STRATEGIC CONTROL PROCESSES IN EPISODIC MEMORY AND BEYOND

Ravi Dev Mill

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2015

Strategic control processes in episodic memory and beyond

Ravi Dev Mill

University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of

PhD in Cognitive Neuroscience at the University of St Andrews

Submitted 21st November 2014

Acknowledgements

I would like to acknowledge the Scottish Imaging Network: A Platform for Scientific Excellence (SINAPSE) for funding the duration of my PhD. I would also like to thank the School of Psychology & Neuroscience for admitting me into the St Andrews PhD programme. Special thanks go to my supervisor, Dr Akira O'Connor, for guidance and support that greatly eased my PhD experience. I would also like to convey my deep gratitude, firstly to my grandmother, Surinder Pal Kaur, for providing steady encouragement over these three (and a bit) years, and lastly to my mother, Shivneet Kaur Mill, without whose inspiration and advice I would neither have begun nor completed this thesis.

Abstract

The evaluation of past experience is influenced both by the strength of retrieved memories and factors in the immediate retrieval environment, including emphasised goals and cued expectations. However, the laboratory study of episodic memory has neglected such environmental influences, despite their overt contribution to real-world decision outcomes. The aim of this PhD thesis was to rectify this neglect, and clarify the interaction of memory evidence and environmental strategies in the service of strategic memory control. A related aim was to investigate whether control processes identified in the isolated domain of episodic memory in fact performed a more general or “cross-domain” function.

An initial series of behavioural experiments (Experiments 1-3) elucidated an overlooked source of strategic bias in the standard recognition environment – implicit goal emphasis imparted by question format. Experiment 4 investigated whether the question bias was commonly enacted across different domains of evaluation, yielding modest evidence in favour of this underlying cross-domain function. Experiment 5 instantiated more explicit manipulation of goal emphasis and cued expectation, and recovered independent and opposing strategic effects of these two environmental factors, emerging across episodic and non-episodic domains. Experiment 6 employed a simultaneous EEG-fMRI approach to elucidate the neural correlates of memory control, identifying a modulation of the late positive event-related potential during the resolution of mnemonic conflict, which was sourced to BOLD variation in regions of the rostral cingulate zone and intraparietal sulcus. Experiment 7 used pupillometry to examine pupil-linked autonomic systems that have also been implicated in memory control, and isolated two distinct components of the dilation response evoked during environmental conflict – an “early amplitude” unexpected familiarity effect and a “trailing slope” uncertainty effect. The findings illuminate the cross-domain underpinnings of an adaptive memory control system, evidenced in behaviour and across different functional neuroimaging modalities, and across episodic and non-episodic domains of evaluation.

Declaration form

1. Candidate's declarations:

I, Ravi Dev Mill, hereby certify that this thesis, which is approximately 74000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September 2011 and as a candidate for the degree of PhD in Cognitive Neuroscience in September 2011 the higher study for which this is a record and was carried out in the University of St Andrews between 2011 and 2014.

Date signature of candidate

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in Cognitive Neuroscience in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date signature of supervisor

3. Permission for publication: (*to be signed by both candidate and supervisor*)

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED COPY
a) No embargo on print copy

ELECTRONIC COPY
a) No embargo on electronic copy

Date signature of candidate signature of supervisor

Table of contents

Table of figures

Table of tables

Chapter 1: General Introduction

1.1. Environmental influences on real-world memory evaluation

Our evaluations of the past are made against a changing milieu of real-world environments, each bringing with them different influences on the evaluative process and consequent behaviour. In assimilating these variable influences, it is likely that episodic memory overlaps with other cognitive systems in having evolved the capacity to adaptively regulate its core mnemonic processing in light of immediate environmental demands. Nevertheless, laboratory memory testing environments are typically static and impoverished, reflecting the field's predominant focus on 'memory-specific' processes, involved in encoding and retrieval, and the related neglect of the constraints imposed on these domain-specific processes by factors within the test environment. This neglect not only instantiates a significant ecological disparity between real-world and laboratory memory use, but also serves to undermine the conception of memory retrieval as an evaluative process, wherein final memory decisions emerge from the interaction of memory-specific and adaptive environmental processes at the time of test. To demonstrate this, two notable aspects of the test environment that influence overt memory evaluation outcomes in the real world will presently be outlined, namely cued expectations and perceived goals.

Environmental factors that cue expectations of what is likely to be remembered serve as an additional source of diagnostic information beyond the assessed strength of the retrieved memories. To emphasise the contribution of expectation to memory evaluation outcomes, consider the following real-world testing scenarios (modified from Mandler, 1980) in which an observer travelling on a bus evaluates whether different co-passengers have been encountered previously (i.e. deemed ‘old’) or not (i.e. deemed ‘new’; see Figure 1.1. for an accompanying schematic). Firstly, consider “passenger A” – an individual who evokes in the observer a vague sense of prior encounter without specific details that would enable the definitive placement of that individual in a particular episode. This subjective memory state has formerly been characterised as reflective of two independent episodic memory systems i.e. a feeling of “familiarity” in the absence of explicit “recollection” of encoding details (Mandler, 1980; Yonelinas, 2002), and alternatively as reflective of weak levels of a single underlying memory strength continuum (Stretch & Wixted, 1998b; Squire, Wixted & Clark, 2007). Demurring from the long-standing debate over which of these processing accounts best characterises the observer’s retrieval state, it should be apparent that the available memory evidence is only weakly diagnostic of the correct memory decision to be made for passenger A. Furthermore, this weak memory evidence is coupled with the lack of any prior expectations of encountering passenger A in the specific environmental setting of the bus. Hence, weak memory strength and the lack of a cued expectation collectively inform an uncertain, possibly inaccurate memory decision.

Compare this with the observer’s evaluation of “passenger B” – a close friend evoking a high level of memory strength derived from repeated prior encounters. Additionally, this close friend regularly travels on the same bus as the observer, enforcing an environmentally cued expectation of encountering him/her on the bus. Hence, the strong memory strength and strong cued expectation combine to elicit the quick and accurate recognition of passenger B (i.e. a certain ‘old’ decision). Consider now the observer’s evaluation of “passenger C” – another friend evoking a similarly high level of memory strength as passenger B, albeit one who rarely catches the same bus as the observer. Despite the absence of an expectation of encountering this individual on the bus, the high memory strength available is sufficiently diagnostic of the identity of passenger C. The high memory strength therefore acts to countermand the unexpectedness of passenger C and leads to an 'old' recognition decision. Finally, consider the observer's evaluation of "passenger D" – an individual who evokes an equivalently indeterminate level of memory evidence as passenger A, but who is distinguished by wearing a bus conductor uniform. This uniformed identity serves to instil a strong environmentally cued expectation for passenger D, given that it is entirely reasonable to expect a bus conductor on a bus. Hence, this memory evaluation unfolds in an inverted fashion to that involving passenger C, such that the strong expectation overrides the weak memory strength evidence to enable the accurate recognition of passenger D. The outlined scenarios should highlight the significance of environmentally cued expectations in dynamically interacting with retrieved memory evidence to provide the diagnostic evidence for reaching memory decisions.

A second notable way in which the test environment influences memory evaluation is by emphasising the wider goals of the evaluative event, and relatedly conveying the anticipated reward and punishments of particular decision outcomes. These emphasised goals impose higher-order constraints on the evaluative process in light of what is most adaptively relevant in the immediate environment. To demonstrate this, consider again an observer evaluating memory in two different testing scenarios (see Figure 1.2. for a companion schematic). In scenario 1 (Figure 1.2a), the observer is evaluating whether someone spotted whilst shopping in a supermarket is a previously met acquaintance. In scenario 2 (Figure 1.2b), the observer is evaluating whether a member of a line-up of suspects is the perpetrator of a crime to which he/she was an eyewitness. These two scenarios share the same underlying recognition memory process (i.e. adjudicating between a recognised ‘old’ and an unrecognised ‘new’ decision, as in the cued expectation examples above) but differ markedly in the environment in which the engagement of this process occurs. If we assume that the individuals under scrutiny in each scenario evoke the identical level of memory strength, it is likely that these differing environments will provoke very different memory decision outcomes.

In the supermarket scenario, all conceivable evaluative outcomes present a limited opportunity for reward or punishment – the worst consequence of an incorrect recognition decision is some degree of social embarrassment. Hence, the observer might require a fairly weak level of memory evidence to recognise the potential acquaintance as ‘old’, and this increases the likelihood of them making an ‘old’ decision. Contrast this with anticipated behaviour in the eyewitness scenario, in which the repercussions of an incorrect recognition decision are substantial, in potentially leading to the wrongful conviction of an innocent person. Hence, the eyewitness observer might demand a fairly strong level of memory evidence to recognise the suspect as ‘old’, and this deters them from making an ‘old’ decision. The shift in overt behavioural tendencies between these two scenarios reflects how environmentally emphasised goals regulate the use of available memory evidence to ensure adaptive decisions.

The above examples should illustrate the formative influences of cued expectations and emphasised goals on the evaluation of memory in the real world. However, the laboratory study of memory has somewhat neglected these environmental influences and it was the overall aim of my PhD to rectify this neglect. This entailed the use of existing laboratory paradigms, and the development of new ones where necessary, to systematically manipulate aspects of the test environment in which memory evaluation is engaged. Behavioural and functional neuroimaging methods as applied in these paradigms aimed to isolate the effects of expectations and goals on memory evaluation, and piece together the dynamic interactions of these processes, both exogenously with memory strength processes as well as endogenously with each other, in determining evaluative outcomes. This implied interactive property raises a related aim – to treat the "immediate environment" as an important cognitive factor in its own right and interrogate the processing of environmental influences on evaluation in cognitive domains beyond memory. Evidence of an overarching system that mediates environmental modulations of core evidence processing across different cognitive domains would have fundamental implications for wider theorisations of cognition and consciousness.

The remainder of this introductory chapter will discuss findings of behavioural and functional neuroimaging research that are relevant to the outlined aims. I will firstly present a review of a conceptual framework that was adopted in targeting the primary objectives of my PhD – cognitive control. In the brief overview of the control literature that follows, the utility of applying this goal-driven, adaptive framework to the study of memory evaluation should be made apparent, especially as pertains to clarifying the integration of higher-order expectations and goals with memory-specific processing.

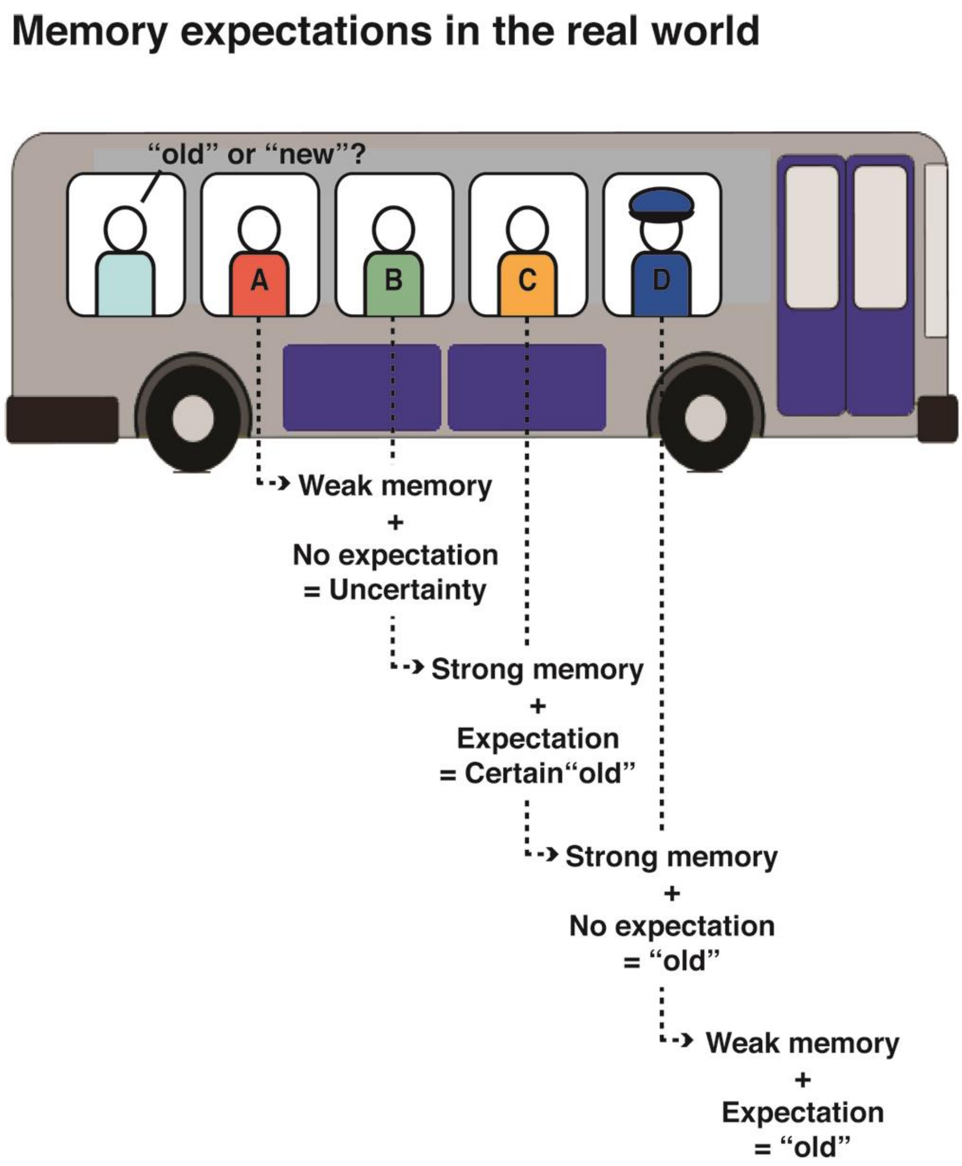


Figure 1.1. Schematic representation of the bus scenario used to demonstrate the influence of environmentally cued expectations on real-world memory evaluation. The passenger in light blue is evaluating recognised 'old' or unrecognised 'new' decisions for passengers A, B, C and D; each of whom is associated with a different combination of evoked memory strength and the presence/absence of expectations of making 'old' decisions, with both factors driving different memory decision outcomes.

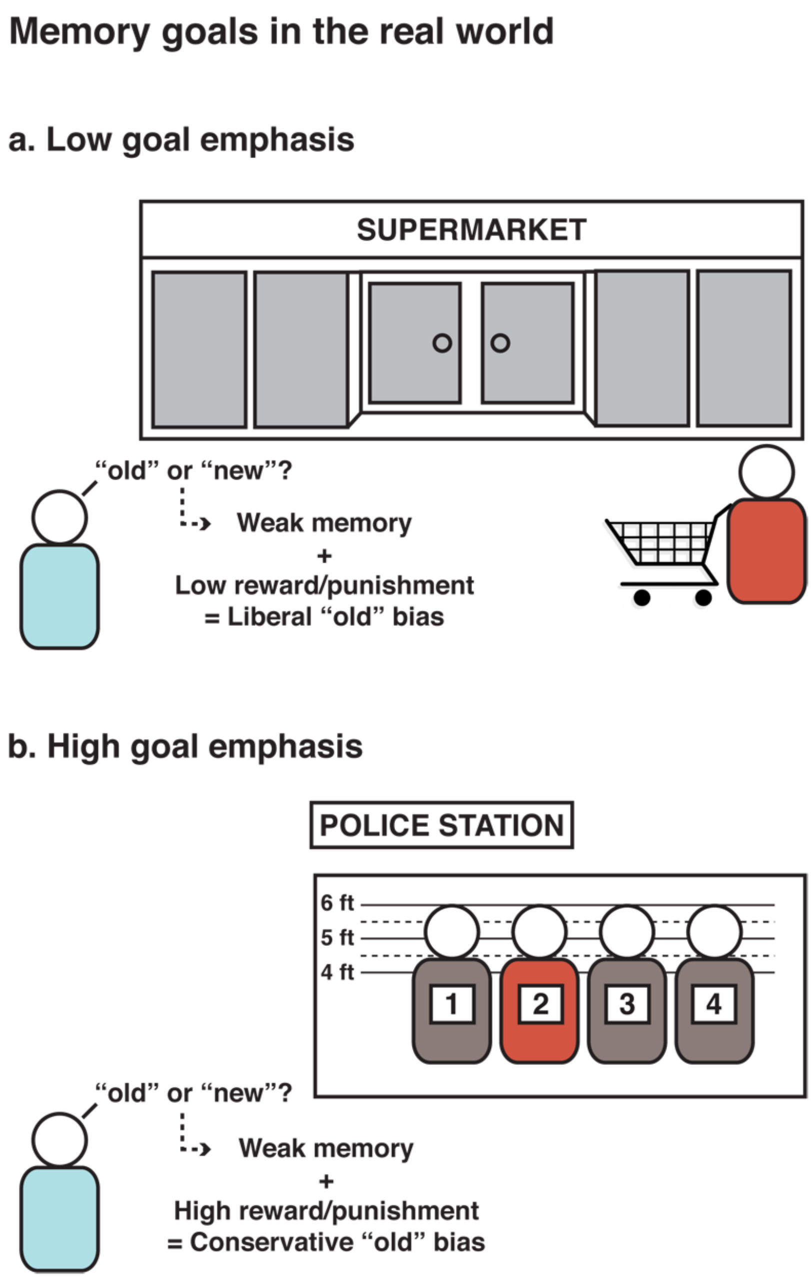


Figure 1.2. Schematic representation of the two scenarios used to demonstrate the influence of environmentally emphasised goals on real world memory evaluation. **a.** Supermarket scenario in which the scope for reward and punishment is reduced (i.e. low goal emphasis). **b.** Eyewitness scenario in which the scope for reward and punishment is increased (i.e. high goal emphasis). In both scenarios, a light blue observer is evaluating a recognised 'old' or unrecognised 'new' decision for an individual in red for whom available memory evidence is equivalently weak. The variation in goal emphasis information across the two scenarios disposes the observer to different decision outcomes – an increased likelihood of responding 'old' in the supermarket scenario ('liberal 'old' bias') and a reduced likelihood of responding 'old' in the eyewitness scenario ('conservative 'old' bias').

