007) performed a similar fMRI study to investigate network structure for task control, using both ROI seed-based correlation analysis and ICA analysis of resting state data to cluster task-positive regions identified in a previous study (Krasnow et al., 2003). Their results were very similar to those of Dosenbach and colleagues, distinguishing what they termed the executive-control network (fronto-parietal) from a salience-processing network (cingulo-opercular). They emphasised that the cingulo-opercular network was more extensively coupled to subcortical areas (including striatum) than the fronto-parietal network.

The fronto-parietal and cingulo-opercular networks identified by Dosenbach and colleagues are illustrated in Figure 5-3 (adapted from Dosenbach et al., 2007).

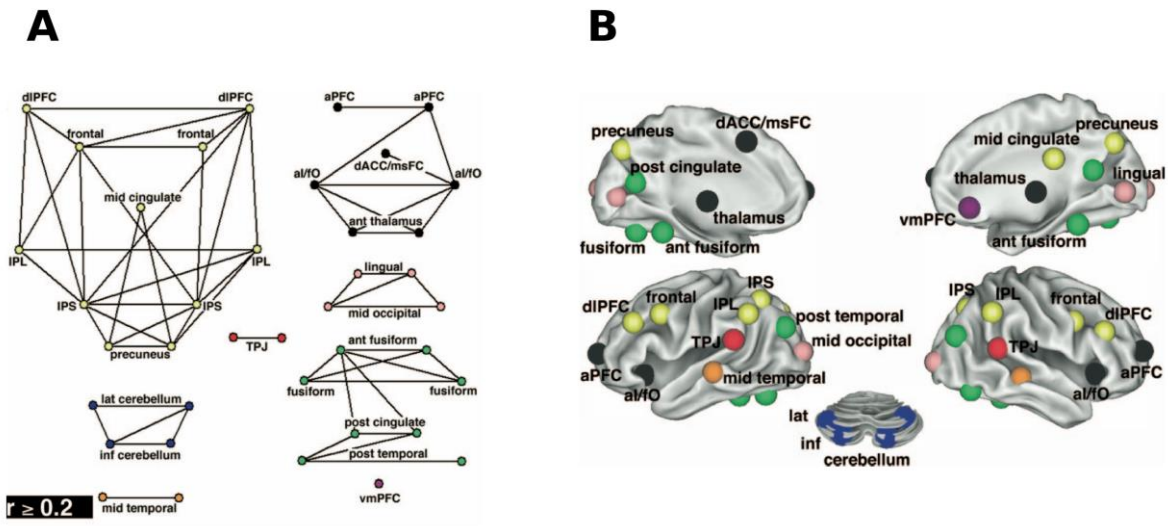


Figure 5-3 **Control network clustering based on resting state functional connectivity MRI (rs-fcMRI).** In **A**, fronto-parietal nodes are shown in yellow, cingulo-opercular nodes in black. *Figure adapted from Dosenbach et al., 2007.*

Fronto-parietal network nodes (yellow, Figure 5-3) include intra-parietal sulcus (IPS), inferior parietal lobule (IPL), precuneus, a precentral region ventral to FEF termed 'frontal'[17], and dorso-lateral prefrontal cortex (dlPFC; or anterior MFG). These regions have all been associated with spatial attention (Corbetta et al., 1998; Gitelman et al., 1999; Szczepanski, Konen, & Kastner, 2010). The cingulo-opercular network (black spheres, Figure 5-3) contains a dorso-medial node loosely termed the dorsal anterior cingulate (dACC[18]), the anterior insula/frontal operculum, the anterior PFC[19], and the thalamus. Seeley et al. (2007) distinguish between the more anterior dACC (MNI [±10,34,24]) and the more posterior pre-SMA (MNI [10,16,50], closer to Dosenbach's

---

17 This region is close to both IFJ and a more posterior precentral region variously termed iFEF (Derrfuss et al., 2012) and precentral cortex (Szczepanski et al., 2010). This region is termed 'precentral cortex' in the context of the meta-analysis results and MEG data, but iFEF where dissociated from IFJ.

18 In the region of SMA/pre-SMA (Sallet et al. 2013; Zhang et al. 2012). The ROI used for MEG analysis is in pre-SMA and that term is adopted below.

19 In the vicinity of lateral frontal pole, Neubert at al. (2014).

dACC coordinate). Both sites are functionally connected with the cingulo-opercular network in their analysis. This anatomical dissociation is in agreement with recent parcellations of the dorsomedial PFC using diffusion-weighted MRI (Sallet et al., 2013) and resting state functional connectivity (S. Zhang et al., 2012). The more posterior SMA ([~MNI ±10,4,59]) is a premotor region directly coupled with motor cortex, whereas the more anterior pre-SMA ([~MNI ±14,23,52]) is transitional between premotor cortex and prefrontal cortex and is coupled with prefrontal and premotor regions (Sallet et al., 2013).

Dosenbach et al. (2006; 2007) proposed that the cingulo-opercular network was responsible for instantiating, maintaining and adjusting 'task set' across a wide range of task domains, based on the observation that the dACC and insula/operculum were particularly active during the initiation and termination of a wide range of tasks, showed a sustained signal during task performance, and activated following errors. They speculated that the fronto-parietal network meets specific task demands within trial, given it exhibits a more phasic signal locked to task events. Seeley et al. (2007) associated the cingulo-opercular network with salience processing and interoception. They emphasise the coupling of the cingulo-opercular network with subcortical sites linked to motivational processing (e.g. striatum), and argue that this network is responsible for maintaining task set in order to best meet homeostatic goals, whilst top-down control of posterior cortex (i.e. operations on perceptual representations) is mediated by the fronto-parietal network. These two accounts are broadly similar: the cingulo-opercular network is concerned with task set, and the fronto-parietal network with top-down control. Whether the relationship between the two networks is hierarchical, with the cingulo-opercular network triggering activity in the fronto-parietal network, was unclear. According to Dosenbach and colleagues, the two networks constitute parallel control mechanisms, addressing different aspects of task

control (Dosenbach et al., 2008; Petersen & Posner, 2012). A more hierarchical view would be in line with the hypothesis that the dACC is responsible for detecting conflict between competing cognitive representations or processes, and for triggering recruitment of the fronto-parietal network to boost top-down control and resolve the conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Yeung, 2013).

Comparing Figure 5-3 (control networks) with Figure 5-2 (meta-analysis results): whilst both precues and retrocues activate fronto-parietal nodes, retrocues additionally recruit the cingulo-opercular network nodes (anterior insula/fO and dorsomedial PFC). Within dorsomedial PFC, the peak retrocue activation was in the pre-SMA (Sallet et al., 2013). Besides this broad pattern, there are further differences. Firstly, in the area of the precentral gyrus, precues recruit FEF, whereas retrocues recruit a more ventral precentral region. Secondly, retrocues, but not precues, recruit the mid-temporal gyrus (MTG) and temporo-parietal junction (TPJ). These latter regions do not fall into either of Dosenbach and colleagues' two control networks (Figure 5-3), in which bilateral TPJ and MTG form isolated node pairs.

To interpret the pattern of activations observed in the meta-analysis from the network perspective, the broad functional proposals about the function of each of the executive networks introduced above (Dosenbach et al., 2008; Petersen & Posner, 2012; Seeley et al., 2007) need to be squared with specific hypotheses about the roles their constituent regions play in preparatory attention and control of working memory.

5.2.3.2 The fronto-parietal network in retrocueing

One school of thought emphasises the role of the fronto-parietal network in retrocueing, and the similarity between the patterns of activation seen in this network following a prospective and a retrospective cue (Nobre et al., 2004). Lateral prefrontal regions (dlPFC, FEF, iFEF, IFJ) are suggested to play a similar role in directing attention both in

the sensory domain and within short-term memory, working in concert with parietal regions (IPS, SPL) to mediate top-down biasing of representations in perceptual cortex (Gazzaley & Nobre, 2012). This top-down influence is held to prioritize the representation of the cued item, either by biasing encoding in the case of precues (Murray et al., 2011) or biasing maintenance in favour of the cued item in the case of retrocues (Lepsien & Nobre, 2006; Nobre et al., 2004).

Within the fronto-parietal network, the inferior frontal junction (a region close to the ventral precentral site identified in the retrocue meta-analysis) is implicated as a critical region in top-down biasing. Zanto and colleagues (Zanto, Rubens, Thangavel, & Gazzaley, 2011) used fMRI to identify frontal regions whose functional connectivity with V4/V5 was specifically increased when subjects were selectively encoding stimuli on the basis of colour/motion (respectively). The right IFJ emerged as common to both feature types, and TMS applied to this region reduced attentional modulation of the P1 EEG component during encoding. Similarly, Kuo and colleagues (Lepsien et al., 2005) ran a psychophysiological interaction (PPI) analysis on fMRI data to look for frontal regions involved in retinotopic modulation of visual cortex, following a retrocue to an item from a spatially distributed memory array. The right inferior frontal sulcus (IFS) in particular (close to Zanto and colleagues' right IFJ region) had stronger functional connectivity with the attended part of visual cortex following a retrocue.

Other fronto-parietal regions besides IFJ, including dlPFC and FEF, have been shown to increase in BOLD coherency with perceptual regions following retrocues (Kuo, Yeh, Chen, & D'Esposito, 2011). Of these sites, the FEF has long been implicated in the control of eye movements but also in covert shifts in spatial attention (Bichot & Schall, 1997; Corbetta et al., 1998; Kodaka, Mikami, & Kubota, 1997). Interestingly, the meta-analysis presented here dissociates FEF and a more ventral precentral region close

to the IFJ, the former activating following precues, and the latter following retrocues. A recent fMRI study that aimed to disentangle different regions in the left inferior frontal sulcus dissociated inferior FEF (iFEF) from the IFJ (Derrfuss, Vogt, Fiebach, Cramon, & Tittgemeyer, 2012). One sub-maxima of the left precentral cluster associated with retrocues in the meta-analysis is very close to their left iFEF coordinate (Derfuss et al., iFEF [-55, 3, 42]; meta-analysis left precentral retrocue activation [-50, -2, 40]), but this cluster in the meta-analysis also extended more ventrally incorporating their IFJ coordinate ([-42,6,33]), which is also close to the IFJ co-ordinate reported by Kuo and colleagues and Zanto and colleagues (discussed above). The meta-analysis activation in bilateral FEF following precues is unsurprising given the FEF's well documented involvement in allocation of covert spatial attention, but the lack of an activation in FEF with retrocueing suggested that some mechanisms recruited with prospective cues – perhaps those particularly involved in covert shift in spatial attention – are not recruited as part of mnemonic selection.

5.2.3.3 Direct role for frontal operculum in top-down biasing

If the fronto-parietal network is the locus of a task-general top-down control system, then the additional cingulo-opercular activation occurring following retrocues seems unlikely to be directly implicated in mediating top-down biasing of perceptual cortex. However, a different perspective emphasises the role of the anterior insula/fO, arguing that this is the critical region responsible for top-down modulation of memory representations following a retrocue. Higo and colleagues (Higo et al., 2011) employed a task in which on each trial, two stimuli were presented in sequence, drawn from two of three categories: images of houses, images of faces, or images of human bodies with the head obscured. A subsequent cue then either allowed subjects to direct their attention towards one or other of the stimuli in memory (selective cue), or gave no information (unselective cue). Shortly afterwards, a probe array appeared, containing three

objects. Subjects had to judge whether the cued item was present in the probe array in selective cue trials, and whether either of the two memory items was present in the array in unselective cue trials. The bilateral frontal operculum (fO) was the only region in which activation was higher with selective as compared to unselective cues. The stimulus categories were chosen because they activated distinct occipito-temporal regions: the parahippocampal place area (PPA), the fusiform face area (FFA), and the extra-striate body area (EBA)[20]. Cueing to a stimulus of a specific category boosted activity in the region preferentially sensitive to that category. A PPI analysis was used to investigate the link between the fO activation and cue induced modulation of activity in functionally specialized occipito-temporal areas. Activity in the fO was found to predict modulation of occipito-temporal regions, but only when the relevant category for that region had been cued. Finally, 15 minutes of 1Hz TMS was used to temporarily deactivate fO prior to an fMRI scan, in which cue-modulation of the category-specific occipito-temporal regions was found to be reduced. On the basis of these data, the authors argued that the fO was directly responsible for top-down biasing of perceptual representations. A subsequent study (Nelissen et al., 2013) by the same group, used MVPA to decode activity in perceptual cortex in a similar paradigm, and found that the fO uniquely correlated with cue-induced changes in visual cortex patterns.

From the dual-network perspective, this direct role for the fO is inconsistent with an account which implicates fronto-parietal regions, and specifically the IFJ/IFS, in top-down modulation of perceptual cortex during mnemonic selection. Direct involvement of the fO in retrocue-induced top-down modulation of perceptual cortex would imply an alternative pattern of involvement of the frontal networks, with the cingulo-opercular network directly mediating the effects of retrocues, and the

[20] The relatively higher sensitivity of these areas to the stimulus classes in question implies some degree of functional specialization, but not that these areas *uniquely* process these categories - the nomenclature is for convenience only.

fronto-parietal nodes involved in some other capacity (perhaps relating to WM load (Higo et al., 2011) or directing preparatory attention for upcoming probe stimuli (Stokes, 2011)).

5.2.3.4 Sequential organization

Both of the accounts discussed so far emphasise top-down biasing of perceptual representations. However, as outlined in chapter 3, downstream processes of output gating associated with the prioritization of a single item for task control might also be critical in mediating the retrocue benefit. One possibility is that the fronto-parietal network is responsible for selecting and manipulating representations, and by contrast the fronto-opercular network has a more downstream role: it is responsible for determining which representations are permitted to guide action. This sequential account diverges from the model by Dosenbach and colleagues, in which the two executive networks operate in parallel (Dosenbach et al., 2008). It is similar to a cascading account of cognitive control proposed by Banich (Banich, 2009), in which dlPFC is suggested to select task-relevant representations, and dACC to select the information that guides responding.

A number of considerations motivate this idea. One is the additional recruitment of the cingulo-opercular network following retrocueing as compared to prospective cueing: if the cingulo-opercular nodes were associated with the prioritization of a single representation for action control, this might explain their selective recruitment by retrocues. A precue allows for preparatory top-down biasing for encoding of a representation, and accordingly recruits fronto-parietal mechanisms involved in top-down selective attention, but there is not yet an encoded representation for which to gate an action link. By contrast, a retrocue allows subjects to not only resuscitate perceptual content associated with the cued item (recruiting the fronto-parietal network

to exert a top-down influence on sensory cortex) but also selectively prioritize the cued item to guide action[21]. As a corollary retrocues allow subjects to prevent accessory items from guiding behaviour (c.f. 'misgating', chapter 3, and its amelioration by retrocues).

The time-course of activation of the two control networks in imaging experiments supports the sequential view. For example, Ploran and colleagues (Ploran et al., 2007) used a very simple decision task to dissociate the time-course of activation of various frontal regions. They slowly revealed an image over 16s, and asked their subjects to press a button as soon as they recognized what was displayed. They reasoned that regions associated with evidence accumulation for the categorization would begin to ramp up early as the image was revealed, but would ramp up more quickly in trials when recognition occurred sooner. By contrast, regions associated with the 'moment of recognition' – i.e. the moment the identity of the item entered cognitive awareness and could be transformed into a response – would show a phasic activation around the time of the button press. They classified time-courses over a large number of spatial ROIs according to these criteria. Fronto-parietal regions, including the ventral precentral and the anterior MFG and various regions in parieto-occipital cortex (including left IPS), were 'evidence accumulators'. By contrast, bilateral frontal operculum and dACC/pre-SMA all had phasic 'moment of recognition' time-courses. Whilst fronto-parietal regions were involved in accumulating evidence for perceptual classification, only once a fully formed item representation was available was the cingulo-opercular network activated, immediately preceding the production of a recognition response.

---

[21] This is not to say that retrocues allow subjects to prepare a *specific* action plan. In the current task, it is impossible for the subject to know which finger to press until the probe stimulus is presented. Rather, prioritization of the cued item is suggested to prepare the ground for the sequence of events following probe presentation (including the comparison of the cued memory representation with the probe stimulus) that lead to a response, by functionally prioritizing the representation of the cued item and perhaps also inhibiting the representations of the uncued items to prevent them from influencing behaviour.

This view of the cingulo-opercular network is consistent with association between dorsomedial prefrontal cortex and action selection – particularly response selection in contexts where pre-potent responses have to be over-ridden (Rushworth, Buckley, Behrens, Walton, & Bannerman, 2007). For example, Crone and colleagues (Crone, 2005) scanned subjects whilst they performed a task in which they responded either to univalent stimuli, which were always associated with the same response, or bivalent stimuli, which were associated with two different responses: for these stimuli, subjects had to use a contextual cue to select the right response. The pre-SMA was activated when switching between bivalent rules. Contrasting activation in response to cues for bivalent and univalent rules revealed a wider network of frontal regions including the MFG and bilateral anterior insula.

Clinical observations also link the dorsomedial PFC to response selection. Lesions of the dorsomedial PFC can in extreme cases give rise to a syndrome of akinetic mutism (Otto et al., 1998), in which there is a complete dearth of self-generated action and speech (without physical paralysis), and milder impairments of spontaneous/internally generated responding are seen following less catastrophic cingulotomy (R. A. Cohen, Kaplan, Moser, Jenkins, & Wilkinson, 1999). Conversely in obsessive compulsive disorder (OCD), in which poorly controlled thought and action generation are the key clinical feature, dACC is hyperactive compared with controls in tasks inducing response conflict, and cingulotomy is an effective treatment of last resort for refractory illness (Milad & Rauch, 2012).

5.2.3.5 Investigating network roles with MEG

Each of the functional proposals discussed above makes a different prediction about the time-course of activation in the frontal networks following a retrocue. fMRI is not suited to testing these differences, because the slow haemodynamic response function

(HRF) obscures fine-grained differences in activation time-courses. In contrast, MEG is ideally suited to this role. In a previous chapter the time-course of reactivation of sensory cortex following a retrocue was established, providing a reference against which to compare frontal activations. To summarize the predicted activation time-courses, the view just outlined (in which the fronto-parietal network mediates top-down biasing, and the cingulo-opercular regions mediate output gating) predicts that fronto-parietal nodes will activate first, in order to retrieve perceptual content associated with the cued item. This activation should be at the same time as, or preceding, sensory reactivation as indexed by alpha lateralization. The cingulo-opercular nodes should activate following sensory reactivation, to prioritize the selected item. By contrast, Higo and colleagues' proposed role for the frontal operculum in top-down modulation of sensory representations would see the fO activating first, to drive top-down modulation of sensory cortex. The fronto-parietal network might activate either simultaneously or subsequently (e.g. directing attention for the upcoming probe item). Finally, a hierarchical view (Botvinick et al., 2001; Yeung, 2013) would see the dorsomedial PFC (and perhaps fO) activating first, which would drive a subsequent fronto-parietal activation. Both network activations should precede the sensory reactivation in this case.

The derivation of MEG ROIs from the meta-analysis activation maps is discussed in section 5.2.4 (**Spatial ROIs for MEG analysis**), and the time-course of induced activations in control-network ROIs is presented in section 5.4 (**MEG results**) and used to inform competing theories about the role of these nodes in mediating the cue benefit.

5.2.4 Spatial ROIs for MEG analysis

The results of the meta-analysis were used to derive a set of spatial ROIs for MEG. ROIs were mirrored in the left and right hemisphere (to permit comparisons of laterality), and sufficiently sparse to sample the regions of interest sensibly, given the limited spatial resolution of MEG. The results from both meta-analyses were first simplified into a sparse set of coordinates, and then mirrored across the hemispheres. These steps are described in more detail in Appendix 9.2. In addition to the regions found using the meta-analysis, an occipital ROI (from Dosenbach et al., 2007) was added as well as an ROI in the hand area of motor cortex (Yousry, 1997).

	MNI coordinates		
	x	y	z
1/2: IPS0	±34	-76	26
3/4: mid IPS	±30	-68	40
5/6: anterior IPS	±39	-48	40
7/8: SPL	±12	-68	60
9/10: FEF	±27	-3	52
11/12: precentral	±46	1	43
13/14: anterior MFG	±40	39	23
15/16: hand area	±35	-20	60
17/18: TPJ	±52	-38	32
19/20: MTG	±62	-40	0
21/22: occipital	±26	-94	0
23/24: anterior insula / fO	±44	13	-4
25: pre-SMA	0	24	42

Table 5 **MNI coordinates of MEG ROIs**

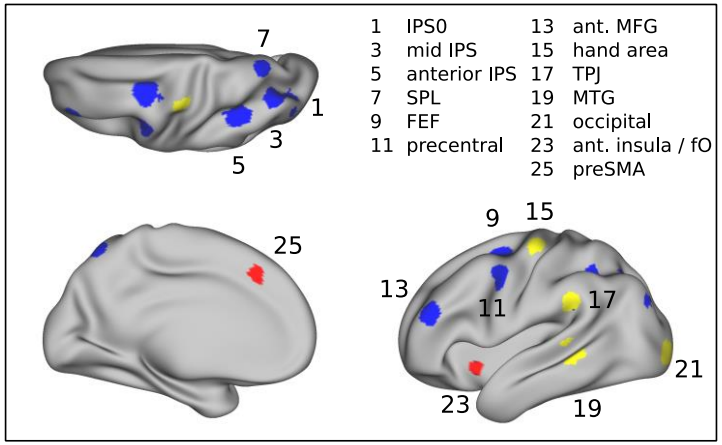


Figure 5-4 **MEG ROIs** shown on the left hemisphere of the CARET brain, as 8mm spheres. Following Dosenbach et al. (2007) on network membership of nodes, ROIs belonging to the dorsal attention network are coloured blue, ROIs belonging to the cingulo-opercular network red, and all remaining ROIs in yellow.

5.3 MEG methods

Experimental procedures for the precue/retrocue task, MEG data acquisition, data preprocessing and the beamforming algorithm are described in detail in chapters 2 and 4.

5.3.1 Time-frequency maps of cue-induced responses

In order to characterise the time-course of activation in each region of interest, a virtual electrode was created at each ROI coordinate using an LCMV beamformer, and the time-frequency representation of the data was computed from each trial at each virtual electrode. This was done separately for two broad spectral bands: a low frequency band between 3 and 30Hz, and a high frequency band between 30 and 120Hz.

The analysis was split into the high and low frequency band for two reasons. Firstly, the beamformer estimates a spatial filter based on the covariance properties of the data. The covariance of the data differs for different spectral bands. The beamformer weights should be over the narrowest practicable spectral window in order to optimise the covariance estimate for the analysed data (Gross et al., 2013 see also chapter 2). However, the tighter the bandwidth chosen, the more beamformer filters are needed to span the frequencies of interest, and the less information available for the estimation of each filter. A compromise is to beamform the low frequencies and the higher frequencies separately, in two broad bands, allowing the covariance estimate to differ between these two bands, but ensuring each analysis has enough information to stably construct a spatial filter.

The optimal algorithm for time-frequency decomposition is different for the low and high frequency ranges. The Hanning taper approach is preferred for the low frequency

band, whereas in order to best control the inherent trade-off between frequency resolution and time resolution and optimise signal-to-noise in the high frequency band, it is preferable to use a multi-taper approach. We used Fieldtrip's Hanning taper algorithm to compute the 3-30Hz TFR, choosing a Hanning window of 4 cycles of the underlying frequency (so that the width of the window in time varies with frequency). We used Fieldtrip's multitaper functionality to compute the 30-120Hz TFR, using 5 Slepian tapers for a frequency resolution of ±10Hz and time resolution of 150ms.

The time-frequency data were averaged across task conditions within subjects, masking out any times associated with eyeblinks (by setting data from these periods to NaN, and using MATLAB's *nanmean*). The condition averages were then subtracted within subjects. These contrasts were then averaged at the group-level to create time-frequency maps of cue-related activity for each ROI. Significance testing was performed by forming clusters in the time and frequency dimension for each ROI and comparing cluster area to one-sample permutation distributions of cluster area (derived by sign-flipping the subtractions for 50% of subjects).

5.3.2 Whole brain analyses

To check whether the ROIs fully captured cue-induced activity, cue-triggered changes in power were computed over the whole brain for the theta, alpha and beta bands. This was done by computing power time-courses for virtual sensors on an 8mm grid spanning the volumetric brain space. A GLM was fitted to these source-space data with a regressor picking out trials from each cue condition, and spatiotemporal cue activation maps were created by contrasts of the precue and retrocue conditions with the neutral cue conditions (analogous to the subtractions described in section 5.3.1). In order to threshold these maps statistically, 4D maps were first collapsed into 3D activation maps

by dividing the analysis epochs into sequential 300ms windows, and averaging the GLM beta values across these windows. This yielded 3D activation maps, which were separately tested for induced synchronization/desynchronization by comparing the volume of observed clusters with a permutation distribution of cluster volume (using FSL's *randomise* function).

5.3.3 Behavioural correlations

To investigate correlations between the activations seen in fronto-parietal and cingulo-opercular control regions with behavioural performance, a time-frequency 'mask' was defined from the group contrasts in the time-frequency domain for each effect of interest. Trial-wise data from individual subjects was averaged over the time-frequency bins that belonged to significant group-level clusters for the ROI in question (precue analysis: IPS, FEF, anterior MFG, pre-SMA; retrocue analysis: IPS, ventral premotor, anterior MFG, anterior insula, pre-SMA). This trial-wise measure was correlated with reaction time within each cue condition separately, for each effect of interest. The behavioural data were also median split within-condition using the trial-wise measures, and accuracy was compared between the two halves of the data. Finally a trial-wise alpha lateralization measure extracted from a pattern classification analysis on sensor space data (chapter 4) was also median split to investigate the link between trial-wise activation in control regions and sensory reactivation.

5.3.4 Orthogonalised power correlation

The meta-analysis results presented above were interpreted in terms of putative control networks inferred from BOLD correlations in fMRI (Dosenbach et al., 2007; Seeley et al., 2007). In order to test whether the MEG signal beamformed at ROI coordinates

derived from the meta-analysis were able to replicate aspects of this network structure, a pair-wise functional connectivity measure, orthogonalised power correlation (Hipp, Hawellek, Corbetta, Siegel, & Engel, 2012), was used to map the functional connectivity between ROIs, and between each ROI and the rest of the brain. This measure was computed for both resting state and task data, and used to replicate functional connectivity discovered using fMRI.

The correlation in the band-limited power between two brain regions gives a spectrally-resolved measure of functional connectivity somewhat analogous to the fMRI BOLD correlation that has been used establish pairwise connectivity in fMRI. However, signal spread in source space (i.e. common signal pickup) dominates the power correlation between ROIs (Schoffelen & Gross, 2009). Orthogonalised power correlation factor out this shared signal by exploiting the fact that if two virtual sensors are measuring the same underlying signal, it must be at identical phase in both sensors (Hipp et al., 2012). The metric is computed for pairs of ROIs. First the power of only that part of the band-limited frequency-domain signal from the first ROI that is orthogonal in the complex plane to the signal from the second is computed. This orthogonalised power timeseries is then correlated with the raw power timeseries from the second ROI. This factors out correlation resulting from the part of the signal with shared phase, and therefore cuts out artefactual power correlation resulting from signal spread.

Resting-state data are commonly used to partition brain regions into networks on the basis of their functional connectivity. A six-minute resting state MEG dataset was collected for each of the 50 subjects tested. These data were used to investigate the functional connectivity between the ROIs derived in section 5.2.4. The data were first pre-processed (pipeline described in more detail in chapter 2). Neuromag's Maxfilter

algorithm was used to de-noise the data, employing the spatiotemporal option which increases the robustness of filtering by using correlations between the inner, transitional and outer spheres to help determine which components to filter out. Segments of data corrupted by artefact were then identified by visual inspection, and eye-blinks were detected and excluded from the analysis using a semi-automatic algorithm. The orthogonalised power correlation was then computed for each ROI pair. Time-domain source-space signals for each virtual sensor were computed using an LCMV beamformer with covariance matrix estimated between 3 and 30Hz, excluding corrupted data segments and blinks from the covariance estimation. These time-domain signals were transformed to the frequency domain between 3 and 30Hz in 1Hz steps, using a Hanning taper spanning 7 cycles at each frequency, stepping the taper through the data in 250ms steps. For each ROI pair, the signals were then orthogonalised as described by Hipp and colleagues (2012), and the power correlation was computed by correlating the orthogonalised power in each frequency band over the whole duration of the resting state scan, excluding artefact-corrupted data segments and segments around eye-blinks. Because behavioural confounds were not relevant for this analysis, and the spatiotemporal Maxfilter option allowed for more robust de-noising, all 50 subjects' data were used.

Analogous functional connectivity measures were derived from the task data (for the 38 subjects for whom task analyses were computed). The data were pre-processed and transformed to the frequency domain as described above (section 5.3.1). The key difference between the resting and task analyses is that the power correlations were computed across trials for the task analysis, as opposed to across time for the resting-state analysis. This was done within condition and was time-resolved, resulting in a condition-wise time-frequency map of power correlation for each task condition, and for each of the analysis epochs (precue, array, retrocue). These data

were collapsed over time and condition to yield overall functional connectivity maps for the task data.

The power correlation analyses yield a matrix of correlations between ROI pairs for both resting and task data, for each frequency band separately. A basic network clustering analysis was performed, similar to those described for fMRI connectivity data (Dosenbach et al., 2007; Salvador et al., 2005). First correlations were separately averaged for the theta band (3-5Hz), the alpha band (8-12Hz) and the beta band (18-30Hz). Within each band, the correlations were converted into a distance measure by z-scoring the correlation values and then subtracting them from [maximum z-scored value +1][22]. An average linkage (Dosenbach et al., 2007; Salvador et al., 2005) hierarchical clustering algorithm (implemented in MATLAB's *linkage* function) was used to group these distance values and identify clusters of particularly correlated regions.

[22] Adding an offset makes sure all distances are positive, but provided the offset renders all distances greater than zero, its magnitude makes no difference to the cluster solution and can be arbitrarily chosen.

5.4 MEG results

5.4.1 ROI analysis of induced responses

5.4.1.1 Low frequency induced responses

The pattern of induced responses to informative cues in the fronto-parietal and cingulo-opercular ROIs was broadly similar in the left and right hemisphere. Where there were significant activations, they were mostly present in both hemispheres (separate time-frequency plots for left and right hemisphere ROIs are given in the appendix). There were two exceptions to this rule. Following precues, there was an activation in the right anterior MFG that was not present on the left, and a desynchronization in the left mid-IPS that was not present on the right. In order to plot the data economically, the time-frequency results were averaged over the left and right hemisphere ROIs, and are shown in Figure 5-5 below, in which the two unilateral effects are marked with an asterisk. These unilateral effects survived averaging over side.

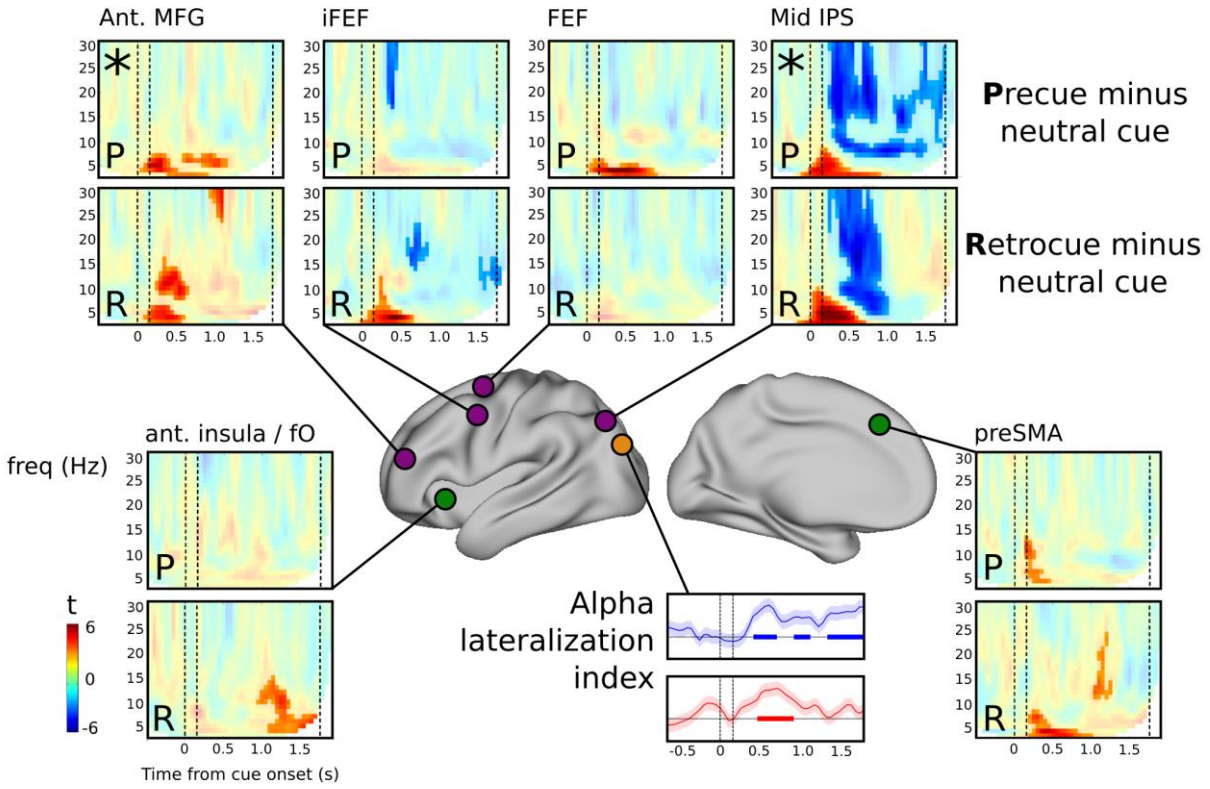


Figure 5-5 **Fronto-parietal and cingulo-opercular induced responses to informative cues.** Contrast [informative cue minus uninformative cue]. Precue effects are shown in the upper plots, retrocue effects in the lower plots. The time-course of 8-12Hz (alpha) lateralization in visual cortex is also shown to facilitate comparison with the time-course of control network activation. Full colours mark significant clusters. The asterisks mark effects that were unilateral (precue ant. MFG activation was RHem only; precue mid IPS desynchronization was LHem only). All other effects were present in both the right and left hemispheres. Cluster-permutation p-values were derived from data averaged across the hemisphere at the subject-level, with the exception of the two unilateral effects for which the p-value for the appropriate side is reported. The pattern of activations across fronto-parietal ROIs (blue markers) and cingulo-opercular ROIs (red markers) was consistent with the fMRI meta-analysis. **Precues** give rise to a sustained alpha/beta desynchronization in the left mid-IPS ($p<0.0002$) lasting until the presentation of the memory array, matching the time-course of alpha lateralization in occipital cortex, consistent with a role for left IPS as a proximal control region for this attentional effect. Right anterior MFG ($p=0.0004$, early cluster; $p=0.029$, late cluster) and bilateral FEF ($p=0.0016$) are activated in the theta band early following the cue. Cingulo-opercular nodes are not activated with the exception of a short-lived activation in the pre-SMA immediately following the cue ($p=0.006$). **Retrocues** gave rise to an alpha/beta desynchronization in the mid IPS ($p<0.0002$) which lasted until ~1s post-cue, matching the time-course of alpha lateralization. Activations in the theta and alpha/beta band in the anterior MFG ($p=0.0054$, $p=0.0004$) preceded the parieto-occipital effects. Consistent with the pattern in fMRI, retrocues did not activate the FEF but did give rise to a bilateral activation in the more ventral precentral ROI ($p=0.0036$). Retrocues also gave rise to activations in the cingulo-opercular nodes. The pre-SMA was activated immediately following the retrocue in the theta band ($p=0.0006$) and also later in the epoch in the beta-band (~1.2s post cue; $p=0.006$). At this later time-point, there was also a bilateral activation in the anterior insula/frontal operculum in the theta/alpha band ($p<0.0002$).

There are several similarities in the pattern of induced responses to precue and retrocues in ROIs associated with the fronto-parietal network. For both precues and retrocues there is a broadband desynchronization in the mid IPS, though this is transient

for retrocues and persists until the memory array for precues (resembling the time-course of alpha lateralization in occipital cortex / IPS0). For both cue types, this desynchronization is accompanied by a synchronization in the theta band, that is similarly more sustained for precues than for retrocues. Precues induced a bilateral theta-band response lasting until around 700ms post cue whereas retrocues induce a similar theta response in the precentral ROIs but not in FEF. This replicates the pattern seen in the meta-analysis, in which precues were found to activate FEF, whereas retrocues instead activated a more ventral region of the precentral gyrus. Both cue types activate the anterior MFG (right side only for precues), though the spectral signature and time-course is different. Precues give rise to a ~6Hz activation that appears immediately following the cue, and later reappears around 1s post-cue. Retrocues give rise to a response that has a theta component, but also a component at 10-15Hz lasting from cue offset until ~0.8s post-cue.

In contrast to the fronto-parietal regions, the full cingulo-opercular network was recruited only by retrocues, replicating the pattern observed in the fMRI meta-analysis. There are two phases of activation in the cingulo-opercular ROIs, one immediately following cue onset, and a second between 1s post-cue, and the probe stimulus (1.76s). The latter activation recruits both bilateral fO and the pre-SMA. The early activation is restricted to the pre-SMA, and is in the theta band. Precues also give rise to a brief activation of the pre-SMA immediately following the cue. The second phase of pre-SMA reactivation following retrocues appears to be restricted to the beta-band, but in fact this is because the ROI was placed on the midline, and the late theta-band activation in the pre-SMA was slightly right-lateralized. This theta activation in pre-SMA can be clearly seen in the whole-brain analysis in Figure 5-9, below (the full set of surface projections and volumetric plot for the whole-brain analyses are shown in the appendix).