1.2. Introducing cognitive control and its relevance to memory

Cognitive control refers to the ability to flexibly adjust thought, emotions and behaviour in accordance with perceived goals and anticipated reward (Baddeley & Della Salla, 1996; Miller & Cohen, 2001; Norman & Shallice, 1986). This construct is central to goal-directed cognition and encompasses a number of sub-processes, including working memory maintenance (Fuster, 1985; Koechlin, Basso, Pietrini, Panzer & Grafman, 1999), selective attention (Banich et al., 2000; Desimone & Duncan, 1995) and action selection (Kerns et al., 2004; Passingham, 1993). The engagement of this panoply of processes is critically dependent on the experience of conflict and subjective uncertainty in the attainment of specified goals (Botvinick, Braver, Barch, Carter & Cohen, 2001; Kahneman & Tversky, 1973), as well as changes to the goals themselves (Monsell, 2003; Kouneiher, Charron & Koechlin, 2009). This is contrasted with conditions of reduced conflict and static goals, wherein pre-potent, reflexive mechanisms of cognition and action dominate (Norman & Shallice, 1986; Shiffrin & Schneider, 1977).

A number of laboratory paradigms have been used to instantiate conflict between possible response options, including the Stroop task (in which participants respond to the relevant perceptual or semantic category of presented stimuli and inhibit the irrelevant one; Stroop, 1935), Flanker task (make a button-press dependent on the direction of a ‘target’ arrow and inhibit the incongruent/interfering direction suggested by other ‘flanker’ arrows in the same array; Eriksen & Eriksen, 1974), go/no-go task (respond on the basis of a frequent/expected cue and inhibit a response for an infrequent/unexpected one; Brutkowski, 1964) and oddball task (respond to infrequent/unexpected ‘target’ stimuli and inhibit a response to frequent/expected ‘lures’; Squires, Squires & Hillyard, 1975). Changes in goal state have also been manipulated in the laboratory (albeit less frequently than response conflict) in paradigms involving rule-learning such as the Wisconsin Card Sorting Test (in which participants sort pictorial cards based on changing rules; Grant & Berg, 1948), task-switching (switch performance between different tasks on the basis of anticipatory cues; Sudevan & Taylor, 1987) and incentivised decision-making (make decisions under variable reward conditions; Damasio, 1994; Small et al., 2005). By engaging under both conditions of response conflict and changes to goal state, control therefore represents an overarching cognitive faculty that integrates variable environmental influences in ongoing decision-making.

At the neural level, it has been suggested that both forms of environmentally-induced control impact via the common instilment of a 'top-down bias', which imposes goal-relevant constraints on ongoing processing (Desimone & Duncan, 1995; Miller & Cohen, 2001). Neurophysiological research in non-human primates has localized this bias to neuronal populations of prefrontal cortex (PFC; Fuster, 1980; Goldman-Rakic, 1987; Duncan & Desimone, 1995), with added functional significance imparted to interactions of PFC and downstream parietal cortex, with which it shares dense anatomical interconnections (Cavada & Goldman-Rakic, 1989; Chaffee & Goldman-Rakic, 1998; Chaffee & Goldman-Rakic, 2000). Evidence from functional neuroimaging research in humans also supports prominent roles for these regions in mediating cognitive control. The use of electroencephalography (EEG) and analysis of the event-related potential (ERP) has delineated temporally precise ERP markers of cognitive control that have been sourced to prefrontal and parietal scalp generators, including the N2 (Näätänen & Gaillard, 1983) and P3 complex (Sutton et al., 1965; Donchin, 1981).

Research involving functional magnetic resonance imaging (fMRI) has also yielded control activations in prefrontal and parietal regions, which have benefited from the method's greater spatial resolution to enable localisation to more precise sub-regions. fMRI activations underlying response conflict have been observed in medial PFC (mPFC; Ridderinkhof, Ullsperger, Crone & Nieuwenhuis, 2004), cingulate cortex (including the anterior cingulate cortex and the rostral cingulate zone, ACC and RCZ respectively; Botvinick, Cohen & Carter, 2004; Braver, Barch, Gray, Molfese & Snyder, 2001; Ridderinkhof et al., 2004), lateral PFC (lPFC; Koechlin, Ody & Kouneiher, 2003; Kouneiher, Charron & Koechlin, 2009), and the regions of the inferior parietal lobule (IPL; Bunge, Hazeltine, Scanlon, Rosen & Gabrieli, 2002; Bunge, Dudukovic, Thomason, Vaidya & Gabrieli, 2002). Similar prefrontal regions have also been linked with controlled processing in response to changes to goal state and anticipated reward, including mPFC (Kouneiher et al., 2009), ACC/RCZ (Botvinick, 2007; Bush et al., 2002) and lPFC (Macdonald, Cohen, Stenger & Carter, 2000). Further, the spatial affordances of fMRI have enabled the identification of sub-cortical brain regions involved in control including the thalamus, which has been linked with the engagement of control under conflict (Aron, Behrens, Smith, Frank & Poldrack, 2007; Macdonald et al., 2000), and the dorsal striatum, which has been linked with the regulation of control under varying goals and associated reward (Delgado, Stenger & Fiez, 2004; Elliott, Friston & Dolan, 2000; Knutson, Fong, Adams, Varner & Hommer, 2001). The highlighted regions often engage in parallel to implement control (e.g. Hazeltine, Poldrack & Gabrieli, 2000; Kerns et al., 2004), and this implied network-level engagement has recently been formalised by functional connectivity methods recovering a distributed fronto-parietal control network

(Vincent, Kahn, Snyder, Raichle & Buckner, 2008) that has been highlighted as one of the core, independent functional networks of the human brain (Power et al., 2011).

It should be noted that the provided overview of the cognitive control literature is general, and has therefore obviated some of the conceptual nuances surrounding the precise engagement of control (e.g. the proposed segregation of control between ‘early’ conflict detection and ‘late’ control allocation sub-processes; Botvinick et al., 2001) and functional neuroimaging attempts at dissociating the sub-components of the associated neural response, both in the EEG (e.g. the proposed early/late segregation of the N2 into the N2b and N2c components, Folstein & van Petten, 2008; as well as the early/late segregation of the P3 into the P3a and P3b; Polich, 2007) and fMRI literatures (with mPFC and cingulate regions linked with conflict detection, and lPFC and IPL linked with control allocation; Macdonald et al., 2000; Bunge et al., 2002; Brass & von Cramon, 2005). Nevertheless, the summarised literature should convey the significance of the core construct of cognitive control, in recruiting a dedicated neural system across different task contexts to promote flexible and adaptive interactions with the environment.

The preceding overview should therefore allude to the utility of applying a cognitive control framework to the study of environmental influences on episodic memory evaluation - the primary focus of my PhD. Indeed, both environmental influences highlighted in the introductory section, namely cued expectations and emphasised goals, are readily accommodated within a control framework. To emphasise this, consider again the previously described bus scenario, wherein the absence of environmentally cued expectations at the time of evaluating passengers A and C led to uncertainty as to the correct memory decisions to be made. This uncertainty could be characterised as a mnemonic form of the response conflict investigated in the control literature, serving to heighten the control with which available memory evidence is evaluated. This controlled memory processing was unsuccessful in diagnosing the identity of passenger A, whereas it was successful in diagnosing the identity of passenger C. Conflict was also instantiated in the evaluation of passenger D (the bus conductor), albeit in inverse fashion, with the lack of diagnostic memory evidence driving the subsequent controlled assessment of the environment. These effortful evaluations under conflict are contrasted with passenger B, for whom high levels of memory strength and strong cued expectations precluded the need for control.

Now also reconsider the second real-world example, involving memory evaluations made in two environments with marked differences in emphasised goals – a supermarket and a suspect line-up. It should be apparent that the described behavioural adjustment in the eyewitness scenario, reflecting the demand for greater memory evidence under a situation of greater goal emphasis and projected consequences, parallels the recruitment of cognitive control in variations to the goal state (especially those contrived in incentivised decision-making). As in other forms of decision-making, memory decision outcomes that carry a greater propensity for reward or punishment are likely to be evaluated more rigorously and this is reflected in control recruitment. The reassessment of the previous real-world examples should emphasise that unexpected events generate response conflict and goal-emphasised events induce changes to goal state, with both processes heightening the cognitive control over memory evaluation.

Attempts to draw such direct parallels with findings and perspectives in the field of cognitive control have been lacking in episodic memory research. Rather, the primary focus of the majority of existing memory models is on characterising the nature and strength of retrieved memories, with a comparative neglect of the environmental constraints imposed in the evaluation of this output. Such ‘retrieval output’ frameworks include global matching models (Gillund & Shiffrin, 1984; Hintzman, 1988), threshold models (Johnson, Kounios & Reeder, 1994) and both single process (Stretch & Wixted, 1998b) and dual process variants of the signal detection model (Yonelinas, 2002). However, a growing body of research interrogating the higher-order strategies engaged at retrieval, both within the confines of the above retrieval output models and in bespoke models centring on strategic processes, implicate similar underlying processes to those emergent from the cognitive control literature. These ‘memory control’ frameworks are summarised below.

1.3. Memory models that implicate aspects of control

One of the first attempts at characterising controlled aspects of memory evaluation was provided by the source monitoring framework (Johnson, Hashtroudi & Lindsay, 1993). The model originally sought to explain how memories are attributed to their previously encoded ‘source’ and relatedly how mistaken source attributions arise of the type reported in fallacious eyewitness testimony (Loftus, 1996; Lindsay & Johnson, 1989). Central to the model is a parcellation of the subjective experience of memory evaluations between ‘automated’ attributions, driven by retrieved source details of sufficient quality, and ‘controlled’ attributions, driven by an effortful construction of source details (Johnson & Raye, 1981). Subsequent iterations have expanded beyond source memory to encompass all instances when retrieval is subjectively characterised as effortful and constructive, which includes old/new recognition decisions (Johnson et al., 1993). Empirical research has shown that the efficacy of source monitoring is to a large extent determined by aspects of encoding, as demonstrated by the difficulties in source monitoring observed for encoding sources that are perceptually similar (Ferguson, Hashtroudi & Lindsay, 1992). Notably, source monitoring has also been shown to be sensitive to aspects of the retrieval environment, with source accuracy improving with a longer response window (facilitating adequate completion of effortful processing; Johnson, Kounios & Reeder, 1994) and with full compared to divided attention at test (consistent with the need to expend focused effort; Jacoby, 1991).

The importance of these environmental influences is further emphasised by application of the source monitoring framework to the Loftus suggestibility procedure – a laboratory analogue of eyewitness testimony (Loftus, 1996). In the task, participants are first shown visual stimuli (e.g. a film of a car crash) and are later presented with a memory test, with verbal questions about the event framed in either 'neutral' (e.g. "Did you see a headlight?") or 'suggestible' terms (e.g. "Did you see *the* headlight?"). In a second and final memory test, the suggestible condition has been found to increase the likelihood of falsely recognising stimuli that were not presented in the original film. Lindsay and Johnson (1989) demonstrated that the suggestibility effect was removed when the format of the final memory test was framed as a source monitoring task (i.e. a "source?" prompt accompanied by "visual", "verbal" or "both" response options), whereas it was preserved if the test was framed as standard old/new recognition. Participants were hence prompted by the source memory format to heighten their control over retrieval, and this parallels behaviour in the real-world eyewitness test scenario described previously, in which greater memory evidence was demanded in the eyewitness environment that framed more prominent goals and decision consequences.

Research conducted under the source monitoring framework has therefore been valuable in highlighting impacts of the test environment on the strategies adopted by memory evaluators to reach final decisions. Further, the implied effortful qualities can be extended to capture the previously described engagement of control in response to greater emphasis on goals and reward/punishment-related information in the test environment, as well as when cued expectations conflict with memory evidence. However, a limitation of this framework lies in its broad scope and associated lack of explanatory specificity. Although encoding and retrieval factors that generate the subjective experience of effort are accommodated in the framework, precisely how these experiences arise from either memory-specific retrieval processes or wider goal-driven processes is unclear.

Indeed, Jacoby, Kelley & McElree (1999) elaborated on the source monitoring framework with the aim of more precisely delineating its constituent control mechanisms. This model more clearly situates itself within the dual process signal detection model of retrieval output (Yonelinas, 2002) to address the above noted ambiguity as to how control precisely operates in source monitoring. Automated processes are now assumed to operate through 'familiarity' and effortful processes through 'recollection', and it is by prioritising the retrieval of recollected content (and associated high levels of memory strength) as the diagnostic basis for making memory decisions that memory control is primarily exerted. Support for this dual process control framework comes from variations of the process dissociation procedure (Jacoby, 1991), wherein participants are asked to either respond 'old' irrespective of encoded source ('inclusion'), or respond 'old' only for items from a specific source ('exclusion'). Correct performance in the inclusion condition is assumed to reflect the contributions of both familiarity and recollection, whereas performance in the exclusion condition reflects recollection alone (given the need to retrieve source details), and modelling the difference in performance between these conditions yields quantifiable estimates of familiarity and recollection. Empirical findings from the process dissociation procedure have confirmed that test manipulations which interfere with source monitoring processes, such as response speeding and dividing attention, impact by selectively reducing recollection estimates (Jacoby, 1991; Jacoby, Toth & Yonelinas, 1993). Conversely, manipulations that enhance automated processing selectively increase levels of familiarity, as demonstrated by the 'false fame' effect (wherein faces that have been repeated to enhance familiarity are more likely to be mistakenly recognised as famous; Jacoby, Kelley, Brown & Jasechko, 1989) and manipulations of perceptual fluency (wherein the ease or fluency with which test items are perceived increases the likelihood of their false recognition; LeCompte, 1995).

Jacoby et al.'s (1999) model therefore makes valuable formalisations of the memory control mechanisms hinted at in the source monitoring framework. In emphasising the role of strategically monitoring for recollection and associatively high memory strength, the model provides valuable insight into precisely how memory evidence is constrained by immediate goals. Indeed, the model goes further in outlining two methods by which this recollective source monitoring strategy can be implemented, either as the 'late correction' of false feelings of familiarity or by the 'early selection' of recollected content at retrieval (also referred to as 'source-constrained retrieval'; Jacoby, Shimizu, Daniels & Rhodes, 2005). By inference, late correction is putatively involved in the 'reactive' heightening of control under response conflict, and early selection is the likely mechanism underlying the 'proactive' heightening of control under goal emphasis, and this reactive/proactive control framework has direct descendants in the cognitive control literature (Braver, 2012).

However, in assuming that memory control impacts *exclusively* by prioritising the recovery of recollection, Jacoby et al.'s framework fails to accommodate decision-making under conditions wherein recollection is not successfully engaged in the service of control. In the precipitant cases of low memory strength and associated uncertainty, environmentally-cued expectations and emphasised goals are likely to introduce a more direct influence on decision outcomes, as illustrated by the expectation-driven identification of the episodically unfamiliar conductor in the earlier bus scenario (see Section 1.1.; see Figure 1.1.). Indeed, control mechanisms in both the Jacoby et al. framework and the original Johnson et al. (1993) source monitoring framework, are more precisely aimed at characterising the control over retrieval output, 'retrieval control', rather than the control over how this retrieval output is used to reach decision outcomes, 'decision control'. This latter specification necessitates a wider understanding of precisely how the environment drives the heightening of memory control, which requires greater appreciation of the adaptive constraints applied to the entire evaluative process beyond the mere output of memory evidence.