The data suggest that following a retrocues there are two distinct phases of activation: first, perceptual content associated with the cued memory item is reactivated, which recruits the fronto-parietal network nodes. Once this top-down reactivation is complete, all three cingulo-opercular network nodes are activated (pre-SMA and bilateral fO). Figure 5-6 collapses the time-courses shown in Figure 5-5 into bars indicating significant activations, and plots them relative to the alpha lateralization in IPS0.

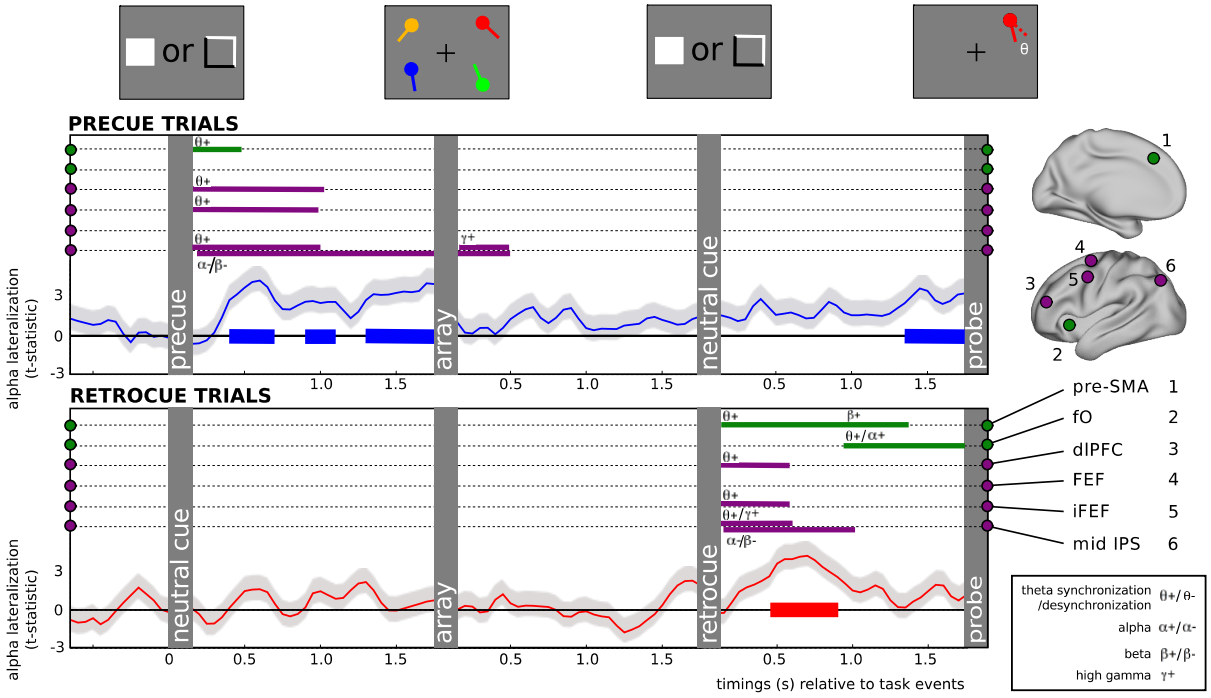


Figure 5-6 **Time-course of alpha lateralization and significant activation in control regions.** Each control region is associated with a horizontal dashed line on which a coloured bar is plotted for times at which there was a significant synchronization or desynchronization in that ROI. The pre-SMA activated in the theta band following retrocues. This activation overlaps with a later activation in the beta-band. The mid-IPS remains desynchronized in the alpha band until around 0.5s into the array epoch. A re-emergence of alpha lateralization prior to probe onset can be seen following precues. This is not accompanied by significant activation in control regions.

Two further ROIs, the middle temporal gyrus (MTG) and temporo-parietal junction (TPJ), were identified in the retrocue fMRI meta-analysis. In the network framework proposed by Dosenbach and colleagues, these nodes do not belong to either the fronto-parietal or cingulo-opercular networks, forming isolated node-pairs. Induced

responses to precues and retrocues in the MTG/TPJ are shown in Figure 5-7. The induced responses in TPJ resemble the induced responses seen in the parietal cortex (and it is possible that there is some signal bleed between these nearby ROIs). The MTG resembles the cingulo-opercular regions, in that it is recruited later in the epoch in the beta band.

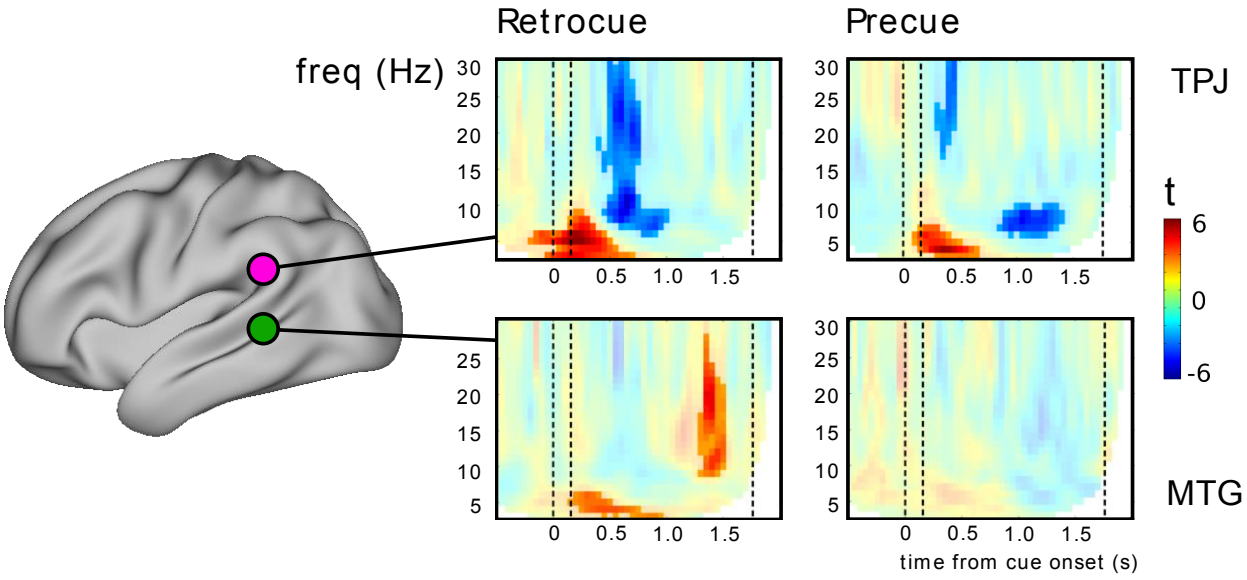


Figure 5-7 **Induced responses to informative cues in MTG/TPJ.** Both precues (RHS panels) and retrocues (LHS panels) activate the TPJ, and in both cases the pattern is similar to the pattern observed in mid-IPS (Figure 5-5). Only retrocues activate MTG, giving rise both to an early theta-band response and a later beta-band synchronization.

5.4.1.2 High frequency induced responses

A separate beamformer and time-frequency analysis was run for the 30 to 120Hz frequency band. The only significant modulations of gamma-band activity by cue were in the occipital cortex and IPS. Gamma band activations for the [precue minus neutral cue] contrast are plotted for both the epoch following the precue, and the epoch following the presentation of the memory array. Gamma band activations for the [retrocue minus neutral cue] contrast are plotted for the retrocue epoch.

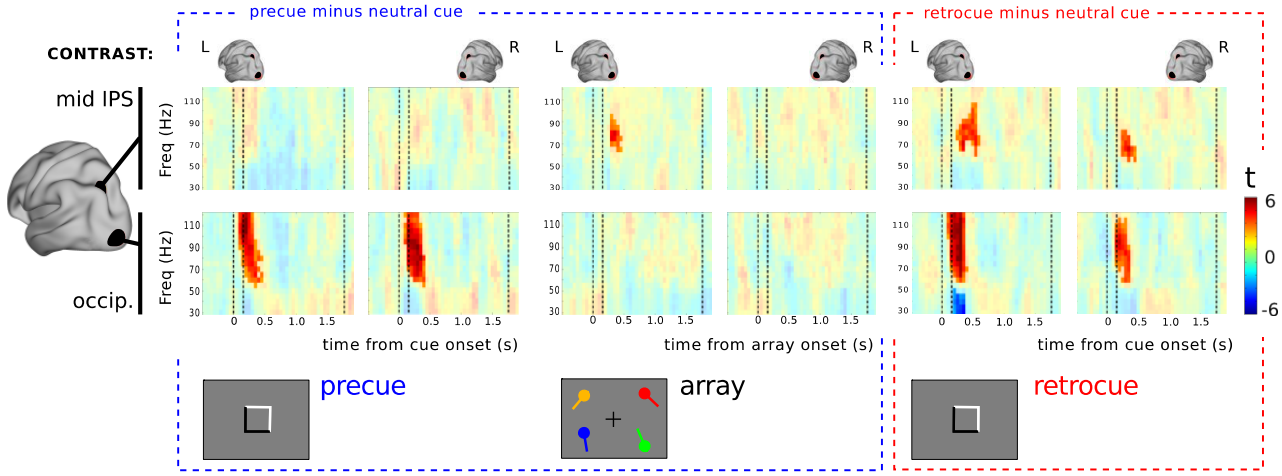


Figure 5-8 **Gamma band induced responses.** Significant activations are marked by full colour saturation. The top row shows modulations of gamma activity in precue trials compared to neutral cued trials. Following a precue (**left**), there are bilateral activations in the high gamma band in occipital cortex (both $p < 0.0002$), reflecting the enhanced ERF observed to informative as compared with uninformative cues (150-400ms post-cue). Following the memory array (**centre**), there is a burst of additional high gamma activity in the left mid IPS ($p = 0.015$) in trials where a precue had been given (250-450ms post-array). Like precues, retrocues (**right**) induce more gamma activity in visual cortex than do neutral cues (150-400ms post-cue, LHem $p < 0.0002$, RHem $p = 0.0014$), but they also induce bilateral activations in mid IPS, following slightly later after the cue (250-550ms, LHem $p < 0.0002$, RHem $p = 0.0172$). The overall pattern suggests that high gamma in mid-IPS is a marker for top-down selection of encoded stimuli, but not preparatory attention in advance of encoding.

A burst of activity in the high gamma band (from 60Hz upwards) in the occipital cortex in response to both precues and retrocues probably reflects the enhanced ERF to informative cue stimuli. There is also a burst of activity between 60 and 100Hz in the mid-IPS following the stimulus array in precue trials only (LHS), and following the retrocue (bilateral, but stronger on the left). High gamma in mid-IPS appears to correspond to the selection of an encoded representation. Following the stimulus array, subjects are able to select one out of the four presented items for encoding in precue trials, but have to encode all items unselectively in neutral cue (and retrocue) trials. Following the retrocue, subjects are able to select and prioritise one of the items already in memory. By contrast, following the precue there is not yet an encoded representation to prioritize and IPS gamma is not significantly increased.

In precue trials there is a double dissociation between occipital and IPS gamma: occipital gamma is increased following the precue, but IPS gamma is unchanged, whereas following the memory array, IPS gamma is increased but occipital gamma is unchanged. This double dissociation in precue trials stands in favour of the idea that the gamma activations in mid IPS were a correlate of selection, as opposed to simply reflecting the heightened ERF associated with processing an informative versus an uninformative cue.

5.4.2 Whole-brain maps

In order to check whether there were activations in regions additional to the ROIs derived from the meta-analysis, we performed a whole brain analysis (5.3.2) for the theta band (3-7Hz), the alpha band (8-12Hz) and the gamma band (60-80Hz). The 4D spatiotemporal map was averaged over successive 300ms windows, and cluster-based permutation tests were performed on the resulting 3D activation maps. Cluster permutation statistics were computed for the informative cue versus neutral cue contrasts with a cluster-forming threshold of 3 (t-statistic). Whole brain activation maps for the full timeseries are shown in Appendix 9.4.

There was a focal activation in the theta band additional to those captured in the ROI analysis, in the left frontal pole. This activation was present in the averaged volume between 900 and 1200ms following a precue in medial frontal pole, and in the volumes between 1200ms and 1500ms, and 1500ms and 1800ms following a retrocue, in lateral frontal pole. These volumes are shown projected onto the CARET brain in Figure 5-9. For precues, in the same 900 to 1200ms time window, there was also an activation in the right temporal lobe[23].

The ERF is associated with widespread deep medial activations that were strongest in the theta-band (volumetric plots, appendix). These deep activations are particularly strong in implausible ventral regions such as the brainstem, and probably reflect a limitation of the source reconstruction when dealing with highly correlated cortical sources. Since the beamformer estimates the weights independently for each voxel, this spurious deep activity should not affect cortical activity estimates. This is discussed further in the chapter 2 (section 2.3.2).

---

[23] This projects poorly onto the CARET surface, but is clear in the volumetric plots in the appendix.

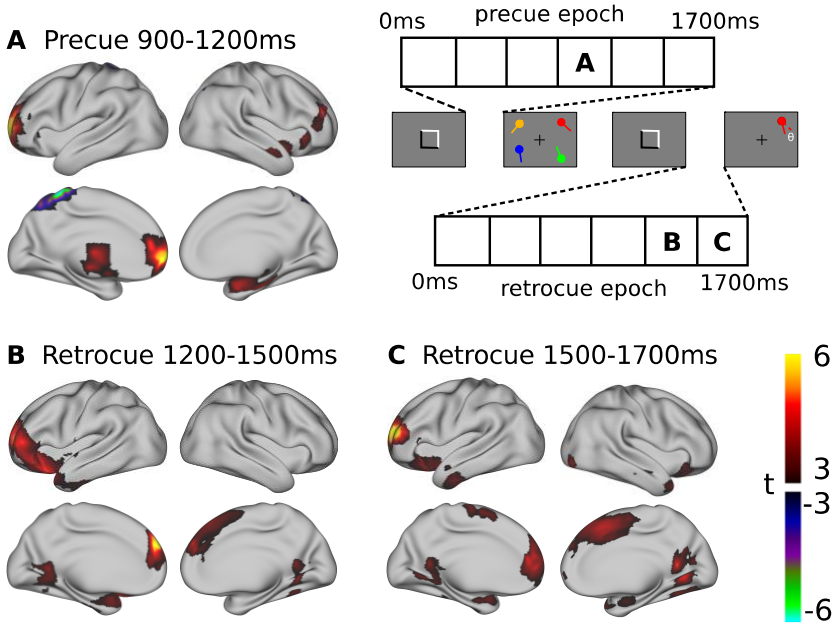


Figure 5-9 **Whole-brain activations showing frontal pole activity.** Only supra-threshold activations (threshold t = 3) belonging to significant clusters are shown. Following both precues and retrocues, but at slightly different latencies, there is an activation in the left medial frontal pole (not captured in the meta-analysis derived ROIs). The peak frontal pole activation is slightly more ventral for precues. For retrocues, activations in the fO and pre-SMA (captured in the ROI analysis) are also present.

5.4.3 Correlation with behaviour and alpha lateralization

For each significant fronto-parietal or cingulo-opercular cluster (Figure 5-5), the time-frequency bins that belonged to the significant group level cluster were used as a mask to extract a trial-by-trial value for that cluster, for each subject. This trial-wise measure was then correlated with reaction time within condition, for each ROI/cluster. The behavioural data were also median split within-condition on the basis of the power for each ROI/cluster, and accuracy was compared between the two halves of the data. A trial-wise alpha lateralization measure extracted from a pattern classification analysis on sensor space data was also median split to investigate the link between trial-wise activation in control regions and sensory reactivation.

No significant trial-wise relationships were found between activation in control regions and either behavioural measures or alpha lateralization.

5.4.4 Functional connectivity

The task analyses described above assume that beamforming activity from the ROIs selected from the fMRI meta-analysis successfully reconstructed the signal from the frontal areas of interest. To validate the sensitivity of the ROIs used in the task analysis to the frontal regions of interest, and investigate the connectivity of those regions in different spectral bands, a functional connectivity measure, orthogonalised power correlation (Hipp et al., 2012), was derived from both resting-state and task MEG data using the same ROIs as were used for the task analysis. This measure was used to characterise the patterns of functional connectivity between MEG ROIs, replicating network structure discovered using fc-fMRI. Whole-brain seed-based correlation maps were also computed for left and right hemisphere ROIs separately. There were collapsed across side: the pattern for right hemisphere ROIs was mirrored in the sagittal plane and averaged with the pattern for left hemisphere ROIs. This results in a set of maps in which ipsilateral correlations are shown in the left hemisphere, and contra-lateral correlations in the right hemisphere. As the different parietal nodes showed similar patterns of connectivity, only mid IPS is plotted here.

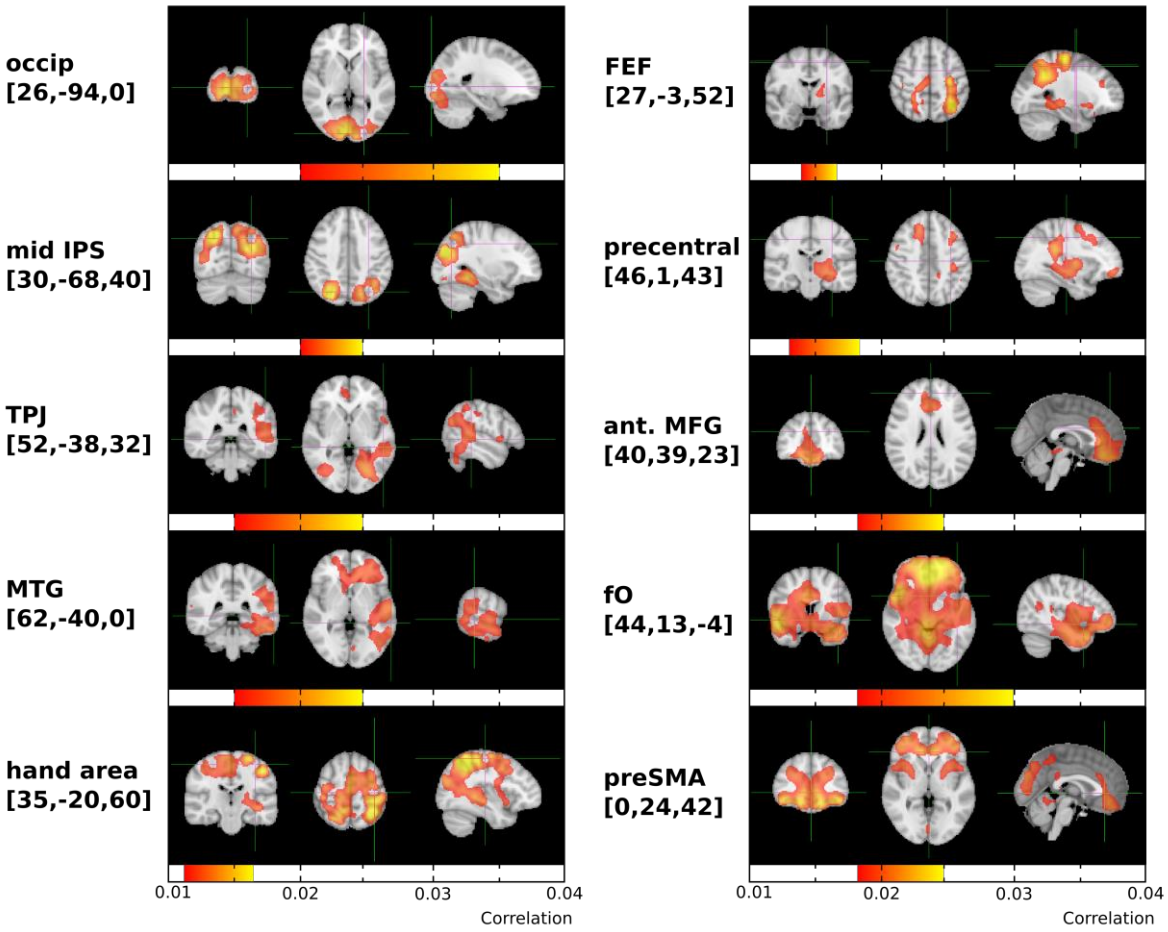


Figure 5-10 **Theta-band (3-5Hz) orthogonalised power correlation.** Power correlation maps for right/left hemisphere ROIs were averaged across hemisphere, yielding maps in which the RHS as plotted (left hemisphere in radiological convention) represents the ipsilateral side to the seed ROI. Maps were thresholded to display the strongest correlations only. Correlations in the theta band were strongest for occipital cortex and anterior frontal regions. The pre-SMA is correlated with the anterior/fO as well as frontal pole, replicating the cingulo-opercular network as seen in fc-fMRI. There is also a band of correlation extending towards anterior MFG (dlPFC). The fO is correlated with contralateral fO as well as medial frontal pole and dorsomedial PFC. MTG is correlated with lateral frontal pole. Other ROIs (e.g. hand area, mid IPS) are correlated with their contralateral counterpart.

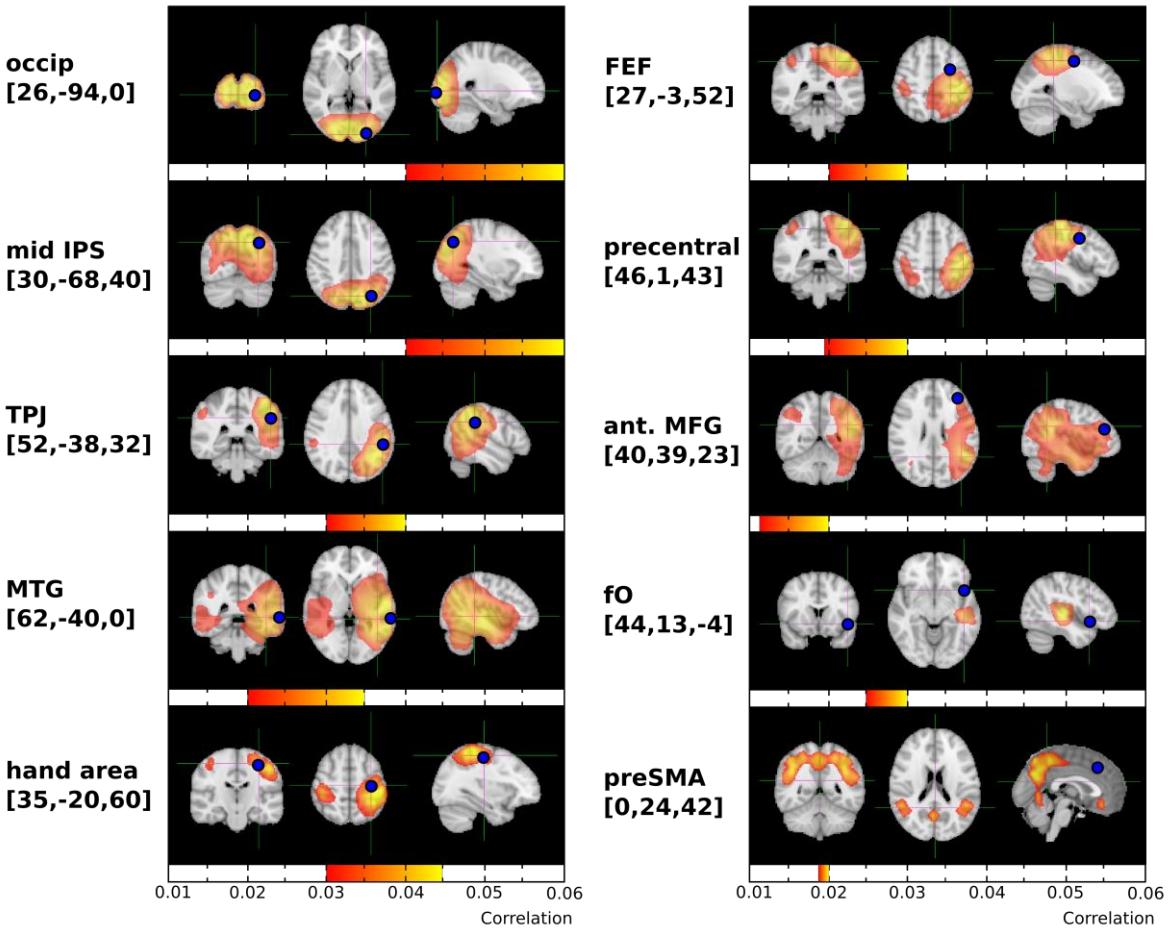


Figure 5-11 **Alpha-band (8-12Hz) orthogonalised power correlation.** Power correlation maps for right/left hemisphere ROIs were averaged across hemisphere, yielding maps in which the RHS as plotted (left hemisphere radiological convention) represents the ipsilateral side to the seed ROI. Maps were thresholded to display the strongest correlations only, and the cuts were chosen to best show the pattern of connectivity. Correlations in the alpha band were strongest in the occipito-parietal ROIs, and thresholded maps for occipital cortex, mid-IPS, and TPJ reveal only local and contralateral connectivity. The correlation pattern for more frontal ROIs reveals long-range connections. Precentral gyrus and FEF are correlated with bilateral parietal sites, as is the anterior MFG (albeit weakly).

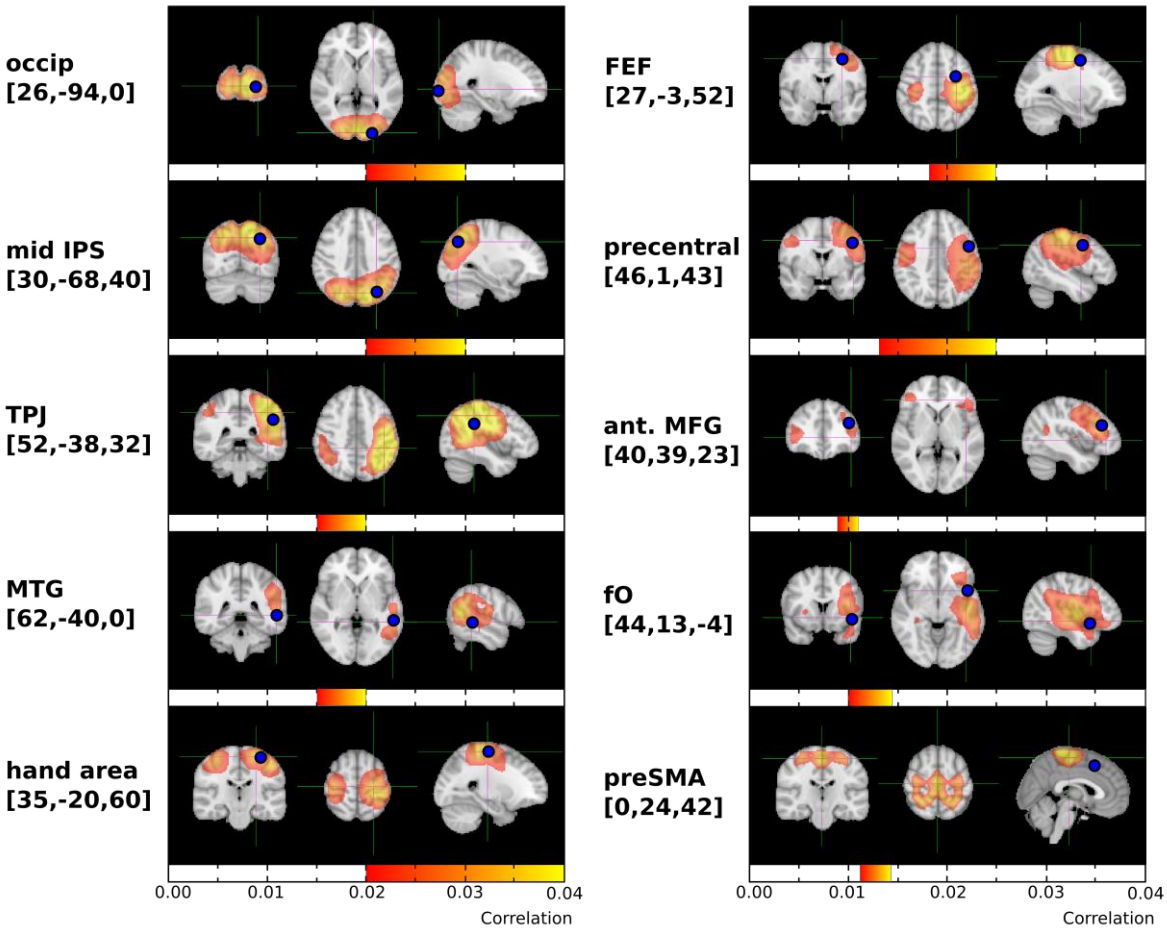


Figure 5-12 **Beta-band (18-30Hz) orthogonalised power correlation.** Power correlation maps for right/left hemisphere ROIs were averaged across hemisphere, yielding maps in which the RHS as plotted (left hemisphere radiological convention) represents the ipsilateral side to the seed ROI. Maps were thresholded to display the strongest correlations only, and the cuts were chosen to best show the pattern of connectivity. Correlations in the beta band were predominantly local or with the same ROI on the contralateral side. Motor cortex was strongly correlated across hemispheres. pre-SMA was correlated with motor cortex.

The theta-band maps are plotted above (Figure 5-10). The pre-SMA and fO are correlated in the theta band, and both are also correlated with frontal pole areas. This is consistent with the cingulo-opercular network pattern derived from fc-fMRI (Dosenbach et al., 2007) and also the pattern of task activations, in which fronto-polar cortex was co-activated with fO and pre-SMA. Interestingly, MTG was also correlated with lateral frontal pole, again consistent with the pattern observed in the task data. Otherwise, several regions were correlated with their contralateral counterpart. ROIs

also tended to correlate most strongly with directly adjacent parts of cortex. This bias for local correlations is consistent with the results of Hipp et al. (2012) who found that seed regions were most correlated with nearby regions in the same hemisphere as the seed, although additional long distance correlations were also present. This could reflect residual zero-lag signal spread (Hipp et al., 2012; Schoffelen & Gross, 2009), or it might reflect a true tendency for nearby brain regions to interact more strongly with one another: connectivity is highest over short distances (peaking at ~3cm) in the human brain, as measured using BOLD correlations or DTI (Kaiser, 2011; Salvador et al., 2005).

The alpha band (Figure 5-11) was heavily correlated within occipito-parietal cortex. Frontal seed ROIs revealed inter-regional connections. FEF, precentral gyrus, and dlPFC were all correlated with bilateral parietal cortex. Interestingly the pre-SMA was also correlated with bilateral parietal cortex and precuneus. The beta-band analysis was less revealing of long-range connectivity. As expected, bilateral motor cortex was strongly correlated in the beta band. The remaining ROIs largely correlated locally, or with their contralateral counterpart. Again pre-SMA was an exception, correlating with motor cortex.

The spatial maps plotted above reveal long-range correlations that replicated the network structure discovered using fc-fMRI. The pre-SMA, fO and frontal pole (cingulo-opercular network areas) were all mutually correlated in the theta band. In the alpha band, whilst the correlations with IPS ROI were dominated by local correlations, the remaining fronto-parietal network regions (FEF, precentral cortex, ant. MFG) all correlated with bilateral IPS.

To check that the task data contained similar correlation patterns to the resting-state data, a similar analysis was performed on task data from the retrocue epoch. The resting-state correlation maps were condensed into a set of pair-wise correlations

between ROI. Power was correlated across trials in the task data, to generate an analogous set of correlation matrices. The correlation matrix between ROI pairs was converted to a distance measure, and hierarchical clustering analysis was performed on the distance measure (Dosenbach et al., 2007) to identify nodes that clustered together. These data are shown below for both resting state data (50 subjects, correlating power over time) and task data (38 subjects, correlating power over trials), for the theta-band analysis.

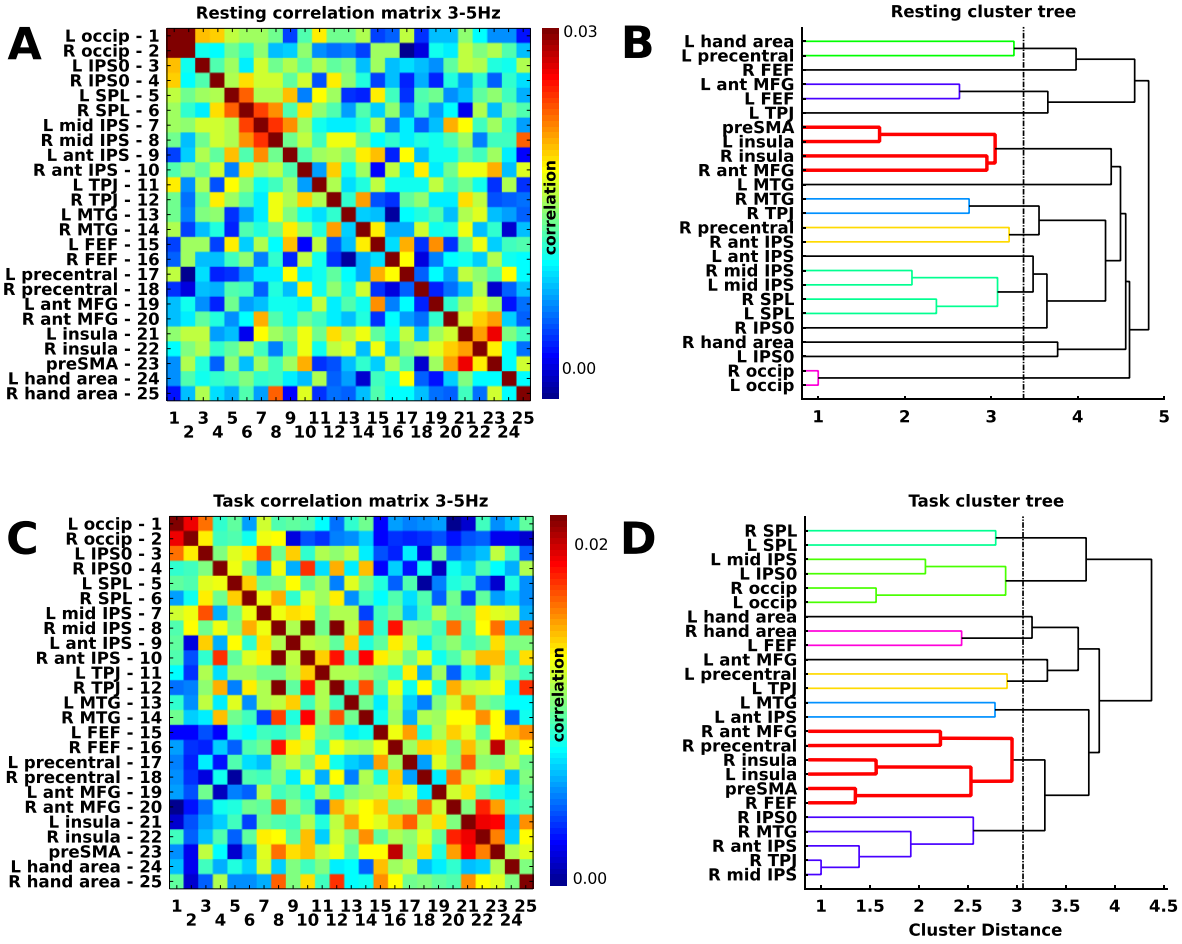


Figure 5-13 **Pair-wise correlations, 3-5Hz, resting state data (A,B) and task data (C,D).** The correlation matrix (LHS) was converted into a distance measure which was used to construct a hierarchical cluster tree (RHS). The tree was arbitrarily cut at 70% of the maximum distance between nodes in order to highlight cluster groupings. In both analyses, pre-SMA and anterior insula/fO are grouped together. This cluster has been coloured red and emboldened in both figures. In the resting state analysis, right ant. MFG is also included in this cluster. In the task analysis, right ant. MFG, right iFEF (precentral) and right FEF are clustered with the cingulo-opercular regions. Other clusters in the theta band largely reflect local or contralateral correlations (e.g. the strong correlation between left and right occipital cortex).

The pairwise correlation structures computed from resting state data resembled the correlation structures computed from the task data (Pearson's **r** between correlation matrices for resting and task analyses: theta band, $r=0.34$, $p<0.0001$; alpha band, $r=0.93$, $p<0.0001$; beta band $r=0.85$, $p<0.0001$). The correlation was weakest for the theta band analysis plotted above. This may in part be because the theta band is sensitive to ERF responses in the task data, which introduce additional correlations. However, inspecting the correlation plots and cluster trees, the cingulo-opercular network is revealed in both analyses. Bilateral fO and pre-SMA are clustered together, although in each case they are joined by a second frontal region: for the resting-state analysis, the right anterior MFG, and for the task analysis, the right FEF/iFEF in addition.

5.5 Discussion

5.5.1 Network activation time-courses

The spatial pattern of induced responses in the MEG data was consistent with the pattern observed in the meta-analysis of fMRI studies. Both precues and retrocues modulated activity in the fronto-parietal network ROIs, including the anterior MFG (dlPFC), precentral regions (FEF for precues, iFEF/IFJ for retrocues), and the IPS. Only retrocues recruited the cingulo-opercular nodes, with the exception of a short-lived activation in the pre-SMA following precues.

The two networks also had dissociable activation time-courses. For both precues and retrocues, the fronto-parietal activations preceded alpha lateralization in occipital cortex, and the desynchronization in the left IPS closely matched the time-course of the alpha effects. Occipital alpha lateralization (chapter 4) likely indexes the allocation of spatial attention in the case of precues, and the reactivation of perceptual content associated with the cued item in the case of retrocues. Conversely, following retrocues, the cingulo-opercular ROIs co-activated only when alpha lateralization was returning to baseline (although the pre-SMA was activated immediately following the cue). Both the fronto-parietal network and the fO have been suggested to mediate top-down modulation of sensory cortex following a retrocue (Higo et al., 2011; Nobre et al., 2004). The MEG time-courses presented here are consistent with a direct role for the fronto-parietal network in mediating top-down control over perceptual cortex. They are not consistent with a direct role for the fO in top-down biasing or top-down reactivation, but they do suggest that the fO is recruited at a later stage following a retrocue, perhaps mediating output gating.

5.5.2 Cingulo-opercular network

Within the cingulo-opercular nodes, there was a further dissociation: pre-SMA activated both immediately following cues (albeit transiently following precues), and later reactivated at the same time as bilateral activation of fO. Pre-SMA has been proposed to drive changes in top-down control by detecting cognitive conflict (Yeung, 2013) putting it upstream of the fronto-parietal network (Botvinick et al., 2001), but others have proposed that the pre-SMA is responsible for high-level action selection (Rushworth et al., 2007) putting it downstream (Banich, 2009). The MEG data are consistent with both hypotheses: pre-SMA activated immediately following both precues and retrocues, but following retrocues it activated again later in the trial along with bilateral fO. The putative roles for pre-SMA need not be mutually exclusive. Speculatively, the pre-SMA might be broadly involved in scheduling cognitive operations, including both top-down control of internal representations (in this case, retrieving a particular working-memory representation) and prioritizing representations to guide action[24].

The anterior insula/fO was only recruited late following retrocues, after alpha lateralization in perceptual cortex had subsided, and yet two fMRI studies (Higo et al., 2011; Nelissen et al., 2013) have implicated this region in top-down control of internal representations following a retrocue. Elucidating the functional role of regions in the vicinity of the anterior insula is complicated by the subdivision of this area into several nearby regions (Cauda et al., 2011; C. Kelly et al., 2012; Neubert et al., 2014). The anterior insula is strongly functionally connected with the pre-SMA (Cauda et al., 2011;

[24] However, these theta activations immediately post-cue need to be treated with caution, as the beamformer may generate spurious activity on the midline when there are highly correlated bilateral sources (as might be associated with the ERF; see chapter 2). This caveat applies mainly to the pre-SMA activation immediately following the precue, which was short-lived. The more sustained activation in the theta-band following the retrocue is less likely to be attributable to a mis-localized ERF.

Dosenbach et al., 2007) whereas the more anterior frontal operculum (which curves out from the insula onto the lateral surface) is not, and is instead functionally connected with area 9 (dorso-medial PFC more anterior than the pre-SMA) and temporal lobe regions (Neubert et al., 2014). The dividing line between these two regions is somewhat unclear (Higo et al., 2011). The ROI derived from the fMRI meta-analysis is in the anterior insula, and is more posterior than the peak fO activation described by Higo and colleagues. It is possible that the activation reported by Higo and colleagues is in a separate and more anterior anatomical compartment to the anterior insula activation in the meta-analysis. However, the most parsimonious account of these data is that a single anterior insula/fO region is activated by retrocues (the same as described by Dosenbach and colleagues).

Higo et al. (2011) and Nelissen et al. (2013) suggest that fO is the critical region mediating selection from within working memory. However, in these studies it was not possible to separate retrocue-triggered memory retrieval from selection processes associated with parsing the probe items and selecting an appropriate response. The probe array in these experiments followed very shortly after the retrocue (1.5-2.5s) compared to the HRF time-course, and three items were presented (in contrast to the single probe item in the current task), introducing additional selection pressure following probe presentation. In retrocue ('selective') trials, subjects were asked to respond only about the memory item to which they had been cued, and to ignore the remaining two items on screen. One of these irrelevant probe items was of the same stimulus category as the uncued memory item, introducing a potential response conflict (which might have required active suppression). Given the complex overlapping demands of retrieval, selective processing of a single probe item, and possibly suppression of an inappropriate response mapping, it is difficult to attribute the fO activation associated with the retrocue/probe array in this task to one or other cognitive

operation. However, if the need to resolve conflict between competing item-response mappings was indeed a major source of cognitive load in this task, then this is consistent with a role for the fO in prioritizing a single item for control of action, and perhaps in suppressing competing item-action associations.

5.5.3 Fronto-parietal network

A prevailing hypothesis is that neural mechanisms of top-down input to sensory cortex are shared between top-down biasing mediating perceptual attention (Desimone & Duncan, 1995) and top-down activation of perceptual cortex for working memory operations (Lepsien et al., 2005) or mental imagery (Stokes et al., 2009a). The neural dissociations between these functions are less well characterized. Within the fronto-parietal network, the MEG data replicated a dissociation seen in the meta-analysis: precues activate FEF whereas retrocues activate a more ventral region of the precentral gyrus, referred to variously as precentral cortex (Szczepanski et al., 2010), IFJ (Lepsien et al., 2005), or iFEF (Derrfuss et al., 2012)[25]. It could have been argued that the dissociation between FEF and iFEF observed in the fMRI meta-analysis was because the retrocue meta-analysis incorporated a subset of studies invoking non-spatial selection, whereas the precue experiments exclusively cued spatial attention. However, the fact that this dissociation is replicated in the MEG data, in a task in which precues and retrocues were associated with the same basic stimulus features and configuration (differing only in time of cue presentation relative to the memory array), suggests that the critical factor distinguishing activation in these regions is the distinction between perceptual attention and internal selection.

[25] Derrfuss and colleagues dissociated the IFJ from the iFEF using fMRI, but these regions are very close to one another anatomically, and the activation observed in the current data might have arisen in either (or both).

The (left) mid-IPS may be a critical parietal control region for top-down modulation of sensory cortex. The time-course of alpha desynchronization in mid IPS closely matched the time-course of alpha lateralization in more posterior regions of visual cortex (Figure 5-7). High-gamma activation (>60Hz) in the mid-IPS also marked moments at which a single item could be selected (but not the allocation of preparatory attention in the absence of encoded representations). A burst of high gamma-band activity occurred in IPS following retrocues, but also following the presentation of the memory array in trials when a precue had been given (i.e. subjects were able to selectively encode one of the items in the array). Importantly, following the memory array, there was no difference in gamma in occipital cortex between precue and neutral cue trials. Conversely, whilst precues (like retrocues) induced a larger gamma response in visual cortex than uninformative cues, they did not significantly boost gamma in mid-IPS. These dissociations rule out a trivial explanation for the increase in IPS gamma in terms of the enhanced ERF to informative as compared to neutral cues. IPS gamma was also dissociated from alpha desynchronization: following a precue, there was a sustained desynchronization in alpha power in mid IPS but no boost in gamma power[26]. The gamma activations observed were between 70 and 90Hz, in the high gamma range. High gamma may reflect a true oscillation (Fries, Reynolds, Rorie, & Desimone, 2001) in which case it would be expected to be band-limited, but the superposition of post-synaptic potentials in activated cortex could also give rise to a broad-band high gamma signal (Uhlhaas, Pipa, Neuenschwander, Wibral, & Singer, 2011). The gamma activations in visual cortex following a cue, which are broad-band (extending in frequency from 60Hz to the top of the analysis window at 120Hz), are more likely to reflect such 'neural hash' than a true oscillation. Gamma activity with similar spectral profile (extending up to ~200Hz) is observed in intracranial recordings

[26] Similarly, alpha power in visual cortex lateralizes following a precue or retrocue, but gamma power does not.

from the brain surface in visual and parietal cortex (Lachaux et al., 2005) following visual stimulation. The gamma responses seen in the IPS appear band-limited between 70 and 90Hz, but this may be because the low signal-to-noise ratio only permits the peak of the gamma response to reach significance. On the assumption that the gamma response in IPS reflects increased neural firing, it is interesting that it is dissociated from desynchronization in the alpha band. It is possible that changes in alpha power in deep layers of occipito-parietal cortex (Buffalo, Fries, Landman, Buschman, & Desimone, 2011) mediate top-down changes in excitability (Saalmann et al., 2012), but this excitability change is sub-threshold in that it does not modulate firing rate in superficial layers (as indexed by gamma power) unless a representation is available for active processing (see discussion, chapter 4).

Gamma-band responses in mid-IPS were stronger in the left hemisphere ROI than the right for both precues and retrocues, as was alpha desynchronization. This left-lateralization is at first sight at odds with the right-hemisphere dominance usually assumed for spatial attention (Corbetta et al., 1993; Mesulam, 1981; Riddoch et al., 2010). However, whilst exogenous attentional capture appears right-lateralized in the parietal cortex, endogenous generation of attentional biases activates parietal cortex more symmetrically (Nobre et al., 1997), and may even be left-lateralized (Gillebert et al., 2012; Kim, 1999).

5.5.4 MTG/TPJ and frontal pole

The analysis and discussion has focussed on core fronto-parietal and cingulo-opercular network regions as identified from resting-state functional connectivity in fMRI (Dosenbach et al., 2008). However, additional regions were also implicated in the fMRI

meta-analysis: the left middle temporal gyrus and the left temporo-parietal junction. These regions do not group with either control network (Dosenbach et al., 2007).

The role of these ventral regions is unclear. Corbetta and colleagues have shown that the TPJ is activated when an unexpected stimulus triggers attentional re-orienting, and have suggested that this region is involved in the disengagement of the attention to permit re-orienting (Corbetta et al., 2000; Corbetta & Shulman, 2002). Consistent with this idea, lesions of the TPJ (and also MTG) are more associated with extinction than neglect (Riddoch et al., 2010). However, the right and not left TPJ has been associated with attentional functions (Corbetta & Shulman, 2002; Riddoch et al., 2010), whilst the TPJ (and MTG) activation associated with retrocueing was in the left hemisphere. Interestingly, the TPJ demonstrated a pattern of activation very similar to that seen in more dorsal IPS following both precues and retrocues[27], whereas the left MTG activated only following retrocues and had a time-course resembling that of the pre-SMA/fO, activating late in the epoch.

A whole brain analysis of induced responses also identified a theta-band activation in the left medial frontal pole, late in the epoch following both precues and retrocues (Figure 5-9). For retrocues, in the whole-brain theta-band analysis with a cluster-forming threshold of t=3, this activation was part of a large significant cluster encompassing the pre-SMA, left frontal pole and left fO. To obtain a peak MNI coordinate for the frontal pole activation, the data were averaged between 1.2 and 1.8s post-cue (collapsing the two time-windows in which the frontal pole cluster appeared in the whole-brain analysis), and the threshold was raised to t=5, which isolated the frontal pole activation as a separate cluster with peak MNI coordinate [-30, 42, 8]. For precues, the frontal pole cluster was significant between 0.9 and 1.2s post-cue, but there

[27] Given the limited spatial resolution of MEG and the strong IPS response, it is possible that reflects signal bleed from more dorsal IPS.

were no pre-SMA or fO activations. The peak MNI coordinate for the frontal pole cluster was more medial and anterior than for retrocues, at [-6, 66, 0]. Bearing in mind the spatial uncertainty in the MEG source localization, the precue and retrocue activations corresponded roughly to medial frontal pole (FPm, BA10) and the recently delineated and uniquely human lateral frontal pole (Neubert et al., 2014). Precues additionally activated the right temporal cortex/MTL in the same 900-1200ms time window as the medial frontal pole was activated, with a local maximum in the MTL at MNI [26,2,-24].

The functional role of the frontal pole regions is not well characterized. Burgess and colleagues have argued that rostral PFC may be involved in switching cognitive processing between the perceptual domain and internal representations (P. W. Burgess, Dumontheil, & Gilbert, 2007). Henseler et al. (Henseler, Krüger, Dechent, & Gruber, 2011) used a task in which attention was either directed externally to upcoming stimulus events, or internally to encoded representations (c.f. the precue/retrocue paradigm used here) and found that medial frontal pole was activated during external attention, whereas lateral frontal pole was activated during attention to internal representations. This would be consistent with the pattern observed in the MEG, in which precues activated medial frontal pole and retrocues lateral frontal pole. Henseler and colleagues found that attention to internal representations additionally recruited the bilateral fO, dlPFC and pre-SMA, consistent with the fMRI meta-analysis and MEG data presented here. Attending to perceptual content additionally recruited the right MTG and also amygdala, consistent with the right temporal lobe / MTL activation observed following precues in the MEG.

These functional associations are at least consistent with the pattern of activity seen in MEG, suggesting that whilst lateral frontal pole may co-activate with fO/pre-SMA late

following retrocues (secondary to fronto-parietal activation), precues may also recruit a 'secondary' network including medial frontal pole and MTG. However, the frontal pole (and precue MTG) activations were not replications of meta-analysis activations (unlike the fronto-parietal and cingulo-opercular activations discussed above). If frontal pole is consistently recruited during attentional cueing tasks, then it is surprising that these activations were not consistently detected in the fMRI studies reviewed above. Nevertheless, anterior PFC is functionally connected with the pre-SMA and fO, forming part of the cingulo-opercular network identified by Dosenbach et al (2007), and the whole brain map of fO and preSMA orthogonalised power correlations in the theta band replicated this finding – both were correlated with frontal pole (Figure 5-10). The role of these frontal pole areas in orienting to perceptual and internal representations remains to be confirmed. On the basis of their late activation in MEG, they appear to play a downstream as opposed to direct biasing role, as argued above for the cingulo-opercular network. A high-level role (for example, a monitoring or evaluative role) would be consistent with their position at the top of the frontal lobe hierarchy (Fuster, 2003).

5.5.5 Conclusions

In this chapter, cue-induced MEG responses in source-space were shown to replicate spatial patterns of activation observed in similar tasks in fMRI. The time-course of these induced responses was used to inform hypotheses about the role of the fronto-parietal and cingulo-opercular networks in cognitive control. In a previous chapter, alpha lateralization in visual cortex was shown to index preparatory attention and internal selection. The fronto-parietal network nodes activated prior to this alpha lateralization, consistent with a direct role for this network in top-down biasing. Conversely, the fO and pre-SMA co-activated only later in the epoch following a retrocue,

consistent with a downstream role in prioritizing cognitive representations for control of action. The pre-SMA also activated immediately following both precues and retrocues, suggesting that it may play a more general role in coordinating cognitive events than the fO. Finally, additional frontal pole regions were activated following both precues and retrocues, and like the cingulo-opercular activations, the timing of these activations suggested a downstream role in a control cascade (Banich, 2009).

The functional roles of the fronto-parietal network in top-down biasing of sensory cortex are well established in relation to attentional control (Mesulam, 1981; Petersen & Posner, 2012; Posner & Petersen, 1990) and this function generalizes easily to top-down control of sensory cortex in the service of cognitive operations (Stokes, 2011). The functions of the cingulo-opercular areas are less well established (Petersen & Posner, 2012), partly because the underlying anatomy is less well delineated, and partly because they mediate functions lying midway between stimulus and response, that are hard to operationalize. The functional attributions suggested above for the cingulo-opercular regions are made on the basis of the 'best fit' between the MEG time-courses and prior hypotheses arising from fMRI experiments. The claim that the cingulo-opercular network is involved in prioritizing one of a number of representations for use in cognition or to guide action (Olivers et al., 2011), perhaps forming the core of the 'global workspace' proposed by Baars (Baars, 2005; Dehaene & Changeux, 2011), is speculative.

In order to map the flexible network interactions that give rise to cognitive control, future work will need to emphasise techniques such as MEG or intra-cranial recording that combine fine spatial and temporal resolution. Cognition is highly dynamic, consisting in internal events that follow one another with a time-course much shorter than the temporal resolving power of fMRI. Discovering the role of different frontal structures in these short-lived cognitive events or 'attentional episodes' (Duncan, 2013)

requires methods that can discriminate activations on a fine-grained timescale. This will be crucial in differentiating the functional contributions of areas that are currently described as domain general or task general (Dosenbach et al., 2006; Fedorenko, Duncan, & Kanwisher, 2013) on the basis of fMRI activations. Breaking down the cognitive events that underlie task execution into their 'atomic' constituents (executed in rapid succession by the brain) and mapping these to transient activations in the frontal cortex will reveal the functional diversity in control areas that activate ubiquitously across many tasks. Like grammatical words in language, which are present regardless of semantic content but nevertheless have specific syntactic roles that depend on their sequential placement in a sentence, frontal control regions may be recruited dynamically across a wide range of task contexts to perform specific cognitive operations on diverse representations.

6 Reinforcement learning and selection for working memory

Chapter abstract

Working memory is capacity limited, and selection mechanisms act to limit encoding only relevant information from the environment. A large body of research has used cues to study strategic control of working memory contents. However, the effect of reward associations on working-memory gating has been little investigated behaviourally, despite the existence of a number of neural models implicating the dopamine system and basal ganglia in control of working memory. Understanding the relationship between reinforcement learning and cognitive control may be important in understanding the pathological mechanisms underlying disorders like schizophrenia, in which dopaminergic dysfunction is associated with impairments in cognitive control. In this chapter, a series of experiments are presented that investigate the potential role of reward associations in determining which items gain access to working memory, in the absence of other top-down biases for encoding. Two hypotheses are considered: one suggests that phasic dopamine release opens the gate to working memory and therefore mediates a time-specific but not item-specific gating function, the other that fronto-striatal circuitry learns to preferentially gate reward-associated items (with dopamine as a teaching signal). The results favour the former hypothesis: reward associations boost encoding, but this boost does not discriminate between items.

6.1 Introduction

Most experiments studying top-down control of working memory use explicit cues to manipulate which items are task relevant (Griffin & Nobre, 2003; McNab & Klingberg, 2007; Nee & Jonides, 2009; Vogel & Machizawa, 2004; Vogel, McCollough, & Machizawa, 2005). People are able to use these cues to control what information is represented in working memory and selected to guide action. Whilst there are cases in everyday life when we make use of imperative cues to guide top-down control (for example, we often use pointing gestures to guide attention when communicating), we probably as often select the contents of working memory based on the inherent goal-relevance of stimuli we encounter. Besides strategic selection biases associated with particular task sets, items can be inherently goal-relevant either by virtue of having primary reward association or by virtue of their association with primary reinforcers (James, 1890). For example, sugary/fatty/salty food, the consumption of which has positive hedonic consequences, is a primary reinforcer. The takeaway menu that constitutes the first step in obtaining the sugary/fatty/salty food is goal relevant by association only: eating the menu itself is unlikely to have positive hedonic consequences.

Working memory is a cognitive resource that allows us to retain goal-relevant representations and chain sub-goals over time gaps to eventually obtain reward. As such, we might expect reward associations to guide the selection of working-memory representations. Known interactions between the dopamine system, basal ganglia and prefrontal cortex have inspired computational models of the role of reward learning in control of working memory (Braver & Cohen, 1999; Hazy et al., 2006). The effects of reinforcement history on selection have also been studied behaviourally in the realm of visual attention, using visual search paradigms in which reaction time is typically the

dependent variable (B. A. Anderson, Laurent, & Yantis, 2011). In this chapter a set of behavioural experiments are presented that were designed to look directly at the effect of reinforcement history on selective encoding for working memory. Specifically, the aim was to establish whether, in the absence of a top-down strategy determining which items should be encoded, items with a higher reward association would have a competitive advantage for encoding into WM.

6.1.1 The role of dopamine and basal ganglia in selection for working memory

Prefrontal cortex is the region most commonly associated with working memory and cognitive control. Jacobsen and Nissen (1937) found that ablating frontal cortex in monkeys impaired working-memory performance, operationalizing working memory using a delayed response task to build on earlier more impressionistic studies by Ferrier and Bianchi. This association began a long line of work on the link between frontal cortex and working memory, now richly investigated in humans using brain imaging (Curtis & D'Esposito, 2003; Passingham & Katsuyuki, 2004). The close anatomical coupling between prefrontal cortex and the dopamine system and basal ganglia has motivated accounts of how these systems might interact in the service of cognitive control.

The importance of dopaminergic innervation of prefrontal cortex was first established by Brozoski, who showed that pharmacologically depleting dopamine in monkey prefrontal cortex drastically impaired performance in a delayed alternation task taken to measure working memory in monkeys (Brozoski et al., 1979). Goldman-Rakic emphasised both the link between frontal cortex and working memory and the role of dopamine in prefrontal function (Goldman-Rakic, 1995; Sawaguchi & Goldman-Rakic,

1991; Williams & Goldman-Rakic, 1993), characterizing dopamine projections to the prefrontal cortex in tracer studies.

Dopaminergic projections to frontal cortex predominantly arise in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNpc) (Williams, 1998). These nuclei also send dopaminergic projections to the striatum in the basal ganglia. The striatum engages in reciprocal interactions with frontal cortex (Alexander et al., 1986; Middleton & Strick, 2000), including lateral prefrontal cortex (BA46), a region associated with working memory (D'Esposito, Postle, & Rypma, 2000).

Dopamine is strongly associated with reward processing and reward learning. Phasic dopamine signals convey reward prediction errors (Schultz, 1998; 2013). During learning, appetitive stimuli initially induce bursts of firing in dopamine neurons, but over time the burst firing transfers to stimuli (secondary reinforcers) that are predictive of the primary reinforcers.

There have been several attempts to explain how frontal cortex, dopamine and the basal ganglia might coordinate in the service of cognitive control, and working memory. Braver and Cohen focussed on dopamine and prefrontal cortex (Braver & Cohen, 1999). Their theory was partly motivated by a clinical link between the dopamine system and cognitive control. Schizophrenia, a psychiatric disorder linked to the dopaminergic dysfunction (Seeman, Lee, Chau-Wong, & Wong, 1976), is characterized by a failure of cognitive control, including working memory (Barch & Smith, 2008).

Braver and Cohen proposed that dopamine projections to frontal cortex might gate working-memory updating (Braver & Cohen, 1999; 2000). There is a trade-off between flexible gating of memory and robust maintenance: working memory must be capable of flexibly adapting its contents according to changing task demands, but it must

equally be capable of robustly retaining information despite distraction. One way to resolve these conflicting demands is to postulate separate mechanisms for maintenance and gating: dedicated neural circuitry maintains representations over time, and separate circuitry actively gates information into memory (or prevents it from entering memory) depending on its behavioural relevance. Braver and Cohen proposed that dopamine projections from the VTA to the PFC carry the gating signal, and since dopamine signals reflect the association of a given stimulus with reward (Schultz, 2013), there is a bias for stimuli that are reward-associated, and therefore goal-relevant, to be encoded into WM (Braver & Cohen, 2000). The hypothesis that dopaminergic mechanisms play an important role in gating working memory has since garnered empirical support. D'Ardenne and colleagues (D'Ardenne et al., 2012) used fMRI to image activity in both dlPFC and SNpc/VTA, contrasting a 'context independent' task condition in which subjects responded simply on the basis of a probe stimulus to a 'context dependent' condition in which subjects had to encode a rule into memory, cued by a preceding stimulus, in order to know how to respond to the probe stimulus. This latter condition gave rise to a phasic increase in activity in both the dlPFC and the VTA/SNpc following the context cue. There was less activity in SNpc/VTA on context-dependent error trials than correct trials, suggesting that activity in the dopaminergic nuclei is important for successful updating of working memory.

Dopamine neurons have diffuse terminal projections within PFC. Besides their lack of target specificity, dopaminergic neurons also seem to respond rather homogeneously: a large percentage of recorded dopamine neurons will respond to reward-associated stimuli, or pause if an expected reward is omitted (Schultz, 2013). This phasic activation is temporally stereotyped. A homogeneous phasic signal would seem well suited for temporal selection, briefly opening the gate to working memory. However, it

is not clear that phasic dopamine could selectively gate a subset of multiple simultaneously presented items.

A related model that does permit item-wise selectivity incorporates the striatum. Hazy, Frank, O'Reilly and Badre (Hazy et al., 2007) have proposed a model in which dopamine is thought of as a teaching signal for striatum, more than a gating signal *per se.* The striatum is suggested to control selective updating of working memory. Previous models have proposed a role for basal ganglia in action selection (Redgrave et al., 1999). In Hazy et al.'s model, action selection is generalized to the cognitive domain: parallel cortico-striatal loops are responsible for selective updating of representations in working memory, and striatal circuitry adapts to gate only goal-relevant representations into memory, using dopaminergic prediction error firing as the teaching signal (Frank et al., 2001; O'Reilly & Frank, 2006). The computational model of O'Reilly et al. further distinguishes two forms of gating: input gating, and output gating. The former refers to control over what is represented in memory, the latter to control over which of those representations are permitted to guide action. Similar cortico-striatal interactions are proposed to give rise to both functions (Badre & Frank, 2012; Hazy et al., 2007). This distinction can be compared to the distinction between perceptual and internal selection (Olivers et al., 2011) discussed elsewhere in the context of retrocueing (chapter 3).

fMRI has provided evidence for striatal involvement in control of working memory. McNab and Klingberg (2007) gave human subjects a task in which they were cued on each trial to remember the location of 3 red items, remember the location of 3 red and 2 yellow items, or remember the location of three red items but ignore a further 2 yellow distractor items presented in the same array. The first two conditions manipulated memory load, whilst the latter condition added the requirement to filter distractors. The BOLD signal following the presentation of the cue revealed activation in basal

ganglia when subjects were preparing to filter out distractors. The magnitude of this activation correlated with working memory capacity across subjects, and negatively correlated with parietal BOLD in the retention interval (where higher activity may reflect unnecessary storage of irrelevant items). Lesion studies have also linked basal ganglia with working-memory performance (Voytek & Knight, 2010), basal ganglia lesions specifically impairing the ability to filter relevant from irrelevant information (Baier et al., 2010). Patients with Parkinson's disease, who have depleted dopaminergic innervation in striatum and prefrontal cortex, have behavioural impairments consistent with the proposed filtering role of dopamine and the striatum. Off medication, hypodopaminergic patients are impaired in updating context in the WM-dependent A-X Continuous Performance Task (AX-CPT), whereas on medication, when dopamine is high, they are less able to ignore distracting information (Moustafa, Sherman, & Frank, 2008). These latter findings are also consistent with Braver and Cohen's original proposal that dopamine itself constitutes the gating signal for working memory.

6.1.2 Behavioural evidence for effects of reinforcement history on selection

The models discussed above, which integrate reinforcement systems and control of working memory, predict behavioural effects of reinforcement history upon working-memory gating. The different models predict different behavioural phenomena. The earlier model of Braver and Cohen invokes a temporally specific but otherwise unspecific gating signal, associated with phasic dopamine firing. Recordings of phasic dopamine signals (Schultz, 1986; 1998) provide some idea of the likely time-course of gating: in macaques phasic dopamine signals ramp up within 100ms of encountering a reward-associated stimulus, and last a further 200ms. Behaviourally,

we might predict that if a number of items were presented simultaneously and one of those items carried a reward association, all presented items might receive an encoding boost for WM. This scenario is common enough within the context of a laboratory working-memory task, in which fixation is typically controlled and stimuli are presented briefly and simultaneously. Outside the lab, this apparently non-specific signal might in fact confer an item-specific encoding advantage, by virtue of being time-locked to changes in attentional fixation. We tend to sample the world sequentially: in the visual modality, this is revealed by the dynamics of eye movements. People saccade 3-4 times per second when going about a typical task, like making tea (Land, Mennie, & Rusted, 1999). Sequential samples lasting ~300ms have a similar time-course to dopamine transients, and these processes could be temporally coupled, rendering dopamine-triggered gating a selective mechanism for encoding into VSTM.

The more recent model of Hazy and colleagues (2007), in which corticostriatal loops carry parallel updating signals for working memory, would support selective encoding even if items were presented simultaneously. Items with higher value association should activate specific cortico-striatal loops with which they are associated, and have a competitive advantage for access to working memory.

There is already a growing body of work investigating the effect of reward associations upon attentional selection. This is mostly restricted to the study of visual selective attention, with RT as the dependent variable. Anderson and colleagues (B. A. Anderson et al., 2011) used a two-part design that aimed to dissociate the effect of stimulus-reward associations from the more general motivational/strategic effects of delivering reward feedback during task performance. In the first part, different stimuli are associated with different levels of reward. In a second part, these reward associations are task irrelevant, but the previously rewarded stimuli are retained. The

second task aims to measure whether and how the reward associations bias behaviour. Anderson and colleagues (2011) first gave subjects a visual search task, in which they searched for a red or green target amongst non-targets in different colours, and reported the orientation of the target. In each trial, feedback indicated either high or low monetary reward. One target colour was predominantly associated with high-reward feedback, and one with predominantly low-reward feedback (mapping of colour to reward value was counter-balanced). Following this reward-training task (1024 trials), the same stimuli were delivered in a test phase in which subjects searched for the singleton in an array of coloured shapes. The colour previously associated with high reward was never the target, but was present in 50% of trials associated with one of the distractors. RT was longer when one of the distractors was rendered in a colour that previously served as a target in reward training; this RT slowing was larger when the distractor was rendered in the colour previously associated with high reward. This effect was replicated with a much shorter reward-training task (240 trials) and was still present when these subjects returned to the lab after several days had elapsed. These data demonstrated suggested that prior reward association drives automatic attentional capture, even when stimuli are known by the subject to be task-irrelevant (and in fact distracting). The effect persists (B. A. Anderson, Laurent, & Yantis, 2012) even when the stimuli in the second task are of very different form to those used in the training task (a flanker-interference task, with letters as opposed to shapes and flanker colour as the manipulated variable, was substituted for the visual search task).

Reinforcement history can also capture attention on the basis of item identity as well as by low-level features (like colour). Della Libera and Chelazzi ran an experiment in which subjects compared a shape on the left side of visual space with a target on the right side. There was also an overlapping distractor shape (rendered in a different colour) on the left side. They found that when the distractor item had previously been

associated with high reward value, it was more difficult to ignore, and RT was slower (Libera & Chelazzi, 2009).

Reaction-time effects such as these are consistent with, but not diagnostic of attentional capture. For example, attention might dwell on more rewarded items for longer, thereby slowing RT to remaining items. However, recent work using eye movements as an overt index of attentional allocation has shown that reward-associated items do indeed attract shifts of attention. More eye movements are made toward reward-associated distractors in a search task (Camara, Manohar, & Husain, 2013; Theeuwes & Belopolsky, 2012).

Finally, one study (Raymond & O'Brien, 2009) has found that reward associations modulate stimulus detection in the attentional blink (AB) task. In the standard AB task, two targets (T1 and T2) are embedded one after the other in a stream of distractor items. If the interval between T1 and T2 is short, then the second target is often not available for report (Shapiro, Raymond, & Arnell, 1997). Raymond and O'Brien used a two-part paradigm in which different faces were first associated with either positive or negative expected value. These face stimuli were subsequently embedded in an attentional blink task, as T2. The first target (T1) was a texture patch on which subjects had to make a simple perceptual discrimination. Subjects were asked at the end of the trial to judge whether the T2 face they had seen was familiar or novel. As expected, when the T1-T2 interval was short, performance (as indexed by d') for this discrimination was lower. However, when the T2 face was associated with high positive value, the attentional blink effect was abolished: subjects were able to say whether the face was old or new equally well as in other trials when the T1:T2 interval was long enough that T2 should not be 'blinked'. The finding that past reward value

interacts with a task indexing the temporal dynamics of attention is intriguing given the 'temporal window' hypothesis for dopaminergic gating discussed above.

Taken together, the findings to date show that involuntary attentional capture can be driven by prior reward associations. However, whilst visuospatial attention and working-memory encoding are certainly linked (Schmidt et al., 2002), additional top-down factors beyond visuospatial attentional capture likely contribute to working-memory gating. Motivated by the plausibility of neural models linking fronto-striatal interactions and the dopamine system to control of working memory, a set of experiments was designed to test whether reward history can bias gating for working memory. Specifically, the prediction was that reward history would be capable of biasing encoding for working memory even when there was no *strategic* reason to select the reward-associated item for encoding (i.e. the probability of a memory item being probed did not depend on reward value, and there was no reward feedback for working memory performance). The primary hypothesis was that when there are multiple items competing for representation in working memory, items with a higher prior association with reward would have a competitive advantage for gating into working memory. This was tested using a two- part reward training / WM design, similar to those described above. To preview the results, in the first two experiments we found that, whilst reward history does modulate WM gating, it does so for all items in a WM array, not just the reward-associated item. This is consistent with the original dopamine gating hypothesis of Cohen et al. (Braver & Cohen, 2000) in which the reward signal modulates gating in a time-specific but not item specific manner, but less consistent with Hazy et al.'s (2007) account of striatal gating of WM. A third experiment was run in which items were presented sequentially, to investigate the temporal gating hypothesis.

6.2 Experiment 1

6.2.1 Overview

A two-part paradigm similar to Andersen et al. (2011) was used. Subjects first complete a reward-training task to establish reward associations, then perform the main working-memory task using reward-associated stimuli. Critically, the learned reward value is completely irrelevant to the demands of the subsequent working-memory task.

In the working-memory task (described in full in section 6.2.2.2), there were three types of memory array: (1) all items had high value, (2) all items had low value, and (3) mixed arrays in which half the items had high and half low value. In the mixed-array condition, the probe stimulus could either be of high or low value. In all conditions, the probe matched an item in the array 50% of the time. The two theories discussed in the general introduction make different predictions about the effect of the reward manipulation. Braver/Cohen's 'gating by phasic dopamine' account predicts that the reward manipulation will affect encoding only in terms of the net value of the array, as the gating signal cannot discriminate between simultaneously presented items. By contrast, Hazy and colleagues' (2007) striatal learning account predicts an item-wise bias. Provided performance is below ceiling, the items in the memory array are competing for representation in memory. Reward-associated items should have a competitive advantage, as their reward value should have potentiated their associated cortico-striatal 'stripe' (Hazy et al., 2007). These two predictions are illustrated below in the context of the current task design (Figure 6-1).

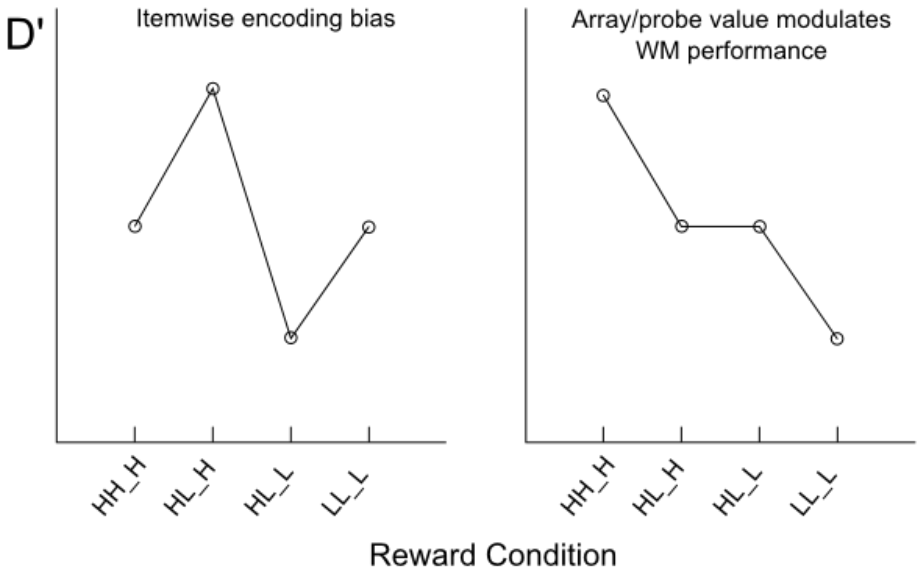


Figure 6-1 – **Predicted patterns, Experiment 1.** Codes on the X-axis: letter pair indicates WM array type (HH, all high value, HL, mixed value, LL all low value). Trailing letter indicated the reward value of the probe item. **Left panel:** predicted result given an item-wise encoding bias due to reward. When all items have the same reward value (HH, LL) each item is equally likely to be gated and performance is dictated by array size only. When some items have high value and some have low value, the high value items are more likely to be encoded and (given a constant capacity) performance is increased for high value items (HL_H) and decreased for low-value items (HL_L). **Right panel**: non-specific facilitation of encoding with higher net reward value of the memory array. **NB** for clarity, in the left panel, it is assumed that the same number of items is encoded regardless of the net value of the array. In contrast, in the right panel, 'effective capacity' is modulated by net reward value. A hybrid between the two patterns is also possible, in which high net reward value boosts overall performance, but high-reward items are at an additional advantage.

6.2.2 Methods

Twenty-three subjects (15 female, 8 male) took part in the experiment. Subjects were recruited using a mailing list. They were aged between 20 and 39 years old (mean age 23.5). Subjects were paid between £11 and £14, depending on their performance in the reward-training task. All subjects had normal or corrected-to-normal visual acuity. Left-handedness did not preclude participation. Ethics approval for the study was granted by the Oxford Central Research Ethics Committee.

The experiment was conducted in a quiet, dimly lit booth. Stimuli were presented on an LCD monitor at a viewing distance of 80cm. During the working-memory task gaze

direction was monitored using an Eyelink 1000 infra-red video eyetracker (binocular). A chinrest was used to stabilize the head.

6.2.2.1 Reward Training

The reward-training task consisted of a sequence of binary choices between novel shapes. The shapes were designed to be easily discriminable, but novel to discourage a verbal strategy (Figure 6-2). On a given trial, a subject was presented with two outline shapes, one on the left and one on the right side of the screen, and selected one or other by pressing a key on the computer keyboard. Reward feedback was given immediately. An on-screen reward bar and a winnings counter incremented, and an auditory signal was given (a cash register sound for high reward, a 'coin drop' sound for low reward, and a low beep for nil reward). Subjects were given no information about the reward value of the unchosen item. The magnitude of the reward feedback depended on the shape chosen, and was a fixed value. Subjects were instructed to win as much as they could over the course of the experiment, and that they would be given whatever they won in cash to take with them at the end of the session. Overall winnings ranged between 11 and 14 pounds (mean £13).