In an attempt to venture beyond the confines of the dual process framework and interrogate these wider constraints, Koriat & Goldsmith (1996) proposed a model in which memory control is mediated by an overarching system with 'metacognitive' properties. This metacognitive system is further deconstructed as comprising a 'monitoring' function, which assesses the quality of retrieval output and associated memory strength, and a 'control' function, which determines whether this memory evidence is acted upon in light of immediate environmental goals. The monitoring component is assumed to underlie the assessment of the strength of retrieved memories, with this information outputted as a probability that the assessment is accurate. This characterisation blurs as to the precise processes contributing to the assessed memory strength (e.g. the relative involvement of familiarity and recollection) and instead emphasises how accurate the overall assessment is in diagnosing decision outcomes (termed 'monitoring effectiveness'). Monitoring effectiveness has been substantiated by empirical dissociation of the accuracy of memory decisions and the subjective confidence in them, with the disparity between the two performance measures enhanced by 'deceptive' stimuli that elicit false memory experiences (Koriat & Goldsmith, 1996). This effect is not anticipated by extant models of retrieval output (including the both variants of the signal detection model) wherein confidence is considered a fine-grained analogue of memory strength and the relationship between retrieval and memory strength is direct.

The second 'control' component of the metacognitive system drives the final decision to act or withhold from acting on the basis of both the monitored memory strength and the wider goals constraining the evaluative event. These goals are indexed by a criterion measure that is adjustable according to the level of memory strength needed to diagnose decisions. The stipulation that control is the critical determinant of action is one of the core features of the Koriat & Goldsmith model, one that captures an ecological disparity between real-world instances wherein memory users have the opportunity to withhold from acting, and the majority of laboratory test formats that force participants to make either an 'old' or 'new' decision. In support of this, empirical findings have highlighted how the provision of a 'withhold' option in 'free report' memory task formats (as opposed to 'forced report' formats that lack this option) can dramatically influence decision outcomes, by improving the accuracy of memory decisions at the expense of their frequency (Koriat & Goldsmith, 1994). Further, this tendency to improve memory quality at the expense of quantity has been shown to increase under conditions of higher incentives (Koriat & Goldsmith, 1996), and is seemingly a reliable behavioural index of memory control. Indeed, similar overt caution was outlined in the earlier eyewitness real-world scenario, wherein heightened goal emphasis led to a similar quality-oriented strategy (see Section 1.1.; see Figure 1.2.).

Koriat & Goldsmith's framework therefore illumines a number of important features of memory control. In highlighting the tendency for research predicated on the various retrieval output models to focus on retrieval quantity rather than quality, the authors illustrate a notable ecological disparity in the laboratory study of memory and the real-world (wherein quality is often the overarching goal; Koriat & Goldsmith, 1994). The demonstration of a reliable behavioural correlate of memory control – namely an increase in memory accuracy combined with a reduction in quantity – is also significant and a mechanistic account of this effect will be delineated in the next section. Collectively, the metacognitive framework emphasises the influence of aspects of the environment (namely task format and incentives) in impacting on memory evaluation, belying their common treatment as nuisance variables in the majority of episodic memory research (Nelson & Narens, 1994; Koriat & Goldsmith, 1996). The model also formalises the previously alluded to distinction between control enacted directly on retrieval (the monitoring sub-function i.e. retrieval control) and control enacted on the output of retrieval (the control sub-function i.e. decision control), with a particular emphasis on the formative influence of the latter form of control. One limitation of the framework is its inadequate delineation of how control is engaged under response conflict (contrasting the considerable detail afforded to goal emphasis). Additionally, empirical substantiations of the model rely on an atypical memory task involving tests of general knowledge, which might recruit a semantic rather than episodic memory system (Tulving, 1972). Nevertheless, aspects of this framework will bear centrally on the remaining introduction and ensuing experimental chapters.

Koriat & Goldsmith's control framework was elaborated upon in two later models, both of which adapted aspects of the original model for investigation in the standard recognition memory task format. Whittlesea's selective construction and preservation of experience model (SCAPE: Whittlesea, 2002) posits a simplified account of memory control, in which the 'direct products' of memory are subjected to a central 'evaluation' function, which both monitors for the strength of the memory products and the wider environmental context. SCAPE shares a similar obviation of the precise mechanisms of retrieval output as the Koriat and Goldsmith model, although it collapses across the monitoring and control sub-functions in favour of a unified central evaluative function. Empirical evidence for this model rests on one experimental paradigm in which expectations cued by aspects of the recognition test environment influence are shown to influence overt behaviour. Specifically, participants make 'old' or 'new' decisions for sentence stems whose potential completing word is either semantically constrained (e.g. "After the accident he was covered in…" generates only a few candidate completing words) or semantically differentiated (e.g. "It was his job to clean out the…" generates a number of candidate completions; Whittlesea, 2002; Whittlesea, 2004). Constrained sentences set up strong expectations of the status of the completing word, which appear embedded in the sentence after a short delay. Expectations instilled by the constrained sentence stems were found to increase the likelihood of recognising the completing word as ‘old’ (relative to unconstrained sentences) regardless of whether this word was actually previously encountered, with this effect attributed to the erroneous evaluation of the subjective ‘sense’ of the semantic expectation as indicative of episodic familiarity.

The expectation effect described by Whittlesea is therefore not entirely reconcilable with those outlined previously in the real-world bus scenario, in that the former effect putatively impacts via retrieval control mechanisms (i.e. on retrieval directly) whereas the latter impacts via decision control mechanisms (i.e. on the products of retrieval). Despite the ambiguous treatment of these two control sub-processes, the conception of the evaluative function as a centralised system that makes attributions across different forms of memory raises intriguing notions of ‘cross-domain’ controlled processing. Exploring the cross-domain aspects of observed memory control processes served as a salient avenue for my PhD research and will be introduced more fully in a dedicated section (see Section 1.6).

Benjamin (2007) proposed a similar centralised control node in his model of memory control, wherein decision outcomes emerge from the interaction of a ‘simple’ memory system (underlying retrieval output) and a ‘complex’ higher-level system sensitive to varying aspects of control. Unlike the frameworks presented above, Benjamin’s control construct is more pervasive in also mediating goal-driven strategies adopted at encoding, as demonstrated empirically by participants’ observed ability to strategically increase study time for items perceived as difficult to encode (Son & Metcalfe, 2000) and those items associated with higher reward (Castel, Benjamin, Craik & Watkins, 2002). A control system that constrains both encoding and retrieval processing in light of immediate goals is consistent with the overarching function suggested by the dedicated cognitive control literature (Miller & Cohen, 2001; Norman & Shallice, 1986).

Further, the influence of this higher-level control system on retrieval is operationalised in analogous fashion to the retrieval control and decision control process dichotomy outlined earlier. Indeed, Benjamin's model coalesces with Koriat & Goldsmith's model in placing particular emphasis on decision control processes as the crucial means by which environmental goals impact on behaviour (unlike retrieval control which is posited to be determined to a large degree by aspects of encoding). A pivotal role is given to the presence of low or non-diagnostic memory strength in driving the need for control, resulting in either a decision to disincline from acting (as proposed by Koriat & Goldsmith, 1996) or to recruit non-episodic inferential strategies as an alternative basis for acting (Reder & Wible, 1984). This latter finding parallels behaviour in the bus real-world scenario in which the identity of 'unfamiliar' passenger D was inferred from his bus conductor uniform to generate the environmentally-cued expectation. Overall, Benjamin's model adds to the emerging emphasis placed on decision control rather than retrieval control as the crucial medium by which adaptive constraints are introduced on memory evaluation. The model also formalises the important role of environmental factors in driving the need for memory control, which was only hinted at by the original Koriat & Goldsmith framework.

In summary, the models outlined above highlight a number of features of memory control that are pertinent to my research aims. Johnson et al.'s (1993) source monitoring framework makes the important distinction between subjective experiences of memory evaluation as being either automated/reflexive or effortful/controlled. Jacoby et al. (1999) illustrate the prioritisation of highly diagnostic memory evidence as a significant method by which control is heightened over memory evaluation. Koriat and Goldsmith (1996) highlight the goal-driven prioritisation of memory quality over quantity as another memory control mechanism, and formalise the important distinction between control processes acting directly on retrieval (defined here as retrieval control) and those that act on the output of retrieval (defined as decision control). Whittlesea (2004) posits a centralised control function, one which putatively overlaps (to some degree) between retrieval and decision control, and is potentially capable of modulating processing in episodic and non-episodic domains. Finally, Benjamin (2007) emphasises the role of the environment in mediating the pervasive involvement of control at both encoding and retrieval in the service of task goals. These extant models therefore provide an equivalent conception of control to that emerging from the dedicated cognitive control literature i.e. that of a cognitive faculty exerting goal- and reward-related modulations of ongoing processing. However, the utility of drawing more direct parallels between these largely separated episodic memory and cognitive control literatures still remains, in that the above models variously focus on explaining mechanisms of control under *either* expectation-driven conflict or variations in goal state. Added reference to the cognitive control literature can accommodate both these control engagements in a unified adaptive framework, in which impediments in attaining task goals (under expectation-driven conflict) and changes to task goals themselves (variations in goal state) both lead to the heightening of a common cognitive control process.

Aspects of the above memory control models that will bear centrally on the remaining introduction and ensuing experimental chapters are worth highlighting. Firstly, a corollary of the emergent distinction between retrieval control and decision control suggests that the latter process is perhaps most amenable to adaptive environmental constraints. In support of this, prior research has highlighted the limits of retrieval control as the observed inability of participants to increase the *quantity* of memories generated at test under incentives (Nilsson, 1987). The inability to generate more memory evidence in response to test incentives is consistent with the notion that even the subjectively experienced constructive/effortful features of retrieval control are crucially predicated by processes engaged at encoding. Conversely, test incentives can affect the *quality* of endorsed memory evidence dramatically, by influencing participants' decision to act or withhold from acting on the basis of their retrieval output (Koriat & Goldsmith, 1996). This pattern of results suggests that control enacted at test is limited in its ability to improve the diagnosticity of memory evidence, but impacts more tangibly on behaviour by regulating how the available memory evidence is used to make adaptive decisions. Decision control is therefore most sensitive to goal-related information imparted by the immediate test environment and exerts a more pronounced influence on behaviour than retrieval control.

The highlighted importance of these decision control processes, combined with their relative neglect in the study of episodic memory, motivated a primary focus on decision control for my PhD research. In this endeavour, I adopted a similar conception of 'memory-specific' retrieval processes as in Koriat and Goldsmith's model (and later adopted by Benjamin), in outputting an assessed level of memory strength that varies in its diagnosticity for ensuing memory decisions. The relative contributions of familiarity or recollection to that assessment was therefore not of primary interest. This strategy conveys the aims of my PhD in not only seeking insights into memory via a greater scrutiny of cognitive control, but also a reciprocal insight into cognitive control via closer scrutiny of memory. Specifically, environmental factors that impact on decision control in the service of episodic memory putatively exert similar influences on decision-making in other domains. Hence, a reliance on memory-specific concepts such as familiarity and recollection in characterising these memory control mechanisms might serve to detract from their potentially wider influence in domains beyond episodic memory.

The conceptual review provided above will now be complemented by a review of empirical findings from behavioural and functional neuroimaging investigations of memory control. The ensuing sections should therefore raise more tangible predictions for behavioural and functional neuroimaging correlates of memory control, which were targeted in my experimental chapters.

1.4. Behavioural correlates of memory control

Research interrogating the behavioural signifiers of memory control has typically employed the single item recognition paradigm, in which participants perform an initial 'study' phase to facilitate encoding of presented stimuli (typically words or images), followed by a 'test' phase in which they decide if stimuli now being presented are 'old' i.e. encountered before at study, or 'new' i.e. not encountered during the experiment. The prevalence of single item recognition in the field of episodic memory rests on its simple task format, which affords taut experimental control over aspects of encoding and retrieval. Indeed, empirical findings from this paradigm have formed the basis for the majority of the retrieval output models alluded to in the previous section. The most dominant of these – the signal detection theory model of recognition (SDT; Green & Swets, 1966; Macmillan & Creelman,

2005) – has been suggested to accommodate aspects of memory control and is hence worthy of more detailed description.

SDT models an individual's memory evaluation as a process of detecting an *old*-item recognition 'signal' embedded against a background of *new*-item 'noise' (see Figure 1.3.). This detection process is predicated on the existence of an underlying continuum of memory strength, upon which the memory strength evoked by *old* and *new* items is separately and normally distributed. The distance between the two distributions reflects the individual's sensitivity in discriminating between *old* and *new* items, and is denoted by the parameter d'. This is calculated as follows:

$$d' = z(H) - z(FA) \tag{1}$$

where z(H) is the normalised hit rate (the proportion of *old* items correctly recognised as 'old') and z(FA) is the normalised false alarm rate (the proportion of *new* items incorrectly recognised as 'old'). Increasingly positive d' values therefore denote a greater separation between the overall levels of memory strength evoked for *old* and *new* items respectively, leading to greater sensitivity in discriminating between them.

Bearing out this association with memory strength, prior recognition research has demonstrated that d' is largely determined by aspects of encoding. For instance, improved sensitivity has been linked to the inherent memorability characteristics of presented stimuli, such as low frequency (Glanzer & Adams, 1985), high concreteness (Groninger, 1976), high imageability (Moeser, 1974) and high orthographic distinctiveness (Zechmeister, 1972). Decreasing the number of stimuli in the study list to effectuate easier encoding episodes has also been shown to improve d' (Ohrt & Gronlund, 1999). Sensitivity also tracks the levels-of-processing achieved by participants at encoding (LOP; Craik & Lockhart, 1972), as illustrated by the increase in *d′* when study items are processed under ‘deep’ relative to ‘shallow’ LOP conditions (Lockhart, Craik & Jacoby, 1976; Maddox & Estes, 1997). The evidenced adjustment of *d′* following various encoding manipulations supports a role for this parameter in indexing the levels of memory strength available at test. Referring back to the retrieval control/decision control dichotomy emerging from the previously summarised memory control models, factors contributing to the assessment of memory strength and the exertion of ‘retrieval control’ over this assessment have been shown to be relatively impervious to test environment manipulations of cognitive control (Nilsson, 1987). Rather it is in the strategic regulation of how the available memory evidence is used in the service of decision control that the test environment most tangibly impacts on behaviour (Koriat & Goldsmith, 1996). Hence, the link between *d′* and memory strength processes render it unlikely as an index of the decision control processes of primary interest to my PhD aims[1].

However, SDT operationalises a second parameter that is more closely linked with aspects of decision control – response criterion or *c*. Criterion acts as a threshold for the amount of memory strength demanded to diagnose an ‘old’ decision, and is calculated as follows:

[1] It is worthwhile mentioning the treatment of old item variance (σ) in different versions of the SDT model. In the simplest version (the equal variance SDT model; Green & Swets, 1966), the variance of old and new item distributions is assumed equal. Subsequent models have modified this assumption so as to provide better fits to recognition data that includes ratings of subjective confidence (Yonelinas, Dobbins, Szymanski, Dhaliwal & King, 1996). The ‘unequal variance’ SDT model fixes the new item variance at 1 and estimates the variance of the old item distribution as a freely varying parameter (Stretch & Wixted, 1998b). Conversely, the dual process SDT model maintains the equal variance assumption and instead includes an additional estimated parameter *R*, which reflects the proportion of *old* items identified with high confidence on the assumed basis of recollection (Yonelinas et al., 1996). The differential treatment of old item variance in these two models lies at the centre of the single vs. dual process debate on the nature of retrieval output. As has been highlighted previously, this debate does not bear centrally on the investigation of decision control processes that are the primary focus of my PhD. Hence, to more adequately address this focus, the ensuing experimental chapters will primarily focus on the simplest equal variance SDT model, and focus on the variation in *d′* and *c* (parameters shared by all the above models) under different control manipulations.

$$c = -0.5 * [z(H) + z(FA)] \tag{2}$$

When criterion is placed at the intersection of the *old* and *new* item distributions, memory evaluation is considered to be statistically ‘optimal’ and unbiased by the influence of higher-order strategies. Strategies are herein defined as any source of information beyond the assessed strength of the available memory evidence that actively contributes to decision outcomes. The adoption of strategies is reflected by criterion shifts away from optimal placement, such that positive shifts indicate a reduced likelihood of responding ‘old’ (commonly referred to as a ‘conservative’ response strategy) and negative shifts indicate an increased likelihood of responding ‘old’ (a ‘liberal’ response strategy; see Figure 1.3.). Hence, memory evaluations under the SDT model are assumed to emerge from the interaction of memory strength and strategic processes, as reflected in estimates of *d′* sensitivity and criterion bias respectively.