There were three possible values for the shape stimuli: nil value, low value (1p), and high value (10p). There were 18 nil value objects, six low value objects, and six high value objects. Over the course of reward training, there were twice as many pairings of a low value object with a nil value object than of either a high with a low value object, or a high with nil value object (Table 6). This was in order to equate the number of times subjects actively chose a high value object and a low value object, to avoid confounding reward value with the number of times an item had been chosen. Trial order was randomized, and the side at which the higher value item appeared was also randomized.

Pairing	Number of trials
High, Low	72
High, Nil	72
Low, Nil	144
(total)	288

Table 6 – **Trial numbers, Experiment 1.** Double the number of nil value vs. low value trials were included than high vs. low of high vs. nil, in order to equate the number of times subjects actively chose high and low value items.

6.2.2.2 Working-memory task

The working-memory task was a 'present/absent' forced-choice task. The working-memory array consisted of either four or two items. After a retention interval, a probe was presented centrally, and subjects were asked to indicate whether the probe item they saw had or had not been present in the memory array. A four-item trial is shown in Figure 6-2.

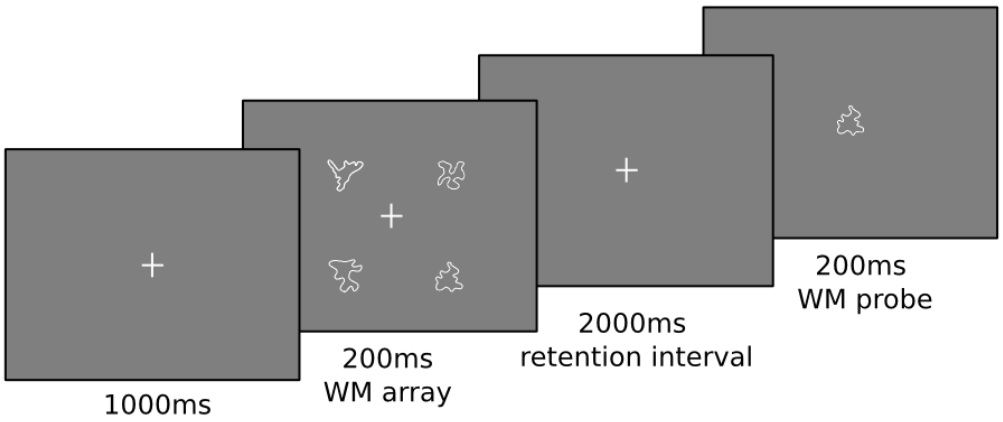


Figure 6-2 – **Trial schematic, Experiment 1.** Subjects were asked to maintain fixation throughout the trial. Following the presentation of the probe item, subjects made a self-timed 'present/absent' response, depending on whether they thought the probe item had been present in the memory array

Following the presentation of the fixation cue, the working-memory array was presented for 200ms, followed by a 2s retention interval. The probe item was then presented for 200ms at fixation, and subjects made a self-timed 'present/absent' response, indicating whether or not the probe item had been a member of the memory array.

Only the set of shapes associated with either a low or high reward value in the reward-training task (6 low value shapes, 6 high value shapes) were used as memory items in the WM task. The different combinations of memory item and probe item values are summarized in Figure 6-3

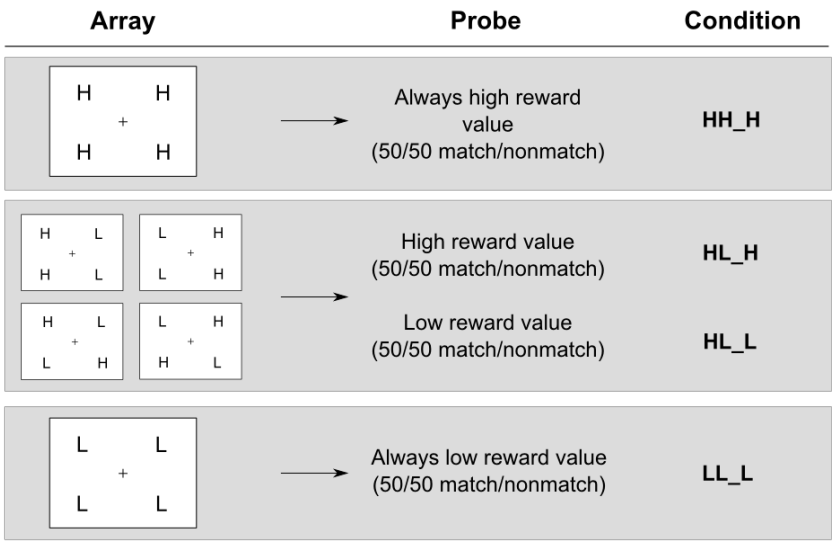


Figure 6-3 – **Task conditions, Experiment 1.** 'H' stands for a stimulus associated with high reward value, and 'L' for a stimulus associated with low reward value. The codes given in the 'condition' column are re-used in results figures, below. The first part gives the properties of the memory array (e.g. 'HL', an array with a mixture of high and low value items) and the second part the value of the probed stimulus from the memory array, on match trials. NB only four item trials are shown here, but in actuality for half of the trials, only two items were presented. In two item trials, the items were always presented on the diagonals.

There were 384 trials in total split evenly between two-item and four-item trials. There were therefore 48 trials for each combination of set size (x2) and reward

condition (x4), consisting of 24 match trials and 24 non-match trials. For set size two, the memory items were presented on the diagonals. In conditions for which the memory items in the array were all of the same value, the probe stimulus in non-match trials was always of the same value as the stimulus array. This was to avoid subjects using the change in reward value to inform their match/non-match decision (as opposed to remembered item identity).

6.2.3 Results

6.2.3.1 Reward Training

Subject's engagement with the reward training task was assessed by dividing trials into the three possible stimulus pair types (High with Low, High with Nil, Low with Nil) and calculating the probability that subjects correctly chose the more valuable item as the task progressed, over a 21 trial moving window. Subjects were generally able to learn to select the more valuable item in all three pair types (Figure 6-4). However, subjects were less likely to choose the low value item in the low-nil pairing than they were to choose the high value item in the high-nil or high-low pairings. This could have been because subjects learned to discriminate the high value items from low/nil value items faster than they learned to discriminate the low value items from nil valued items. However, the behaviour of a subset of subjects, who tended to actively choose the nil valued item in the low/nil pairings (2 subjects, both $p < 10^{-6}$; exact binomial test over the last 40 low/nil trials), suggests that strategic/exploratory choices could have been driving this difference. When presented with one item that is known to have low value, and another item for which the value is not known, there is an incentive to choose the unknown item both because it may be a high value item, and because if it is a nil value item, this is worth learning in order to avoid it in future. Once all item values were

fully learned this would be a suboptimal strategy, but given the large difference between the value of the low and high valued items, it can be an effective strategy during the learning phase[28].

Any subject who failed to learn to choose low value items in preference to nil valued items by the end of the reward training was excluded from the analysis of the WM data. An exact binomial test was used to find the probability of the proportion of choices each subject made over the last 40 trials of the reward training, under the null hypothesis of random choice. Two subjects were excluded because they actively chose the nil valued item in low reward / nil reward pairings. A criterion p-value of $p <= 0.01$ (two-tailed) was used to exclude subjects who did not show evidence of learning *per se*. Four further subjects were excluded on this basis. All other subjects were choosing the low valued items in preference to the nil valued items by the end of training (all $p < 0.0011$ under the null hypothesis of random choice).

The grand averaged learning curves for each choice type, excluding data from the 6 outlier subjects described above, are shown in Figure 6-4.

[28] The reward-training task was modified for experiment 2, so that the value of both items was displayed after subjects made their choice, in order to remove the exploratory incentive.

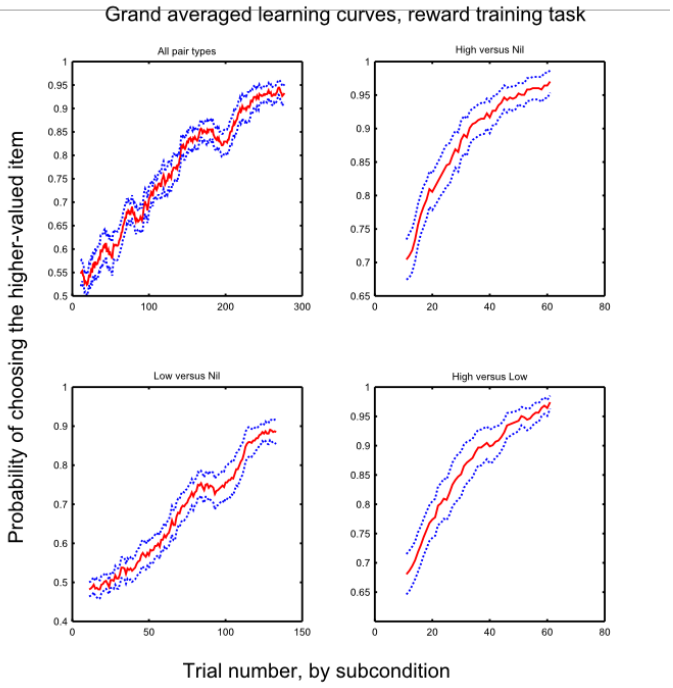


Figure 6-4 – **Learning curves in the reward-training task, Experiment 1.** These averages exclude data from the six subjects who failed to learn to discriminate between nil and low value shapes. Error bands are ±1 SEM.

The ratio (# times high value item chosen) / (# times low value item chosen) was slightly greater than 1 (1.23, std. 0.19) over the course of the reward training, as the value of low value options was learnt more slowly than that of high value options[29]. However, the bias was small, and towards the end of the training session (over the last 40 trials for each sub-condition) the probability of choosing a low value item in preference to a nil value item was similar to the probability of choosing a high value item in preference to a nil value item (0.918 and 0.961 respectively), both being near ceiling.

[29] The closer this ratio to 1, the less chance that a difference in choice history as opposed to reward association *per se* could have driven difference in the memory task.

6.2.3.2 Working-memory task

Trials in which reaction time was longer than 5 seconds or shorter than 0.3 seconds were excluded from the analysis, as they likely reflected lapses in task engagement or premature responding, respectively (on average 2.5/384 trials were excluded).

Data were analysed using repeated-measures ANOVAs with factors for number of items (2 vs 4) and array/probe reward condition (4 levels: HH_H, HL_H, HL_L, LL_L; see Figure 6-3).

The raw hit rate and false-alarm rate are given in Figure 6-5.

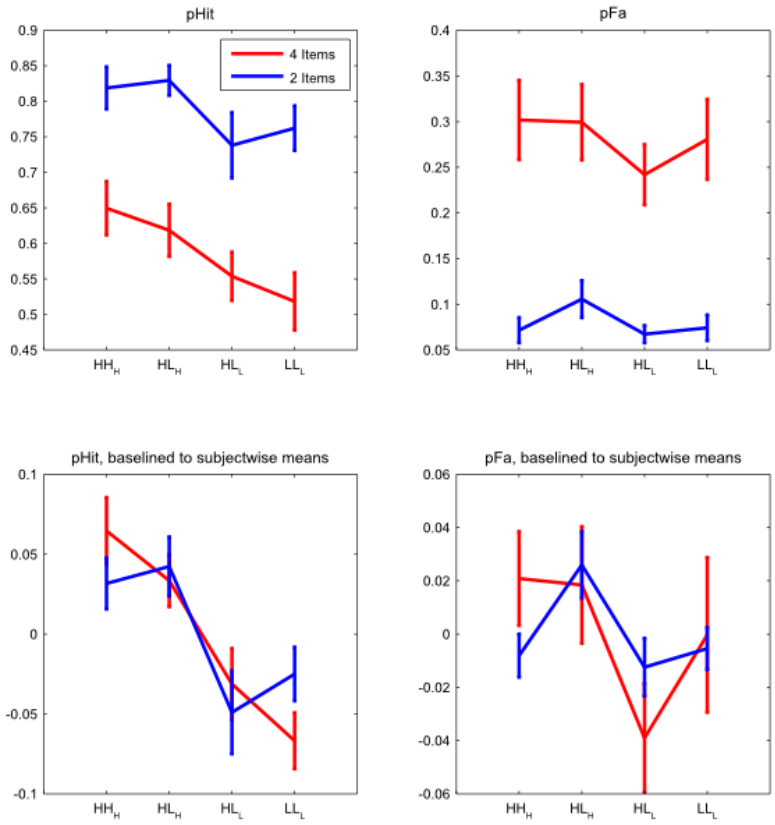


Figure 6-5 – **Hit rate and false alarm rate for the working memory task, Experiment 1.** Error bars are ±1 SEM. Upper panels show group-averaged raw data. The lower panels show the same data baselined (by subtraction) to the subject-wise mean across all reward conditions (but within array size conditions), before group averaging. This puts two- and four-item hit rates on a common scale, and removes between-subject variance from the SEM.

There was a main effect of reward condition upon hit rate ($F(3,51) = 9.06$, $p < 0.0005$). Hit rate was elevated for items of higher value. Additionally, there was a main effect of set size ($F(1,16) = 82.24$, $p < 0.0005$; performance higher for set size 2). There was no evidence for an interaction between reward condition and set size ($F(3,48) = 1.14$, $p = 0.342$). The hit rate was higher for high reward items, but this effect could have been driven by a reward-induced response bias (driven by the reward value of the probe item as opposed to an encoding bias for memory) as the false-alarm rate is elevated for high reward probe items in the non-match trials. Measures from signal-detection theory can be used to distinguish changes in response bias from changes in sensitivity. Parametric signal detection theory measures, d' and criterion c[30], are plotted in Figure 6-6.

[30] Criterion is expressed relative to the 'neutral point' (the value of the hypothetical SDT internal variable which has equal likelihood under the noise and signal distribution), such that positive values imply a bias towards 'no' responses and negative values a bias towards 'yes' responses.

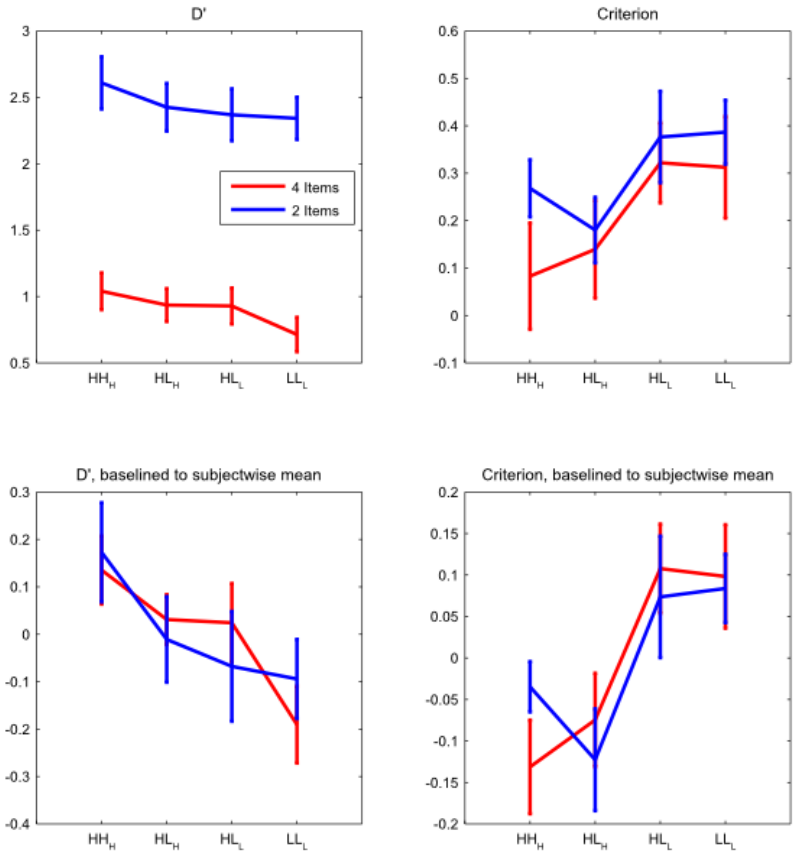


Figure 6-6 – **SDT measures, Experiment 1.** Error bars are ±1 SEM. Top panels are raw data. Bottom panels show the same data baselined (by subtraction) to the subject-wise mean across all reward conditions (but within array size conditions), before group averaging. The left hand panels show D' increasing in step with the net reward association of the working-memory array. The right hand panels show criterion is lowered – i.e. subjects becoming more liberal in their responding – when the probe item is of high value relative to when it is of low value.

Subjects are more liberal in their responding (criterion is lower) when the probe value is high (main effect of reward condition, $F(3,51) = 4.56$, $p = 0.007$). In addition to this response bias, there was a significant main effect of reward condition upon D' ($F(3,51) = 3.28$, $p = 0.029$). D' was elevated for conditions in which the array items had higher net reward. There was no evidence for an interaction between set size and reward condition either for D' ($F(3,48) = 0.311$, $p = 0.817$) or for criterion ($F(3,48) = 0.627$, $p = 0.601$). D' baselined to the subjectwise means is shown with the load conditions collapsed together in Figure 6-7 – D' increases with the net value of the stimulus array.

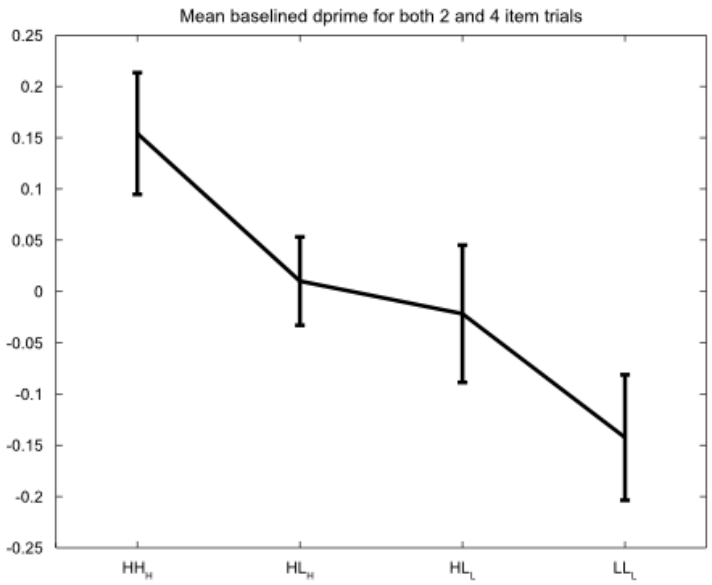


Figure 6-7 – **Baselined D' collapsed over the 2 and 4 memory item trials, Experiment 1.** Error bars are ±1 SEM. D' is proportional to the net reward value of the memory array.

Reward condition also influenced reaction time, but for 'stimulus present' (match) trials only, in which subjects responded more quickly when the memory-matching probe item was of higher value (Figure 6-8). As described above for the accuracy data, RT was first baselined within subject and within set-size, and then collapsed across set size. These data were analysed using a repeated-measures ANOVA with factors Match/Non-match trial (2 levels) and Reward condition (4 levels). There was a significant interaction between factors Match/Non-Match and Reward Condition, $F(3,48) = 2.88$, $p = 0.045$). In non-match trials, a correct response corresponds to an exhaustive comparison of the probe item with the items in memory having yielded no match, whereas in match trials a positive correspondence between the probe item and a memory item could drive an immediate response, so it is reasonable to model RT for match and non-match trials separately. In match trials, high reward value probes were associated with faster RT (separate ANOVA for match trials only, main effect of reward condition $F(3,48) = 4.86$, $p = 0.005$). In non-match trials there was no main effect of reward condition (ANOVA for non-match trials only, main effect of reward condition, $F(3,48) = 0.70$, $p = 0.56$).

Differences attributable to response bias and encoding bias are not easily separable in the reaction time data, and RT is not analysed further here, though these data serve to confirm that our reward manipulation succeeded in influencing behaviour in the WM task. The increase in speed for high-reward match trials could be attributable to an increase in signal strength for a match due to better encoding of the high value item, or a faster response due to faster attentional capture by the probe item.

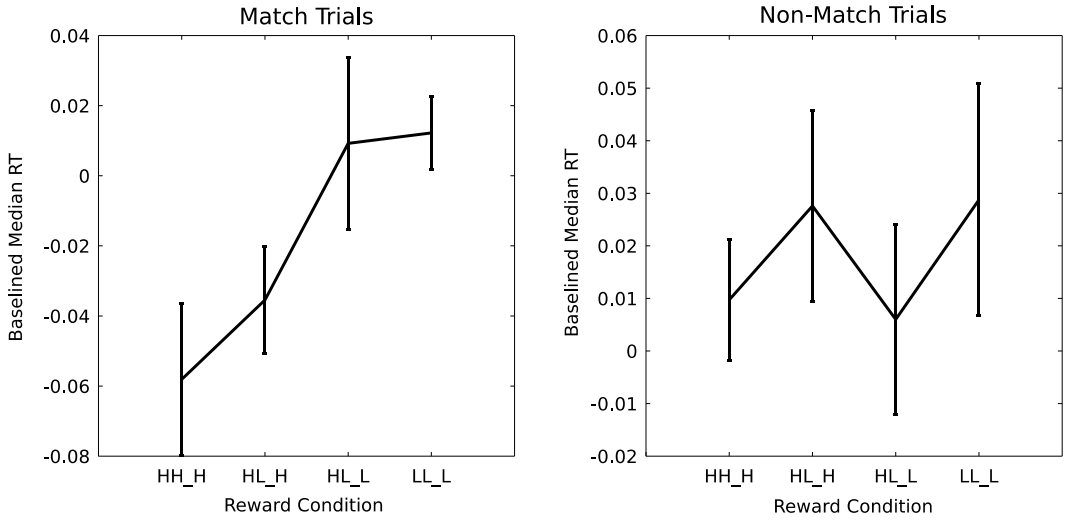


Figure 6-8 **Baselined median reaction time for match and non-match trials** (collapsed across set size; the baseline is common across match/non-match trials, hence the data in the individual plots is not mean-centred). The reward association of the probed item affects reaction time on match trials only: subjects respond more quickly when the probe stimulus is of higher value.

6.2.4 Interim discussion (1)

The pattern of results predicted on the basis of each of the two theoretical accounts discussed in the introduction was illustrated in Figure 6-1. The more recent account of Hazy et al (2007) predicted an item-wise encoding bias in favour of high-reward value items, whereas the Braver/Cohen dopamine gating account predicts an overall facilitation of encoding with higher net reward value. The pattern of results observed was consistent with this second account; compare Figure 6-7 with the right-hand panel of Figure 6-1. There was no evidence for an item-wise bias, but the net value of the

memory array modulated the likelihood that an item will be encoded (regardless of the specific value of that item). The possibility that reward-driven gating is time-specific, but not item specific, is investigated further with a sequential-presentation working-memory task in a subsequent experiment (experiment 3, section 6.4).

Whilst D' did not indicate an item-wise encoding bias in mixed-reward arrays, there was an item-wise difference in hit rate (Figure 6-5), which appeared to be driven by a change in criterion (i.e. response bias) that could have been driven purely by the reward value of the probed item. In Experiment 1 the probe stimulus was presented centrally. Subjects were therefore obliged to perform a search through the memory array to establish a match or non-match. This may have induced a response bias if more highly rewarded items were more likely to be erroneously recalled (an 'intrusion') during search.

Partly for this reason, another version of this experiment was run, in which the probe stimulus was presented in the same location as one of the memory items, and subjects were asked to make a same/different judgement about that specific item. Spatially-tagged retrieval should reduce the memory-search demand and therefore also the scope for reward effects during search (e.g. causing intrusions). In this version of the experiment, the non-match probe stimulus was allowed to be either of high or of low reward value, even for the cases in which the entire memory array was of high or low value. The probed stimulus could therefore either be of high or low value in all conditions, but also either congruent or incongruent with the value of the probed item. This gives a means to test whether subjects were able to encode separately the identity and value of an item, and potentially use the value information to facilitate comparison with the probe stimulus. This design has the further advantage that it should help curb any tendency for subjects to plan in advance which items to encode on a given trial.

For this version of the experiment, trial-wise information was added about both items presented in each trial of the binary choice reward-learning task, to reduce exploratory choice behaviour.

6.3 Experiment 2

6.3.1 Overview

The rationale for this second experiment is discussed in relation to experiment 1 (section 6.2.4, above). In brief, the aim was to replicate the results of the prior experiment, but with greater control over the reward-driven response bias observed in experiment 1.

6.3.2 Methods

Twenty-five subjects (14 female, 11 male) took part in this experiment. Subjects were recruited using a mailing list. They were between 20 and 39 years old (mean age 24.3). Subjects were paid between £13 and £15, depending on their performance in the reward-training task. All subjects had normal or corrected-to-normal visual acuity. Left-handedness did not preclude participation. Ethics approval for the study was granted by the Oxford Central Research Ethics Committee.

The experiment was conducted in a quiet, dimly lit booth. Stimuli were presented on an LCD monitor at a viewing distance of 80cm. During the working-memory task gaze direction was monitored using an Eyelink 1000 infra-red video eye-tracker (binocular). A chinrest was used to stabilize the head.

6.3.2.1 Reward Training

Reward training was as described for the previous experiment, with a single modification. Immediately following each choice, subjects were informed both of the value of the item they chose, but also of the value of the un-chosen item, to reduce exploratory choices.

6.3.2.2 Working-memory task

The working-memory task required an item-specific change/no-change forced-choice decision (Figure 6-9). The set size was restricted to four items, as in Experiment 1 there was no interaction between set size and reward value, and this allowed within-condition trial numbers to be increased, in order to collect an adequate number of reward-congruent / reward-incongruent non-match trials.

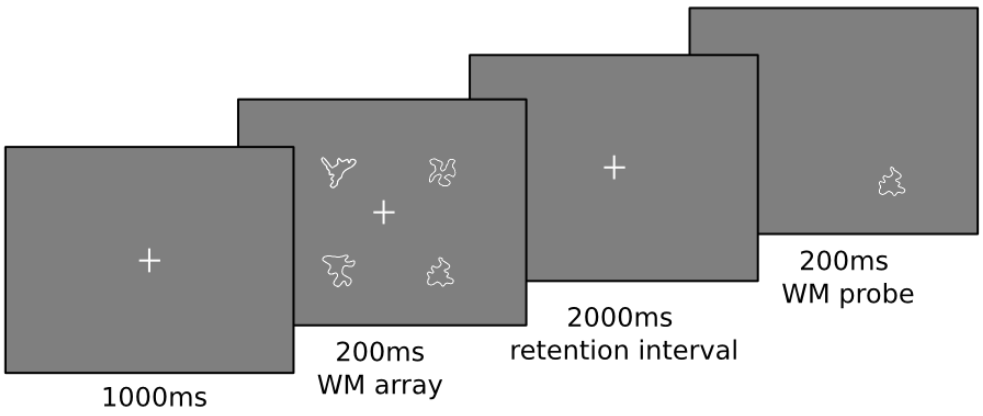


Figure 6-9 **Trial schematic, Experiment 2.** The working memory task was very similar to the task used in experiment 1, with identical stimuli and task timings, except that the probe stimulus was item-specific by virtue of its location.

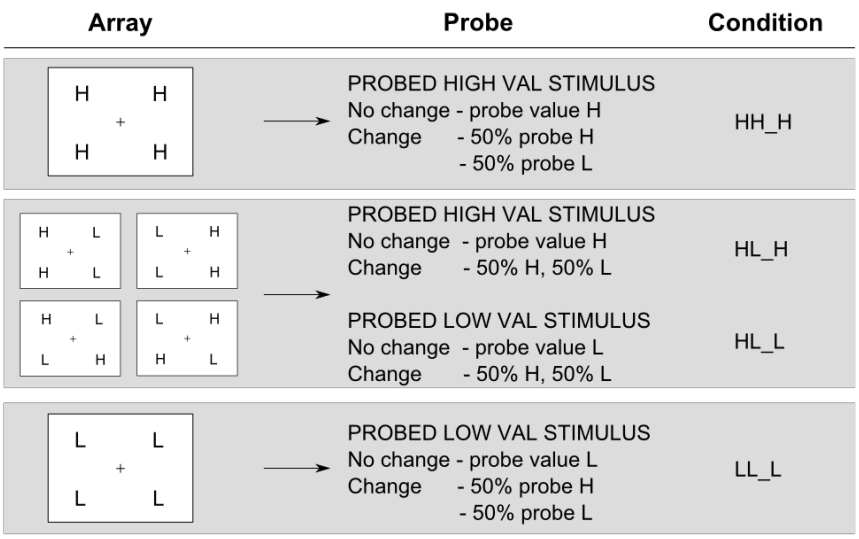


Figure 6-10 **– Task conditions, experiment 2.** The first part of the condition code indicates the type of memory array (e.g. 'HL' is an array containing both high and low reward value items). The second part indicates the value of the probed stimulus from the memory array (NB on change trials, this may not be the same as the value of the probe stimulus itself).

A schematic of the task conditions is shown in Figure 6-10. There were 48 match trials in each of the four conditions, and 48 non-match trials of which 24 had a probe stimulus with the same value association as the probed memory item, and 24 a probe stimulus with a different value association (termed 'reward congruent' and 'reward incongruent' below).

6.3.3 Results

6.3.3.1 Reward Training

In Experiment 1 nil value items were sometimes chosen in favour of low value items. In experiment 2, information was provided about the un-chosen item to reduce the motivation for this strategy. In the current experiment, one subject actively chose nil value items in favour of low value items. This subject was excluded from further analyses. All other subjects successfully learned to perform the task for all pair-types.

The grand averaged learning curves for each choice type, excluding data from the outlier subjects described above, are shown in Figure 6-11.

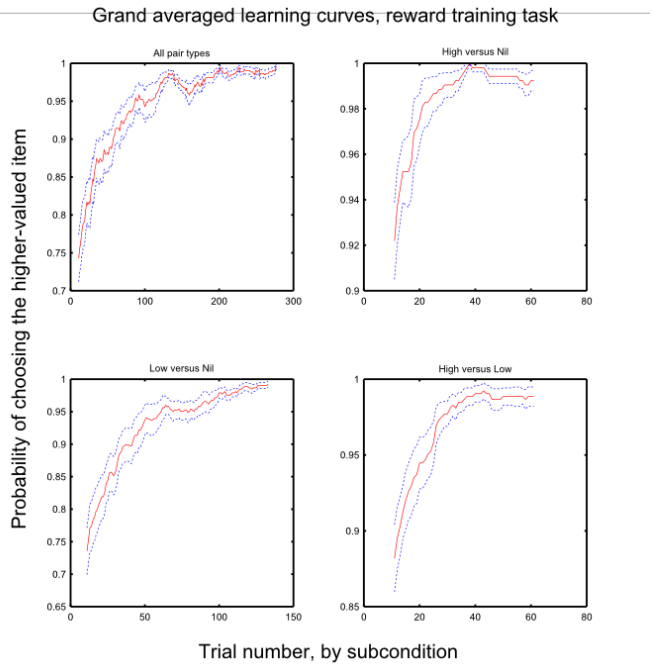


Figure 6-11 **Group-averaged learning curves, reward-training task, Experiment 2.** Performance reached ceiling for all pair types. Error bands are ±1 SEM

The ratio (number of times high value item chosen) / (number of times low value item chosen) was 1.05 (std 0.06) over the course of training. Towards the end of training (last 40 trials per condition) the probability of choosing a low value item in preference to a nil value item was 0.99 and the probability of choosing a high value item in preference to a nil value item also 0.99; both categories were fully learnt, performance being at ceiling.

6.3.3.2 Working memory task

Trials in which reaction time was longer than 5 seconds or shorter than 0.3 seconds were excluded from the analysis, as they likely reflected a lapse in task engagement or premature responding, respectively (on average 1.5/384 trials were excluded).

The hit rate was higher for probed items with a higher reward value (Figure 6-12).

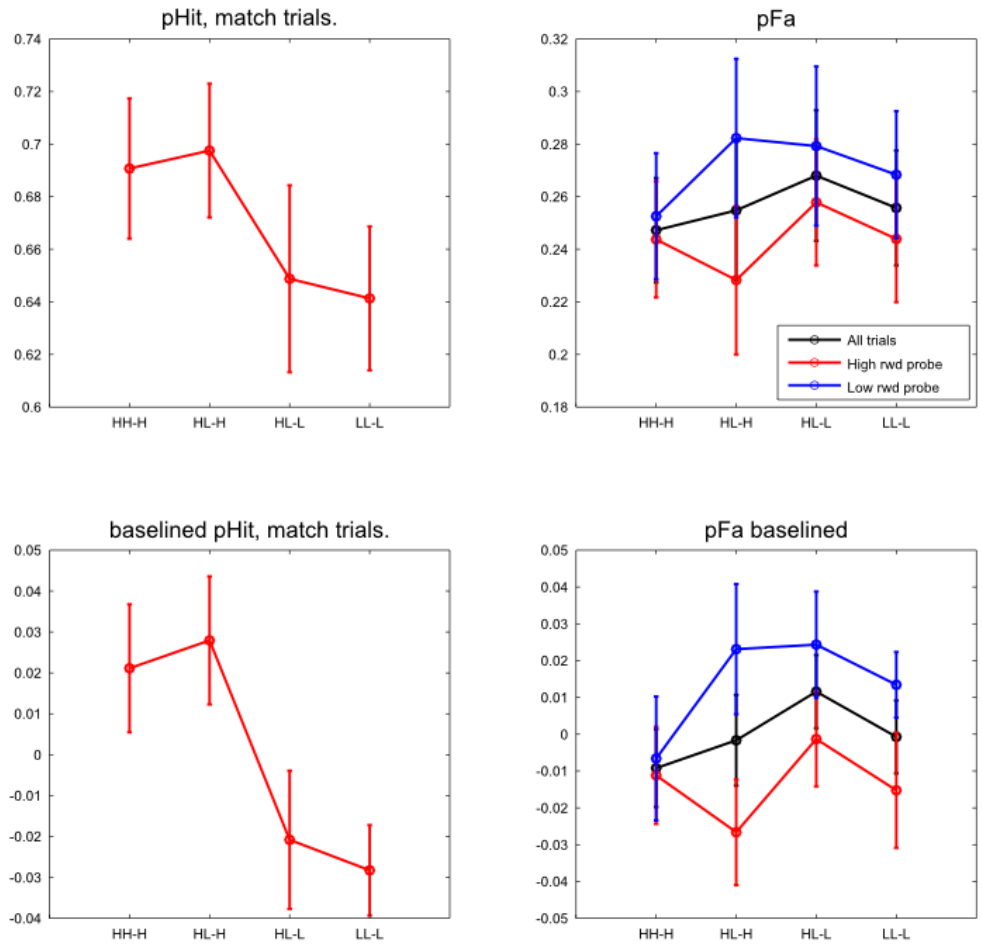


Figure 6-12 **Hit and false-alarm rate for the working-memory task in experiment 2.** The false-alarm rate is shown separately for trials in which the probe stimulus was of high and low reward value. To facilitate comparison with Figure 6-5, bottom panels show data baselined (by subtraction) to the subject-wise mean across all reward conditions, before group averaging. Error bars are ±1 SEM

These data were analysed using a repeated-measures ANOVA with a single factor of reward condition (four levels). There was a significant main effect of reward condition upon hit rate ($F(3,69) = 2.74$, $p = 0.05$). Although the non-match trials contained an equal number of trials in which the probe was of high- or low-reward value, for each condition, the value of the probe in the match trials was necessarily the same as the probed stimulus (i.e. high for the HH_H and HL_H conditions and low for the HL_L and LL_L conditions). A response bias driven by probe value only could therefore still have potentially accounted for the hit rate effect (if subjects were more likely to indicate a

match with a high-reward item). However, if that were the case, the false-alarm rate would be expected to be higher in the non-match trials in which a high value probe stimulus was presented. In fact, the opposite pattern emerged: false-alarm rate was higher when the probe stimulus was of low value. In a repeated-measures ANOVA with factors probe value (2 levels) and reward condition (4 levels) there was a main effect of probe value on false alarm rate ($F(1,69) = 4.81$, $p = 0.039$). There was no main effect of reward condition ($F(3,69) = 0.467$, $p = 0.706$) or interaction between reward condition and probe value ($F(3,69) = 0.671$, $p = 0.573$).

For the primary hypothesis that reward association should confer an item-wise bias for working-memory gating, the key contrast was between HL_H and HL_L conditions, as a difference here could only be driven by an item-wise bias to encode items associated with higher reward value (as the memory arrays were of equal value). A paired-samples t-test was therefore used to test for a difference in the mean between these two conditions. This difference was not significant ($p = 0.12$). There was no evidence for an item-wise bias, although the trend would be consistent with a bias towards encoding high value items.