Early recognition research conducted under the SDT model treated criterion as a nuisance variable relative to memory strength processes of primary interest (e.g. Peterson, Birdsall & Fox, 1954; Blackwell, 1963). However, subsequent research has highlighted the functional significance of this parameter in indexing various higher-order strategies that contribute significantly to decision outcomes. For instance, Hirshman (1995) found that subjects adjusted their criterion at test in accordance with the inferred strength of encoding achieved in the prior study phase. The experiment involved two study-test recognition phases: one in which greater time was afforded to encode items (encouraging the perception of ‘strong’ encoding) and one in which lesser encoding time was afforded (encouraging the perception of ‘weak’ encoding). Criterion placement at test was more liberal following the perceived weak compared to the perceived strong encoding phase, reflecting a strategic enhancement of the tendency to respond ‘old’ to counteract the inferred weakness of the preceding encoding phase. Criterion adjustment has also been linked with inferential strategies emerging solely in the test phase, such as those instilled by manipulations of the base rate proportion of *old* to *new* items in the test list. For example, Ratcliff, Sheu and Grondlund (1992) varied the base rate to affect a selective overabundance of *old* items or *new* items in separate recognition test phases, and observed shifts in criterion that reflected efforts to make the overabundant response more often. Adjustments in criterion can therefore serve as a valuable index of higher-order strategies impacting on the use of available memory evidence.

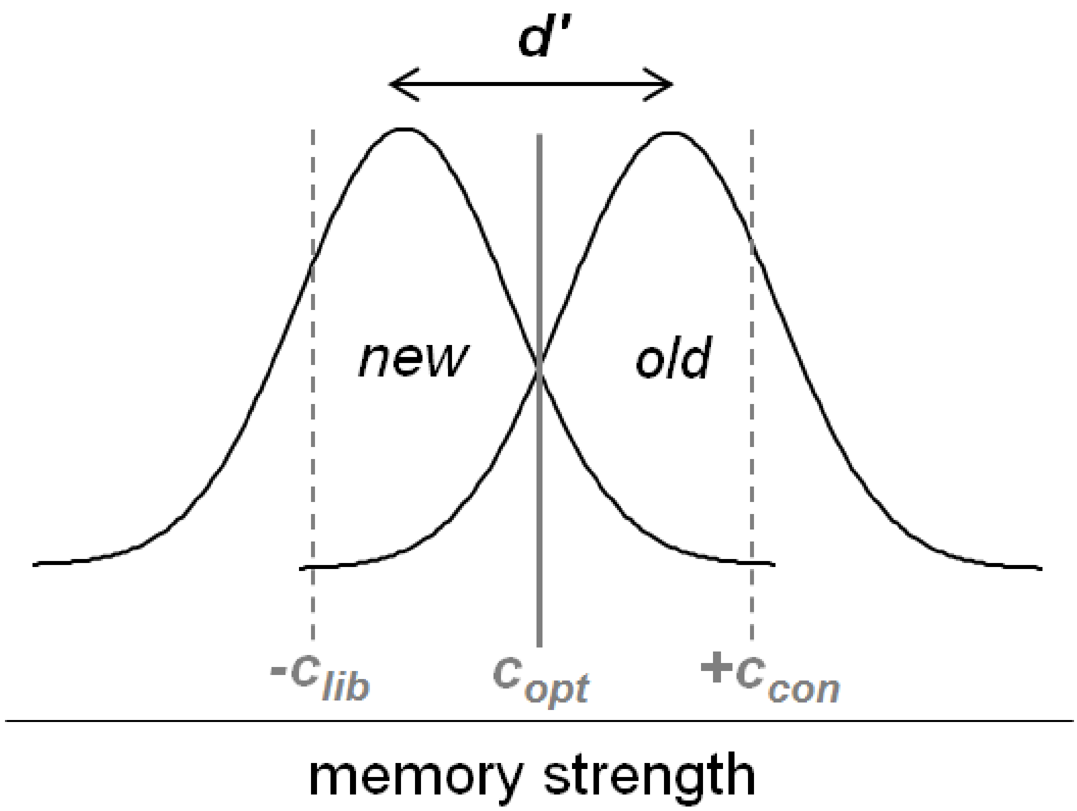


Figure 1.3. Schematic depiction of the signal detection model of retrieval output, with *new* and *old* item normal distributions overlaid on an underlying continuum of memory strength. *d'* = discrimination sensitivity, as indexed by the distance between the two item distributions. c_{opt} = optimal i.e. unbiased criterion placement at the intersection of the two item distributions; $-c_{lib}$ = negative i.e. liberal criterion placement, increasing the likelihood of responding 'old'; $+c_{con}$ = positive i.e. conservative criterion placement, reducing the likelihood of responding 'old'.

Other studies that have failed to elicit strategic criterion shifts serve to emphasise the important mediating influence of participants' willingness to adopt higher-order strategies in their memory evaluations. This willingness depends on the precise nature of the manipulations enacted, in terms of the perceived effort required in attending to them as higher-order sources of information and the perceived utility in doing so to improve recognition performance. For instance, follow-up studies to Hirshman's (1995) encoding strength effect have found that intermixing the strength manipulations within the same test block (as opposed to Hirshman's blockwise manipulation) fails to induce shifts in response criterion, even when differentially strengthened items are explicitly cued by different presentation colours at test (e.g. red presentation colour suggests strongly encoded test items, and green presentation colour denotes weakly encoded test items; Stretch & Wixted, 1998a; Morell, Gaitan & Wixted, 2002). This null effect is partially attributable to the high cognitive effort required to attend to perceived strength as a source of strategic information on a trial-by-trial as opposed to a blockwise basis.

Equivocal effects of base rate manipulations on criterion placement have also been linked to differences in how aware subjects are made to these manipulations. For instance, base rate manipulations act to shift criterion only when the actual proportion of *old* to *new* items is explicitly provided to participants (Ratcliff et al., 1992), with vague mention of the manipulation that lacks explication of the precise proportions found to be ineffective (Herron, Quayle & Rugg, 2003). Estes and Maddox (1995) highlighted performance feedback as another means of heightening strategic awareness, by observing criterion shifts in response to base rate manipulations only in the presence of trial-by-trial feedback. Similar criterion shifts have been evidenced following feedback that bears no relation to actual performance (i.e. false feedback; Han & Dobbins, 2009). Both the mediating effects of the inferred effort and utility in strategically shifting response criterion are accommodated by the metacognitive aspects of the previous memory control models (e.g. Koriat & Goldsmith, 1996). Specifically, if subjects perceive the use of higher-order strategies as difficult or extraneous to the overarching goal of making quick and accurate memory decisions, then they will revert to the assessed memory strength as the sole source of diagnostic evidence.

Indeed, participants are more willing to adopt higher-order strategies if these are more *explicitly* suggested by aspects of the test environment and therefore require less of an inferential process when being integrated into memory evaluation. Re-examining Stretch and Wixted's (1998a) inability to shift criterion despite providing coloured cues at test, it should be apparent that a high degree of inference was required to utilise these cues effectively, given that they in of themselves only probabilistically suggest the strength of encoding for the *old* items presented in that colour, and therefore have to be further translated into a usable likelihood of encountering *old* or *new* items to guide performance. Subsequent research has shown that cues which more directly denote the likely old/new status of presented test items can lead to reliable criterion shifts even on a trial-by-trial basis and in the absence of feedback.

For example, Rhodes and Jacoby (2007) provided old/new likelihood information in the form of probabilistic associations between presentation location and old/new status. Participants were informed that test items appearing on the left side of the screen had a 67% chance of being *old*, whereas those appearing on the right side of the screen had a 67% chance of being *new*. These explicit cues set up expectations of encountering *old* and *new* items on the left and right sides of the screen respectively, which were reflected in opposing criterion shifts for each presentation location condition that acted to maximise endorsement of the expected response (i.e. negative criterion placement to maximise ‘old’ responding for words presented on the left and a positive criterion to maximise ‘new’ responding for words presented on the right). The explicit cueing of old/new status has been shown to shift criterion across a number of cue types, such as presentation colour (e.g. ‘red words are likely old’; Aminoff et al., 2012), affiliation to particular semantic categories (e.g. ‘insect-related words are likely old’; Benjamin & Bawa, 2004) and anticipatory text cues (e.g. ‘likely old’ presented prior to the appearance of test items; O’Connor, Han & Dobbins, 2010; Jaeger, Cox & Dobbins, 2012).These latter findings highlight a link between criterion placement and cued expectation - the first key environmental influence on memory evaluation highlighted in the opening section. Indeed, the summarised laboratory effects of explicit cueing manipulations on criterion hint at a mechanistic explanation for behaviour in the real-world bus scenario presented earlier, in which environmentally cued expectations were used strategically to guide memory evaluation outcomes.

A link between criterion placement and the second key environmental influence – variations in goal and reward-related information – is suggested by studies that provided monetary incentives based on aspects of recognition performance. However, incentives in recognition research have thus far been primarily employed to affect extreme criterion via asymmetric payoffs, such that selectively incentivising correct ‘old’ or correct ‘new’ decisions leads to extreme criterion biases that maximise endorsements of the incentivised response. These extreme biases have been instilled primarily to adjudicate between the predictions of differing retrieval output models (Healy & Kubovy, 1978; van Zandt, 2000), and their wider relevance in clarifying the influence of testing environments on real-world memory evaluation has been overlooked. Further, whilst memory evaluation scenarios involving such extreme reward contingencies sometimes present in the real world, it is more common to confront subtler forms of goal and reward-related information, such as the goal-driven caution highlighted in the eyewitness real-world scenario (Section 1.1.; see Figure 1.2.). To revisit the scenario, the greater consequences associated with making incorrect memory decisions in the eyewitness compared to the supermarket recognition scenarios led to greater caution when making an 'old' decision in the former setting. Whether this heightened caution is reflected in strategic adjustments to criterion placement is an important empirical question for my PhD research.

A related question is whether similar goal-driven mechanisms operate in the absence of explicit mention of reward in the test environment. The tendency for participants to infer goals from aspects of the test environment has been alluded to by both the source monitoring (Johnson et al., 1993) and metacognitive (Koriat & Goldsmith, 1996) memory control frameworks. Indeed, subtle variations of test format have been demonstrated to introduce pronounced effects on memory evaluation. As mentioned previously, Lindsay and Johnson (1989) found that framing memory evaluation as a source monitoring task led to more accurate performance than a standard old/new recognition task format. Further, Koriat and Goldsmith (1996) demonstrated marked differences in performance between free and forced report recognition test formats. Hicks and Marsh (1999) explicitly linked test format influences with the adoption of strategic biases, reporting an increase in the tendency to respond 'old' in a recognition test phase comprising three response options with two 'old' subcategories ('remember old'/'know old'/'new'), compared to a more standard test phase with two equally weighted response options (‘old’/’new’, followed by “remember”/“know”). The pattern of findings suggests that test format can bear pronounced influences on memory evaluation, and might be underpinned by the differential recruitment of higher-order strategies. A limitation of the above studies is the reliance on raw proportion measures that conflate sensitivity and criterion bias (Macmillan & Creelman, 2005). Hence, whether such subtle test format influences are reflected in SDT criterion placement served as another important avenue for my PhD research.

The behavioural findings presented thus far favour a functional role for SDT criterion in indexing the influence of higher-order strategies over memory evaluation. However, the influence of environmental factors in driving the adaptive *need* to invoke these higher-order strategies in the service of memory control has been less clearly demonstrated. Doing so necessitates greater integration with the cognitive control literature summarised previously (see Section 1.2.), and how control is engaged under expectation-induced response conflict and variations in goal state. In both cases, memory control is putatively heightened when endorsing the decision outcome that runs contrary to the instilled strategic bias. For biases associated with cued expectations, memory control is heightened when expectations conflict with available memory evidence, requiring the controlled countermanding of these expectations via access to confirmatory memory analyses or other environmental sources of strategic information. For biases instilled by goal and reward-related information, memory control is heightened when making the decision associated with greater rewards and punishment (i.e. the decision associated with greater ‘goal emphasis’). In comparison, when the decision suggested by a strategic bias is matched by the available evidence, this putatively expedites the process of making an accurate memory decisions.

To summarise, the proposed interplay between cognitive systems respectively underpinning the assessment of retrieved memory strength and the instilment of environmental strategic biases is likely a crucial regulator of memory control. It was a primary aim of my PhD to substantiate the dynamic interplay between these two systems, and elucidate the role of response criterion as a behavioural index of their interaction. Whilst strategic adjustments in criterion under expectation-induced conflict have been previously documented, criterion adjustment in response to environmentally emphasised goals has received less empirical scrutiny. Another of my PhD objectives was therefore to provide behavioural evidence of this newly proposed link between criterion and goal emphasis, and how this role is accommodated with its established relationship with cued expectation.

It is also worthwhile highlighting that despite the primary focus on response criterion as the behavioural correlate of memory control, a more cohesive understanding of the underlying control mechanisms required complimentary analyses of a number of different measures, such as reaction time, response confidence and measures collapsing across subjective decision rather than objective item categories (Duncan, Sadanand & Davachi, 2013; see Section 2.2.1.4. for further details on these measures). This insight at the behavioural level was further complimented by functional neuroimaging investigation of the neural substrates underpinning various aspects of memory control. The ensuing section reviews relevant functional neuroimaging research into recognition memory that informed these investigations.

1.5. Functional neuroimaging correlates of memory control

Early functional neuroimaging studies of recognition memory, like their behavioural counterparts, also focused primarily on memory strength processes. One early finding that has garnered particular research interest is the observed increase in brain activation for correct ‘old’ decisions (hits) compared to correct ‘new’ decisions (correct rejections; CRs). Evidence for this “retrieval success” effect first came from ERP studies reporting scalp voltage fluctuations that distinguished hits from CRs (Sanquist, Rohrbaugh, Syndulko & Lindsley, 1980; Warren, 1980). Subsequent research has separated this late positivity into two topographically distinct sub-components: an early frontal-parietal negativity peaking around ~300-500 Ms post-item onset (the N400) and a late left parietal positivity emergent ~400-800 Ms post-item onset, extending into an ensuing slow-wave component (the P600 or ‘late positive component’; Rugg & Curran, 2007). A functional differentiation for these early and late sub-components has also been proposed, as respective neural markers of familiarity and recollection in accordance with the dual process model of retrieval output. This is supported by the observed selective engagement of each ERP component to the presence of either familiarity or recollection in self-rated assessments of these two processes (Duzel, Yonelinas, Mangun, Heinze & Tulving, 1997; Woodruff, Hayama & Rugg, 2006).

Rugg and colleagues (Rugg, Mark, Schloerscheidt, Birch & Allan, 1998) dissociated these two processes on more objective grounds, by varying the encoding task between deep and shallow levels-of-processing to affect strong and weak overall levels of memory strength at test. ERP analyses at the test phase recovered an N400-like component sensitive to the shallow hit > shallow Miss (i.e. an *old* item incorrectly deemed ‘new’) contrast, which was considered a marker of the use of familiarity to make correct ‘old’ decisions under memory strength conditions unconducive to the use of recollection. A P600-lke component was recovered for the deep Hit > shallow Hit contrast, which was interpreted as reflecting the successful use of recollection under conducive high levels of memory strength.

The retrieval success contrast has also been widely interrogated in fMRI research, which has yielded an increase in BOLD activation in a fairly replicable network of brain regions (see Figure 1.4a for a typical fMRI retrieval success activation map, taken from O’Connor et al., 2010). This network includes prefrontal regions, such as dorsolateral PFC, medial PFC, regions of the cingulate cortex, including anterior and rostral cingulate cortices, as well as parietal regions, including the inferior and superior parietal lobules (de Zubicaray, McMahon, Eastburn, Finnigan & Humphreys, 2005; Kim, 2013; Konishi, Wheeler, Donaldson & Buckner, 2000; Spaniol et al., 2009). Recent research has also observed retrieval success effects in increasing pupil size (Goldinger & Papesh, 2011; Vo et al., 2008), which is considered a reliable physiological index of autonomic nervous system activity implicated in cognitive processing (Aston-Jones & Cohen, 2005; Nieuwenhuis, Geus & Aston-Jones, 2011). Hence, findings from a number of imaging modalities suggest that the correct identification of ‘old’ items is associated with an elevated neural response.

However, whether this neural response is *exclusively* driven by processes involved in successfully retrieving memory content is unclear. Rather, it has recently been conjectured that unconstrained memory control processes might also contribute to the retrieval success response. The potential involvement of memory control is illustrated by reconsidering Rugg et al.’s (1998) ERP results. The N400 ‘familiarity’ effect was predicated on the shallow Hit > shallow Miss contrast, which essentially isolates activity underlying the correct identification of *old* items in conditions of weak and non-diagnostic memory strength. Correct performance under this condition might therefore also result from higher-order strategies recruited in light of the uncertain memory strength conditions (the behavioural correlates of which have been outlined in previous sections e.g. Hirshman, 1995; Reder & Wible, 1984). This provides an alternative interpretation of the early mid-frontal ERP as reflecting the engagement of memory control under conflict, rather than the successful use of familiarity. A control reinterpretation can also be applied to Rugg et al.'s P600 'recollection' effect predicated on the deep Hit > shallow Hit contrast. Rather than exclusively reflecting the successful use of recollection, the observed activation increase might also be reward-related, in signalling the higher goal salience typically imparted to the detection of *old* over *new* items by aspects of the test format (Neville, Kutas, Chesney & Schmidt, 1986). This alternative interpretation suggests the engagement of control under variations in environmentally emphasised goals.

In further support of these control reinterpretations, the neural correlates of retrieval success overlap considerably with those associated with wider aspects of control in the dedicated cognitive control literature, as highlighted by the similarity of the left parietal ERP reported by Rugg et al. to the late positive component of the P3 complex (associated with the general allocation of control; Polich, 2007), and the recurrent report of similar fMRI regions in non-episodic control tasks (Botvinick, 2007; Kouneiher et al., 2009; Ridderinkhof et al., 2004). A related issue is the regular absence of medial temporal lobe (MTL) activations in the fMRI retrieval success effect (Kim, 2013), despite the established linkage between these regions and episodic retrieval in both animal neurophysiological (e.g. Wood, Dudchenko & Eichenbaum, 1999) and human neuropsychological studies (e.g. Corkin, 2002). These re-interpretations should not suggest that retrieval output processes are not prominent contributors to the retrieval success effect, but rather that a ‘process pure’ interpretation of the effect as *solely* driven by retrieval output is an oversimplification. Rather, it is suggested that the effect also involves the contributions of empirically unconstrained memory control processes of the kind recurrently highlighted during this chapter.

Recent attempts at elucidating the precise underpinnings of the retrieval success effect have identified three likely constituents: processes involved in the output and maintenance of retrieval (i.e. those implied by the original retrieval success interpretation), control processes that contribute to effortful aspects of retrieval, and strategic control processes that constrain eventual decision outcomes (Cabeza, Ciaramelli, Olson & Moscovitch, 2008; Kim, 2013; Wagner, Shannon, Kahn & Buckner, 2005). These constituents are equivalent to the memory strength, retrieval control and decision control processes outlined previously, and hence clarifying their respective neural correlates is an important aim for research into memory evaluation. To this end, the operationalization of retrieval control in the functional neuroimaging literature raises clear parallels with the source monitoring framework (Johnson et al., 1993; Jacoby et al., 1999), in capturing the effortful, goal-directed modulation of the act of retrieval itself (Buckner, 2003). Control processes are therefore potentially a salient contributor to the retrieval success effect, in mediating the effortful *attempt* to recover recollective content, which may or may not ultimately lead to *success* in doing so.

In support of this view, fMRI studies that have manipulated the need for effortful source monitoring, either by instantiating encoding tasks that lead to weak memory strength (Wheeler & Buckner, 2003) or demanding increased contextual details to diagnose memory decisions (Ranganath, Johnson & D’Esposito, 2000), have observed activation increases in prefrontal and parietal retrieval regions irrespective of accuracy. Dobbins and colleagues (Dobbins, Foley, Schacter & Wagner, 2002) emphasised the distinction between source monitoring and retrieval success interpretations, by explicitly showing that regions previously associated with retrieval success were in fact sensitive to source monitoring manipulations i.e. showing increased activation for source memory > item recognition tasks, that crucially was insensitive to decision accuracy. More recent fMRI studies employing functional connectivity methods have highlighted the increased coupling between frontal and medial temporal lobe regions as a neural substrate of source monitoring (Barredo, Oztekin & Badre, 2013). This connectivity dynamic can be interpreted as prefrontal regions enforcing a 'top-down' bias on the memory strength processing mediated by the MTL, raising clear parallels with neural mechanisms highlighted in the dedicated cognitive control literature (e.g. Miller & Cohen, 2001). Overall, functional neuroimaging investigations of source monitoring have highlighted the prominent influence of effortful retrieval control processes in driving brain activation during memory evaluation.

Compared to the above retrieval control processes, functional neuroimaging investigations of decision control have suffered from a similar neglect as highlighted in the previous summary of behavioural research (see Section 1.4.). However, recent years have witnessed a burgeoning of interest, motivated by reports of the sensitivity of retrieval success activations to higher-order strategic processes. For instance, Finnigan and colleagues (Finnigan, Humphreys, Dennis & Geffen, 2002) uncovered a link between ERP retrieval success correlates and general monitoring/control processes, by reporting the modulation of the P600 component to both the CRs > FAs contrast (i.e. correctly identified *new* items > incorrectly identified *new* items) and the more anticipated hits > Misses contrast (i.e. correctly identified *old* items > incorrectly identified *old* items). Given that the correct identification of *new* items is unlikely to reflect the presence of recollection (as assumed by the retrieval success effect; Rugg & Curran, 2007), the elevation of the parietal P600 to CRs instead suggests a link with decision factors associated with correct responding irrespective of item status. Of these, the sensitivity of the P600 to subjective assessments of confidence has been demonstrated, in the form of the increased P600 positivity for high compared to low confidence decisions that presented for both hits and CRs (i.e. occurred irrespective of *old/new* item status; Rubin, van Petten, Glisky & Newberg, 1999). This pattern of findings support the stipulations of memory control models (Koriat & Goldsmith, 1996) in demonstrating that metacognitive assessments of performance are a prominent contributor to retrieval activations.

Further, manipulation of the base rate proportion of *old* and *new* items has highlighted the sensitivity of retrieval success activations to strategic processes that overtly impact on decision outcomes (as summarised in Section 1.4.). Herron and colleagues (2003) reported a split of the parietal P600 retrieval success ERP into a left parietal-located positivity sensitive to retrieval success (i.e. the hits > CRs contrast) and a right parietal-located positivity sensitive to the detection of low probability items (i.e. low probability > high probability items, collapsed across *old/new* status). This latter component parallels the elevated parietal positivity observed in oddball paradigms following detection of low probability stimuli (Polich, 2007). Indeed, even the left parietal P600 component linked with retrieval success was not found to be insensitive to the base rate manipulation, given that the greatest P600 amplitude was observed for low probability hits. With reference to the strategic processes outlined in the previous section, Herron et al.'s results suggest a neural marker of the engagement of control to overcome the strategic bias towards making the high probability response.

An fMRI recognition study in which base rate was similarly manipulated also showed a split amongst retrieval success regions by their sensitivity to the probability manipulation (Herron, Henson & Rugg, 2004). This sensitivity to base rate was most prominent in frontal sub-regions, including anterior and dorsolateral PFC, but also encompassed parietal sub-regions, such as the superior parietal lobule. These observations of the active modulation of neural correlates of memory evaluation by decision control processes validate the importance ascribed to these processes in this chapter. However, a limitation of the above base rate studies is their apparent failure to induce a behavioural effect on response criterion (Herron et al., 2003; Herron et al., 2004). Given the robust association between adopted higher-order strategies and adjustments to criterion summarised in Section 1.4., these null effects call into question the actual efficacy of the base rate manipulation in heightening decision control. The null effect on criterion is likely due to the vague information about the base rate manipulation given to participants in both studies (i.e. lacking in explication of the actual proportions), which prior behavioural research suggests would have impaired awareness of base rates as a source of strategic information (Estes & Maddox, 1995; Rhodes & Jacoby, 2007).

However, a handful of functional neuroimaging studies have followed the emerging trend in behavioural research of employing more explicit strategic manipulations, and have thereby enabled more reliable interrogations of memory control processes. For instance, O'Connor and colleagues (2010) provided trial-by-trial textual cues of the form 'likely old' and 'likely new' that preceded the appearance of items at a recognition test phase. Participants were explicitly told that the cues were 75% valid in predicting the status of upcoming items, and this led to their perception as a reliable source of strategic information (as validated behaviourally by an improvement in accuracy for valid cue trials). fMRI results revealed an activation increase in a number of regions previously linked with 'retrieval success' when participants correctly countermanded cues relative to correctly endorsing valid cues (i.e. for the contrast invalid > valid; see Figure 1.4b for O'Connor et al.'s fMRI invalid cue effect). This 'invalid cueing' response manifested for both *old* and *new* items, thereby challenging the *old* item specificity implied by retrieval success accounts. Rather, the effect is consistent with the recruitment of control to resolve the conflict generated by the cued expectation and the item-provoked memory strength analysis. This link with control is strengthened by the fact that regions within this invalid cueing network, such as medial prefrontal cortex, lateral prefrontal cortex and the inferior parietal lobule, are all recurrently activated in non-episodic control tasks (Bunge et al., 2002a; Ridderinkhof et al., 2004).

To further validate their control interpretations, O'Connor et al. also analysed an independent fMRI dataset of standard uncued recognition data, and revealed a positive correlation between the activity of invalid cueing regions and participants' response criterion (i.e. increasing regional activity for an increased bias against responding 'old'). Countermanding criterion-indexed biases against responding 'old' is likely to necessitate similar recruitment of control under conflict (as described in Section 1.4.), and this is borne out by the activation increase in the invalid cueing regions. Subsequent follow-ups have shown that the invalid cueing network is amenable to further functional segregation, both based on the precise control mechanisms underlying the response (i.e. early conflict detection and late control allocation; Mill & Connor, 2014, *under review*) as well as the items driving the response (i.e. with partially separable activation patterns for the detection of unexpected *old* and unexpected *new* items; Jaeger, Konkel & Dobbins, 2013). Hence, O'Connor et al.'s likelihood cueing paradigm is well-suited to the study of memory control mechanisms recruited under environmental conflict. This is one of the two key environmental precipitants of memory control highlighted in this chapter (see Section 1.1. and 1.2.), and hence the paradigm featured prominently in my PhD research.

The second of these environmental cues to decision control – variation to goal and reward-related information – was targeted in an fMRI study by Han and colleagues that also questioned retrieval success (Han, Huettel, Raposo, Adcock & Dobbins, 2010). The authors were motivated by prior suggestion that a prominent contributor to the retrieval success response was the preferential goal salience imparted to the detection of *old* over *new* items by implicit aspects of the test format (Neville et al., 1986). To directly interrogate this, a 'mixed incentives' procedure was applied in separate recognition test phases that served to differentially emphasise the detection of hits (hit incentive condition: +1 token for Hit, -1 token for FA) and the detection of CRs (CR incentive condition: +1 for CR, -1 for Miss) as the higher-order goals of memory evaluation. The strategic influence of the incentive manipulation was behaviourally validated by an observed reduction in confidence for decisions that conflicted with the available incentives. A non-significant incentive effect on criterion was also observed that was nevertheless in a direction reflective of heightened caution towards the emphasised decision (e.g. a positive criterion shift in the hit incentive condition), as predicted by previously discussed memory control models (Koriat & Goldsmith, 1996; see Section 1.3.). fMRI analyses of the non-incentivised control condition recovered a typical network of regions activated by the 'retrieval success' contrast. Subsequent analyses revealed that a number of these regions were in fact preferentially sensitive to the incentives manipulations, including caudate nucleus, lateral PFC, medial PFC, cingulate cortex and lateral parietal cortex. Of these, the caudate nucleus was shown to preferentially track incentive status over and above item status, such that its retrieval success effect was reversed in the CR-incentive condition (i.e. caudate activation in the CR-incentive condition was greater for CRs than hits). Further, the presence of performance feedback served to enhance the incentive effects, thereby paralleling behavioural findings of enhanced strategic processing in the presence of feedback (e.g. Estes & Maddox, 1995).

The observed involvement of the caudate nucleus is consistent with its role in processing goal- and reward-related information highlighted in the dedicated cognitive control literature (Delgado et al., 2004; Elliot et al., 2000; Knutson et al., 2001). Overall, Han et al.'s findings highlight the prominent influence of unconstrained environmental sources of goal emphasis in cueing strategies that differentially bias the rendering of 'old' and 'new' decisions. That these strategies are reflected in the modulation of brain activations linked with retrieval further illustrates the importance of these decision control processes in memory evaluation.

To summarise, functional neuroimaging research in recognition memory increasingly testifies to the importance of retrieval control and decision control processes, which interact with core memory strength processes in determining the outcomes of memory evaluation. The influence of environmental factors on retrieval activations should serve to emphasise the need for tighter experimental control over these factors, so as to enable more valid functional inferences of neuroimaging data. Attempts at isolating these environmental factors would benefit from greater efforts to behaviourally validate the implied strategic influence and verify the underlying neural mechanisms across different imaging techniques and cognitive domains beyond episodic recognition memory. To this end, the final section will outline another important question that was interrogated during the course of my PhD – are memory control processes memory-specific or cross-domain in nature?

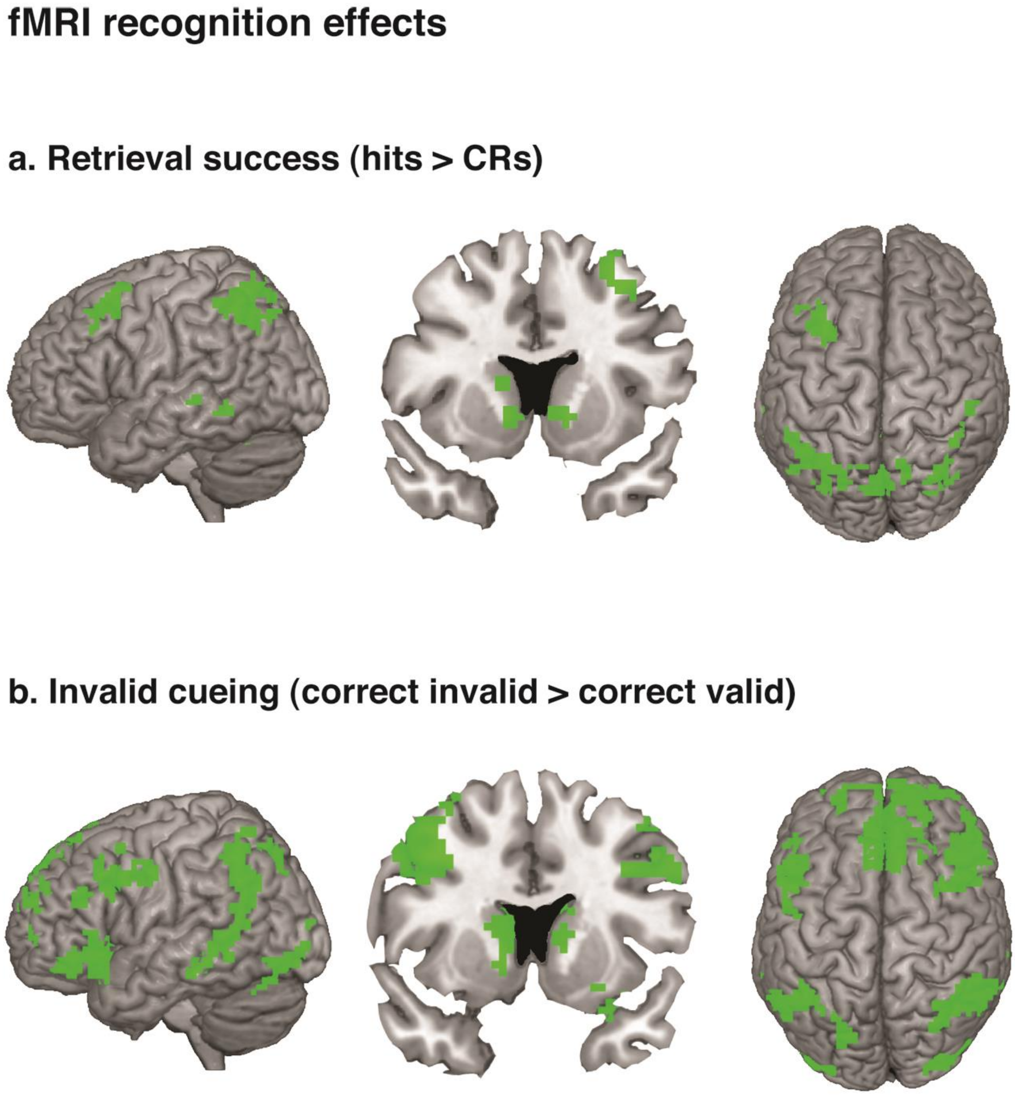


Figure 1.4. fMRI activation maps of the two recognition memory tasks effects highlighted in the main text (Section 1.5.): **a.** the retrieval success effect (correct ‘old’ i.e. hits > correct ‘new’ i.e. correct rejections, CRs) and **b.** the invalid cueing effect (correct decision after invalid cue > correct decision after valid cue). Both maps are rendered using data from O’Connor, Han & Dobbins (2010) provided by Akira O’Connor, at a significance threshold of $p < .001$, 5 contiguous voxels. Substantial overlap between the two activation maps is evident in left inferior parietal, left prefrontal and bilateral striatal regions.