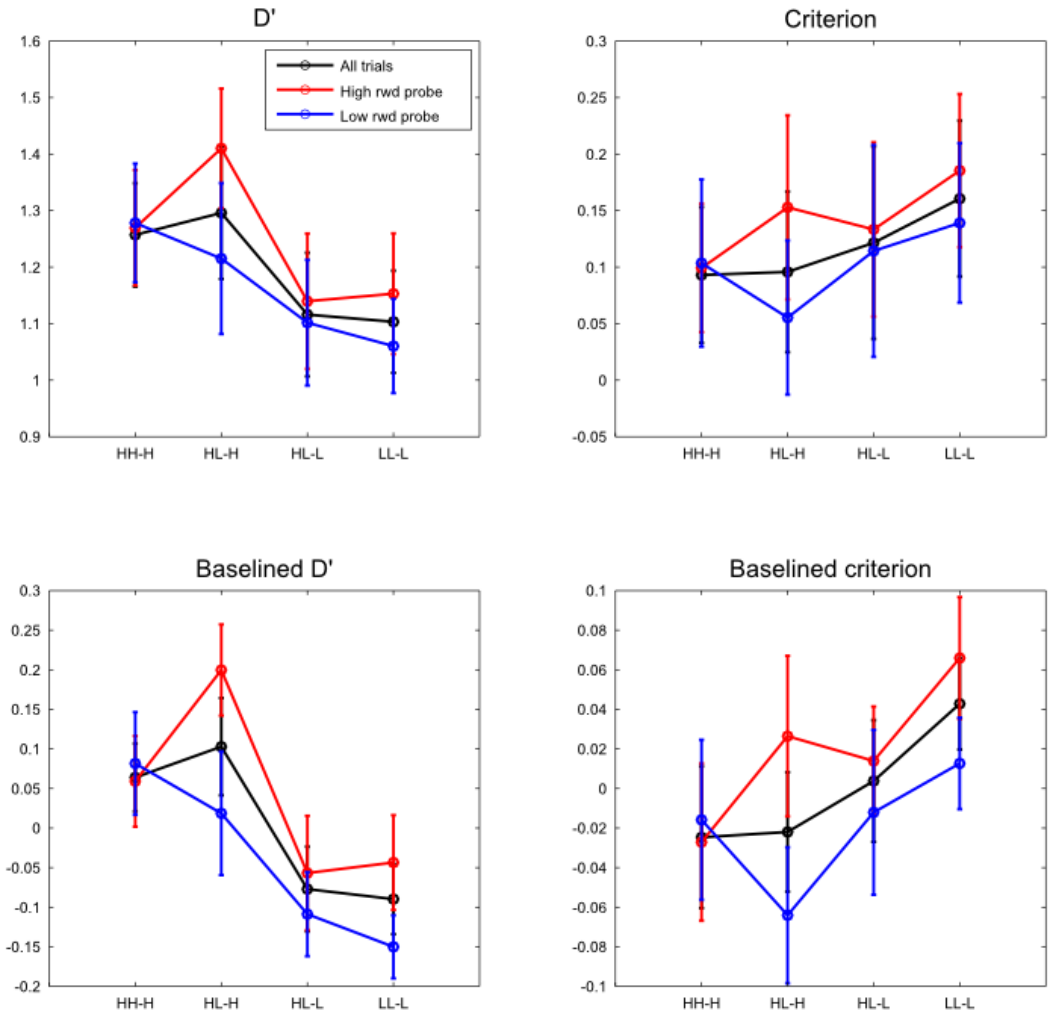


Figure 6-13 **Signal detection measures for the task in Experiment 2.** D' and criterion (c) are plotted separately for an analysis in which only the high-reward probe non-match trials were included, and an analysis including only low-reward non-match trials (c.f. Figure 6-12) as well as the analysis including all the data. Bottom panels show data baselined (by subtraction) to the subject-wise mean across all conditions, before group averaging. Error bars are ±1 SEM.

In line with the approach taken for experiment 1, a signal detection analysis is shown in Figure 6-13. A separate repeated-measures ANOVA with factor reward condition (4 levels) was used to analyse D' and criterion. For D', there was a trend for a main effect of reward condition although this did not reach significance (Greenhouse-Geisser corrected $F(2.09,69) = 2.74$, $p = 0.072$). There was no evidence for a main effect of reward condition upon criterion ($F(3,69) = 0.8$, $p = 0.5$)

There were no significant effects of reward condition upon reaction time in this experiment.

6.3.4 Interim discussion (2)

In both experiment 1 and experiment 2, performance in the working-memory task was influenced by the reward associations instantiated in the prior reward-learning task (as was RT in experiment 1), even though reward was task irrelevant in this subsequent task. The results observed support the idea that prior reward association influences working-memory encoding, but the data did not furnish evidence for an item-wise encoding benefit associated with high reward value. Instead the results from experiment 1 suggest a generalized facilitation of encoding dependent on the net reward value of the array. Any item from a memory array with a high net value was more likely to be encoded. The pattern of results in experiment 2 was qualitatively similar, although in this case, for the mixed-value arrays, hit rate / d' were higher when a high reward item was probed than when a low reward item was probed. This is suggestive of an item-wise bias. However, this difference was not statistically significant.

Across the two experiments, response bias was influenced by reward condition. In experiment 1, there was a robust criterion shift, which appeared to track the reward value of the probe stimulus. Where the probe stimulus was high in value, subjects were more liberal in their responding, tending to make more false alarms. In experiment 2, in which subjects were probed about a specific item in the memory array (because the probe appeared in a particular location), the reward value of the probe stimulus had the opposite effect on response bias: people made more false alarms when the probe stimulus was low in value. It is difficult to reconcile these patterns, save to say that the change from a central to an item-specific probe stimulus clearly changes

subjects' response strategy (likely due to the elimination of search through the memory array). Regardless, d-prime should be robust to this change in strategy.

The pattern of results observed in experiments 1 and 2 suggested two possibilities. Firstly, the experiments may have been insufficiently sensitive to detect an item-wise bias. In experiment 2 there was a trend for an item-wise encoding bias. Secondly, there was evidence, particularly from experiment 1, for a general (not item-specific) facilitation of working-memory performance by reward history. As discussed above, this could be a consequence of a reward-mediated gating mechanism that is temporally specific, but not item specific.

The latter possibility was addressed in a further experiment in which WM stimuli were presented sequentially to render our task sensitive to a temporally specific gating mechanism. This should permit an item-wise encoding bias to be expressed if, as predicted, reward associations trigger a brief window in which WM encoding is boosted. By presenting items in rapid sequence, we aimed to establish the width of this putative window. We predicted that we might see not only a boost for rewarded items, but also an encoding boost for subsequent items in the sequence.

6.4 Experiment 3

6.4.1 Overview

The overall design was the same as in experiments 1 and 2: a reward-training task was completed first, followed by a working-memory task in which the previously reward-associated features were present, but reward was task irrelevant. In the training session, two colours were associated with different levels of monetary reward. After a short break, subjects performed an orientation working-memory task in which some items were of the reward-associated colour. Working-memory items were presented sequentially, and subjects responded by attempting to match the orientation of an item in memory.

6.4.2 Methods

Subjects were recruited though a mailing list. Twenty-two subjects (14 female, 8 male) completed 3 sessions. Subjects were aged between 19 and 28 (mean age 21). Subjects were paid between £15 and £17 per session, depending on their performance in the reward task. All subjects had normal or corrected-to-normal visual acuity. Left-handedness did not preclude participation. Ethics approval for the study was granted by the Oxford Central Research Ethics Committee.

The experiment was conducted in a quiet, dimly lit booth. Stimuli were presented on an LCD monitor at a viewing distance of 80cm. During the working-memory task gaze direction was monitored using an Eyelink 1000 infrared video eyetracker (binocular). A chinrest was used to stabilize the head. Each subject performed three sessions of the

two-part experiment (reward-training and working memory task). The sessions were performed on three separate days.

6.4.2.1 Reward training

The aim of the reward-training task was to associate different colours with different levels of reward. Colours associated with differential reward were equated for presentation frequency during reward training, and also for the number of times subjects directed a choice towards an item with that colour. Red and green were adopted as the rewarded colours. The association of these two colours with high and low reward was counterbalanced across subjects.

The task was framed as a game. Subjects were presented with a square 'arena' in which three coloured patches moved unpredictably, but with smooth trajectories. The subjects controlled the location of a cross-hair with a trackball mouse (Logitech Marble mouse). On any given trial, one out of the three patches would be either red or green (i.e. a rewarded, 'target' colour) and the remaining two patches were drawn from a set of four colours not associated with reward (blue, purple, orange, yellow). This design equates the number of presentations of all colours. Subjects were asked to 'chase' the reward-associated patch with the cross-hairs, and click the mouse button to 'catch' it. The arena within which the patches moved subtended 12 degrees of visual angle, and each patch subtended 0.6 degrees visual angle.

At the beginning of the first session, subjects were given 120 trials of training (attempting to catch the reward-associated colours, but without any reward feedback). This was to familiarize them with the mouse. There was no time limit on responses at this point. Following this training session, subjects performed the task again, but this time with reward feedback. A time limit of 3 seconds was imposed. If subjects failed to respond before this time limit, the trial would time out and the next trial would be

presented. This time-out both kept subjects engaged and permitted staircasing of task difficulty. A staircasing procedure ran continuously with the aim of equating the money won by different subjects, despite differences in the ease with which different subjects performed the task. Every 20 trials, the time available to catch a patch before the task timed out was changed in 100ms steps, decreasing if the subject was more than 90% accurate on the previous 20 trials, and increasing if they were less than 90% accurate. There were 360 trials in total, divided into blocks of 40 trials interspersed with self-timed rest breaks.

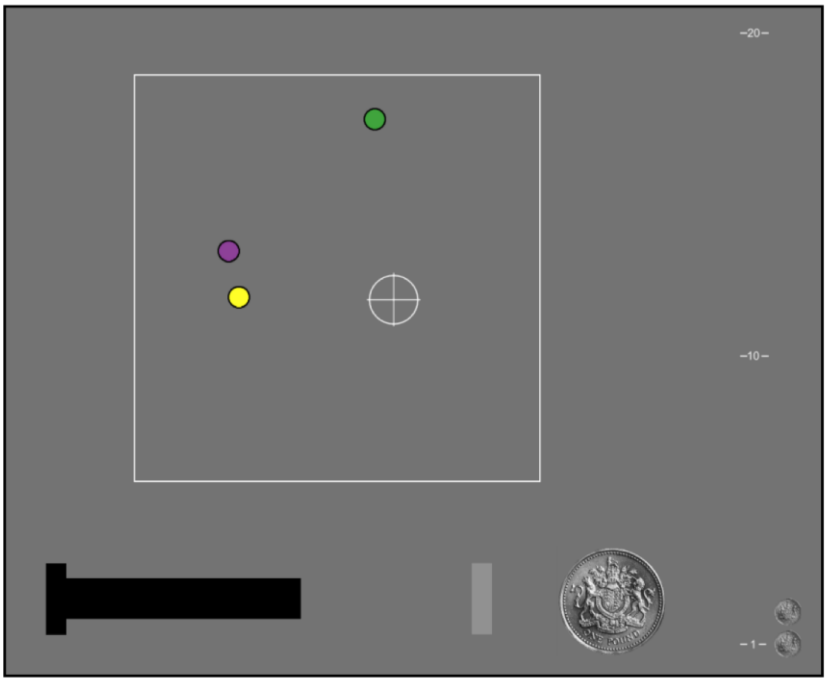


Figure 6-14 – **Screenshot of the reward-training task.** The coloured patches are in constant unpredictable motion. Subjects catch the red or green patch by moving the crosshair over it and pressing the mouse button.

Reward feedback was given visually in the form of a reward bar at the bottom of the screen that grew incrementally from the left towards the right of the screen (Figure 6-14). When this bar hit a target bar at the right hand side of the screen, £1 was added to the subject's winnings. The overall winnings were represented by a coin stack on the right-hand side of the screen. Auditory feedback was also given, in the form of a

'kerr-ching' cash register sound when subjects caught a high reward item, the sound of a single coin dropping into a can when subjects caught a low reward item, and a low tone when subjects missed the target item, or timed out. Subjects were informed before the beginning of the reward training that the money they won was real and would be given to them at the end of the experiment. The high-reward colour earned subjects 9 pence if caught, the low reward colour 1 pence. Subjects earned between £15 and £17 pounds per session.

6.4.2.2 Working memory task

We employed a precision/capacity type of orientation memory task (Bays & Husain, 2008; W. Zhang & Luck, 2008). Subjects were presented with a sequence of symmetrical 'UFO' stimuli, and were asked to recall the orientation of one of the stimuli after a short interval, by rotating a probe item to match the remembered orientation. The memory stimuli were presented sequentially and in the centre of the screen. There were three coloured memory stimuli, always followed by a white distractor stimulus that subjects were informed they would never be asked about (but that they were to view). This stimulus was included to ensure the last item was not encoded in a qualitatively different manner to the preceding items (e.g. as an after-image). After a delay, one of the coloured stimuli reappeared with a random orientation. Subjects moved the trackball of the mouse horizontally to rotate the probe stimulus until it matched the remembered orientation for the remembered stimulus of the matching colour, at which point they pressed the mouse button. Subjects were asked to make an attempt to match the orientation even if they felt they were completely guessing. A trial schematic with trial timings is given in Figure 6-15.

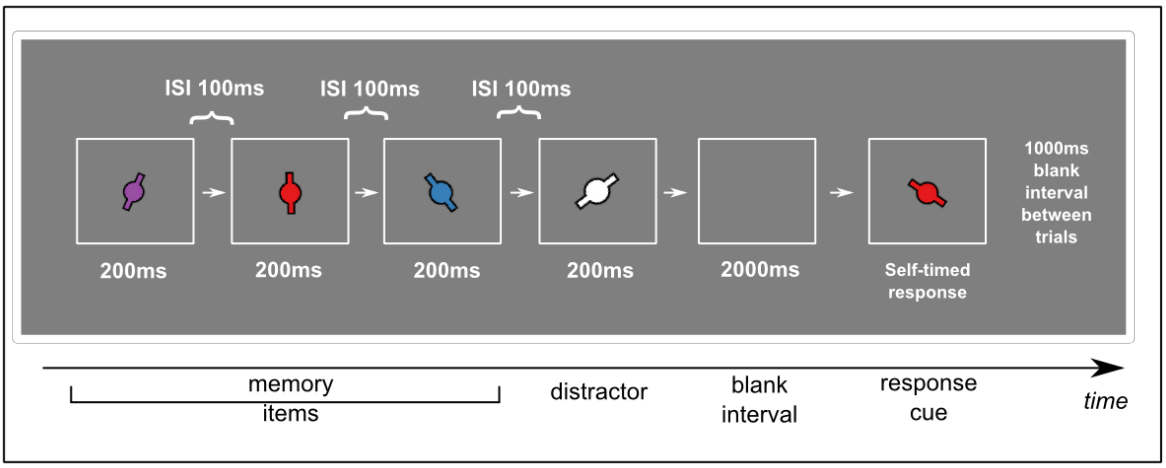


Figure 6-15 **Schematic of the sequential precision/capacity task.** Items were presented in rapid succession. A white to-be-ignored item always finished the sequence, acting as a mask to reduce potential after-image benefits for the final item in the sequence. Subjects were able to rotate the response item to best match the remembered orientation of the stimulus of matching colour, and then press the mouse button to register their response.

The task was divided into 12 blocks of 30 trials each (360 trials), with rest breaks between blocks. On each trial, one of the three stimuli was a previously rewarded colour, with 50% of trials harbouring the high reward associated colour, and 50% the low reward associated colour. The remainder of the stimuli were rendered in the unrewarded colours from the reward-training task. The previously rewarded colour could appear in any of the three sequence locations with equal probability. Each of the three sequence locations had an equal probability of being probed.

6.4.3 Results

The data from the working-memory task were analysed using a mixture model (W. Zhang & Luck, 2008). The response from each trial is expressed as the difference in response orientation from the orientation of the memory item (response error, bounded [-π: π] in radians). Provided subjects are doing better than chance, this distribution has a peak around zero error. Representative data from one subject is shown in Figure 6-16.

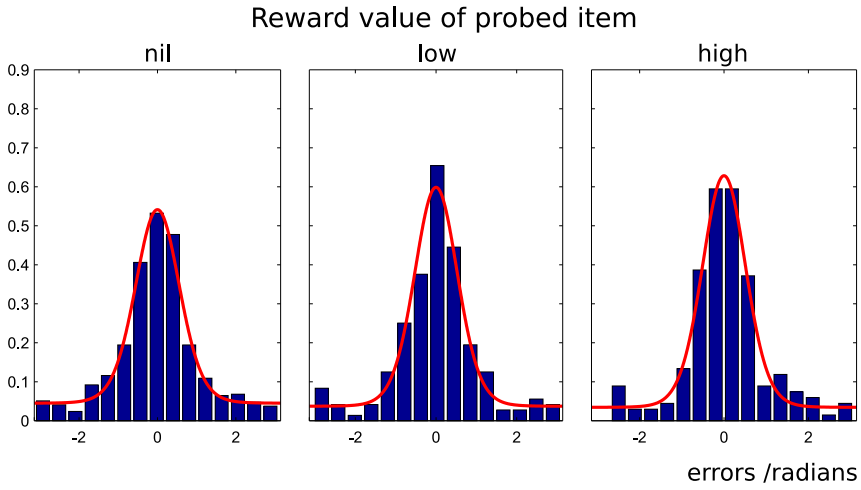


Figure 6-16 **Illustrative response-error histograms for one subject**. This figure collates data from across all three sessions, but splits the data by reward value of the probed item. The blue bars represent the histogram of response errors. The red lines are the mixture distributions.

The mixture model fits the distribution of response errors as a mixture of a Von Mises distribution, representing the precision with which items are stored in memory provided they have been encoded, and a uniform guess distribution for trials in which the probed item was not encoded. The fit is captured in two parameters: the probability of guessing on any given trial (p(guess)) and the precision of the Von Mises distribution (kappa). The model was fitted using maximum likelihood estimation constrained by prior distributions on the two parameters, the priors ensuring stable fits for conditions with smaller numbers of trials (see Appendix 9.1 for further details).

Trials were binned by reward value of the probed item (nil, low, high), and the experimental session from which they came (#1, #2, #3). As the sequence location of the probed item and the sequence location of the rewarded item were randomly and independently assigned, the nil-reward item was probed on 2/3 of trials. Nil reward trials were therefore randomly subsampled before model fitting to ensure the same number of trials were fitted for each reward/session bin. Precision values for two subjects were zero for the majority of bins, and inspecting the error histograms confirmed that these subjects were performing at chance (flat error distributions).

They were therefore excluded from the analysis. Precision and guess rate averaged over the remaining subjects (N=20) are plotted in Figure 6-17.

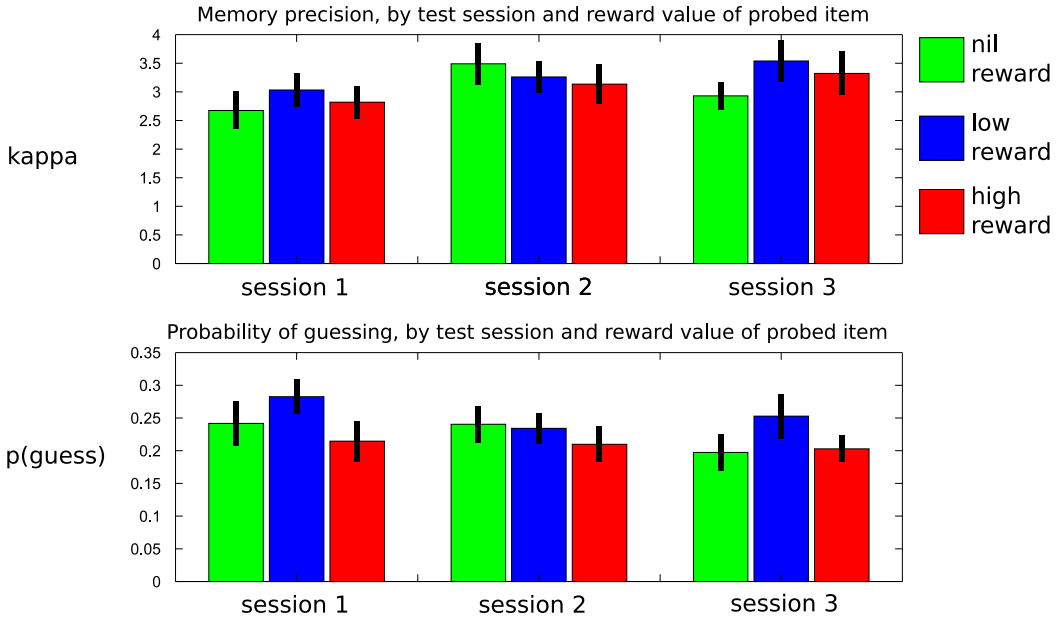


Figure 6-17 **Precision and guess rate by session and reward value of probed item**. Guess rate is consistently slightly lower for high reward value items than for the low reward value items across all three sessions. There is no consistent effect of reward value upon precision. Error bars are ± 1 SEM.

Nil reward items differ from the high and low reward items in task relevance in the reward training task, whereas the rewarded items were equated for task relevance and differed only in their reward association. The analysis was therefore restricted to high and low value items only. A repeated-measures ANOVA with factors of Reward (two levels) and Session (three levels) revealed a main effect of reward upon guess rate ($F(1,19)=7.95$, $p = 0.011$). There was no main effect of Session ($F(1.65,38) = 0.758$ $p = 0.454$), or Session*Reward interaction ($F(1,38) = 0.382$, $p = 0.685$). A separate ANOVA was run for precision, and there was no evidence for a main effect of either factor, or an interaction between the factors.

The mean difference in guess rate and kappa between high and low reward items is plotted in Figure 6-18.

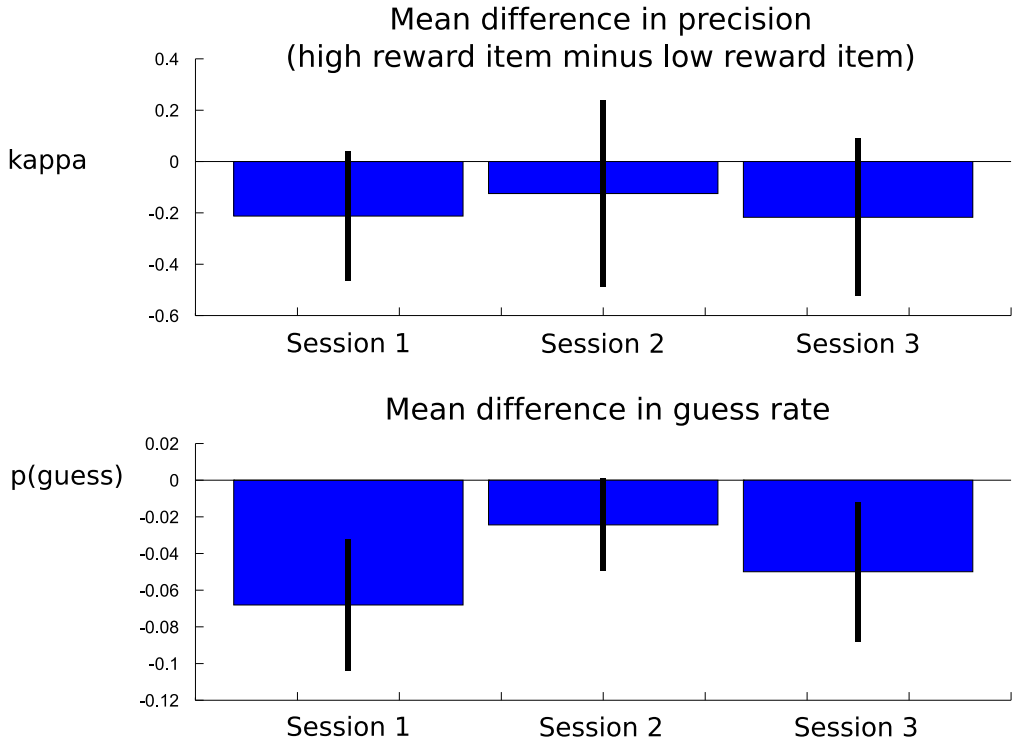


Figure 6-18 **Mean of subject-wise differences in precision and guess rate, low and high reward value probed item.** The guess rate was consistently lower for high reward value probed items than for low value probed items. Error bars are ± 1 SEM.

In each session the probability of guessing is slightly lower when subjects were probed on high reward value items than when they were probed on low reward value items. This would be consistent with an item-wise bias for working memory encoding, but given the sequential nature of the task, is also consistent with a purely temporal selection mechanism (discussed above).

If the effect were dependent on a brief window of enhanced encoding into WM, driven by the reward association of the sequence item, this might 'spill over' to the subsequent item, and facilitate the encoding of that item also. This was tested by analysing only trials in which the second and third sequence locations were probed. These were split into two groups: trials in which the probed item immediately followed the presentation of a low value item (at the first or second sequence location) and trials in which the probed item immediately followed the presentation of a high value item.

These two groups were equated for mean sequence location of the probed item (50% second location, 50% third location) and differed only in the reward value of the preceding item (which was in both cases a target item during reward training). The mean of subject-wise differences in kappa and guess rate are plotted in Figure 6-19.

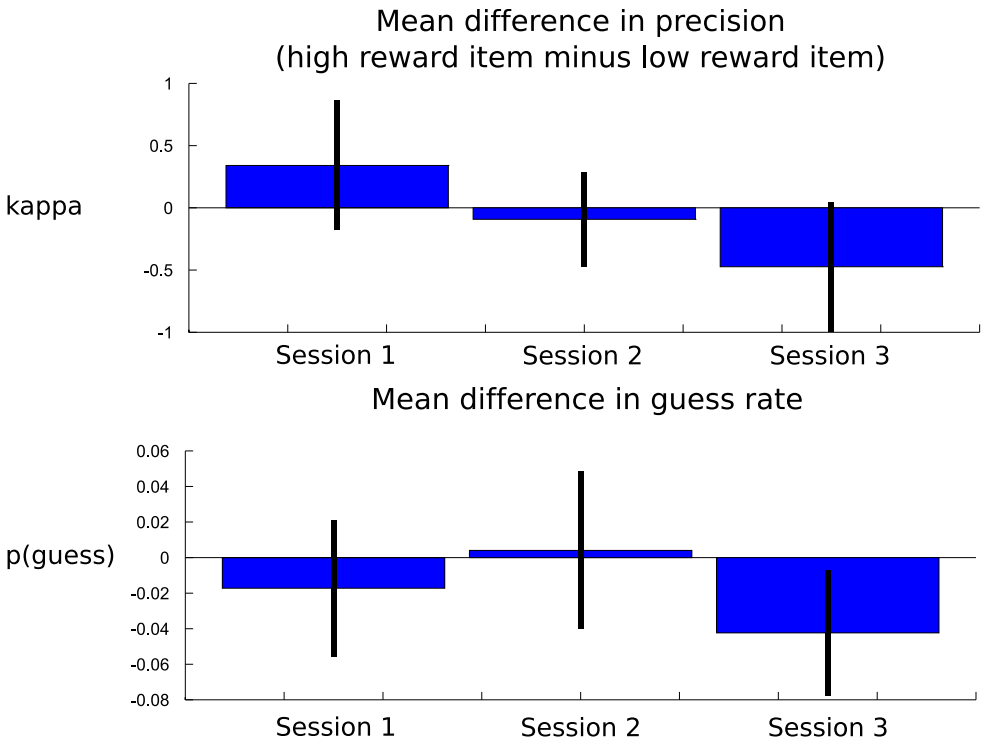


Figure 6-19 **Mean of subject-wise difference in guess rate for probed items immediately following a high or low value item.** There was no consistent effect of prior item value on either parameter. Error bars are ± 1 SEM.

A repeated-measures ANOVA (factors of Session, 3 levels, and Reward, two levels) revealed no main effects of session or reward value of the preceding item in the sequence (or interaction) for either kappa or guess rate.

6.4.4 Discussion

In experiment 3, items were temporally segregated at encoding. The stimuli were presented in rapid succession (SOA 300ms). This speed of presentation was chosen because it is close to the time-course of phasic dopamine (Schultz, 1998). The effect of

the reward value of the memory items was investigated both for encoding of the item that carried the reward association, but also for encoding of the subsequent item, to see whether a time-specific encoding 'boost' triggered by reward associated item might carry over to the subsequent item in the sequence.

The data were consistent with an item-wise bias for encoding of high-reward items, but there was no evidence for carry-over of the reward effect to the subsequent item. This suggests that, with temporal segregation, presentation of a reward-associated stimulus can boost encoding of that stimulus, but that this effect is short lived, and does not carry over to the subsequent item presented 300ms later. This temporal specificity of the encoding bias is interesting given the effects observed in experiment 1 and 2 – it suggests that the boost of encoding seen for high reward value arrays in those experiments was not simply due to a general alerting or motivating effect associated with high reward trials (boosting performance through greater effort/engagement on those trials), but might correspond to a temporally specific effect at encoding. The data from all three experiments presented here are discussed further below (section 6.5).

6.5 Discussion

The results of experiment 1 and 2 were consistent with the idea that encountering an item with a prior reward association facilitates working memory encoding, but in an otherwise unselective manner. This benefit might have reflected a mechanism that briefly facilitates encoding when a rewarded item is encountered, opening an 'updating window' for working memory (Braver & Cohen, 2000). In experiment 3, this idea was investigated by presenting working memory items in rapid succession. There was a modest benefit in encoding for reward associated items: subjects guessed less often when they were probed on an item carrying the high-reward associated colour. However, at the fairly short SOA used here (300ms) there appears to be little carry-over of the reward benefit to encoding from one stimulus to the next in the sequence. If the reward association confers an encoding advantage to rewarded items by virtue of facilitating encoding over a brief time window, and item selectivity is gained by virtue of this gating event being tied to sequential sampling of the environment, we might expect the time window over which encoding is facilitated to be quite short (on the order of the time spent fixating between saccades when viewing scenes). Consistent with this model, the data presented here suggest an encoding advantage lasting at most a few hundred milliseconds.

Phasic dopamine signals ramp up 50-110ms after the trigger stimulus, and last another 100-200ms (Schultz, 1998). This time-course is of the same order as the time-course of the attentional blink (Shapiro et al., 1997), which is most effective at reducing T2 visibility when the 'blinked' target follows the first target at 200-300ms, and is not very effective when the T1:T2 interval is less than 100ms. As discussed above, Raymond et al. have shown that rewarded stimuli can escape the attentional blink (Raymond &

O'Brien, 2009), and Slagter et al. (2012) have linked attentional blink magnitude to dopamine receptor binding of [F-18]fallypride in the striatum, using PET. Interestingly, rodent studies have revealed that an oscillation at 4Hz synchronizes VTA, PFC and hippocampus (Fujisawa & Buzsáki, 2011). It is tempting to speculate that the dopamine system, striatum and prefrontal cortex might run on a cycle with a period of a few hundred milliseconds, which gates items into working memory in time with sensory sampling.

This hypothesis goes well beyond the scope of the preliminary data presented here, which are at most consistent with a sequential gating mechanism. In order to better characterise this mechanism behaviourally, experiment 3 could be repeated with a much faster rate of stimulus presentation, in order to characterise the time-course over which reward associations facilitate encoding. This would likely require switching back to an object-memory task, as used in experiment 1, as even at load three many subjects found the orientation memory task difficult, guessing on a large proportion of trials. A whole-report rapid serial visual presentation (RSVP) task using shape stimuli might be suitable (see below).

There was no statistical evidence for an item-wise bias in experiments 1 or 2, but the data from experiment 2 were suggestive of an item-wise bias at a trend level. It is possible that this experiment simply lacked the power to detect an item-wise effect with 20 subjects. Because of the multiple sessions of data acquired for each subject for experiment 2 (in order to boost the within-subject potency of the manipulation, and allow characterization of any slow learning effects across sessions) this experiment was rather time-consuming to run, but the lack of an effect of session suggests that at least with regard to learning, multiple sessions of reward training were not necessary. In order to address (a) the time-course of the encoding benefit, (b) the possibility that

there was insufficient power to detect an item-wise bias in experiments 1 and 2, further work should combine serial/parallel presentation, and present items with a rapid time-course. An RSVP task in which either one or two items are presented in each frame would allow for this: rapid serial presentation would allow the time-course of reward biasing of encoding to be characterized, and having some displays in which items are shown in parallel would permit a specific test of whether the encoding bias was item-wise or purely temporal. Whole report for the sequence would be an appropriate dependent measure. A natural further step would be to investigate the interaction between reward biasing of encoding and cued biasing of encoding. In the experiments presented here, the task-irrelevance of reward value in the working memory task was relied upon to prevent strategic use of reward value to bias encoding. This is a weak control over strategic biasing: subjects might conceivably consciously bias encoding toward reward associated items, even though there is no strategic reason to do so. To obtain better control over strategic influences, the interaction between item-wise reward history and prospective/retrodictive cues for encoding could be investigated. Probabilistic cues indicating which item is likely to be probed could be either set against or in concert with prior reward associations, trial by trial. If the reward value of uncued items still boosted encoding of those items despite a cue biasing encoding towards other items, this would strengthen the case that reward history biases encoding by a non-strategic mechanism.

However, the pattern of results observed in the experiments presented here was not suggestive of an item-wise bias, instead providing empirical support for a brief but non-specific boost of encoding following presentation of a reward-associated item. This latter account is incompatible with strategic top-down biasing of encoding based on reward value, as this would be expected to affect only reward associated items, whereas

the temporal mechanism benefits both high and low value items indiscriminately, if the net reward value of an array is high.

The experiments presented here establish preliminary support for a time-specific gating effect of reward-associated stimuli. The work was motivated by theories of dopaminergic and striatal function, but behavioural experiments cannot directly inform the neural hypothesis. Event-related fMRI focussing on striatal, SNpc/VTA and prefrontal activations during the working memory phase of our experiments would be the most straightforward way to investigate the neural underpinnings of the effect of reinforcement on encoding biases for working memory in human subjects. M/EEG would provide complementary evidence on the precise latency and localization of cortical responses modulated by reward value, which would help establish the degree to which encoding benefits might be driven by changes in low-level visual salience (Hickey, Chelazzi, & Theeuwes, 2010), or by additional downstream fronto-striatal mechanisms controlling encoding for working memory, as hypothesised here. However, more robust behavioural effects of reward history would be required for such experiments to be viable. Finally, in non-human primates and rodents, the activity of dopamine neurons can be recorded directly. It would be interesting to investigate the temporal coupling between phasic dopamine release and exploratory eye movements in primates, or sniffing behaviour in rodents, given the possibility that dopaminergic gating is chained to the time-course of active sampling of the environment.

In summary, a series of experiments was performed to behaviourally test the hypothesis that prior reward history can alter the probability that items are encoded into working memory. Based on two mechanistic theories, there were two alternative hypotheses as to the nature of this effect. One theory (Braver & Cohen, 2000) predicts a temporally specific, but not item specific, encoding boost associated with reward. A second theory

(Hazy et al., 2007) predicts item-specific biases for encoding, dependent on the reward history of those items. The results were broadly consistent with the first, but not the second account. However, the effects observed were statistically modest, and the trend for an item-specific bias in experiment 2 makes it difficult to firmly conclude that the second account was unsupported. Further work is needed to build on the experiments presented here.

7 General Discussion

Chapter Abstract

This general discussion outlines the implications of the current work and the potential for future work, around three themes.

The first theme is that working memory control is dynamic: internal gating mechanisms continually reconfigure memory by selecting and deselecting representations to support moment-by-moment task demands. This is in contrast to a 'standard model' of working memory that emphasises the role of sustained attentional control in maintenance and treats working memory as a static snapshot of perception. In chapter 3 behavioural data were presented that support an output gating account of the retrocue benefit: rather than changing *what* is represented retrocues help efficiently select an item to guide behaviour. In chapter 4, complementary MEG results showed that retrocues only transiently modulate activity in perceptual cortex, as opposed to the more sustained shift in top-down biasing predicted in the standard model.

The second theme is the role of frontal and parietal networks in cognitive control. In chapter 5, the temporal dynamics of network activation were investigated, dissociating the role of two major functional networks in control of working memory. Further investigating the temporal dynamics of network activations and interactions should help establish the functional architecture of cognitive control.

The final theme is the role of basal ganglia and the dopamine system in cognitive control. In chapter 6 this was investigated indirectly, via a reward manipulation, but the behavioural experiments were motivated by neural models of the interactions between dopamine/basal ganglia and PFC. This topic has been explored theoretically, but not empirically. The behavioural data presented in chapter 6 provide preliminary support for a temporal gating hypothesis of working memory encoding, which may be an interesting direction for future work.

7.1 Dynamic access in working memory

7.1.1 Input and output gating as separable factors

The 'standard model' of working memory control sees maintenance as internally directed attention, analogous to perceptual attention (Cowan, 2000), and selection within memory as a shift in that attentional set. A key contention of this thesis has been that mechanisms acting to prioritise an item in memory, allowing it to guide action or on-going cognition (i.e. mediating 'output gating'), are equally important in determining task performance (LaRocque et al., 2014; Olivers et al., 2011). These factors have been studied under the banner of task-switching: Garavan (1998), McElree (2006) and Oberauer (2012) have characterised the reaction time costs of switching between items/tasks. However, these tasks were quite different from the simple feature-memory tasks typical of the visual working memory literature, involving multiplexed mental counters, recall of word-lists, and arithmetic calculations, respectively. It was not clear that the same output-gating considerations would apply to tasks that emphasise storage and comparison of perceptual features. In chapter 3, an in-depth behavioural analysis of the simple visual precision-capacity working memory task used in the MEG study suggested that the same factors do apply. Retrocues altered the probability that an item would be recalled but not the precision with which items were represented, altered reaction time by an additive factor, and eliminated responses based on the wrong memory item. These effects taken together are more consistent with an output gating account of the retrocue benefit than an explanation in terms of biasing competition between memory items (Murray et al., 2013) or protecting memory items from decay (Pertzov et al., 2012).

Working memory capacity is of considerable interest as a metric (Cowan, 2000), given it is strongly correlated with intelligence (Fukuda, Vogel, Mayr, & Awh, 2010b) and scholastic aptitude (Gathercole, Pickering, Knight, & Stegmann, 2004). The memory task used in chapter 3 is typical of the working memory tasks that are used in the literature to establish capacity estimates (Bays & Husain, 2008; W. Zhang & Luck, 2008). The results presented here suggest that such tasks systematically underestimate capacity, under a 'buffer' model of working memory in which capacity strictly means *representational capacity* (i.e. how much information is in principle maintained in mind, ignoring output gating factors). A different perspective would be to view measured capacity as an emergent property of the system, incorporating output-gating (retrieval) factors. This may be a more realistic framework for future work, as contemporary models of working memory maintenance increasingly downplay the role of persistent delay activity in favour of synaptic maintenance accounts (Buonomano & Maass, 2009; Mongillo et al., 2008) that blur the distinction between short-term and long-term storage, and re-emphasise the importance of cues and contextual bindings in retrieval (Nairne, 2002; Oberauer, 2013).