1.6. Cross-domain properties of memory control processes

An emerging theme from the theoretical and empirical review in this introductory chapter is that the control processes that constrain memory might impose similar constraints on other cognitive domains. Such speculations on the 'cross-domain'[2] nature of control have arisen in both the dedicated cognitive control literature (Ridderinkhof et al., 2004; Polich, 2007) and the memory control literature (Jacoby et al., 1999; Whittlesea, 2002). Findings and theoretical perspectives in other cognitive fields also connote similar cross-domain features of processing. One early example is the 'general capacity' theory of processing load (Kahneman, 1973; Moray, 1967), which posits the existence of a centralized cognitive system capable of integrating processing across different task domains. This view was predicated on the findings of dual-task paradigms in which subjects are required to learn to integrate across two tasks performed sequentially (e.g. Task 1: 'press right index finger when you hear a tone', Task 2: 'press left index finger when a light is flashed) so as to flexibly process their conjunction (e.g. upon simultaneous presentation of the tone and the light, subjects learn to press both the left and right index fingers to receive a reward; Colavita, 1971; Kahneman, 1973). The ability to integrate processing across different tasks to form unified rules and strategies implies that part of the processing system has access to both implicated domains (i.e. visual perception and motor response).

A similar overarching control system is implied by an influential neuroscientific model of attention (Corbetta & Shulman, 2002), which proposes that prefrontal and parietal regions affiliated with two distinct functional networks environment – the dorsal

[2] For the benefit of this chapter and the remainder of this thesis, a distinction should be made here between my use of the terms 'modalities' and 'domains' – both of which have often been used interchangeably in previous research. The former pertains to the *sensory* form in which to-be-evaluated stimuli are presented (e.g. verbal, pictorial, auditory etc.), whereas the latter pertains to the *cognitive* operations undertaken when attempting to discriminate the status of to-be-evaluated stimuli (e.g. episodic memory, semantic memory, visual perception etc.).

attention network and the ventral attention network – facilitate orienting to diverse forms of motivationally relevant stimuli in the environment. The dorsal attention network, which prominently comprises regions of the intraparietal sulcus and dorsal prefrontal cortex, is implicated in providing 'top-down' biases on the basis of prior expectations and current goals that facilitate attending to different perceptual features, such as spatial location (Corbetta, Kincade, Olinger, McAvoy & Shulman, 2000) and object shape (Wojciulik & Kanwisher, 1999). These same fronto-parietal regions are also implicated in maintaining 'top-down' expectations in non-visual domains, such as in auditory and tactile discrimination (Downar, Crawley, Mikulis & Davis, 2000) and motor response tasks (Kawashima, Roland & O'Sullivan, 1995).

The capacity for this dorsal fronto-parietal network to allocate attentional resources across different visual modalities, as well as across perceptual and motor domains, renders it as a promising neural substrate for cross-domain processing. Indeed, this network has also been suggested to mediate attentional demands in episodic memory tasks, by imposing 'top-down' strategic biases on memory strength processes (Cabeza et al., 2008). Whilst these models nominally focus on the effect of environmental factors on attention, it is likely that these factors impact similarly on cognitive control, given that attention is often considered a constituent of control (Banich et al., 2000; Miller & Cohen, 2001). Indeed, activation in both the dorsal attention network and the ventral attention network has been observed to underlie a 'reorienting' response subsequent to violations of prior expectations (Corbetta, Patel & Shulman, 2008), which was previously highlighted as a key precipitator of cognitive control (see Section 1.2.). Empirical evidence of a cross-domain control system remains largely speculative, however the sparse behavioural and functional neuroimaging research presently available is summarised below.

Tentative behavioural support for cross-domain control comes from reports of common manipulations impacting similarly on measures of performance across different task domains. For instance, the degree to which individuals' confidence ratings correlate with their actual decision accuracy has been shown to be reliable across two different visual detection tasks (orientation discrimination and contrast discrimination; Song et al., 2011). This finding is consistent with the underlying engagement of a 'metacognitive' system (that conceptually overlaps with Koriat & Goldsmith's, 1996, memory control model described in Section 1.3.) capable of performing an overarching monitoring/control role, albeit confined in this demonstration to two tasks that differed in *modalities* within the same visual cognitive *domain*. With regard to measures more directly determinative of decision outcomes, environmentally-cued expectations have been found to reliably increase accuracy and reduce reaction time across perceptual (Posner, Snyder & Davidson, 1980), motor (Abrams & Jonides, 1988) and episodic memory tasks (O'Connor et al., 2010). Conversely, instances where expectations conflict with available item evidence have been shown to reliably increase reaction time and reduce accuracy across cognitive domains (O'Connor et al., 2010; Posner et al., 1980), supporting the role of these raw performance measures as behavioural correlates of cross-domain control processes.

In a recent study involving more refined behavioural measures, White and Poldrack (2014) probed the cross-domain correspondence of two response bias parameters estimated from the drift diffusion model (Ratcliff, 1978). The diffusion model differentiates between biases that impact directly on how stimuli are processed (termed 'stimulus bias') and those that impact on the eventual responses made (termed 'response bias'), with each form of bias indexed by a separate model parameter. This differentiation parallels the retrieval control/decision control dichotomy outlined previously in this chapter (see Section 1.3.). The authors instantiated independent manipulations of these forms of bias in a perceptual task (involving ‘bigger’/‘smaller’ judgements of line length stimuli relative to a presented exemplar) and a standard recognition task (involving ‘old’/‘new’ decisions in a standard recognition format). Stimulus bias was manipulated by instructing participants to demand differing levels of diagnostic evidence when making decisions in the respective tasks. Response bias was manipulated by varying the base rate proportion of respective items in each task (which reliably affects strategic bias, as illustrated previously). The results revealed striking influences of these manipulations in affected concordant shifts in both stimulus bias and response bias estimates across both task domains. White and Poldrack’s findings therefore provide direct behavioural evidence in support of a unified control system that mediates the use of similar higher-order strategies across different task domains.

Further support for this cross-domain system is provided by functional neuroimaging studies. This is in largely speculative form in the ERP control literature, with the N2b/N2c complex associated with response conflict (Kok, 2001; Ridderinkhof et al., 2004) and the P3a/P3b/late positive complex signalling expectancy violations (Polich, 2007; Squires, Wickens, Squires & Donchin, 1976) both suggested to have cross-domain properties that have remained largely empirically unsubstantiated. fMRI research has affected more direct interrogations, with recent functional connectivity findings suggesting that regions affiliated with a large-scale fronto-parietal control network dynamically interact with other domain/modality-specific functional networks to facilitate the upregulation of control across a number of task domains (Cole et al., 2013). A more anatomically derived proposal is that of the hierarchical organisation of prefrontal cortex (Badre & D'Esposito, 2009), with caudal regions participating in domain/modality-specific processes and rostral regions involved in abstracted, integrative functions linked with control. A similar hierarchical functional organisation was suggested to underlie episodic memory processing (Buckner, 2003), with caudal frontal regions involved in the retrieval of modality-specific information and rostral frontal regions involved in control exerted over this retrieved information.

Fleck and colleagues (Fleck, Daselaar, Dobbins & Cabeza, 2006) directly interrogated the cross-domain properties of prefrontal and parietal regions recurrently linked with aspects of memory control. Participants performed discrimination tasks in two different domains – perceptual (decide which of 'upper' and 'lower' coloured segments covered a larger onscreen area) and episodic (standard 'old/new' recognition) – and provided ratings of subjective confidence. Brain regions in which activation increased for low compared to high confidence decisions were assumed to underpin the engagement of control under uncertainty. Conjunction analyses revealed regions that were commonly engaged by the low > high confidence contrast across both task domains, which included regions of the cingulate cortex, dorsolateral prefrontal cortex and superior parietal lobule. Fleck et al.'s findings therefore provide clear evidence of the cross-domain properties of memory control regions, challenging prior suggestion of their specificity to the episodic domain (Henson, Rugg, Shallice & Dolan, 2000).

Overall, the findings summarised above provide incipient behavioural and functional neuroimaging evidence for the existence of a cross-domain control system. Further elucidation of these control processes would benefit from greater attempts to subject the same sample of participants to common control manipulations enacted in different task domains (as effectuated by White & Poldrack, 2014, and Fleck and colleagues, 2006). Behavioural work to this effect was conducted in the ensuing experimental chapters, by following up White and Poldrack's demonstration of the cross-domain correspondence of expectation-induced bias with examination of the cross-domain correspondence of goal- and reward-related biases. The need for more explicit manipulations of cross-domain control is especially important in the aim of elucidating the underlying neural correlates. This would prevent an over-reliance on confidence as an inferred index of decision control (as in Fleck et al., 2006), which is potentially problematic given the ambiguous nature of the measure e.g. heightened activation in low confidence trials might reflect participants' overall lack of motivation to engage in the task, or reflect the amalgamated recruitment of likely functionally distinct retrieval and decision control processes. A related aim is to utilise multi-modal imaging methods to probe the coherence between markers of cross-domain control evoked in the ERP and fMRI literatures. An improved understand of such cross-domain features would likely yield insight into the flexible and abstractive cognitive functions that are most intimately tied to our conscious experience.

1.7. Conclusion

The literature review provided in this introductory chapter should serve to highlight the conceptual and empirical motivations for my research into memory control. The first section described real-world scenarios in which memory evaluation is overtly influenced by specific aspects of the test environment, namely cued expectations and emphasised goals. With reference to the dedicated cognitive control literature, these two environmental influences were characterised as provoking the engagement of control under expectation-induced response conflict and variations in higher-order decision goals respectively. The putative electrophysiological and hemodynamic neural correlates of these control processes were then sourced to the co-ordinated activity of prefrontal, parietal and sub-cortical brain regions. This was followed by a description of extant episodic memory models, in which more precise mechanisms by which cognitive control impacts on mnemonic processing were outlined. One key feature emerging from this section was the proposed dichotomy between control processes that act on retrieval itself (‘retrieval control’) or on the output of retrieval (‘decision control’), with the latter process comparatively neglected in the field despite it being more amenable to the influence of environmental factors.

Nascent attempts at elucidating the behavioural correlates of these memory control processes were then summarised, with particular emphasis placed on environmentally cued biases that affect the criterion parameter of the SDT model of retrieval output. The role of the interplay between these strategic biases and available memory strength in regulating levels of memory control was also highlighted. Functional neuroimaging investigations of episodic memory also increasingly bear out the importance of factoring in memory control processes, as evidenced by the progressive refinement of activation patterns originally linked with the success of retrieval output to accommodate their modulation by both retrieval control and decision control processes. The final section formalised an emergent question as to whether the highlighted memory control processes were indeed memory-specific, or instead capable of influencing processing in other cognitive domains.

The primary avenues of investigation that emerged from this review and that were empirically interrogation during my PhD are as follows. Firstly, the literature and the depicted real-world scenarios highlight the need to more fully elucidate aspects of the test environment capable of influencing participants' strategic biases and relatedly provoking the need for memory control. Secondly, the potential cross-domain properties of these environmental influences and their precipitant control mechanisms warrant systematic investigation. Finally, the utilisation of both behavioural and functional neuroimaging methods would aid in conclusively isolating these memory control processes. These avenues were pursued in the ensuing experimental chapters. Chapter 2 presents behavioural evidence for a novel aspect of the test environment capable of engaging memory control via variation in goal emphasis – the wording of the test question. Chapter 3 interrogates the cross-domain properties of the observed question emphasis effect, in a paradigm involving episodic and non-episodic forms of evaluation. Chapter 4 probes the interaction of the two highlighted decision control mechanisms, in a paradigm involving more explicit manipulation of both goal emphasis (via mixed incentives) and textually cued expectations, and examines their persistence across episodic and semantic processing domains. Chapter 5 details a multi-modal neuroimaging study involving simultaneous EEG-fMRI which sought a comprehensive insight into the neural correlates of decision control recruited under response conflict in episodic and semantic domains. Chapter 6 directly scrutinised the adaptive basis of these decision control mechanisms, by assessing the effects of expectation-induced memory conflict on autonomic nervous system activity measured by pupillometry. The experimental chapters collectively evidence an attempt to elucidate the precise behavioural and functional neuroimaging correlates of memory control. The observed findings serve to align the laboratory study of memory evaluation more closely with the overarching adaptive context in which it takes place in the real world.

<u>Chapter 2:</u> Question format as an implicit source of goal emphasis in recognition memory

2.1. General Introduction

The preceding review chapter highlighted the capacity for the test environment to encourage the use of higher-order strategies in memory evaluation. However, precisely which aspects of the test environment are capable of provoking such strategic influences is in need of further clarification. This arises as laboratory investigations of environmentally-provoked strategies have thus far primarily focused on one manipulation – decision cueing. Cueing the likely old/new status of an upcoming test item has been found to instil a strategic bias towards making the cued response (Aminoff et al., 2012; O'Connor, Han & Dobbins, 2010; Rhodes & Jacoby, 2007; see Section 1.4 for further detail on this effect). This bias has been shown to be behaviourally indexed by signal detection criterion estimates (Macmillan & Creelman, 2005). For example, being cued that the status of an upcoming test item is likely to be *old* leads to a negative shift in criterion, reflecting the lower level of memory strength required for an "old" decision (e.g. Rhodes & Jacoby, 2007). The observed confirmatory bias therefore bears out the influence of environmentally cued expectations on the strategies adopted in memory evaluation highlighted in the previous chapter. However, the exclusive research focus on cueing manipulations has led to a neglect of other environmental factors capable of influencing evaluative strategies. The aim of this chapter was to systematically interrogate one such neglected factor, pertaining to how aspects of the environment serve to emphasise the goals and reward associated with particular decision outcomes. The findings have contributed to a published article in a peer-review journal (Mill & O'Connor, 2014; see Appendix A for the full manuscript).

Variation in environmental "goal emphasis" information can have pronounced effects on the outcomes of memory evaluation, as demonstrated in the real-world scenario presented in the previous chapter involving recognition decisions in eyewitness test and supermarket environments (see Section 1.1.; see Figure 1.2.). The greater consequences of making incorrect "old" decisions in the eyewitness compared to the supermarket environment led to a hypothesised reduction in the likelihood of responding "old", consistent with the use of a more cautious evaluative strategy under heightened goal emphasis. Despite this real-world significance, the laboratory study of goal and reward-related influences on memory evaluation have thus far been confined to extreme manipulations of payoff incentives. For example, rewarding correct "old" decisions (i.e. hits) in the absence of any other payoff contingencies serves to increase the likelihood of responding "old" (Healy & Kubovy, 1978; van Zandt, 2000). This confirmatory bias is in an identical direction to that observed following the above cueing manipulations and hence may reflect the adoption of the same underlying strategy. Neither manipulation is likely to provide a laboratory analogue of the goal-driven caution evidenced in the eyewitness scenario, which putatively yields a strategic bias in a converse disconfirmatory direction, reflecting a reduced likelihood of endorsing the decision category emphasised as a higher-order goal.

In search of this laboratory analogue, prior work in the dedicated cognitive control literature suggests that mixed incentives might elicit similar disconfirmatory strategies in non-episodic forms of decision-making. Unlike the extreme reward-only payoffs provided in the incentivised recognition studies mentioned above (e.g. Healy

& Kubovy, 1978), mixed incentives involve the provision of *both* reward for correct performance and punishment for incorrect performance for one of two available decision options. These payoff contingencies are commonly applied in go/no-go tasks, which require the commission of button-press responses (i.e. “go” responses) to one class of stimuli and inhibition of button-press responses (i.e. “no-go” responses) to another class of stimuli (Brutkowski, 1964). The selective provision of mixed incentives for “go” responses (i.e. +1 incentive for correct “go” and -1 incentive for incorrect “go” responses, with no incentives associated with “no-go” responses irrespective of correctness) motivates a bias against making the “go” response (Newman, Widom & Nathan, 1985; Hartung, Milich, Lynam & Martin, 2002). The disconfirmatory direction of this bias is consistent with the instilment of strategic caution towards making the “go” decision associated with a higher propensity for reward and punishment. Hence, mixed incentive manipulations serve to heighten goal emphasis in non-episodic forms of decision-making, raising the possibility of exerting similar influences on episodic decision-making and providing a laboratory analogue of the strategic behaviour described in the eyewitness real-world scenario. Indeed, these explicit forms of heightening memory control under goal emphasis are examined in greater detail in Chapter 4.

However, the aim of the present chapter was to examine the influence of more implicit sources of goal emphasis, as imparted by the format of the memory test itself. This follows from reported effects of test format on response strategies in the applied domains of questionnaire design (‘yea-responding’; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) and eyewitness testimony (‘leading questions’; Loftus, 1996). Test format influences have been comparatively neglected in memory research, despite the stipulations of extant memory control models that participants routinely infer the wider goals of memory evaluation from subtle aspects of the presented format (Koriat & Goldsmith, 1996). In support of this, strategies that serve to improve performance have been evidenced after framing a memory test as a source monitoring compared to a standard item recognition task (Lindsay & Johnson, 1989), as well as for free (as opposed to forced) report formats (Koriat & Goldsmith, 1996). These test format manipulations affect memory evaluation in a "global" way, by emphasising the need for *overall* accuracy when making both "old" and "new" decisions. Such global goal emphasis mechanisms putatively contribute to the overarching strategic caution described in the eyewitness real-world scenario.