The attraction of the classic capacity concept was its generality: a single number might capture a property that explains cognitive performance across a wide range of domains (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). However, more usefully generalizable measures might be obtained by breaking down the different selection factors determining working memory performance (which all load on the classic capacity estimate). Such measures should be more robust to changes in task (although their contribution to performance may vary depending on task). For example, a distinction has recently emerged between the *precision* of stored memories and the *number* of such representations that can be recalled (Bays & Husain, 2008; W. Zhang & Luck, 2008). In terms of individual differences, these factors appear to be independent:

Awh et al. (2007) and Fukuda et al. (Fukuda et al., 2010b) both found that the quantity and resolution of representations in memory were independent across subjects, with Fukuda and colleagues (2010b) additionally finding that only the quantity of objects that could be retained correlated with fluid intelligence.

This thesis presented evidence that output-gating factors play an important role in determining the (supposed) capacity limit in simple working memory tasks. Efficacy of selection at encoding (input gating) and output gating might constitute dissociable factors loading on memory performance. The current study was not sufficiently powered for a full factor analysis, but a large-scale study of individual differences in the efficacy of prospective and retrospective cueing using a task similar to the one presented in this thesis[31] might allow for such an analysis. It would be particularly interesting to also gather data from a paradigm such as the classic stop-signal task indexing inhibitory motor control (Verbruggen & Logan, 2008), or to design and implement a novel motor task indexing efficacy of action selection, to see whether action control and output gating of working memory load on a common factor, as might be predicted by fronto-striatal models of WM control (Chatham et al., 2014; Hazy et al., 2007).

7.1.2 Re-interpreting neural correlates of mnemonic selection

The functional interpretation of the retrocue benefit defended in this thesis requires a re-appraisal of neural evidence that common control networks are recruited by perceptual attention and mnemonic selection (Nee & Jonides, 2009; Nobre et al., 2004), and that both prospective and retrospective cues modulate activity in perceptual cortex (Kuo et al., 2014; Munneke et al., 2012; Sligte et al., 2009). These observations have in the past been thought to imply that perceptual memories are maintained in perceptual

[31] Or a set of tasks spanning modalities – auditory, visuospatial, linguistic – but with similar cueing manipulations, to help factor out modality-specific capacity limits and isolate the executive factors of interest.

cortex (Pasternak & Greenlee, 2005) and that top-down attentional control acts on these memories just as it would act on perceptual input (Gazzaley & Nobre, 2012). Recently, a new view has emerged: only the activated or 'focussed' item in memory drives top-down signalling (Olivers et al., 2011), such as can be decoded from the fMRI signal using MVPA (LaRocque et al., 2014). Rather than constituting maintenance, this activation in fact reflects active representation of the item in the focus of attention. The results from the MEG experiment presented in this thesis (chapter 4) suggest an even more conservative view than this. Whilst perceptual cortex is recruited in an item-specific manner when a memory item is cued, this recruitment is transient. Lasting ~500ms in the experiment presented here, it declines to baseline about 500ms before the memory probe is presented. This observation is most consistent with a 'refreshing' function, briefly reviving perceptual content associated with the cued item in memory (Raye et al., 2008).

What function does this refreshing event serve? It is conceivable that the activation was epiphenomenal - subjects automatically visualized the cued stimulus after a retrocue, but this mental imagery did not in of itself facilitate the subsequent comparison with the probe item. This thesis did not employ a causal manipulation (such as TMS) that could address this question directly. However, as well as correlating with cue direction, alpha power in sensory cortex correlated with subsequent behavioural performance. To quote Edward Tufte, "Correlation is not causation but it sure is a hint".

Discovering the role of this reactivation amounts to discovering how the perceptual comparison between the memory item and probe item is coded. Simple comparisons of this type are a subset of the much wider and richer set of operations we can perform on seemingly image-like cognitive representations. Two related proposals consistent

with *transient* refreshing activity are considered here. One is a 'flexible workspace' theory: it suggests that special neural resources and codes exist in the brain to represent cognitive content in a format that can be flexibly manipulated (Baars, 2005; Danilova & Mollon, 2003). Refreshing is part of the transformation of representations encoded in the labelled-lines of perceptual cortex from a retinotopic representation to this more flexible format. The second proposal - the 'processing rule' or 'connectivity state' theory – suggests that refreshing helps set up new task-specific input-output mappings in the brain, in preparation for the probe stimulus (Stokes et al., 2013).

This second hypothesis is based on the idea that rather than encoding persistent 'snap-shot' activity states of previous perceptual input, the function of working memory is to remap future responses based on prior input (Stokes & Duncan, 2013). Taking an example from the MEG task, imagine an identical probe stimulus presented on two different trials. On one trial, the corresponding memory stimulus was slightly clockwise of the probe stimulus and the appropriate response is therefore to press the left-hand button, and on another trial it was slightly anticlockwise and the appropriate response is to press the right-hand button. The contention is that working memory consists in a temporary change in network *connectivity* that encodes this mapping rule – 'switching the points' on new input, rather than storing a static representation of the memory item. The representation can be inferred from the new input-output mapping. This is related to the idea that memory is stored 'silently' as temporary synaptic weight changes (Mongillo et al., 2008)[32]. With four items in mind, there are four possible input-output mappings, depending on the location of the probe item. Refreshing a single item following a retrocue might be part of a reconfiguration of temporary network connectivity in preparation for the probe item, setting up a single unambiguous

[32] But is worth noting that neither idea implies the other. Maintained activation in context units could switch input-output mappings, as proposed by Mante et al (2013). Conversely, synaptic storage could mediate representational memory, as is commonly assumed to be the case for long-term memory.

input-output mapping that will lead to the appropriate behavioural response when the probe item is presented. This could involve changes in network connectivity within frontal cortex (Buschman, Denovellis, Diogo, Bullock, & Miller, 2012), but also changes in connectivity between flexible prefrontal resources (Duncan, 2001) and labelled lines in perceptual cortex. The idea is attractive because it accounts not only for the representational format of the memory, but also the mechanism by which the comparison with the probe item is performed and the behavioural output is produced.

A case can also be made for a special representational format or central code in the brain dedicated to flexible cognitive manipulation (the 'flexible workspace' theory). This is similar to the above account in that it proposes that refreshing allows a representation access to a prefrontal resource, but rather than reconfiguring input-output mappings (Mante, Sussillo, Shenoy, & Newsome, 2013; Stokes et al., 2013), the prefrontal buffer is suggested to explicitly *represent* the properties of the item in memory. For example, Danilova and Mollon (2003) lean on the simple case of comparing two simultaneously-presented on-screen items (Gabor stimuli) to make the following argument:

1. Empirical data show that there is little or no accuracy or reaction time deficit when the on-screen distance between items to be compared (on spatial frequency, orientation or contrast) is increased.
2. Whilst comparisons between stimuli very close together on the screen could plausibly be based on the output from local comparators within visual cortex (i.e. a third neuron directly comparing activity in two labelled-line feature-selective neurons), the number of comparators required to connect feature-neurons for every possible feature comparison between all locations in the visual field would be prohibitively large.

3. Comparison must therefore depend on some sort of central code that can represent object features independent of 'labelled lines' in perceptual cortex.

This central buffer might be extremely capacity limited. For example, it might correspond to the active memory item proposed by Olivers et al. (Olivers et al., 2011). With four items in memory, a retrocue might allow one of those items to be 'uploaded' to the central buffer. This might involve reactivating the labelled-line sensory neurons associated with that item, in order to access its features, explaining the transient retinotopic re-activation of sensory cortex observed in the MEG data. Whilst the idea of 'free-floating' representations in prefrontal cortex runs counter to our intuitions about how the brain anchors representations in terms of receptive fields, this coding scheme is how the internet works – one wire can carry an infinite variety of signals depending on addressing and header information associated with information packets. Danilova and Mollon (2003) coin the term 'cerebral bus' to refer to such a coding scheme as it might exist in the brain. This suggestion bears comparison with the idea of a 'global workspace' through which privileged representations could gain access to a wide range of brain systems during cognitive manipulation (Baars, 2005).

How could these ideas be tested with available methods? Multivariate pattern analysis is a powerful technique that when applied to high-temporal resolution MEG data (which also has rich spatial content) could help elucidate the role of sensory cortex in working memory operations (LaRocque et al., 2014). Activity patterns in sensory cortex can be decoded by comparing them to training patterns established by presenting physical stimuli. A simple version of this approach was used in the current thesis, in which the response to a probe item was used to train a classifier that could decode the direction of attention. With the appropriate training set, the orientation of the grating stimuli themselves can be successfully decoded from the MEG signal. By decoding orientation

information from visual cortex before and after cues, and in the run-up to probe stimuli, the temporal dynamics of top-down reactivation or attentional templates could be established. There are many conceivable experimental designs that could exploit this technique; one example is outlined here.

In order to establish whether top-down reactivation reflects prior perceptual input or preparation for upcoming probe stimuli (or both) the relationship between memory item and probe stimulus can be manipulated. In delayed paired-associate tasks, a memory stimulus instructs the observer that a pre-learnt paired associate will be the subsequent target item (in contrast to the classic delayed-match-to-sample task, in which the cue and target item are one and the same). These tasks have been used to reveal prospective coding for the upcoming paired-associate item in prefrontal cortex (around principal sulcus) in monkeys (Rainer, Rao, & Miller, 1999; Stokes et al., 2013). Adapting such a paradigm for use in conjunction with a decoding approach might allow the role of visual cortex in human working memory performance to be better characterized. A classic paired associate task would be restricted to looking at categorical decoding (e.g. Lewis-Peacock & Postle, 2012), but a variation in which a transformation on a parametric memory dimension (e.g. orientation) is cued between memory array and probe might be more powerful. For example, imagine using the same task as presented in this thesis, but with a retrocue that either instructs subjects to prepare to compare a probe stimulus to the cued item, or to compare a probe stimulus to the cued item, plus a transformation (e.g. +45 degrees rotation). If the refreshing event observed in this thesis reflects pure retrieval (and perhaps representation in a prefrontal buffer as suggested above), then only the original stimulus orientation should be decodable from sensory cortex during refreshing. If it represents the configuration of a new input-output mapping, the refreshing event might allow the new (transformed)

orientation to be decoded, or it might even be possible to track the time-course of a transformation from the original to the expected orientation (Shepard & Metzler, 1971).

7.2 Control networks

In chapter 5 an fMRI meta-analysis identified a set of frontal and parietal regions involved in retrocueing that in aggregate constitute much of the 'multiple demand' network as described by John Duncan (Duncan, 2010; 2013). Duncan notes that this set of regions is ubiquitously involved in a wide range of cognitive tasks (Dosenbach et al., 2006; Fedorenko et al., 2013), and suggests that they are responsible for coordinating the sequential attentional episodes that make up cognition (Duncan, 2013). This high-level view does not discriminate between sub-networks or regions, but in chapter 5 an fMRI meta-analysis demonstrated that precues and retrocues produce different patterns of activation within the multiple demand network. This dissociation between cue types mapped to a proposed distinction between a fronto-parietal and a cingulo-opercular control network, made by Dosenbach and colleagues (Dosenbach et al., 2007; 2008) on the basis of resting state fMRI correlations between ROIs capturing task-general brain regions (Dosenbach et al., 2006). Broadly, precues activated the fronto-parietal network whilst retrocues additionally activated the cingulo-opercular network (with some additional differences in the dorsal/ventral location of the activations in premotor cortex). This dual-network spatial template was used to characterise the time-course of control network activation in source space MEG induced-response data. The MEG replicated the predicted spatial dissociations between precues and retrocues, but also furnished their activation time-courses. Following retrocues, the fronto-parietal and cingulo-opercular time-courses were dissociated: the fronto-parietal sites activated immediately following the cue, whereas

the cingulo-opercular sites co-activated later in the trial. Within the cingulo-opercular network, the pre-SMA was (tentatively) dissociated from the bilateral fO, in that it also showed activations immediately post-cue.

These data motivated the following hypothesis: whilst fronto-parietal sites will be recruited whenever representations in sensory cortex are activated, biased or retrieved as part of task performance (a function spanning preparatory attention and working memory retrieval, and which would likely also encompass imagery and cognitive manipulation such as mental rotation) the cingulo-opercular sites have a 'downstream' role and are instead involved in gating representations to guide action, particularly in the presence of competing candidates (Banich, 2009). It will be interesting to see whether this proposed functional segregation bears up in future work.

The approach used in the current thesis (ROI analysis of source-space MEG, combined with cross-temporal sensor-space decoding to establish state time-courses) may be useful to clarify the role of frontal sites in future work. As argued in chapter 5, since cognitive processing seems likely to be highly dynamic, consisting of transient sequential episodes (Duncan, 2013), methodologies combining the ability to spatially dissociate activation in different prefrontal sites and map these activations with high temporal resolution will be necessary to discover the structure of cognitive control. Besides MEG, intra-cranial EEG or electrocorticography (ECoG; only available with the participation of neurological patients) has these properties, and will be useful to validate findings from MEG. The target will be to establish typical patterns of network activation or regional involvement, in time as well as space, which correspond to control events. For example, in the current thesis there was some (weak) evidence that the pre-SMA had a more general role in cognitive control than did the fO, responding

bi-phasically following retrocues[33]. In the past this region has been associated with the detection of cognitive conflict (Yeung, 2013), and with triggering appropriate responses in other control sites. A more general form of this hypothesis is that the pre-SMA coordinates activations in other control sites: in our task, perhaps first triggering fronto-parietal activation and then driving an activation in fO. The prediction of this (illustrative) hypothesis would be that pre-SMA activations should generally precede other control network activations, and the pre-SMA should respond particularly strongly following cues that trigger a new sequence of operations, as a correlate of sequence-planning.

In developing new analysis tools to address these questions, improving temporal resolution will be an important goal. MEG has very high inherent temporal resolution, but it is still limited by the problem of averaging over within- and between-subject variability in timing. A group-level analysis identifies the 'peaks' of the temporally overlapping events across subjects, by analogy with the way that spatial activation clusters across subjects in fMRI can be quite variable in location, forming a 'Venn diagram' pattern. Taking this cross-subject variability in time-course into account would increase both the temporal specificity of the observed activations, allowing more fine-grained temporal order effects to emerge, and also increase the sensitivity of analyses probing interactions between different brain regions[34]. This kind of analysis would involve first characterising differences in subjectwise time-courses. For example, Hidden-Markov Models (HMMs) are a useful tool to identify the evolution of

---

[33] And also responding weakly following precues. A caveat is that activations on the midline at the time of event-related responses need to be treated with caution as they may reflect the mislocalization of highly correlated lateralized signals (see chapter 2).

[34] A number of interaction analyses were attempted as part of the work performed in this thesis (e.g. task-locked changes in coherence; task-locked changes in orthogonalised power correlation; lagged power correlations; task-locked Granger causal interactions) but were not reported as they did not yield convincing measures of inter-regional interactions that could then be compared between task conditions. Only the 'static' orthogonalised power correlation measure yielded meaningful interaction measures, and this approach collapsed over time and task condition.

sequential states in brain data (Baker et al., 2014). One proposed method uses a variant on the HMM to infer a 'canonical' time-course across a range of noisy observations (Listgarten, Neal, Roweis, & Emili, 2004). Individual time-series can then be warped back onto this canonical time-series. Developing such a method for MEG data to improve the temporal resolution of the subject- and group-level results was beyond the scope of this thesis, but it should in principle be a way to further capitalize on the temporal resolving power of MEG.

A separate point raised in chapter 5 (but not discussed in depth) was the association of *increases* in low-frequency (alpha/beta band) spectral power in frontal regions with cue-induced activations. This contrasts with the prevailing assumption that in parietal and sensory cortex, *decreases* in spectral power in the alpha/beta range are associated with active processing, and increases are more likely to be observed in the theta/gamma bands (Hanslmayr et al., 2012). This has also been suggested to apply for frontal cortex – for example, Hanslmayr and colleagues (2011) used combined EEG-fMRI to show that in a long-term memory paradigm, beta-frequency power decreases in left inferior prefrontal cortex during encoding were predictive of better subsequent memory performance, and correlating the BOLD signal with beta power showed that there was more BOLD activity in PFC when beta power was lower. This seems inconsistent with the beta-frequency induced responses associated with activation in prefrontal cortex in the current thesis. However, the literature as a whole is not consistent on this point. Recordings from monkey PFC contained increases in both beta and alpha power associated with cue-triggered rule implementation and inhibition (respectively) in a contextual rule-switching task (Buschman et al., 2012). Future work will need to clarify whether low frequency oscillations have the same relationship to cortical activation in frontal as in parieto-occipital regions (e.g. by running tasks evoking robust frontal activations in both MEG and fMRI, and comparing patterns of oscillatory

responses in the MEG to BOLD activation patterns in the fMRI). One important consideration is that the transient bursts of activation seen in the low frequency range in frontal cortex in the current study might not have reflected true oscillations. Any perturbation in the time-domain signal, including evoked-responses[35] will have an associated pattern in the frequency domain, with a frequency profile depending on its time-course. The induced activations observed in frontal sites in chapter 5 may have reflected such 'bumps' in activation, as opposed to true changes in oscillatory power. The spectral variability of the responses (e.g. dlPFC activated at alpha following retrocues, but at theta following precues) is consistent with this idea.

7.3 Role of basal ganglia and dopamine in memory control

Chapter 6 investigated behavioural predictions of theoretical models of the putative memory-gating interaction between dopamine, basal ganglia and PFC (Braver & Cohen, 2000; Hazy et al., 2007). This was addressed by examining the effects of reward associations of memory stimuli, which both theories predict should modulate gating at encoding. The results supported a temporal-gating account in which observing an item with a high reward association briefly boosts encoding for working memory. This was in line with an older theory in which phasic dopamine controls working memory gating (Braver & Cohen, 2000) but not the theory that cortico-striatal loops can preferentially gate specific items into working memory (Hazy et al., 2007). A sequential version of the working memory task suggested the gating window did not last more than 300ms, which is consistent with the time-course of phasic dopamine (Schultz, 2013). It was proposed that whilst in the rather artificial lab context in which fixation was controlled

[35] Ignoring for the moment the hypothesis that evoked-responses consist in phase-alignment of on-going oscillations (Klimesch, Sauseng, Hanslmayr, Gruber, & Freunberger, 2007).

and stimuli were presented only briefly a temporal updating window did not yield item-specific updating, when paired with sequential shifts in attention this might reflect an effective mechanism for controlling access to working memory. However, as is discussed in detail in chapter 6, the effect sizes associated with reward associations were modest, and further work is necessary to replicate and extend these preliminary findings.

Rodent data suggests that VTA, PFC and hippocampus are synchronized at 4Hz (Fujisawa & Buzsáki, 2011) whilst rats make memory-guided decisions. Recently, Guitart-Masip and colleagues (2013) showed using MEG that MTL and PFC were synchronized in humans at 4Hz during a decision making task based on episodic memory. During working memory tasks, prominent theta oscillations are often observed on the frontal midline in EEG (Hsieh & Ranganath, 2014), but their functional role remains a mystery. An influential suggestion is that they coordinate nested gamma bursts representing different memory items (Lisman & Idiart, 1995). Rather than just coordinating maintenance, a theta rhythm in VTA, hippocampus and PFC might also coordinate gating. A speculative extension of the temporal gating hypothesis would be that working memory updating runs to a slow (theta) clock. This might help coordinate activity in anatomically disparate dopaminergic, striatal and prefrontal sites involved in working memory gating.

7.4 Conclusions and future directions

The common thread in this thesis has been the importance of the temporal dimension in cognitive control. Retrocues are effective because they permit the segregation of mnemonic access and probe processing (chapter 3), access to perceptual cortex is transient (chapter 4), control networks activate in a sequential cascade (chapter 5), and

interactions between the dopamine system and prefrontal cortex may mediate temporal gating of working memory contents (chapter 6). The temporal dynamics of control are rapid: several sequential cognitive operations can succeed one another over the course of a second or two. fMRI has been an extremely useful tool to map out control networks, but it is ill-suited to mapping temporal activation patterns. Methodologically, human cognitive neuroscience will likely make greater use of techniques such as MEG and ECoG that combine high spatial and temporal resolution in order to explore the temporal structure of cognitive control. Multivariate analysis approaches such as cross-temporal pattern analysis (King & Dehaene, 2014) and Hidden Markov Modelling to infer hidden brain states (Baker et al., 2014) will be a useful part of our armoury in analysing these temporally-resolved data.

At a theoretical level, we are likely to move from the 'static' theories that currently dominate the attentional and working memory literature positing sustained top-down biases (Desimone & Duncan, 1995) and static patterns of maintenance activity (X.-J. Wang, 2001), towards more dynamic models (Duncan, 2013; Stokes et al., 2013). In tandem with the need to understand the temporal dynamics of cognitive control, we need to better integrate subcortical structures such as the striatum (Hazy et al., 2007), thalamus (Saalmann et al., 2012) and dopaminergic nuclei (Braver & Cohen, 2000) into currently cortico-centric (Fedorenko et al., 2013) models. Here too temporal dynamics may be important – a speculative hypothesis is that slow theta oscillations coordinate activity in these anatomically disparate sites (Fujisawa & Buzsáki, 2011; Guitart-Masip et al., 2013).

*"**Question**: If Control's control is absolute, why does Control need to control?*

***Answe**r: Control… needs time."*

William S Burroughs, 'Ah Pook'

8 References

Adrian, E. D., & Matthews, B. (1934). The Berger rhythm: potential changes from the occipital lobes in man. *Brain*.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annual Review of Neuroscience*, *9*(1), 357–381. doi:10.1146/annurev.ne.09.030186.002041

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. doi:10.1073/pnas.1104047108

Anderson, B. A., Laurent, P. A., & Yantis, S. (2012). Generalization of value-based attentional priority. *Visual Cognition*, *20*(6), 647–658. doi:10.1080/13506285.2012.679711

Anderson, D. E., Vogel, E. K., & Awh, E. (2013). A common discrete resource for visual working memory and visual search. *Psychological Science*, *24*(6), 929–938.

Astle, D. E., Summerfield, J., Griffin, I., & Nobre, A. C. (2011). Orienting attention to locations in mental representations. *Attention, Perception, & Psychophysics*, *74*(1), 146–162. doi:10.3758/s13414-011-0218-3

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal*.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual Working Memory Represents a Fixed Number of Items Regardless of Complexity. *Psychological Science*, *18*(7), 622–628. doi:10.1111/j.1467-9280.2007.01949.x

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208. doi:10.1016/j.neuroscience.2005.08.023

Baars, B. J. (2005). Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. In *Progress in Brain Research* (Vol. 150, pp. 45–53). Elsevier. doi:10.1016/S0079-6123(05)50004-9

Baddeley, A. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society*, *4*(05), 523–526.

Baddeley, A. D., & Hitch, G. J. (1975). Working memory. *The Psychology of Learning and Motivation*.

Badre, D., & Frank, M. J. (2012). Mechanisms of Hierarchical Reinforcement Learning in Cortico-Striatal Circuits 2: Evidence from fMRI. *Cerebral Cortex*, *22*(3), 527–536. doi:10.1093/cercor/bhr117

Baier, B., Karnath, H. O., Dieterich, M., Birklein, F., Heinze, C., & Muller, N. G. (2010). Keeping Memory Clear and Stable--The Contribution of Human Basal Ganglia and Prefrontal Cortex to Working Memory. *Journal of Neuroscience*, *30*(29), 9788–9792. doi:10.1523/JNEUROSCI.1513-10.2010

Baker, A. P., Brookes, M. J., Rezek, I. A., Smith, S. M., Behrens, T., Smith, P. J. P., et al. (2014). Fast transient networks in spontaneous human brain activity. *eLife*, *3*. doi:10.7554/eLife.01867

Banich, M. T. (2009). Executive Function: The Search for an Integrated Account. *Current Directions in Psychological Science*, *18*(2), 89–94. doi:10.1111/j.1467-8721.2009.01615.x

Barch, D. M., & Smith, E. (2008). The Cognitive Neuroscience of Working Memory: Relevance to CNTRICS and Schizophrenia. *Biological Psychiatry*, *64*(1), 11–17. doi:10.1016/j.biopsych.2008.03.003

Barnes, G. R., Hillebrand, A., Fawcett, I. P., & Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, *23*(2), 120–127. doi:10.1002/hbm.20047

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *321*(5890), 851–854. doi:10.1126/science.1158023

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. doi:10.1167/9.10.7

Beckmann, C. F., DeLuca, M., Devlin, J. T., & Smith, S. M. (2005). Investigations into Resting-State Connectivity Using Independent Component Analysis. *Philosophical Transactions: Biological Sciences*, *360*(1457), 1001–1013.

Beckmann, M., Johansen-Berg, H., & Rushworth, M. F. S. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. *Journal of Neuroscience*, *29*(4), 1175–1190. doi:10.1523/JNEUROSCI.3328-08.2009

Bell, C. C. (1989). Sensory coding and corollary discharge effects in mormyrid electric fish. *Journal of Experimental Biology*, *146*(1), 229–253.

Bichot, N. P., & Schall, J. D. (1997). Dissociation of visual discrimination from saccade programming in macaque frontal eye field. *Journal of Neurophysiology*.

Bisley, J. W., Zaksas, D., Droll, J. A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *Journal of Neurophysiology*, *91*(1), 286–300.

Blakemore, C., & Tobin, E. A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research*, *15*(4), 439–440.

Bonnefond, M., & Jensen, O. (2012a). Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. *Current Biology*, *22*(20), 1969–1974. doi:10.1016/j.cub.2012.08.029

Bonnefond, M., & Jensen, O. (2012b). Alpha Oscillations Serve to Protect Working Memory Maintenance against Anticipated Distracters. *Current Biology*, *22*(20), 1969–1974. doi:10.1016/j.cub.2012.08.029

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. doi:10.1016/j.tics.2008.02.009

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. doi:10.1037/0033-295X.108.3.624

Braver, T. S., & Cohen, J. D. (1999). Dopamine, cognitive control, and schizophrenia: the gating model. *Progress in Brain Research*, *121*, 327–349.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. *Control of Cognitive Processes: Attention and Performance XVIII*, 713–737.

Brodmann. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth.

Brookes, M. J., Stevenson, C. M., Barnes, G. R., Hillebrand, A., Simpson, M. I., Francis, S. T., & Morris, P. G. (2007). Beamformer reconstruction of correlated sources using a modified source model. *NeuroImage*, *34*(4), 1454–1465.

Brozoski, T. J., Brown, R. M., Rosvold, H. E., & Goldman, P. S. (1979). Cognitive deficit caused by regional depletion of dopamine in prefrontal cortex of rhesus monkey. *Science*, *205*(4409), 929–932.

Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). Laminar

differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, *108*(27), 11262–11267.

Buonomano, D. V., & Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience*, *10*(2), 113–125. doi:10.1038/nrn2558

Burgess, P. W., Dumontheil, I., & Gilbert, S. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends in Cognitive Sciences*, *11*(7), 290–298. doi:10.1016/j.tics.2007.05.004

Buschman, T. J., Denovellis, E. L., Diogo, C., Bullock, D., & Miller, E. K. (2012). Synchronous Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex. *Neuron*, *76*(4), 838–846. doi:10.1016/j.neuron.2012.09.029

Callicott, J. H. (1999). Physiological Characteristics of Capacity Constraints in Working Memory as Revealed by Functional MRI. *Cerebral Cortex*, *9*(1), 20–26. doi:10.1093/cercor/9.1.20

Camara, E., Manohar, S., & Husain, M. (2013). Past rewards capture spatial attention and action choices. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *230*, 291. doi:10.1007/s00221-013-3654-6

Cauda, F., D'Agata, F., Sacco, K., Duca, S., Geminiani, G., & Vercelli, A. (2011). Functional connectivity of the insula in the resting brain. *NeuroImage*, *55*(1), 8–23. doi:10.1016/j.neuroimage.2010.11.049

Chafee, M. V., & Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *Journal of Neurophysiology*, *79*, 2919–2940.

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal Output Gating during Selection from Working Memory. *Neuron*, (81), 930–942.

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

Chun, M. M. (2011). Visual working memory as visual attention sustained internally over time. *Neuropsychologia*, *49*(6), 1407–1409. doi:10.1016/j.neuropsychologia.2011.01.029

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585–1599. doi:10.1146/annurev.neuro.20.1.25

Cohen, R. A., Kaplan, R. F., Moser, D. J., Jenkins, M. A., & Wilkinson, H. (1999). Impairments of attention after cingulotomy. *Neurology*, *53*(4), 819–819.

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. doi:10.1016/j.intell.2008.01.002

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, *10*(9), 910–923.

Cooper, N. R., Croft, R. J., Dominey, S. J. J., Burgess, A. P., & Gruzelier, J. H. (2003). Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International Journal of Psychophysiology*, *47*(1), 65–74. doi:10.1016/S0167-8760(02)00107-1

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences*, *95*(3), 831–838.

Corbetta, M., & Shulman, G. L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature Reviews Neuroscience*, *3*(3), 215–229.

doi:10.1038/nrn755
Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*(4), 761–773.
Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292–297. doi:10.1038/73009
Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *J Neurosci*, *13*(3), 1202–1226.
Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*(01), 154–176.
Crone, E. A. (2005). Neural Evidence for Dissociable Components of Task-switching. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(4), 475–486. doi:10.1093/cercor/bhi127
Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423. doi:10.1016/S1364-6613(03)00197-9
D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research*, *133*(1), 3–11. doi:10.1007/s002210000395
Danilova, M. V., & Mollon, J. D. (2003). Comparison at a distance. *Perception*, *32*(4), 395–414. doi:10.1068/p3393
David, O., Kiebel, S. J., Harrison, L. M., Mattout, J., Kilner, J. M., & Friston, K. J. (2006). Dynamic causal modeling of evoked responses in EEG and MEG. *NeuroImage*, *30*(4), 1255–1272. doi:10.1016/j.neuroimage.2005.10.045
de Haan, B., Morgan, P. S., & Rorden, C. (2008). Covert orienting of attention and overt eye movements activate identical brain regions. *Brain Research*, *1204*, 102–111. doi:10.1016/j.brainres.2008.01.105
Dehaene, S., & Changeux, J.-P. (2011). Experimental and Theoretical Approaches to Conscious Processing. *Neuron*, *70*(2), 200–227. doi:10.1016/j.neuron.2011.03.018
Derrfuss, J., Vogt, V. L., Fiebach, C. J., Cramon, von, D. Y., & Tittgemeyer, M. (2012). Functional organization of the left inferior precentral sulcus: Dissociating the inferior frontal eye field and the inferior frontal junction. *NeuroImage*, *59*(4), 3829–3837.
Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.
Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, *12*(3), 99–105. doi:10.1016/j.tics.2008.01.001
Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 11073–11078. doi:10.1073/pnas.0704320104
Dosenbach, N. U. F., 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*(5), 799–812. doi:10.1016/j.neuron.2006.04.031
Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473. doi:10.1111/1467-9280.00290
Downing, P., & Dodds, C. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689–703. doi:10.1080/13506280344000446
Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, *2*(11), 820–829. doi:10.1038/35097557

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. doi:10.1016/j.neuron.2013.09.015

D'Ardenne, K., Eshel, N., Luka, J., Lenartowicz, A., Nystrom, L. E., & Cohen, J. D. (2012). Role of prefrontal cortex and the midbrain dopamine system in working memory updating. *Proceedings of the National Academy of Sciences*, *109*(49), 19900–19909. doi:10.1073/pnas.1116727109/-/DCSupplemental

Economo, von, C. F., & Koskinas, G. N. (1925). Die cytoarchitectonic der Hirnrinde des erwachsenen Menschen. *Springer*.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., Compte, A., & Romo, R. (2009). Mechanism for Top-down Control of Working Memory Capacity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6802–6807.

Egner, T., Monti, J. M. P., Trittschuh, E. H., Wieneke, C. A., Hirsch, J., & Mesulam, M. M. (2008). Neural Integration of Top-Down Spatial and Feature-Based Information in Visual Search. *Journal of Neuroscience*, *28*(24), 6141–6151. doi:10.1523/JNEUROSCI.1262-08.2008

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, *30*(9), 2907–2926. doi:10.1002/hbm.20718

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment : Abstract : Nature. *Nature*, *392*(6676), 598–601. doi:10.1038/33402

Erickson, M. A., Maramara, L. A., & Lisman, J. (2010). A single brief burst induces GluR1-dependent associative short-term potentiation: a potential mechanism for short-term memory. *Journal of Cognitive Neuroscience*, *22*(11), 2530–2540.

Ester, E. F., Serences, J. T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *J Neurosci*, *29*(48), 15258–15265.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16616–16621. doi:10.1073/pnas.1315235110

Ferrera, V. P., Rudolph, K. K., & Maunsell, J. H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *J Neurosci*, *14*(10), 6171–6186.

Foxe, J. J., Simpson, G. V., & Ahlfors, S. P. (1998). Parieto-occipital ~10Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport*, *9*(17), 3929–3933.

Frank, M. J., & Badre, D. (2012). Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(3), 509–526. doi:10.1093/cercor/bhr114

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of Oscillatory Neuronal Synchronization by Selective Visual Attention. *Science, New Series*, *291*(5508), 1560–1563.

Fries, P., Womelsdorf, T., Oostenveld, R., & Desimone, R. (2008). The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *J Neurosci*, *28*(18), 4823–4835.

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi:10.1016/j.tics.2009.04.005

Fujisawa, S., & Buzsáki, G. (2011). A 4 Hz Oscillation Adaptively Synchronizes Prefrontal, VTA, and Hippocampal Activities. *Neuron*, *72*(1), 153–165. doi:10.1016/j.neuron.2011.08.018

Fukuda, K., Awh, E., & Vogel, E. K. (2010a). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177–182. doi:10.1016/j.conb.2010.03.005

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010b). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679. doi:10.3758/17.5.673

Fuster, J. M. (1990a). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology*, *64*(3), 681–697.

Fuster, J. M. (1990b). Prefrontal Cortex and the Bridging of Temporal Gaps in the Perception-Action Cycle. *Annals of the New York Academy of Sciences*, *608*(1), 318–336.

Fuster, J. M. (2003). *Cortex and mind: Unifying cognition.* Oxford University Press.

Fuster, J. M. (2008). *The Prefrontal Cortex*. The prefrontal cortex: anatomy.

Fuster, J. M., & Alexander, G. E. (1971). Neuron Activity Related to Short-Term Memory. *Science, New Series*, *173*(3997), 652–654.

Fuster, J. M., & Jervey, J. P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science*, *212*(4497), 952–955.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition*, *26*(2), 263–276. doi:10.3758/BF03201138

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*(1), 1–16.

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

Gibson, J. J. (1977). The concept of affordances. *Perceiving*.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*(3), 496–512. doi:10.1016/S1053-8119(03)00162-9

Gillebert, C. R., Caspari, N., Wagemans, J., Peeters, R., Dupont, P., & Vandenberghe, R. (2012). Spatial Stimulus Configuration and Attentional Selection: Extrastriate and Superior Parietal Interactions. *Cerebral Cortex*, (23), 2840.

Gitelman, D. R., Nobre, A. C., Parrish, T. B., LaBar, K. S., Kim, Y.-H., Meyer, J. R., & Mesulam, M. M. (1999). A large-scale distributed network for covert spatial attention Further anatomical delineation based on stringent behavioural and cognitive controls. *Brain*, *122*(6), 1093–1106.

Goldman-Rakic, P. S. (1995). Cellular Basis of Working Memory. *Neuron*, *14*, 477–485.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic Updating of Working Memory Resources for Visual Objects. *Journal of Neuroscience*, *31*(23), 8502–8511. doi:10.1523/JNEUROSCI.0208-11.2011

Gould, I. C., Rushworth, M. F., & Nobre, A. C. (2011). Indexing the graded allocation of visuospatial attention using anticipatory alpha oscillations. *Journal of Neurophysiology*, *105*(3), 1318.