However, more "selective" mechanisms that impart goal emphasis to *either* "old" or "new" decisions are also likely to be involved. Reconsidering the eyewitness scenario, "old" decisions might be selectively emphasised if the observer perceives the consequences of making an incorrect "old" decision as greater than making an incorrect "new" decision (in that the former decision connotes a committed action that results in the wrongful conviction of an innocent individual, whereas the latter maintains the status quo of the ongoing suspect search and is comparatively less consequential). Similar selective emphasis for "old" decisions might even be imparted in the supermarket scenario comparatively lacking in global emphasis, in that making an incorrect "old" decision for a stranger is likely to lead to an awkward and potentially embarrassing social encounter (whereas an incorrect "new" decision maintains the status quo of supermarket shopping). Selective emphasis imparted by both test environments might therefore lead to heightened strategic caution in making "old" decisions, as a parallel to selective emphasis imparted to "go" decisions in the mixed incentive research alluded to previously (e.g. Newman et al., 1985).

A potential implicit source of such selective goal emphasis in laboratory test environments is the format of the test question. Question format in single item recognition typically comprises the question "old?" with "yes" and "no" keypress response options (indeed, item recognition is often termed yes/no recognition; Donaldson, 1996). If this format implicitly imparts emphasis to "old" decisions, and this implicit emphasis serves to introduce a strategic bias on memory performance, then this bias is putatively engaged in covert form in a large number of laboratory recognition studies. Hence, this first chapter describes a series of three recognition experiments in which we provide behavioural evidence that question format does indeed serve as an environmental source of implicit goal emphasis. The strategic nature of this bias in inducing caution was indexed across all experiments by effects on various estimates of SDT criterion (a measure of bias which bins responses by "objective" *old* or *new* item status) and 'decision accuracy' (which bins responses by "subjective" decision status; Duncan, Sadanand & Davachi, 2012).

2.2. Introduction to Experiment 1

The first experiment examined whether question formats adopted as standard in laboratory recognition tests are capable of influencing participants' strategies. Following an incidental encoding procedure, the wording of the test question was varied in separate blocks according to two factors: the decision emphasised by the question ("question emphasis": old or new emphasis) and the number of decisions mentioned in the question ("question dimension": single ("old" or "new?") or composite dimensions ("old/new?" or "new/old?")). To clarify, I speculated that emphasis would be imparted in single dimension questions to the decision presented in isolation (e.g. “old?” emphasises “old” decisions), whereas it would be imparted to the decision appearing first in composite dimension questions (e.g. “old/new?” emphasises “old” decisions). The capacity for question emphasis to shift criterion in “disconfirmatory” goal-driven directions (i.e. reducing the likelihood of endorsing the emphasised decision) as described in the introduction was hence explored in both these question dimension conditions.

2.2.1. Method

2.2.1.1. Participants. The sample comprised 30 self-reported native English speakers (14 female, mean age = 24.0; age range = 18-39 years), who all passed a minimum performance threshold of d' sensitivity > 0.1 across all question format conditions. Informed consent was obtained in accordance with the University Teaching and Research Ethics Committee at the University of St Andrews (UTREC) and all participants were compensated for their time at the rate of £5/hr.

2.2.1.2. Stimuli. For each participant, a different set of words was randomly sampled from a pool of 2199 singular, common nouns from the English Lexicon Project (Balota et al., 2007), following the removal of low frequency words (Hyperspace Analogue to Language HAL frequency high-pass cut-off: 7.70). This served to exclude highly distinctive items, which could lead to extreme variations in memory strength and consequently attenuate the strategic bias effects of primary interest (final word list characteristics: mean HAL frequency = 8.98, mean word length = 7.24, mean number of syllables = 2.43). Each of the four presented study-test blocks comprised 100 words, with 50 words presented at both study and test (*old* words) and 50 words presented at test only (*new* words). The experiment was presented using PCs running MATLAB (The MathWorks Inc., Natick, MA, 2000) and Psychophysics Toolbox (Brainard, 1997).

2.2.1.3. Procedure. Participants were presented on-screen instructions and a practice task prior to completing four self-paced study-test blocks (see Figure 2.1a for a design schematic). The study task across all blocks required participants to count the syllables of single words appearing onscreen (i.e. prompted by the question “syllables?” with keyboard response options “1” through “4+”). A test phase immediately followed each study phase, wherein participants decided whether serially presented words were *old* (i.e. seen before at study) or *new* (i.e. not seen before during the experiment) for a test list comprising 50 words of each item category. Crucially, the format of the test question was varied in blockwise fashion according to the decision emphasised (old or new) and the dimensions of the question i.e. the number of decisions presented (single or composite), leading to four questions being presented in separate blocks: “old?” and “new?” (i.e. single old and single new emphasis questions, both with “yes” and “no” response options assigned to the “1” and “2” number keys), as well as “old/new?” and “new/old?” (i.e. a composite old emphasis question with “old” and “new” assigned to “1” and “2” number keys, and a composite new emphasis question with “new” and “old” assigned to the “1” and “2” keys).

For all question types, participants also made a confidence rating on a three-point scale (i.e. prompted by “confidence?” with “low”, “medium” and “high” rating options), which appeared 0.25s after making an old/new response. Each study and test trial was preceded by a 0.5s fixation cross. Across participants, the study-test block order was pseudo-randomised, such that the single dimension question blocks always preceded the composite dimension blocks, with question emphasis fully randomised within this static ordering of dimension conditions. This led to a total of 4 run orders to which participants were randomly assigned.

2.2.1.4. Calculation. The primary analyses across all three experiments were conducted on sensitivity (*d'*) and response criterion (*c*) parameters estimated from the equal variance signal detection model (EV SDT; Green & Swets, 1966; Macmillan & Creelman, 2005; see Section 1.4. for more detailed explanation of these measures). A correction for errorless responding was made in accordance with Snodgrass and Corwin (1988), by taking the numbers of hits (Hn; the number of old items correctly judged “old”), misses (Mn; the number of old items incorrectly judged “new”), correct rejections (CRn; the number of new items correctly judged “new”) and false alarms (FAn; the number of new items incorrectly judged “old”), and calculating adjusted hit (H') and false alarm rates (FA') as follows:

$$\mathbf{H'} = \frac{\mathbf{Hn+0.5}}{\mathbf{Hn+Mn+1}} \tag{1}$$

$$\mathbf{FA'} = \frac{\mathbf{FAn+0.5}}{\mathbf{FAn+CRn+1}} \tag{2}$$

These adjusted measures were then used to compute *d'* and *c*:

$$\mathrm{d'} = \mathbf{z}(\mathbf{H'}) - \mathbf{z}(\mathbf{FA'}) \tag{3}$$

$$c = -\mathbf{0.5} \times [\mathbf{z}(\mathbf{H'}) + \mathbf{z}(\mathbf{FA'})] \tag{4}$$

Eliciting response confidence enabled supplementary analyses of equivalent sensitivity and bias parameters estimated from the unequal variance signal detection model (UEV SDT; Macmillan & Creelman, 2005). This differs from the equal variance model in that the variance of the *old* item distribution is allowed to vary whilst the *new* item distribution variance remains fixed at 1. Hence, UEV d' and UEV criterion (c_a), along with old distribution variance (σ), were estimated for each subject using nonlinear regression and a least squares criterion (iteratively minimizing the difference between actual hit/false alarm rates and those estimated from the varying d', c_a and σ parameters), as implemented in the Solver function in Excel (Harris, 1998; O'Connor, Guhl, Cox, & Dobbins, 2011). Bias as indexed by the c_a parameter was then adjusted to take into account the variation in optimal criterion placement engendered by the freely varying old item variance parameter (as described in O'Connor et al., 2011). Optimal criterion in the UEV model was therefore calculated as:

$$\textbf{optimal } c = \frac{\sqrt{d'^2\sigma^2+2\sigma^4\ln(\sigma)-2\sigma^2\ln(\sigma)}-d'}{\sigma^2-1} \tag{5}$$

The distance between c_a and optimal c (i.e. c_a minus optimal c) was termed c_{rel} and used as the criterion estimate in the UEV analyses. Fitting to both versions of the signal detection model ensured that any observed question format effects were not confined to the equal variance SDT model, and rather persisted even in the unequal variance SDT model that attempts to account for added response variability associated with confidence ratings. For both equal and unequal variance models, larger d' parameters indicate greater sensitivity (better discrimination of *old* from *new*), and larger c parameters indicate a more conservative bias (reduced tendency towards responding “old”).

An additional analysis examined the proportion correct out of all “old” or “new” decisions made by each participant (old_{corr} and new_{corr} respectively). These “decision accuracy” measures bin responses according to “subjective” decision status (i.e. the total number of “old” or “new” decisions made, which varied across subjects), and are hence not to be confused with the hit and correct rejection rates used for the SDT measures, which bin responses according to “objective” item status categories (i.e. the total number of *old* or *new* items actually present in the test list, which remained fixed across subjects). The decision accuracy measures were calculated for each participant as follows:

$$\text{oldcorr} = \frac{\textbf{number of correct } \text{"old"} \textbf{ decisions}}{\textbf{number of correct } \text{"old"} \textbf{ decisions+number of incorrect "old" decisions}} \tag{6}$$

$$\text{newcorr} = \frac{\textbf{number of correct } \text{"new"} \textbf{ decisions}}{\textbf{number of correct } \text{"new"} \textbf{ decisions+number of incorrect } \text{"}\textbf{new}\text{"} \textbf{ decisions}} \tag{7}$$

These measures complemented the question-driven biases revealed in the primary SDT analyses with insight into how these biases impacted on the accuracy of memory decision-making, thereby serving to elucidate their underlying strategic nature.

2.2.2. Results and Discussion

2.2.2.1. Equal variance sensitivity and bias. Means and standard deviations for all three experiments presented in this chapter are provided in Table 2.1. The effects of question format on sensitivity were firstly investigated in a 2(question dimension: single or composite) x 2(question emphasis: old or new) repeated measures ANOVA on equal variance (EV) estimates of *d′*. This revealed a significant main effect of question dimension such that *d′* was higher for single (M = 1.75, SD = 0.52) compared to composite questions (M = 1.50, SD = 0.49), $F(1,29)$ = 10.61, p = .003, η_p^2 = .27. The reduced sensitivity for composite questions is likely due to these questions being presented after both single question types across all run orders, rendering this test condition more susceptible to deleterious effects of fatigue and greater elapsed time from the study phase. There was neither a significant main effect of emphasis nor an interaction, $F(1,29) = 1.74$, $p = .198$, $\eta_p^2 = .06$ and $F < 1$ respectively. The primary question format manipulation of goal emphasis therefore did not influence participants' sensitivity at test.

The same 2(question dimension) x 2(question emphasis) ANOVA was conducted on EV criterion, which yielded neither a significant main effect of dimension nor an interaction effect, both *Fs* < 1. A trend main effect of emphasis was observed, such that *c* was higher in the old ($M = 0.05$, $SD = 0.33$) compared to the new emphasis condition ($M = -0.04$, $SD = 0.32$; see Figure 2.1b), $F(1,29) = 2.98$, $p = .095$, $\eta_p^2 = .09$. Although non-significant, the direction of this criterion trend is reflective of question emphasis reducing the likelihood of endorsing the decision emphasised by the question. This bias direction is contrary to that observed following cueing manipulations, which act to increase endorsement of the cued decision (e.g. O'Connor et al., 2010). Rather these findings reflect a different strategic mechanism, wherein the test question serves to emphasise one memory decision as a higher-order goal and provokes heightened caution as to how that emphasised decision is rendered.

2.2.2.2. Unequal variance sensitivity and bias. The correspondence of the above effects was probed with parameters estimated from the unequal variance SDT model, which takes into account variation in response confidence. A 2(question dimension) x 2(question emphasis) ANOVA on UEV d' revealed a significant main effect of dimension, with UEV d' higher in the single ($M = 2.17$, $SD = 0.98$) compared to the composite condition ($M = 1.75$, $SD = 0.68$), $F(1,29) = 9.69$, $p = .004$, $\eta_p^2 = .25$. There was neither a main effect of emphasis nor an interaction, both *F*s < 1. The same 2 x 2 ANOVA on UEV c_{rel} yielded neither a main effect of dimension nor an interaction, both *F*s < 1. As before, there was a trend towards a main effect of emphasis, such that c_{rel} was higher for old (*M* = -0.07, *SD* = 0.36) compared to new emphasis questions (*M* = -0.18, *SD* = 0.38), *F*(1,29) = 3.20, *p* = .084, η_p^2 = .099 (see Table 2.1. for means). Collectively, the UEV analyses accord with the EV analyses in recovering a trend effect of question emphasis on criterion bias (with *d'* sensitivity measures unaffected by emphasis).

2.2.2.3. Decision accuracy. To elucidate the strategic nature of the question format bias, we also analysed how the accuracy of "old" and "new" decisions (old_{corr} and new_{corr} respectively) varied across question conditions. A 2(question dimension) x 2(question emphasis) ANOVA for old_{corr} revealed a main effect of dimension, such that "old" decisions were more accurate in the single (*M* = .81, *SD* = .08) compared to the composite condition (*M* = .77, *SD* = .09), *F*(1,29) = 10.14, *p* = .003, η_p^2 = .26. This is consistent with the SDT sensitivity effects detailed above and is hence also attributable to the delayed presentation of the composite question blocks relative to the single question blocks. There was also a trend main effect of question emphasis such that the accuracy of "old" decisions was higher in the "old" emphasis (*M* = .80, *SD* = .08) compared to the "new" emphasis condition (*M* = .77, *SD* = .09), *F*(1,29) = 3.79, *p* = .061, η_p^2 = .12. There was no dimension by emphasis interaction, *F* < 1. The effects on old_{corr} highlight that the bias induced by question emphasis is strategic in nature and serves to improve the accuracy of the emphasised decision.

The same 2 x 2 ANOVA conducted on new_{corr} revealed an analogous order-driven main effect of dimension, such that "new" decisions were more accurate in the single (*M* = .80, *SD* = .08) compared to the composite condition (*M* = .77, *SD* = .09), *F*(1,29) = 9.96, *p* = .013, η_p^2 = .19. Interestingly, both the main effect of emphasis and the emphasis by dimension interaction were clearly non-significant, both *F*s < 1 (see Figure 2.1d). The clear lack of an emphasis effect on new_{corr} suggests that the question emphasis-provoked strategies might preferentially impact on “old” decision evaluations, a finding that is considered in greater detail in the general discussion (see Section 2.5.).

Overall, the findings from Experiment 1 provide preliminary evidence for question formats presented as standard in the field of recognition memory acting as an implicit source of goal emphasis in the test environment. The goal emphasis imparted by these questions selectively influenced criterion estimates (with sensitivity comparatively unaffected) to elicit a disconfirmatory bias that reduced the likelihood of endorsing the decision emphasised by the question. Further, the decision accuracy analyses highlighted the strategic nature of this bias, with question emphasis also improving the accuracy of emphasised decisions when rendered. However, the non-significant trends evidenced in the analyses prevent strong inferences based on these data. The lack of significance is perhaps due to the reduced power of Experiment 1, derived from the relatively low number of trials per condition and the small sample collected. This interpretation was supported by an independent study in which single dimension questions of varying emphases were presented to a large sample, revealing analogous effects of question emphasis on criterion and decision accuracy that reached conventional significance[3] (Mill & O’Connor, 2014; see Appendix A).

[3] This experiment was not included in my thesis as it was programmed in JavaScript by my supervisor, Akira O’Connor, and not by me.

Table 2.1. Design key and descriptive statistics for all experiments, including equal variance (EV) and unequal variance (UEV) estimate of sensitivity (d') and criterion (c and c_{rel}), as well as decision accuracy for "old" (old_{corr}) and "new" decisions (new_{corr}).

		Experiment 1				Experiment 2				Experiment 3			
Encoding task		syllables				case		pleasantness		case			
Question format		old?	new?	old/new?	new/old?	old?	new?	old?	new?	old?	new?	old?	new?
Response format		-	-	-	-	-	-	-	-	yes	yes	no	no
Emphasis		**old**	**new**	**old**	**new**	**old**	**new**	**old**	**new**	**old**	**new**	**new**	**old**
EV d'	*M*	1.81	1.69	1.54	1.47	1.37	1.28	2.83	2.57	1.26	1.28	1.26	1.18
	SD	0.55	0.64	0.54	0.58	0.70	0.87	0.63	0.54	0.55	0.82	0.80	0.66
EV c	*M*	0.05	-0.03	0.04	-0.05	0.26	0.07	-0.09	-0.13	0.12	0.00	-0.07	0.08
	SD	0.35	0.25	0.38	0.43	0.41	0.26	0.34	0.32	0.38	0.34	0.30	0.35
UEV d'	*M*	2.24	2.10	1.70	1.79	1.55	1.45	3.62	3.18	1.52	1.42	1.32	1.44
	SD	1.23	1.17	0.72	0.86	0.82	1.00	0.99	1.24	1.06	0.68	0.74	0.86
UEV c_{rel}	*M*	-0.07	-0.15	-0.07	-0.22	0.07	-0.09	-0.31	-0.26	-0.03	-0.24	-0.22	-0.08
	SD	0.49	0.32	0.38	0.56	0.50	0.35	0.51	0.44	0.45	0.41	0.39	0.47
old_{corr}	*M*	.82	.79	.78	.76	.78	.73	.90	.87	.74	.71	.70	.72
	SD	.09	.09	.09	.11	.11	.12	.07	.07	.11	.12	.12	.12
new_{corr}	*M*	.81	.80	.77	.77	.71	.71	.92	.91	.70	.71	.72	.68
	SD	.08	.10	.10	.10	.10	.11	.07	.07	.08	.10	.12	.10

Note: *M*, mean; *SD*, standard deviation.