Griffin, I. C., & Nobre, A. C. (2003). Orienting Attention to Locations in Internal Representations. *Journal of Cognitive Neuroscience*, *15*(8), 1176–1194. doi:10.1093/brain/117.3.553

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *NeuroImage*, *65*, 349–363. doi:10.1016/j.neuroimage.2012.10.001

Guitart-Masip, M., Barnes, G. R., Horner, A., Bauer, M., Dolan, R. J., & Düzel, E. (2013). Synchronization of medial temporal lobe and prefrontal rhythms in human decision making. *The Journal of Neuroscience*, *33*(2), 442–451. doi:10.1523/JNEUROSCI.2573-12.2013

Haber, S. N. (2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy*, *26*(4), 317–330. doi:10.1016/j.jchemneu.2003.10.003

Hadjipapas, A., Hillebrand, A., Holliday, I. E., Singh, K. D., & Barnes, G. R. (2005). Assessing interactions of linear and nonlinear neuronal sources using MEG beamformers: a proof of concept. *Clinical Neurophysiology*, *116*(6), 1300–1313. doi:10.1016/j.clinph.2005.01.014

Haegens, S., Händel, B. F., & Jensen, O. (2011a). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience*, *31*(14), 5197–5204. doi:10.1523/JNEUROSCI.5199-10.2011

Haegens, S., Luther, L., & Jensen, O. (2012). Somatosensory anticipatory alpha activity increases to suppress distracting input. *Journal of Cognitive Neuroscience*, *24*(3), 677–685. doi:10.1162/jocn_a_00164

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011b). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19377–19382. doi:10.1073/pnas.1117190108

Haegens, S., Osipova, D., Oostenveld, R., & Jensen, O. (2010). Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. *Human Brain Mapping*, *31*(1), 26–35. doi:10.1002/hbm.20842

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, *6*. doi:10.3389/fnhum.2012.00074

Hanslmayr, S., Volberg, G., Wimber, M., Raabe, M., Greenlee, M. W., & Bäuml, K.-H. T. (2011). The relationship between brain oscillations and BOLD signal during memory formation: a combined EEG-fMRI study. *Journal of Neuroscience*, *31*(44), 15674–15680. doi:10.1523/JNEUROSCI.3140-11.2011

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

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *362*(1485), 1601–1613. doi:10.1098/rstb.2007.2055

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, *139*(1), 105–118. doi:10.1016/j.neuroscience.2005.04.067

Händel, B. F., Haarmeier, T., & Jensen, O. (2011). Alpha oscillations correlate with the successful inhibition of unattended stimuli. *Journal of Cognitive Neuroscience*, *23*(9), 2494–2502. doi:10.1162/jocn.2010.21557

Henseler, I., Krüger, S., Dechent, P., & Gruber, O. (2011). A gateway system in rostral PFC? Evidence from biasing attention to perceptual information and internal representations. *NeuroImage*, *56*(3), 1666–1676.

doi:10.1016/j.neuroimage.2011.02.056
Herrmann, C. S. (2001). Human EEG responses to 1-100Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, *137*(3-4), 346–353. doi:10.1007/s002210100682
Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward Changes Salience in Human Vision via the Anterior Cingulate. *Journal of Neuroscience*, *30*(33), 11096–11103. doi:10.1523/JNEUROSCI.1026-10.2010
Higo, T., Mars, R. B., Boorman, E. D., Buch, E. R., & Rushworth, M. F. S. (2011). Distributed and causal influence of frontal operculum in task control. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(10), 4230–4235. doi:10.1073/pnas.1013361108
Hikosaka, O. (2005). Basal Ganglia Orient Eyes to Reward. *Journal of Neurophysiology*, *95*(2), 567–584. doi:10.1152/jn.00458.2005
Hilgetag, C. C., Théoret, H., & Pascual-Leone, A. (2001). Enhanced visual spatial attention ipsilateral to rTMS-induced "virtual lesions" of human parietal cortex. *Nature Neuroscience*, *4*(9), 953–957. doi:10.1038/nn0901-953
Hillebrand, A., & Barnes, G. R. (2002). A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, *16*(3), 638–650.
Hillebrand, A., & Barnes, G. R. (2005). Beamformer Analysis of MEG Data. In *Magnetoencephalography*. International review of neurobiology.
Hipp, J. F., Hawellek, D. J., Corbetta, M., Siegel, M., & Engel, A. K. (2012). Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, *15*(6), 884–890. doi:10.1038/nn.3101
Hollingworth, A., & Maxcey-Richard, A. M. (2013). Selective maintenance in visual working memory does not require sustained visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 1047–1058. doi:10.1037/a0030238
Houtkamp, R., & Roelfsema, P. R. (2006). The effect of items in working memory on the deployment of attention and the eyes during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 423–442. doi:10.1037/0096-1523.32.2.423
Hsieh, L.-T., & Ranganath, C. (2014). Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. *NeuroImage*, *85*(P2), 721–729. doi:10.1016/j.neuroimage.2013.08.003
Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, *15*(3), 470–476. doi:10.1038/nn.3017
Jacobsen, C. F., & Nissen, H. W. (1937). Studies of cerebral function in primates. IV. The effects of frontal lobe lesions on the delayed alternation habit in monkeys. *Journal of Comparative Psychology*, *23*(1), 101.
James, W. (1890). *The Principles of Psychology* (1st ed.). New York: Henry Holt and Co, Inc.
Johnson, M. K., Mitchell, K. J., Raye, C. L., & Greene, E. J. (2004). An Age-Related Deficit in Prefrontal Cortical Function Associated with Refreshing Information. *Psychological Science*, *15*(2), 127–132.
Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., & Anderson, A. W. (2003). fMRI evidence for an organization of prefrontal cortex by both type of process and type of information. *Cerebral Cortex*, *13*(3), 265–273.
Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. A., & Sanislow, C. A. (2005). Using fMRI to investigate a component process of reflection: Prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective, &*

Behavioral Neuroscience, *5*(3), 339–361. doi:10.3758/CABN.5.3.339
Johnson, M. R., & Johnson, M. K. (2009). Top–Down Enhancement and Suppression of Activity in Category-selective Extrastriate Cortex from an Act of Reflective Attention. *Journal of Cognitive Neuroscience*, *21*(12), 2320–2327. doi:10.1162/jocn.2008.20094
Johnson, M. R., Mitchell, K. J., Raye, C. L., D'Esposito, M., & Johnson, M. K. (2007). A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. *NeuroImage*, *37*(1), 290–299. doi:10.1016/j.neuroimage.2007.05.017
Kaiser, M. (2011). A tutorial in connectome analysis: Topological and spatial features of brain networks. *NeuroImage*, *57*(3), 892–907. doi:10.1016/j.neuroimage.2011.05.025
Kanwisher, N., McDermott, J., & Chun, M. M. (2002). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Et Al., Foundations in Social Neuroscience, MIT Press, Cambridge, MA*.
Kapur, S. (2003). Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia. *American Journal of Psychiatry*, *160*(1), 13–23. doi:10.1176/appi.ajp.160.1.13
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.
Kelly, C., Toro, R., Di Martino, A., Cox, C. L., Bellec, P., Castellanos, F. X., & Milham, M. P. (2012). A convergent functional architecture of the insula emerges across imaging modalities. *NeuroImage*, *61*(4), 1129–1142. doi:10.1016/j.neuroimage.2012.03.021
Kelly, S. P., Lalor, E. C., Reilly, R. B., & Foxe, J. J. (2006). Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *Journal of Neurophysiology*, *95*(6), 3844–3851.
Kiebel, S. J., Garrido, M. I., Moran, R., Chen, C. C., & Friston, K. J. (2009). Dynamic causal modeling for EEG and MEG. *Human Brain Mapping*, *30*(6), 1866–1876.
Kim, Y. (1999). The Large-Scale Neural Network for Spatial Attention Displays Multifunctional Overlap But Differential Asymmetry. *NeuroImage*, *9*(3), 269–277. doi:10.1006/nimg.1999.0408
King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. doi:10.1016/j.tics.2014.01.002
Kiyonaga, A., & Egner, T. (2012). Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review*, *20*(2), 228–242. doi:10.3758/s13423-012-0359-y
Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. doi:10.1016/j.tics.2012.10.007
Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews*, *31*(7), 1003–1016.
Kodaka, Y., Mikami, A., & Kubota, K. (1997). Neuronal activity in the frontal eye field of the monkey is modulated while attention is focused on to a stimulus in the peripheral visual field, irrespective of eye movement. *Neuroscience Research*, *28*(4), 291–298. doi:10.1016/S0168-0102(97)00055-2
Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181–1185.
Krasnow, B., Tamm, L., Greicius, M. D., Yang, T. T., Glover, G. H., Reiss, A. L., & Menon, V. (2003). Comparison of fMRI activation at 3 and 1.5 T during perceptual, cognitive,

and affective processing. *NeuroImage*, *18*(4), 813–826.
Kubota, K., & Niki, H. (1971). Prefrontal cortical unit activity and delayed alternation performance in monkeys. *Journal of Neurophysiology*, *34*(3), 337–347.
Kuo, B.-C., Stokes, M. G., Murray, A. M., & Nobre, A. C. (2014). Attention Biases Visual Activity in Visual STM. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn_a_00577
Kuo, B.-C., Yeh, Y.-Y., Chen, A. J.-W., & D'Esposito, M. (2011). Functional connectivity during top-down modulation of visual short-term memory representations. *Neuropsychologia*, *49*(6), 1589–1596.
Lachaux, J.-P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., et al. (2005). The many faces of the gamma band response to complex visual stimuli. *NeuroImage*, *25*(2), 491–501. doi:10.1016/j.neuroimage.2004.11.052
Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., et al. (2009). ALE Meta-Analysis Workflows Via the Brainmap Database: Progress Towards A Probabilistic Functional Brain Atlas. *Frontiers in Neuroinformatics*, *3*. doi:10.3389/neuro.11.023.2009
Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(11), 1311–1328. doi:10.1068/p2935
Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*(2), 149–164. doi:10.1016/S0042-6989(02)00402-9
LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*, 5. doi:10.3389/fnhum.2014.00005
LaRocque, J. J., Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2013). Decoding attended information in short-term memory: an EEG study. *Journal of Cognitive Neuroscience*, *25*(1), 127–142. doi:10.1162/jocn_a_00305
Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–135). New York: Wiley.
Lepsien, J., & Nobre, A. C. (2006). Attentional Modulation of Object Representations in Working Memory. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(9), 2072–2083. doi:10.1093/cercor/bhl116
Lepsien, J., Griffin, I. C., Devlin, J. T., & Nobre, A. C. (2005). Directing spatial attention in mental representations: Interactions between attentional orienting and working-memory load. *NeuroImage*, *26*(3), 733–743.
Lepsien, J., Thornton, I., & Nobre, A. C. (2011). Modulation of working-memory maintenance by directed attention. *Neuropsychologia*, *49*(6), 1569–1577.
Lewis, S. J., Dove, A., Robbins, T. W., Barker, R. A., & Owen, A. M. (2004). Striatal contributions to working memory: a functional magnetic resonance imaging study in humans. *European Journal of Neuroscience*, *19*(3), 755–760. doi:10.1111/j.1460-9568.2003.03108.x
Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia*, *50*(4), 470–478. doi:10.1016/j.neuropsychologia.2011.11.006
Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. doi:10.3758/BF03193587
Libera, Della, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, *20*(6), 778–784.
Lisman, J. E., & Idiart, M. A. P. (1995). Storage of 7 ± 2 Short-Term Memories in Oscillatory Subcycles. *Science, New Series*, *267*(5203), 1512–1515.
Listgarten, J., Neal, R. M., Roweis, S. T., & Emili, A. (2004). Multiple alignment of

continuous time series. *Advances in Neural Information Processing Systems*, 817–824.
Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity:from psychophysics and neurobiologyto individual differences. *Trends in Cognitive Sciences*, *17*(8), 391–400. doi:10.1016/j.tics.2013.06.006
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.
MacDonald, A. W. (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
Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychonomic Bulletin & Review*, *19*(3), 470–476. doi:10.3758/s13423-012-0235-9
Malmivuo, J., Suihko, V., & Eskola, H. (1997). Sensitivity distributions of EEG and MEG measurements. *Biomedical Engineering, IEEE Transactions on*, *44*(3), 196–208.
Mangun, G. R., Hopfinger, J. B., & Buonocore, M. H. (2000). The neural mechanisms of top-down attentional control - Nature Neuroscience. *Nature Neuroscience*, *3*(3), 284–291. doi:10.1038/72999
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
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*.
Matsukura, M., Luck, S. J., & Vecera, S. P. (2007). Attention effects during visual short-term memory maintenance: Protection or prioritization? *Perception & Psychophysics*, *69*(8), 1422–1434. doi:10.3758/BF03192957
Maxcey-Richard, A. M., & Hollingworth, A. (2013). The strategic retention of task-relevant objects in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 760–772. doi:10.1037/a0029496
McElree, B. (2006). Accessing Recent Events. In *Psychology of Learning and Motivation* (Vol. 46, pp. 155–200). Elsevier. doi:10.1016/S0079-7421(06)46005-9
McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology. General*, *118*(4), 346–373. doi:10.1037/0096-3445.118.4.346
McNab, F., & Klingberg, T. (2007). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107. doi:10.1038/nn2024
Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, *10*(4), 309–325. doi:10.1002/ana.410100402
Mesulam, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology*, *28*(5), 597–613. doi:10.1002/ana.410280502
Middleton, F. A., & Strick, P. L. (2000). Basal Ganglia Output and Cognition: Evidence from Anatomical, Behavioral, and Clinical Studies. *Brain and Cognition*, *42*(2), 183–200. doi:10.1006/brcg.1999.1099
Milad, M. R., & Rauch, S. L. (2012). Obsessive-compulsive disorder: beyond segregated cortico-striatal pathways. *Trends in Cognitive Sciences*, *16*(1), 43–51. doi:10.1016/j.tics.2011.11.003
Miller, E. K., & Buschman, T. J. (2013). Cortical circuits for the control of attention. *Current Opinion in Neurobiology*, *23*(2), 216–222. doi:10.1016/j.conb.2012.11.011

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. doi:10.1146/annurev.neuro.24.1.167

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci*, *16*(16), 5154–5167.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci*, *13*(4), 1460–1478.

Mink, J. W. (1996). The basal ganglia: focused selection and inhibition of competing motor programs. *Progress in Neurobiology*, *50*(4), 381–425.

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, *331*(6151), 68–70. doi:10.1038/331068a0

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science*, *319*(5869), 1543–1546.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. doi:10.1016/S1364-6613(03)00028-7

Mosher, J. C., Baillet, S., & Leahy, R. M. (2003). Equivalence of linear approaches in bioelectromagnetic inverse solutions. *Ieee*, 294–297.

Moustafa, A. A., Sherman, S. J., & Frank, M. J. (2008). A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia*, *46*(13), 3144–3156. doi:10.1016/j.neuropsychologia.2008.07.011

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. doi:10.1016/j.tics.2011.07.011

Munneke, J., Belopolsky, A. V., & Theeuwes, J. (2012). Shifting attention within memory representations involves early visual areas. *PLoS ONE*, *7*(4), e35528. doi:10.1371/journal.pone.0035528

Munneke, J., Heslenfeld, D. J., & Theeuwes, J. (2010). Spatial working memory effects in early visual cortex. *Brain and Cognition*, *72*(3), 368–377. doi:10.1016/j.bandc.2009.11.001

Murray, A. M., Nobre, A. C., & Stokes, M. G. (2011). Markers of preparatory attention predict visual short-term memory performance. *Neuropsychologia*, *49*(6), 1458–1465. doi:10.1016/j.neuropsychologia.2011.02.016

Murray, A. M., Nobre, A. C., Astle, D. E., & Stokes, M. G. (2012). Lacking Control over the Trade-Off between Quality and Quantity in Visual Short-Term Memory. *PLoS ONE*, *7*(8), e41223. doi:10.1371/journal.pone.0041223.s004

Murray, A. M., Nobre, A. C., Clark, I. A., Cravo, A. M., & Stokes, M. G. (2013). Attention Restores Discrete Items to Visual Short-Term Memory. *Psychological Science*.

Nairne, J. S. (2002). Remembering Over The Short Term: The Case Against the Standard Model. *Annual Review of Psychology*, *53*(1), 53–81. doi:10.1146/annurev.psych.53.100901.135131

Nee, D. E., & Brown, J. W. (2013). Dissociable Frontal-Striatal and Frontal-Parietal Networks Involved in Updating Hierarchical Contexts in Working Memory. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(9), 2146–2158. doi:10.1093/cercor/bhs194

Nee, D. E., & Jonides, J. (2009). Common and distinct neural correlates of perceptual and memorial selection. *NeuroImage*, *45*(3), 963–975. doi:10.1016/j.neuroimage.2009.01.005

Neisser, U. (1967). *Cognitive psychology.* East Norwalk, CT, US: Appleton-Century-Crofts.

Nelissen, N., Stokes, M., Nobre, A. C., & Rushworth, M. F. S. (2013). Frontal and Parietal Cortical Interactions with Distributed Visual Representations during Selective Attention and Action Selection. *Journal of Neuroscience*, *33*(42), 16443–16458. doi:10.1523/JNEUROSCI.2625-13.2013

Neubert, F.-X., Mars, R. B., Thomas, A. G., Sallet, J., & Rushworth, M. F. (2014).

Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex. *Neuron*, *81*, 700–713.
Nobre, A. C., Coull, J. T., Maquet, P., Frith, C. D., Vandenberghe, R., & Mesulam, M. M. (2004). Orienting attention to locations in perceptual versus mental representations. *Journal of Cognitive Neuroscience*, *16*(3), 363–373.
Nobre, A., & Mesulam, M. M. (2014). Large-Scale Networks for Attentional Biases. In A. Nobre & S. Kastner (Eds.), *Oxford Handbook of Attention* (pp. 1–56). Oxford University Press.
Nobre, A., Sebestyen, G. N., Gitelman, D. R., Mesulam, M. M., Frackowiak, R., & Frith, C. D. (1997). Functional localization of the system for visuospatial attention using positron emission tomography. *Brain*, *120*(3), 515–533. doi:10.1093/brain/120.3.515
Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, *48*(22), 3637–3652. doi:10.1088/0031-9155/48/22/002
Norman, D. A., & Shallice, T. (1986). Attention to action. In *Action research*. Springer.
O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of "mudsplashes." *Nature*, *398*(6722), 34–34. doi:10.1038/17953
O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.
Oberauer, K. (2013). The focus of attention in working memory—frommetaphors to mechanisms, 1–16. doi:10.3389/fnhum.2013.00673/abstract
Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, *21*(3), 164–169.
Offen, S., Schluppeck, D., & Heeger, D. J. (2009). The role of early visual cortex in visual short-term memory and visual attention. *Vision Research*, *49*(10), 1352–1362. doi:10.1016/j.visres.2007.12.022
Ole Jensen, A. M. (2010). Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition. *Frontiers in Human Neuroscience*, *4*. doi:10.3389/fnhum.2010.00186
Olivers, C. N., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334.
Otto, A., Zerr, I., Lantsch, M., Weidehaas, K., Riedemann, C., & Poser, S. (1998). Akinetic mutism as a classification criterion for the diagnosis of Creutzfeldt-Jakob disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*(4), 524–528.
Parkin, A. J. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society*, *4*(05), 518–522.
Pashler, H., & Shiu, L.-P. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin & Review*, *6*(3), 445–448.
Passingham, R. E., & Katsuyuki, S. (2004). The prefrontal cortex and working memory: physiology and brain imaging. *Current Opinion in Neurobiology*, *14*, 163–168.
Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), 97–107. doi:10.1038/nrn1603
Payne, L., Guillory, S., & Sekuler, R. (2013). Attention-modulated Alpha-band Oscillations Protect against Intrusion of Irrelevant Information. *Journal of Cognitive Neuroscience*, *9*(25), 1463–1476.
Pertzov, Y., Bays, P. M., Joseph, S., & Husain, M. (2012). Rapid Forgetting Prevented by Retrospective Attention Cues. *Journal of Experimental Psychology: Human Perception*

and Performance. doi:10.1037/a0030947
Petersen, S. E., & Posner, M. I. (2012). The Attention System of the Human Brain: 20 Years After. *Annual Review of Neuroscience*, *35*(1), 73–89. doi:10.1146/annurev-neuro-062111-150525
Pfurtscheller, G., Stancák, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band--an electrophysiological correlate of cortical idling: a review. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *24*(1-2), 39–46.
Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290. doi:10.3758/BF03203943
Ploran, E. J., Nelson, S. M., Velanova, K., Donaldson, D. I., Petersen, S. E., & Wheeler, M. E. (2007). Evidence Accumulation and the Moment of Recognition: Dissociating Perceptual Recognition Processes Using fMRI. *Journal of Neuroscience*, *27*(44), 11912–11924. doi:10.1523/JNEUROSCI.3522-07.2007
Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. doi:10.1146/annurev.ne.13.030190.000325
Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in Neurobiology*, *23*(2), 223–228. doi:10.1016/j.conb.2012.12.009
Rainer, G., Rao, S. C., & Miller, E. K. (1999). Prospective coding for objects in primate prefrontal cortex. *J Neurosci*, *19*(13), 5493–5505.
Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.
Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. A., & Greene, E. J. (2002). Neuroimaging a single thought: dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage*, *15*(2), 447–453. doi:10.1006/nimg.2001.0983
Raye, C. L., Mitchell, K. J., Reeder, J. A., Greene, E. J., & Johnson, M. K. (2008). Refreshing One of Several Active Representations: Behavioral and Functional Magnetic Resonance Imaging Differences between Young and Older Adults. *Journal of Cognitive Neuroscience*, *20*(5), 852–862. doi:10.1016/j.neuroimage.2003.07.016
Raymond, J. E., & O'Brien, J. L. (2009). Selective Visual Attention and Motivation The Consequences of Value Learning in an Attentional Blink Task. *Psychological Science*, *20*(8), 981–988.
Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, *89*(4), 1009–1023.
Rerko, L., Souza, A. S., & Oberauer, K. (2014). Retro-cue benefits in working memory without sustained focal attention. *Memory & Cognition*. doi:10.3758/s13421-013-0392-8
Riddoch, M. J., Chechlacz, M., Mevorach, C., Mavritsaki, E., Allen, H., & Humphreys, G. W. (2010). The neural mechanisms of visual selection: the view from neuropsychology. *Annals of the New York Academy of Sciences*, *1191*(1), 156–181.
Rihs, T. A., Michel, C. M., & Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *The European Journal of Neuroscience*, *25*(2), 603–610. doi:10.1111/j.1460-9568.2007.05278.x
Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. *Attention and Performance XV*, *15*, 231–265.
Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *J Neurosci*, *31*(40), 14076–14084.

Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *J Neurosci*, *30*(25), 8692–8697.

Roth, J. K., Johnson, M. K., Raye, C. L., & Constable, R. T. (2009). Similar and dissociable mechanisms for attention to internal versus external information. *NeuroImage*, *48*(3), 601–608. doi:10.1016/j.neuroimage.2009.07.002

Roux, F., Wibral, M., Singer, W., Aru, J., & Uhlhaas, P. J. (2013). The Phase of Thalamic Alpha Activity Modulates Cortical Gamma-Band Activity: Evidence from Resting-State MEG Recordings. *Journal of Neuroscience*, *33*(45), 17827–17835.

Rowe, J. B., & Passingham, R. E. (2001). Working Memory for Location and Time: Activity in Prefrontal Area 46 Relates to Selection Rather than Maintenance in Memory. *NeuroImage*, *14*(1), 77–86. doi:10.1006/nimg.2001.0784

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S., & Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science*, *288*(5471), 1656–1660.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applies statistician. *The Annals of Statistics*, 1151–1172.

Rushworth, M. F. S., Buckley, M. J., Behrens, T. E. J., Walton, M. E., & Bannerman, D. M. (2007). Functional organization of the medial frontal cortex. *Current Opinion in Neurobiology*, *17*(2), 220–227. doi:10.1016/j.conb.2007.03.001

Rushworth, M. F. S., Kolling, N., Sallet, J., & Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Current Opinion in Neurobiology*, *22*(6), 946–955. doi:10.1016/j.conb.2012.04.011

Saalmann, Y. B., Pinsk, M. A., Wang, L., Li, X., & Kastner, S. (2012). The Pulvinar Regulates Information Transmission Between Cortical Areas Based on Attention Demands. *Science*, *337*(6095), 753–756. doi:10.1126/science.1223082

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature Neuroscience*, *5*(5), 479–484. doi:10.1038/nn846

Sallet, J., Mars, R. B., Noonan, M. P., Neubert, F. X., Jbabdi, S., O'Reilly, J. X., et al. (2013). The Organization of Dorsal Frontal Cortex in Humans and Macaques. *Journal of Neuroscience*, *33*(30), 12255–12274. doi:10.1523/JNEUROSCI.5108-12.2013

Salmelin, R., Kringelbach, M., & Hansen, P. (Eds.). (2010). *MEG: An introduction to methods* (1st ed.). Oxford: Oxford University Press.

Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. D. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, *15*(9), 1332–1342.

Sanides, F. (1964). The cyto-myeloarchitecture of the human frontal lobe and its relation to phylogenetic differentiation of the cerebral cortex. *Journal Für Hirnforschung*, (7), 269.

Sauseng, P., Klimesch, W., Gerloff, C., & Hummel, F. C. (2009a). Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex. *Neuropsychologia*.

Sauseng, P., Klimesch, W., Heise, K. F., Gruber, W. R., Holz, E., Karim, A. A., et al. (2009b). Brain Oscillatory Substrates of Visual Short-Term Memory Capacity. *Current Biology*, *19*(21), 1846–1852. doi:10.1016/j.cub.2009.08.062

Sawaguchi, T., & Goldman-Rakic, P. S. (1991). D1 Dopamine Receptors in Prefrontal Cortex: Involvement in Working Memory. *Science, New Series*, *251*(4996), 947–950.

Schmidt, B. K., Vogel, E. K., Woodman, G. F., & Luck, S. J. (2002). Voluntary and automatic attentional control of visual working memory. *Perception & Psychophysics*, *64*(5), 754–763.

Schoffelen, J.-M., & Gross, J. (2009). Source connectivity analysis with MEG and EEG. *Human Brain Mapping*, *30*(6), 1857–1865. doi:10.1002/hbm.20745

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of Neurophysiology*, *56*(5), 1439–1461.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, *23*(2), 229–238. doi:10.1016/j.conb.2012.11.012

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*(9), 2349–2356. doi:10.1523/JNEUROSCI.5587-06.2007

Seeman, P., Lee, T., Chau-Wong, M., & Wong, K. (1976). Antipsychotic drug doses and neuroleptic/dopamine receptors. *Nature*, *261*(5562), 717–719.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214. doi:10.1111/j.1467-9280.2009.02276.x

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. doi:10.1016/S1364-6613(97)01094-2

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701–703. doi:10.2307/1731476?ref=search-gateway:64648421b5d02bab035c4098da5800a7

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron*, *60*(4), 709–719. doi:10.1016/j.neuron.2008.09.010

Slagter, H. A., Tomer, R., Christian, B. T., Fox, A. S., Colzato, L. S., King, C. R., et al. (2012). PET Evidence for a Role for Striatal Dopamine in the Attentional Blink: Functional Implications. *Journal of Cognitive Neuroscience*, *24*(9), 1932–1940.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are There Multiple Visual Short-Term Memory Stores? *PLoS ONE*, *3*(2), e1699. doi:10.1371/journal.pone.0001699.g008

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2009). V4 Activity Predicts the Strength of Visual Short-Term Memory Representations. *Journal of Neuroscience*, *29*(23), 7432–7438. doi:10.1523/JNEUROSCI.0784-09.2009

Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed Sensory Memory, Sloppy Working Memory. *Frontiers in Psychology*, *1*. doi:10.3389/fpsyg.2010.00175

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, Involuntary Top-Down Guidance of Attention From Working Memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248–261. doi:10.1037/0096-1523.31.2.248

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Stokes, M. G. (2011). Top-down visual activity underlying VSTM and preparatory attention. *Neuropsychologia*, *49*(6), 1425–1427. doi:10.1016/j.neuropsychologia.2011.02.004

Stokes, M. 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

Stokes, M., & Duncan, J. (2013). Dynamic Brain States for Preparatory Attention. In A. C. Nobre & S. Kastner (Eds.), *The Oxford Handbook of Attention*. Oxford University Press.

Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009a). Top-Down Activation of Shape-Specific Population Codes in Visual Cortex during Mental Imagery. *Journal of Neuroscience*, *29*(5), 1565–1572. doi:10.1523/JNEUROSCI.4657-08.2009

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009b). Shape-Specific Preparatory Activity Mediates Attention to Targets in Human Visual Cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574.

Super, H., Spekreijse, H., & Lamme, V. A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*(5527), 120–124. doi:10.1126/science.1060496

Szczepanski, S. M., Konen, C. S., & Kastner, S. (2010). Mechanisms of Spatial Attention Control in Frontal and Parietal Cortex. *Journal of Neuroscience*, *30*(1), 148–160. doi:10.1523/JNEUROSCI.3862-09.2010

Taulu, S., Kajola, M., & Simola, J. (2004). Suppression of interference and artifacts by the Signal Space Separation Method. *Brain Topography*, *16*(4), 269–275.

Theeuwes, J., & Belopolsky, A. V. (2012). Reward grabs the eye: Oculomotor capture by rewarding stimuli. *Vision Research*, *74*, 80–85. doi:10.1016/j.visres.2012.07.024

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. *Journal of Neuroscience*, *26*(37), 9494–9502. doi:10.1523/JNEUROSCI.0875-06.2006

Tomassini, V., Jbabdi, S., Klein, J. C., Behrens, T. E. J., Pozzilli, C., Matthews, P. M., et al. (2007). Diffusion-Weighted Imaging Tractography-Based Parcellation of the Human Lateral Premotor Cortex Identifies Dorsal and Ventral Subregions with Anatomical and Functional Specializations. *Jneurosci.org*.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, *16*(3 Pt 1), 765–780.

Uhlhaas, P. J., Pipa, G., Neuenschwander, S., Wibral, M., & Singer, W. (2011). Progress in Biophysics and Molecular Biology. *Progress in Biophysics and Molecular Biology*, *105*(1-2), 14–28. doi:10.1016/j.pbiomolbio.2010.10.004

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785. doi:10.1073/pnas.1117465109/-/DCSupplemental/pnas.201117465SI.pdf

van Ede, F., de Lange, F., Jensen, O., & Maris, E. (2011). Orienting Attention to an Upcoming Tactile Event Involves a Spatially and Temporally Specific Modulation of Sensorimotor Alpha- and Beta-Band Oscillations. *Journal of Neuroscience*, *31*(6), 2016–2024. doi:10.1523/JNEUROSCI.5630-10.2011

Van Veen, B. D., & Buckley, K. M. (1988). Beamforming: A versatile approach to spatial filtering. *IEEE Assp Magazine*, *5*(2), 4–24.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1998). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *Biomedical Engineering, IEEE Transactions on*, *44*(9), 867–880.

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424. doi:10.1016/j.tics.2008.07.005

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

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal

individual differences in controlling access to working memory. *Nature*, *438*(7067), 500–503. doi:10.1038/nature04171

Voytek, B., & Knight, R. T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proceedings of the National Academy of Sciences*, *107*(42), 18167–18172.

Waldhauser, G. T., Johansson, M., & Hanslmayr, S. (2012). Alpha/Beta oscillations indicate inhibition of interfering visual memories. *J Neurosci*, *32*(6), 1953–1961.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, *24*(8), 455–463. doi:10.1016/S0166-2236(00)01868-3

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*. doi:10.3389/fninf.2013.00014

Williams, S. (1998). Widespread origin of the primate mesofrontal dopamine system. *Cerebral Cortex*, *8*(4), 321–345. doi:10.1093/cercor/8.4.321

Williams, S. M., & Goldman-Rakic, P. S. (1993). Characterization of the dopaminergic innervation of the primate frontal cortex using a dopamine-specific antibody. *Cerebral Cortex*, *3*(3), 199–222.

Wilson, K. D., Woldorff, M. G., & Mangun, G. R. (2005). Control networks and hemispheric asymmetries in parietal cortex during attentional orienting in different spatial reference frames. *NeuroImage*, *25*(3), 668–683. doi:10.1016/j.neuroimage.2004.07.075

Woldorff, M. G., Hazlett, C. J., Fichtenholtz, H. M., Weissman, D. H., Dale, A. M., & Song, A. W. (2004). Functional parcellation of attentional control regions of the brain. *Journal of Cognitive Neuroscience*, *16*(1), 149–165.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., & Fries, P. (2007). Modulation of Neuronal Interactions Through Neuronal Synchronization. *Science*, *316*(5831), 1609–1612. doi:10.1126/science.1139597

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363–377. doi:10.1037/0096-1523.33.2.363

Woodman, G. F., Carlisle, N. B., & Reinhart, R. M. G. (2013). Where do we store the memory representations that guide attention? *Journal of Vision*, *13*(3), 1–1. doi:10.1167/13.3.1

Woolrich, M., Hunt, L., Groves, A., & Barnes, G. (2011). MEG beamforming using Bayesian PCA for adaptive data covariance matrix regularization. *NeuroImage*, *57*(4), 1466–1479.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific-band electroencephalography increases over occipital cortex. *J Neurosci*, *20*, 1–6.

Yeh, Y.-Y., Kuo, B.-C., & Liu, H.-L. (2007). The neural correlates of attention orienting in visuospatial working memory for detecting feature and conjunction changes. *Brain Research*, *1130*(1), 146–157. doi:10.1016/j.brainres.2006.10.065

Yeung, N. (2013). Conflict monitoring and cognitive control. In K. Ochsner & S. M. Kosslyn (Eds.), *Oxford Handbook of Cognitive Neuroscience*. Oxford University Press.

Yi, D.-J., Turk-Browne, N. B., Chun, M. M., & Johnson, M. K. (2008). When a thought equals a look: refreshing enhances perceptual memory. *Journal of Cognitive Neuroscience*, *20*(8), 1371–1380. doi:10.1162/jocn.2008.20094

Yousry, T. (1997). Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain*, *120*(1), 141–157. doi:10.1093/brain/120.1.141

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. doi:10.1038/nn.2773

Zhang, S., Ide, J. S., & Li, C.-S. R. (2012). Resting-state functional connectivity of the medial superior frontal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(1), 99–111. doi:10.1093/cercor/bhr088

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. doi:10.1038/nature06860

Zokaei, N., Manohar, S., Husain, M., & Feredoes, E. (2014). Causal Evidence for a Privileged Working Memory State in Early Visual Cortex. *Journal of Neuroscience*, *34*(1), 158–162.

9 Appendix

9.1 Mixture modelling of behavioural data

A precision/capacity working memory task is used twice in the work presented in this thesis. Two forms of the task are used. In its forced choice variant (chapter 3), subjects are presented with a memory array of oriented items. Following a retention interval, a probe item appears that is rotated either clockwise or anticlockwise relative to the memory item. Subjects have to judge in which direction they believe the item has rotated relative to what they retain in memory. Several different magnitudes of probe rotation are used, and the dependent variables are the probability of correctly judging the angle of rotation for each magnitude of rotation, and the reaction time. In the continuous or 'dial-up' variant (chapter 6), subjects rotate a probe stimulus until it matches the orientation they remember. In this case the dependent variable is the error in judgement (difference between memory item orientation and the chosen probe orientation) on each trial.

The data from both variants are analysed by fitting a model to the data. A standard model for these data is the mixture model of Zhang et al. (W. Zhang & Luck, 2008). Based on the 'slots' view of short-term memory, this model assumes that on any given trial, the probed item is categorically present or absent in memory: non-guess and guess trials. For the subset of trials in which the item is present in memory, the internal representation of the item is considered to be noisy (imprecise) and to follow a distribution, specifically the Von Mises distribution (which is the circular analogue of the Gaussian distribution). For the subset of trials in which the item is not present in memory, the response is assumed to be completely random with respect to the memory

item (i.e. a uniform distribution over the circle). This model can be used to fit either 'dial-up' or forced choice data. In the former case, the mixture distribution (uniform + Von Mises, in a ratio determined by the guess rate) directly corresponds to the response error distribution. In the latter case, the Von Mises distribution is integrated from -∞ up to the magnitude of the orientation shift on a given trial, giving the probability that the internal representation appeared less shifted than the probe stimulus (and therefore the probability of a correct response on that trial).

The model is fitted using a maximum likelihood approach in both cases. The best fitting parameters are those under which the observed behavioural data have the highest likelihood (P(data|parameters)). The value of the best fitting parameters can be found either by evaluating the likelihood at all possible parameter values over the parameter space, and finding the maximum likelihood within that space, or by finding the maximum likelihood using a simplex search algorithm implemented in MATLAB (*fminsearch*). The former is computationally demanding but allows the behaviour of the model to be checked by plotting the likelihood map over the parameter space (see below). The latter approach is much faster and was used to analyse the data presented elsewhere in this thesis.

The model fit can be unstable over a relatively small number of trials: implausible parameter values may have a high likelihood. This can be seen by evaluating the likelihood at all points in the two-dimensional parameter space (kappa, pGuess), and plotting the resulting likelihood map. In Figure 9-1 the left hand column shows a single subject's likelihood fit for the neutral-cue condition in the task used in the MEG experiment. The model favours a very high precision with a very high guess rate. This parameter pairing is not plausible when compared with the distribution of

parameter values in other subjects, where the precision values cluster around a much lower average.

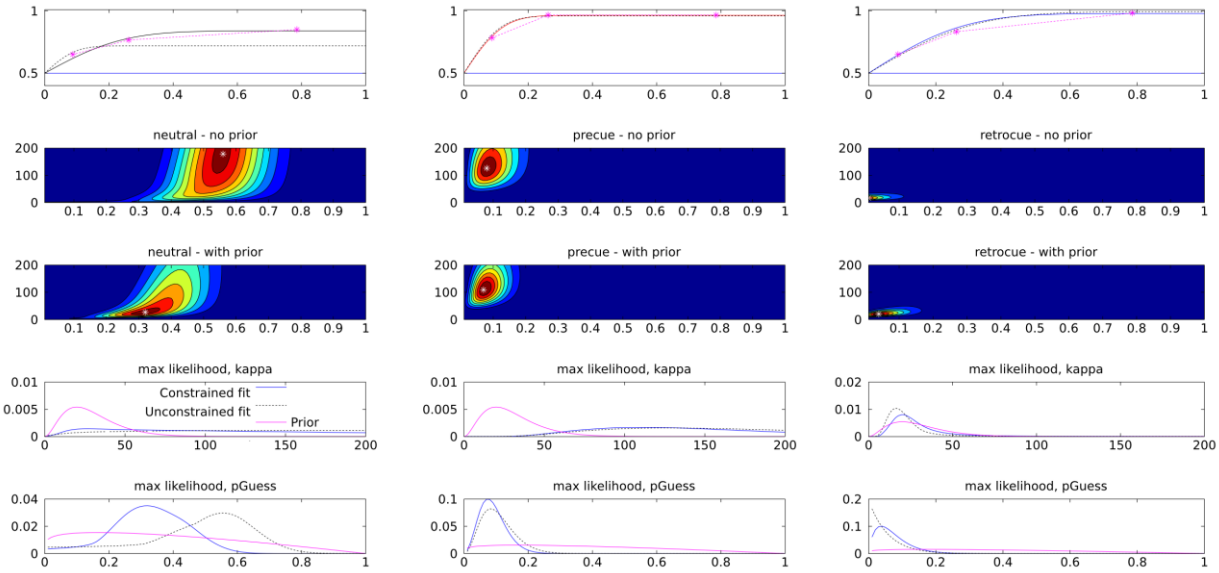


Figure 9-1 **Constraining the mixture model fit with priors**. The heat maps show the negative log likelihood of different kappa/pGuess parameter pairings (all pairs of values over the grid with {0 < kappa < 200} and {0 < pGuess < 1} were evaluated). Data shown here are from a subject for whom the unconstrained fit was poor for the neutral condition: the kappa parameter has been estimated almost at the ceiling of the grid (kappa = 200). This is not the case for the precue or retrocue analyses, for which the peak of the unconstrained likelihood map gives reasonable values. Imposing priors in the neutral condition has the desirable effect of pulling the parameter estimates towards more reasonable values: this can be seen in the top left panel which shows the unconstrained (black dotted curve) fit alongside the constrained (black full curve) fit, with the behavioural data points in pink. In contrast, where the data already constrain the fit, imposing priors has rather little effect on the parameter estimates (centre and right hand columns). The priors therefore prevent failures of the fitting routine without dominating the fit when the data already constrain the most likely parameters.

The model fit can be constrained by adding a prior distribution for the values of each parameter, embodying our prior beliefs about what parameter values are reasonable. If the same parameters are used for all subjects and conditions, then provided the priors chosen are reasonable and serve to constrain the model fits, the exact choice of prior distributions is not too important as differences between best-fitting parameters between subjects/conditions must still be driven by the data. A beta distribution, characterized by parameters α and β, which is bounded between 0 and 1, was chosen as the prior for the guess rate parameter in the mixture model. A gamma distribution, bounded between 0 and +infinity, characterized by parameters κ and θ, was chosen as a

prior for the kappa parameter. Parameter values for each distribution were chosen by approximately matching the median and the inter-quartile range (IQR) of the group-level unconstrained fits to the median and IQR of the prior distributions. The exact parameter values do not matter too much as long as the priors sensibly constrain the model fits, as the priors were common across all subjects and conditions so condition differences could not be driven by the specification of the priors, and when the data already constrained the parameter values effectively, the priors have relatively little influence (Figure 9-1, precue and retrocue conditions).

9.2 Deriving ROIs for MEG analysis from meta-analysis ROIs

The meta-analysis yielded a set of activation clusters, described in chapter 5. In order to identify the activated regions within and across the precue/retrocue meta-analyses from which a set of ROIs could be derived, local maxima from the clusters surviving permutation testing were used. All clusters from the precue and retrocue analyses were used, with the exception of the occipital clusters from the precue analysis (activations inferior and posterior to the trans-occipital sulcus). These were replaced with the occipital ROIs from (Dosenbach et al., 2007).

The results from the ALE meta-analyses were processed in two steps. First, for the precue and retrocue analyses separately, the local maxima of significant clusters were consolidated into a more sparse set of activations that characterized activated regions. Second, these activations were compared across the two meta-analyses to identify regions co-activated by precues and retrocues.

Where a set of evenly spaced activations occur, they may all merge into a single large cluster. For example, the cluster in the right MFG in the retrocue ALE analysis (cluster #1) is spatially extended and contains three local maxima at [-50, -2, 40], [-50, 22, 28] and [-54, 12, 34]. On the other hand, the ALE algorithm permits local maxima to be close together in space (<10mm apart), and local maxima within clusters may also represent the same underlying activation (but sampled with spatial noise across studies). In order to simply the cluster map into a sparse set of activation coordinates, clusters and their local maxima were compared to the Harvard-Oxford cortical structural atlas, which identifies gyral ROIs based on the average of an automated gyral segmentation. Across both analyses, there was only one cluster in which the local

maxima spanned more than one labelled grey matter region. This was the right MFG cluster in the retrocue meta-analysis for which the local maxima are listed above, which spanned left pre-central gyrus and MFG. Only the local maximum lying firmly within the precentral gyrus was kept for this cluster ([50, -2, 40]). For all other clusters, the locations of all local maxima within the cluster were averaged. This resulted in spatially sparse sets of coordinates for each meta-analysis, with no pair of coordinates within either analysis being fewer than 16mm apart (minimum separation 17.8mm, mean separation 76.4mm).

These sets of coordinates from the precue and retrocue meta-analyses were then compared to establish common activation sites. Activations separated by fewer than 16mm were considered to be common to the two analyses, by which criterion there were five common activations. This step yielded a set of 18 locations, shown along with their MNI coordinates in Figure 9-2. These activations were still asymmetric across the hemispheres. These ROIs were further simplified to yield the bilaterally symmetric set of ROIs used for the MEG analysis.

Several IPS foci were present in the combined meta-analysis, but there was some hemispheric asymmetry in the activated regions. Therefore, for the purposes of the MEG ROIs, these IPS/SPL activations were simplified into a bilaterally symmetric set of four locations per side, in order to capture the coarser anatomical subdivisions. Specifically, the activations lying within IPS0/V7 were retained, but the left hemispheric ROI was moved slightly more inferiorly, giving symmetric ROIs at [±34,-76,26], labelled 'IPS0'. Moving anteriorly, the meta-analysis activations distinguished between a medial (SPL) and a more lateral (IPS1/2) site at y = -68, but only in the left hemisphere. Two ROIs per side were placed at this y-coordinate, at [±30,-68,40] (around IPS1/2; labelled 'mid_IPS') and [±12,-68,60] ('SPL'). Finally, the common activation on the left

side at [-39,-48,40], located in the most anterior part of the IPS before it merges with the post-central sulcus (IPS4/5), was mirrored on the right side at [39,-48,40].

A lateral occipital ROI was placed in each hemisphere, at [±27,-89,3]. The left-hemisphere TPJ activation for retrocueing was mirrored in the right hemisphere [±52,-38,32], and the left hemisphere MTG activation was also mirrored [±62,-40,0].

Symmetrical ROIs were placed in the insula [±44,13,-4] (meta-analysis activations at[-40,14,-4] and [48,12,-4]). The anterior MFG activations at [-40,33,26] and [39,48,20] were consolidated into ROIs at [±40,39,23] ('anterior_MFG'). The pre-central activations at [-45,-3,44] and [48,6,42] were consolidated into ROIs at [±46,1,43]. The FEF activations from the precue meta-analysis were [±27,-3,52]. The dACC activation was placed on the midline [0,24,42] (originally [-1,24,42]).

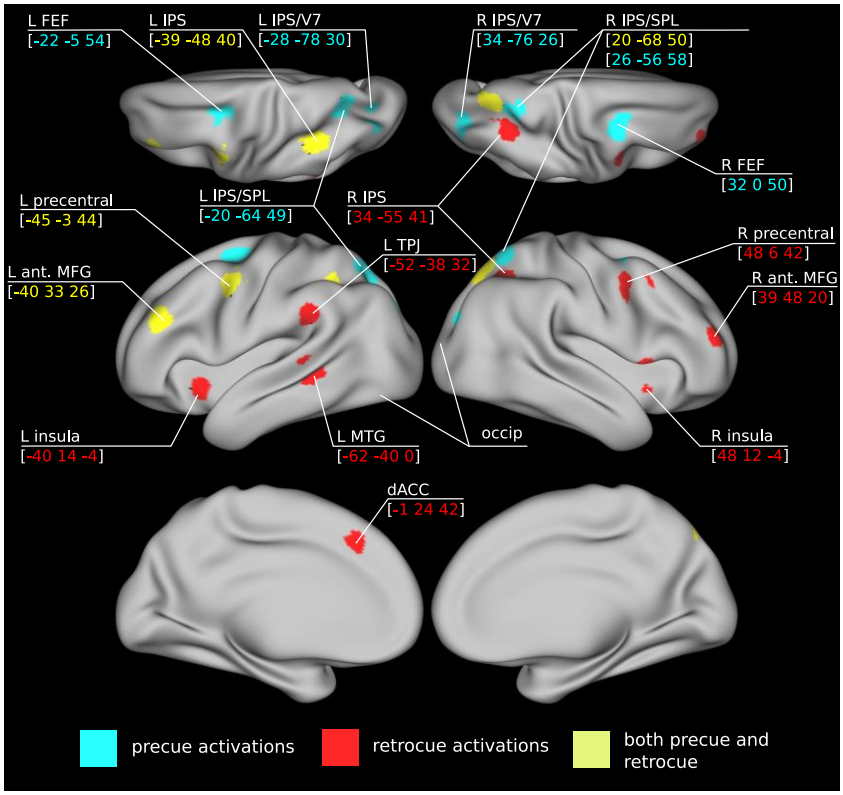


Figure 9-2 **Condensed set of common and unique activations derived from local maxima in the precue and retrocue meta-analyses.** These were simplified to a bilaterally symmetric set of ROIs, as shown in chapter 5.

9.3 ROI time-frequency analyses

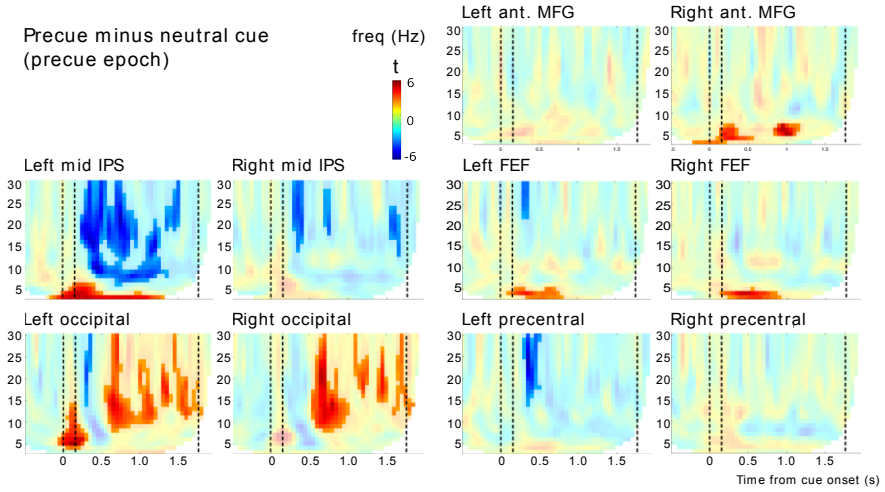


Figure 9-3 **Fronto-parietal induced responses following a precue.** Full colours mark significant clusters. A sustained alpha/beta desynchronization in left mid IPS ($p < 0.0002$) lasts until the presentation of the working memory array, with a concurrent synchronization in the theta band ($p = 0.004$). Sustained theta activations are seen in the left ($p = 0.0042$) and right ($p = 0.0058$) FEF, but not precentral gyrus. There are two significant clusters for the right MFG, one following the cue ($p = 0004$) and one around 1s post-cue (0.029). The sustained activations throughout the precue interval are consistent with sustained maintenance of preparatory bias until the presentation of the memory array, and the sustained lateralization of alpha power in visual cortex following a precue.

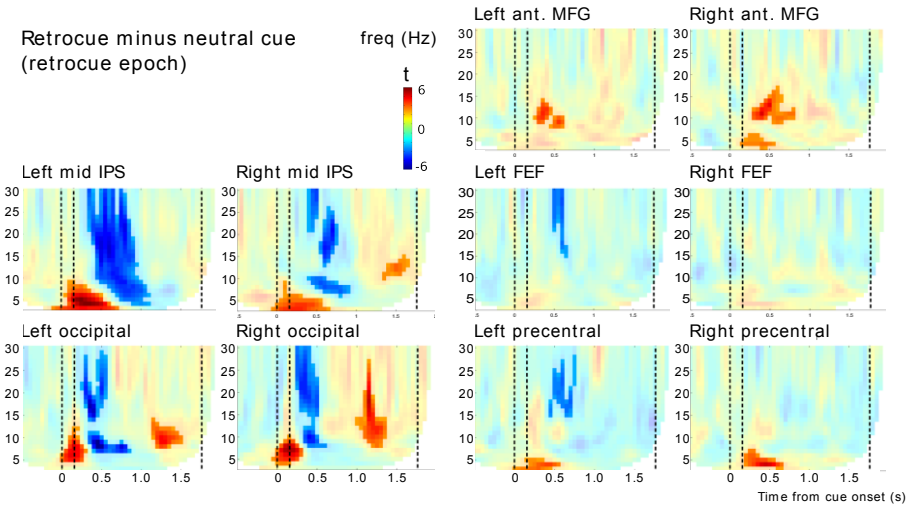


Figure 9-4 **Fronto-parietal induced responses following a retrocue.** Full colours mark significant clusters. A punctate broadband desynchronization in the left mid IPS lasts between 300 and 1000ms post cue ($p < 0.0002$). There are bilateral activations in the theta band in precentral gyrus (left $p = 0.02$, right $p = 0.009$) but not in FEF. There are bilateral activations in the MFG in the alpha/beta band (left $p = 0.005$, right $p = 0.002$) and also a theta band activation in the right MFG ($p = 0.029$). No fronto-parietal node is activated after 1000ms post-cue. This punctate pattern of activation is consistent with the short-lasting alpha lateralization observed in visual cortex following a retrocue.

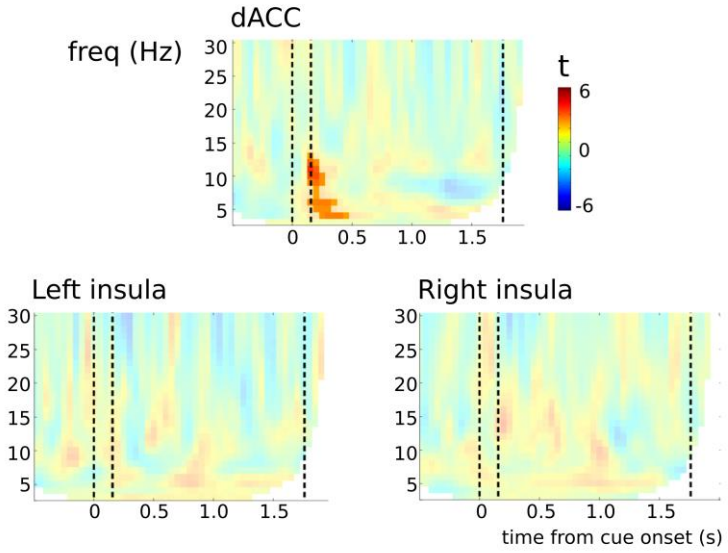


Figure 9-5 **Induced responses in cingulo-opercular ROIs following a precue**. Full colours mark significant clusters. The only significant modulation of activity in the cingulo-operular ROIs following a precue is in the dACC, immediately following cue presentation (p=0.004).

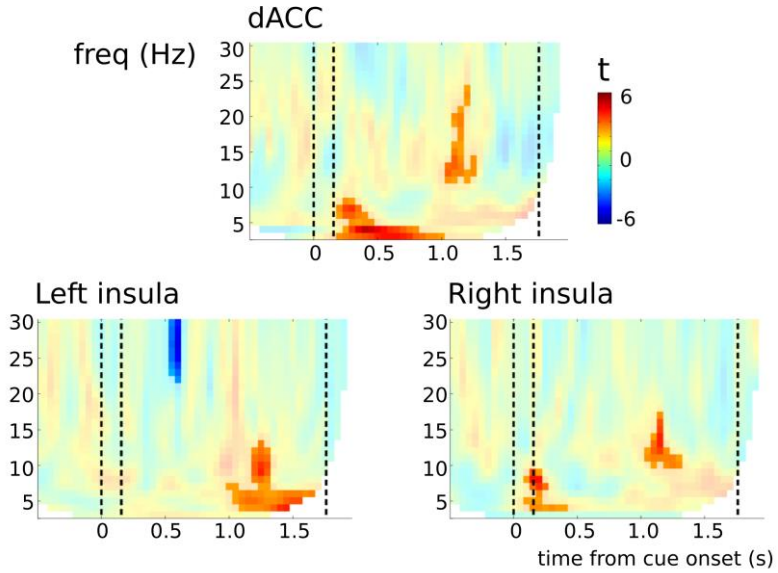


Figure 9-6 **Induced responses in cingulo-opercular ROIs following a retrocue.** Full colours mark significant clusters. From 1s post-cue onwards, all three cingulo-opercular ROIs are more active in the retrocue trials compared to neutral cue trials. Power is increased in both the theta and alpha/beta bands, with significant clusters in theta/alpha ($p < 0.0002$) for the left insula, alpha/beta ($p = 0.04$) for the right insula, and the beta band ($p = 0.008$) for the dACC. Inspecting nearby areas of the spectrograms that did not reach significance, the spectral pattern is fairly similar across the three ROIs. These late activations begin only after the fronto-parietal activations have fallen back to baseline, suggesting that the fronto-parietal and cingulo-opercular networks activate sequentially following a retrocue. The dACC is additionally activated earlier in the trial, in the theta band ($p = 0.002$).

9.4 Whole-brain induced responses

Maps are shown projected onto the CARET cortical surface to facilitate comparison with meta-analysis figures, but as this projection gives a poor representation of ventral/deep activations, the same data is also shown plotted over FSL's MNI152 standard brain, as slices through the brain volume.

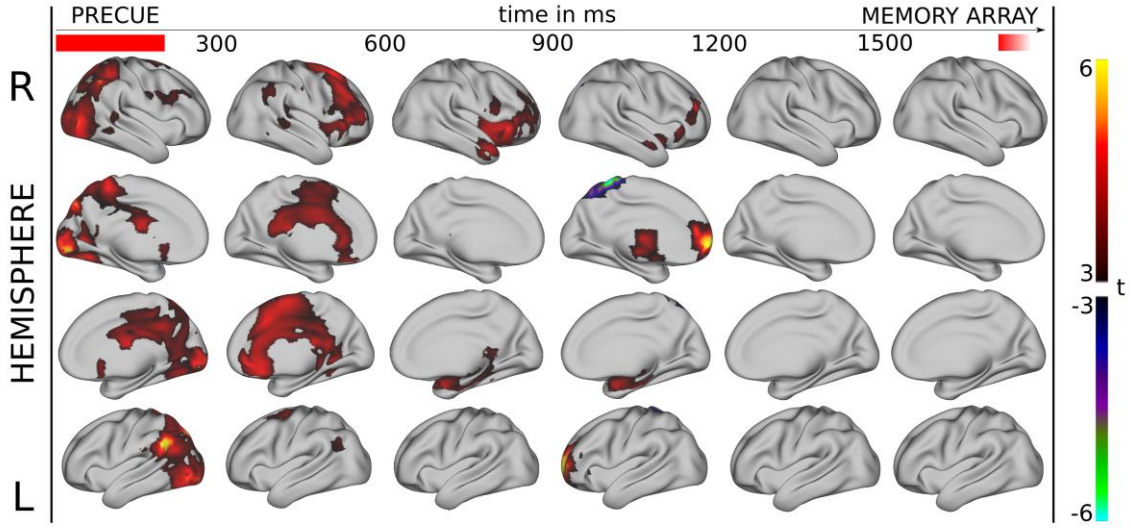


Figure 9-7 **3-7Hz (theta band) whole brain analysis**, [precue minus neutral cue] in the precue epoch, on the CARET cortical surface. Between cue onset and 600ms post-cue, there is a forward-propagating pattern of activity that is particularly strong in deep medial regions. This activation may reflect the ERF. Activations in the region of FEF can be seen on the lateral surfaces between 300 and 600ms post-cue. An additional activation in the medial frontal pole is apparent between 900 and 1200ms post-cue.

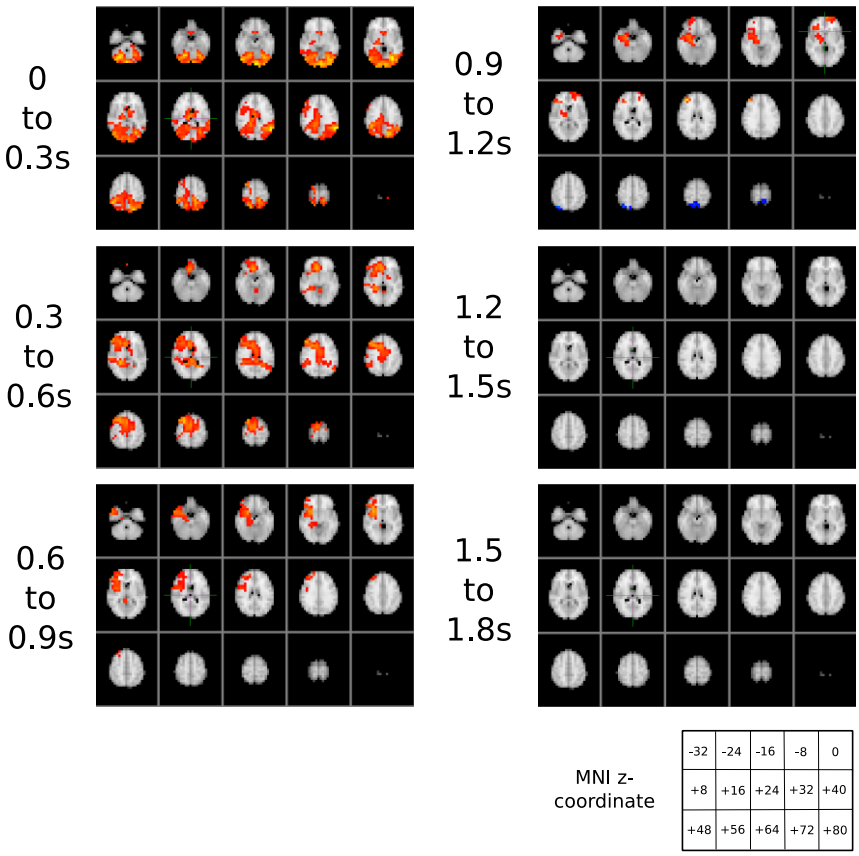


Figure 9-8 **3-7Hz (theta band) whole brain analysis**, [precue minus neutral cue] in the precue epoch, in FSL's MNI152 standard brain space. Only activations that are part of a significant cluster are shown. Power is thresholded between t-stats of [+3,+6] and [-3,-6]. Deep medial activations probably reflecting the heightened ERF to the cue stimulus are visible in first and second time-periods.

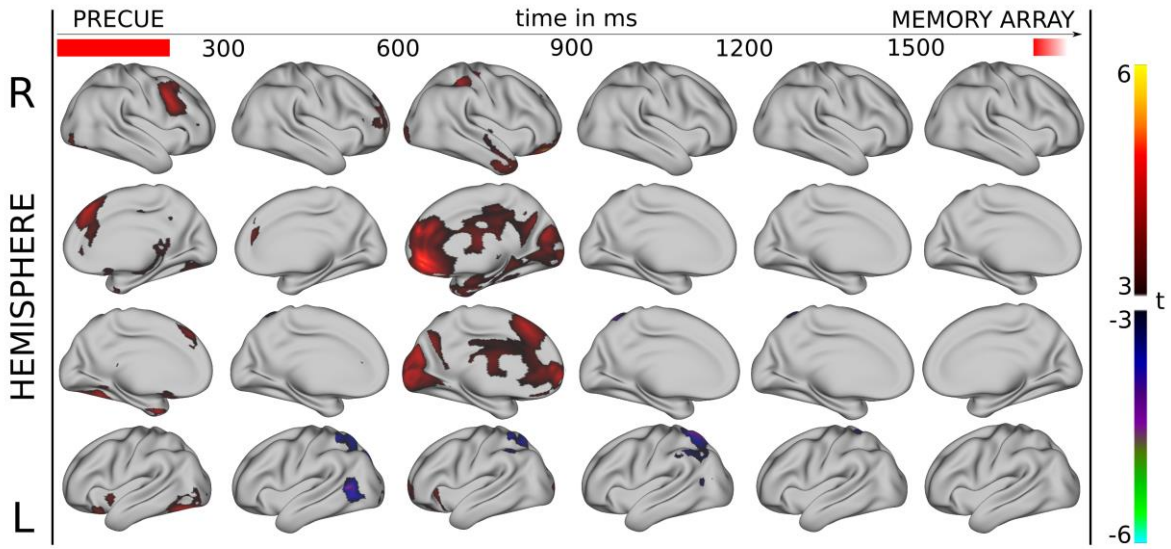


Figure 9-9 **8-12Hz (alpha band) whole brain analysis**, [precue minus neutral cue] in the precue epoch, on the CARET cortical surface. A focal deactivation in the left mid-IPS persists from 300ms post-cue until the presentation of the memory array. An activation is seen between 600 and 900ms post-cue in the frontal pole, possibly corresponding to the slightly later frontal pole activation in the theta band.

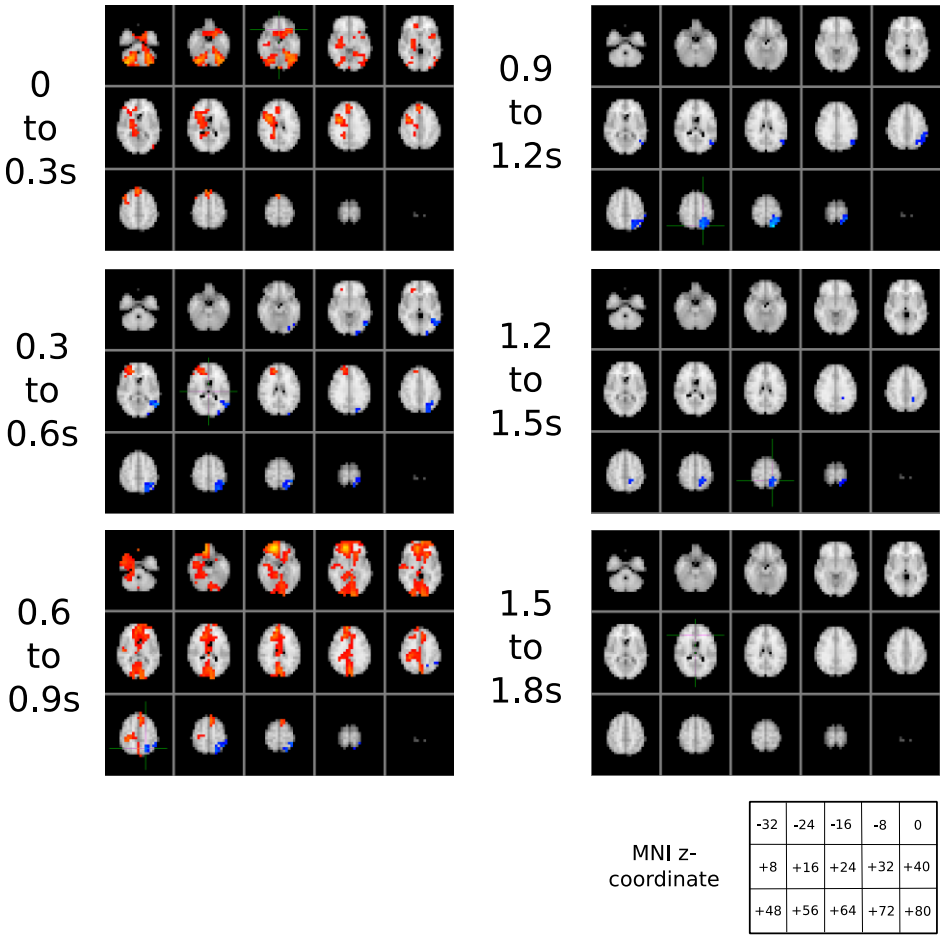


Figure 9-10 **8-12Hz (alpha band) whole brain analysis**, [precue minus neutral cue] in the precue epoch, in FSL's MNI152 standard brain space.

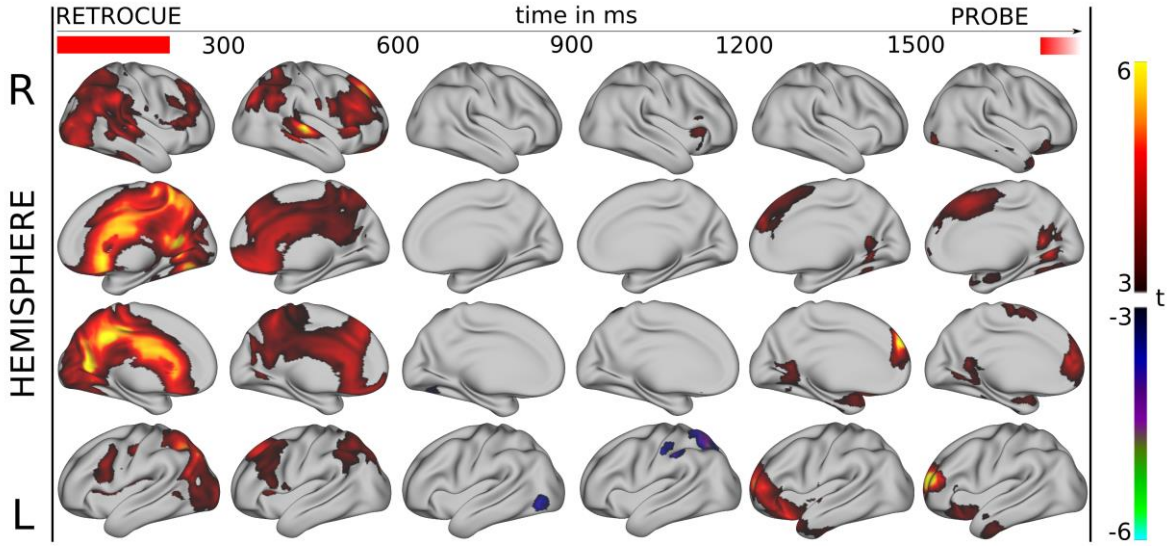


Figure 9-11 **3-7Hz (theta band) whole brain analysis**, [retrocue minus neutral cue] in the retrocue epoch, on the CARET cortical surface. There is a forward propagating activation in the first two time intervals, particularly in deep/medial voxels, that may reflect the ERF. There is an activation in the frontal pole between 1200ms and the presentation of the probe item that is not captured in the ROI analysis. At the same time, there are activations in the dorsomedial prefrontal cortex and insula/operculum.

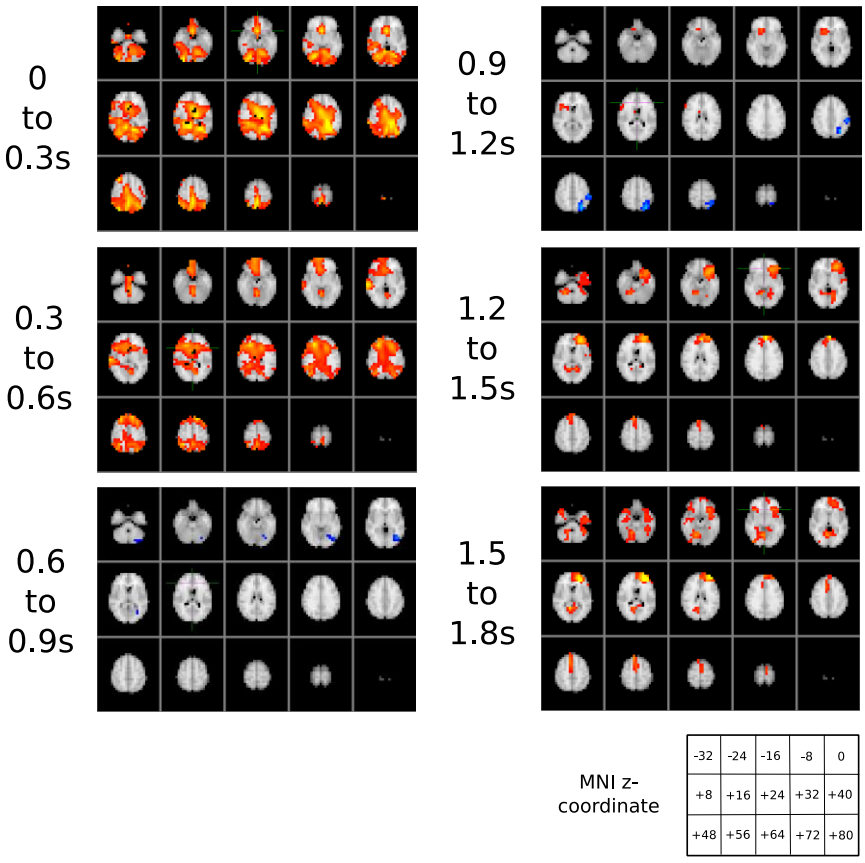


Figure 9-12 **3-7Hz (theta band) whole brain analysis**, [retrocue minus neutral cue] in the retrocue epoch, in FSL's MNI152 standard brain space.

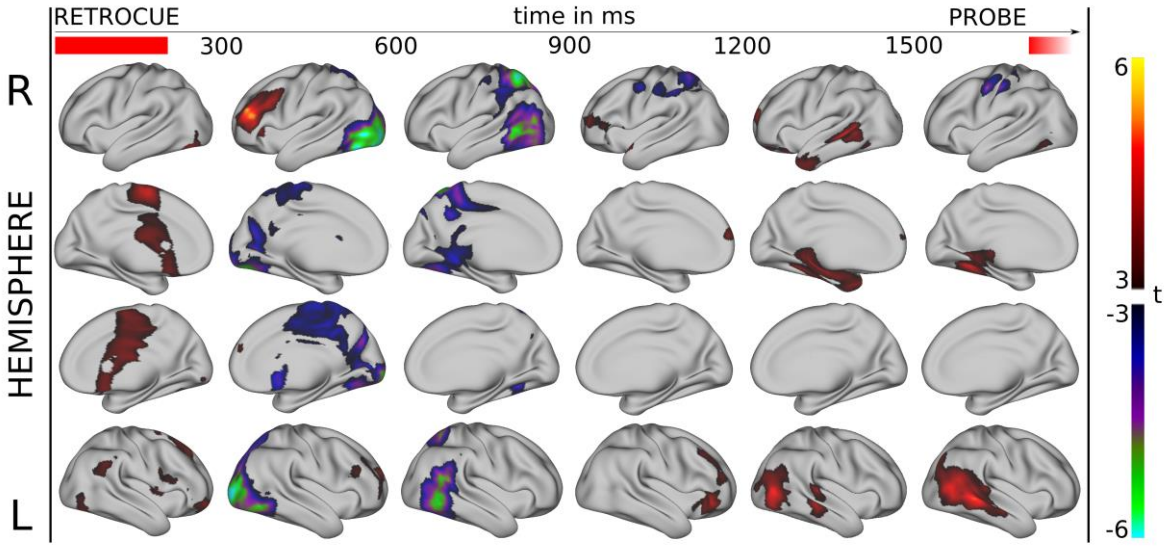


Figure 9-13 **8-12Hz (alpha band) whole brain analysis**, [retrocue minus neutral cue] in the retrocue epoch, on the CARET cortical surface. The focal activation in left MFG observed in the ROI analysis can be seen between 300 and 600ms post-cue. Separate desynchronizations are apparent in lateral visual cortex (bilaterally, around LOC) and in the IPS.

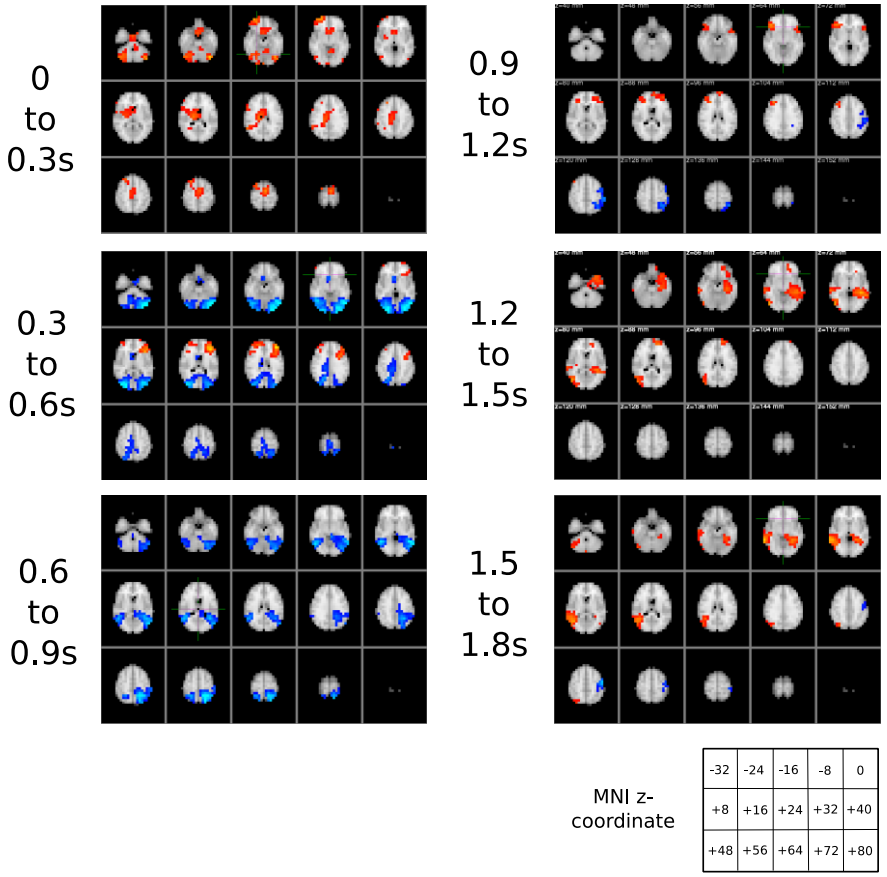


Figure 9-14 **8-12Hz (alpha band) whole brain analysis**, [retrocue minus neutral cue] in the retrocue epoch, in FSL's MNI152 standard brain space