EXPERIMENT 1

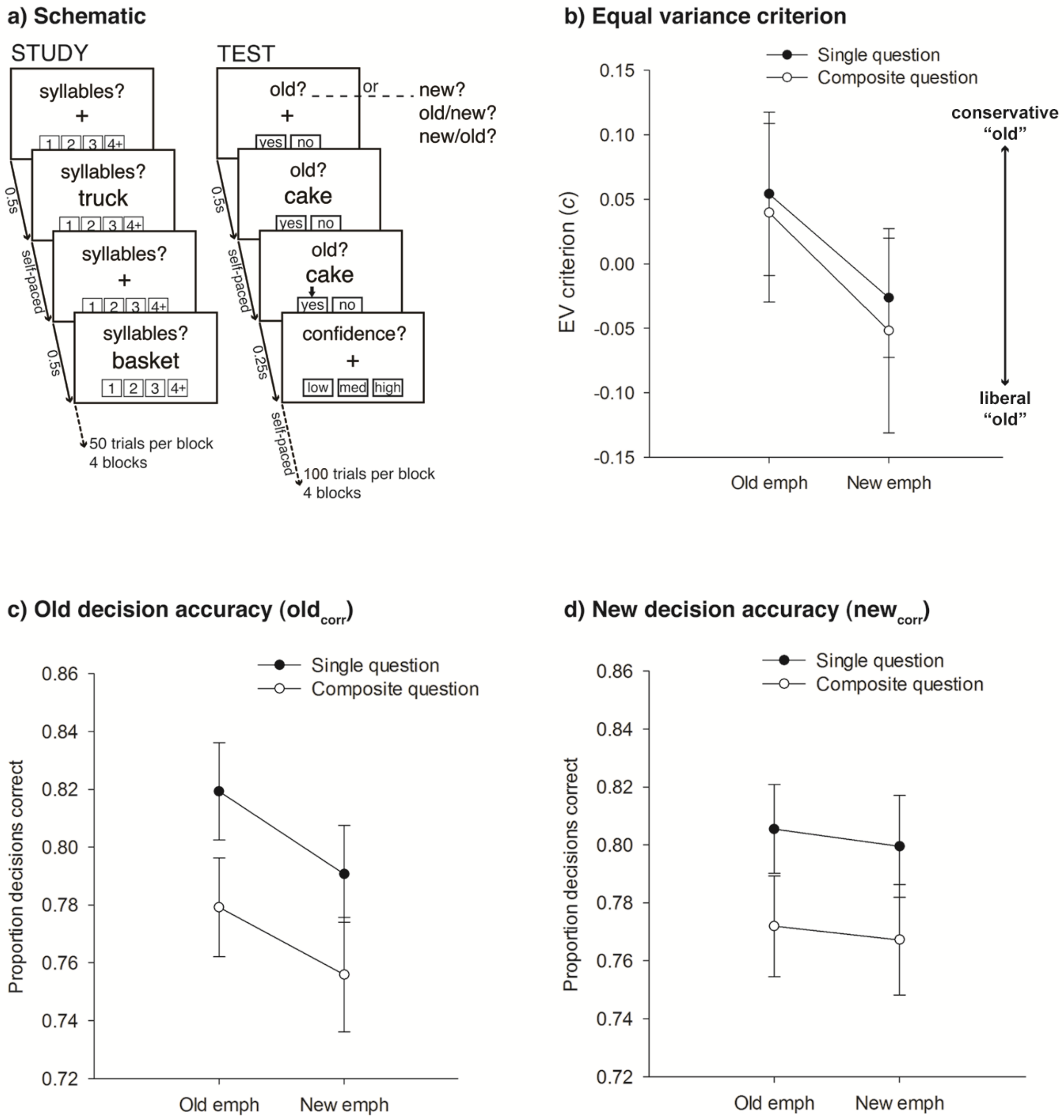


Figure 2.1. Design and results of Experiment 1. **a)** Design schematic showing the study phase (which was the same for all test phases) and the test phase for the single dimension "old" emphasis question ("old?"). Note that the response options for the single dimension "new" emphasis question ("new?") was identical to that presented above. Response options for composite dimension questions were as follows: "old/new?" question with the "old" response to the left of the "new" response, and "new/old?" question with "new" to the left of the "old" option. The remaining graphs depict question emphasis effects on **b)** equal variance criterion (EV *c*), **c)** old decision accuracy (old_{corr}), and **d)** new decision accuracy (new_{corr}). Note that the arrowheads to the right of panel **b)** illustrate the bias tendencies

indexed by criterion placement (for this and subsequent criterion figures), such that increasing *c* values reflect a reduced likelihood of responding “old” (i.e. conservative “old” bias) and decreasing values reflect a greater likelihood of responding “old” (i.e. liberal “old” bias). Separate lines in panels **b-d** represent question dimension conditions and error bars represent standard error of the mean.

2.3. Introduction to Experiment 2

Another potential contributor to the numerically small question effects of Experiment 1 was the reasonably high levels of memory strength available at test, as predicated on the "deep" levels-of-processing (LOP; Craik & Lockhart, 1972) achieved by the syllable counting task. Sufficiently high memory strength reduces the need to recruit higher-order strategies to accurately diagnose memory decisions (Jacoby et al., 1999; Koriat & Goldsmith, 1996). To investigate this mediating influence, memory strength was systematically manipulated in Experiment 2 by presenting different tasks at study – a pleasantness judgement task leading to "deep" LOP and a case judgement task leading to "shallow" LOP. This study manipulation was crossed with analogous manipulation of question emphasis as in the single dimension question condition in Experiment 1. To clarify, the case judgement task was expected to instil a lower level of overall memory strength than the syllable-counting task in Experiment 1 (i.e. expected LOP: pleasantness > syllable counting > case judgement), and therefore enhance the emphasis-driven shifts in criterion bias and decision accuracy. Such enhancement would further indicate that the question effect is rooted in higher-order strategies that are bear a greater influence on performance in situations of low overall memory strength and associated uncertainty (Benjamin, 2007; Kahneman, Slovic & Tversky, 1982).

2.3.1. Method

2.3.1.1. Participants. The sample comprised 29 native English speakers who reached the minimum performance threshold (20 female; mean age = 22.1; age range = 19 to 28) from a total of 31 recruited. Two participants were excluded for poor performance based on the same threshold as in Experiment 1 (i.e. $d' < 0.1$). Informed consent and participant compensation were identical to Experiment 1.

2.3.1.2. Stimuli. Word stimuli were randomly drawn from the same pool used in Experiment 1. Each of the four study-test blocks constituted 120 words (60 *old* and *60* new at each test phase), which were presented on PCs running MATLAB and Psychophysics Toolbox as before.

2.3.1.3. Procedure. After the presentation of on-screen instructions and a practice phase, participants completed four self-paced study-test blocks comprising blocked combinations of study LOP (shallow, deep) and test question emphasis ("old?", "new?"; see Figure 2.2a). The two shallow LOP study phases involved a case judgement task for 60 words presented serially in either uppercase or lowercase (30 of each case type) below the prompt "uppercase?" and with "yes" and "no" response options assigned to the keyboard. The two deep LOP study phases involved a pleasantness judgement task, wherein lowercase words appeared beneath a "pleasant?" question with "yes" and "no" response options. A test phase immediately followed each study phase and comprised 60 *old* and 60 *new* words. Participants made old/new recognition decisions with test question emphasis varied analogously to the single dimension question condition in Experiment 1 i.e. with "old?" and "new?" questions presented in separate test blocks (with "yes" and "no" response options available throughout). As before, three-point confidence ratings were solicited 0.25s after each old/new decision. A 0.5 s fixation cross preceded each trial across all study and test blocks. Across participants, the study-test block order was pseudo-randomised such that only one level of a factor would change in any block transition. Hence, while it was possible for a participant to experience a shallow-"old?" to shallow-"new?" block transition, it was not possible for a participant to experience a shallow-“old” to deep-“new” block transition. This led to four orders in total to which participants were randomly assigned.

2.3.2. Results and Discussion

2.3.2.1. Equal variance sensitivity and bias. To confirm that the LOP manipulation affected memory strength as anticipated, a 2(LOP: shallow or deep) x 2(question emphasis: “old?” or “new?”) repeated measures ANOVA was conducted on EV estimates of *d′.* The ANOVA revealed an anticipated main effect of LOP, such that sensitivity was lower in the shallow LOP (M = 1.33, SD = 0.73) than the deep LOP condition (M = 2.70, SD = 0.52), $F(1,28)$ = 109.14, p = .001, η_p^2 = .80. There was also an unexpected main effect of question emphasis such that *d′* was higher in the old (M = 2.10, SD = 0.55) than the new emphasis condition (M = 1.92, SD = 0.57), $F(1,28)$ = 5.84, p = .022, η_p^2 = .17. There was no significant LOP x emphasis interaction, $F(1,28)$ = 1.06, p = .311, η_p^2 = .04. The unexpected main effect of emphasis on *d′* suggests that emphasis impacted on memory strength processes and this finding is elaborated on in ensuing sections.

The primary analysis on EV c was carried out using a 2(LOP) x 2(question emphasis) repeated measures ANOVA. A main effect of question emphasis was observed, with significantly higher c estimates in the old (M = 0.09, SD = 0.24) than the new emphasis condition (M = -0.03, SD = 0.32), $F(1,28)$ = 6.65, p = .015, η_p^2 = .19 (see Figure 2.2b). Although there was no significant interaction effect, $F(1,28)$ = 3.13, p = .088, η_p^2 = .101, a numerical trend was observed, and planned pairwise comparisons revealed that the disconfirmatory question bias achieved statistical significance in the shallow LOP condition but not in the deep LOP condition, $t(28)$ =

3.05, p = .005, d = 0.62 and $t(28)$ = .70, p = .487, d = 0.13 respectively. These findings replicate the disconfirmatory emphasis bias observed in Experiment 1, and support our prediction of its enhancement under conditions of high overall memory strength.

From the same ANOVA on c, we found a main effect of LOP on criterion placement, such that c estimates were significantly lower in the deep (M = -0.11, SD = 0.29) than the shallow LOP conditions (M = 0.17, SD = 0.30), $F(1,28)$ = 23.66, p = .001, η_p^2 = .46. Participants hence adopted a liberal “old” response strategy in the condition of higher memory strength, and this finding is reminiscent of prior reports of criterion sensitivity to metacognitive inferences of overall memory strength (as alluded to in the previous chapter, see Section 1.4.; Glanzer & Adams, 1985; Hirshman, 1995).

2.3.2.2. Unequal variance sensitivity and bias. As in Experiment 1, we probed the correspondence between the above equal variance SDT effects in equivalent unequal variance analyses. A 2(LOP) x 2(question emphasis) repeated measures ANOVA on UEV d' recovered a main effect of LOP, such that d' was significantly higher in the deep (M = 3.40, SD = 0.89) than the shallow LOP condition (M = 1.50, SD = 0.85), $F(1,28)$ = 88.89, p = 0.001, η_p^2 = 0.76. A trend main effect of emphasis was also observed, with sensitivity once again higher under old emphasis (M = 2.59, SD = 0.67) compared to new emphasis (M = 2.32, SD = 0.89), $F(1,28)$ = 3.16, p = 0.086, η_p^2 = 0.10. The LOP by emphasis interaction was non-significant, $F(1,28)$ = 1.62, p = 0.214, η_p^2 = 0.06.

A 2(LOP) x 2(question emphasis) ANOVA on c_{rel} revealed a main effect of LOP, such that c_{rel} was lower in the deep (M = -0.28, SD = 0.40) than the shallow LOP conditions (M = -0.01, SD = 0.36), F(1,28) = 12.23, p = 0.002, η_p^2 = 0.304. There was neither a main effect of emphasis nor a significant interaction, F(1,28) < 1 and F(1,28) = 2.73, p = 0.110, η_p^2 = 0.089 respectively. However, planned pairwise comparisons revealed a trend towards a significant disconfirmatory bias in the shallow LOP condition, t(28) = 1.87, p = .072, d = 0.36, with a higher c_{rel} under old compared to new emphasis, but a clearly non-significant difference in c_{rel} in the deep LOP condition, t(28) = 0.60, p = .556, d = 0.11 (see Table 2.1. for means). Overall, the effects of question emphasis on unequal variance estimates of d' and c_{rel} are broadly consistent with those observed in the equal variance estimate analyses. The weakened effects of question emphasis on c_{rel} likely reflect the conservative nature of the UEV model, in that it aims to account for added response variability associated with assessments of confidence.

2.3.2.3. Decision accuracy. As before, the accuracy of “old” and “new” decisions under varying question emphases was analysed, separately for shallow and deep LOP conditions. A 2(decision type) x 2(question emphasis) repeated measures ANOVA for the shallow LOP condition revealed a significant main effect of decision type, F(1,28) = 9.77, p = .004, η_p^2 = .26 (see Figure 2.2c). Participants’ “old” decisions were characterised by greater overall accuracy (M = .76, SD = .10) than their “new” decisions (M = .71, SD = .09). Similar disparities in the response profiles of “old” and “new” decisions have been reported previously (e.g. Jaeger, Cox & Dobbins, 2012). No main effect of question emphasis was observed, F(1,28) = 1.46, p = .237, η_p^2 = .050. Crucially, the decision type x question emphasis interaction was significant, F(1,28) = 9.04, p = .006, η_p^2 = .244. This suggests that question emphasis improved the accuracy of endorsing the emphasised decision i.e. “old?” improved the accuracy of “old” decisions and “new?” improved the accuracy of “new” decisions (see Table 2.1. for means). Post-hoc comparisons revealed a significant difference in old_{corr} across emphasis conditions and a nonsignificant difference for new_{corr}, $t(28) = 2.23$, $p = .034$, $d = 0.41$ and $t(28) = 0.42$, $p = .679$, $d = 0.07$ respectively. The greater influence of emphasis on old_{corr} compared to new_{corr} parallels the decision accuracy analyses of Experiment 1 in suggesting that the strategies instilled by the question format manipulations might selectively improve the accuracy of "old" decision evaluations. This possibility is discussed in greater detail in the General Discussion in Chapter 7 (Section 7.3.).

A 2 x 2 ANOVA was also conducted for decision accuracy measures in the deep LOP condition, yielding a nonsignificant main effect of decision type and a nonsignificant interaction, $F(1,28) = 3.80$, $p = .061$, $\eta_p^2 = .12$ and $F(1,28) = 1.00$, $p = .326$, $\eta_p^2 = .03$ respectively. A significant main effect of question emphasis was observed, with the question "old?" ($M = .91$, $SD = .05$) leading to greater overall decision accuracy than the question "new?" ($M = .89$, $SD = .05$), $F(1,28) = 5.65$, $p = .025$, $\eta_p^2 = .17$ (see Figure 2.2d). Considered with the observed improvement in *d′* for "old?" questions in the same deep LOP condition, these findings raise the possibility that the question "old?" instilled a more rigorous source monitoring strategy that prioritised the recovery of recollected content (Johnson, Hashtroudi & Lindsay, 1993). In conditions of high overall memory strength (as in the deep LOP condition), this directly improved performance as a higher proportion of recollected *old* items were likely encountered at test. This contrasts the lack of a main effect of emphasis on decision accuracy or *d′* in the shallow LOP condition, where *old* items were less likely to be associated with recollected content and hence the adoption of a monitoring strategy that exclusively prioritised recollection exerted a less beneficial influence on performance. As such, these question format effects putatively impact more directly on memory strength and related retrieval control processes, and hence reflect different strategic influences to the decision control processes of primary interest (see Chapter 7, Section 7.2. for further discussion).

To summarise, the results of Experiment 2 provided further evidence for question format serving as an implicit source of goal emphasis in the test environment. The trend effects observed in Experiment 1, namely the criterion bias reflecting a reduced likelihood of endorsing emphasised decisions coupled with improved accuracy of those decisions when made, were found to be enhanced in the shallow LOP condition as predicted. This is consistent with the heightened engagement of goal-driven strategies to counteract weakly diagnostic memory evidence and ensure accurate decision outcomes.

EXPERIMENT 2

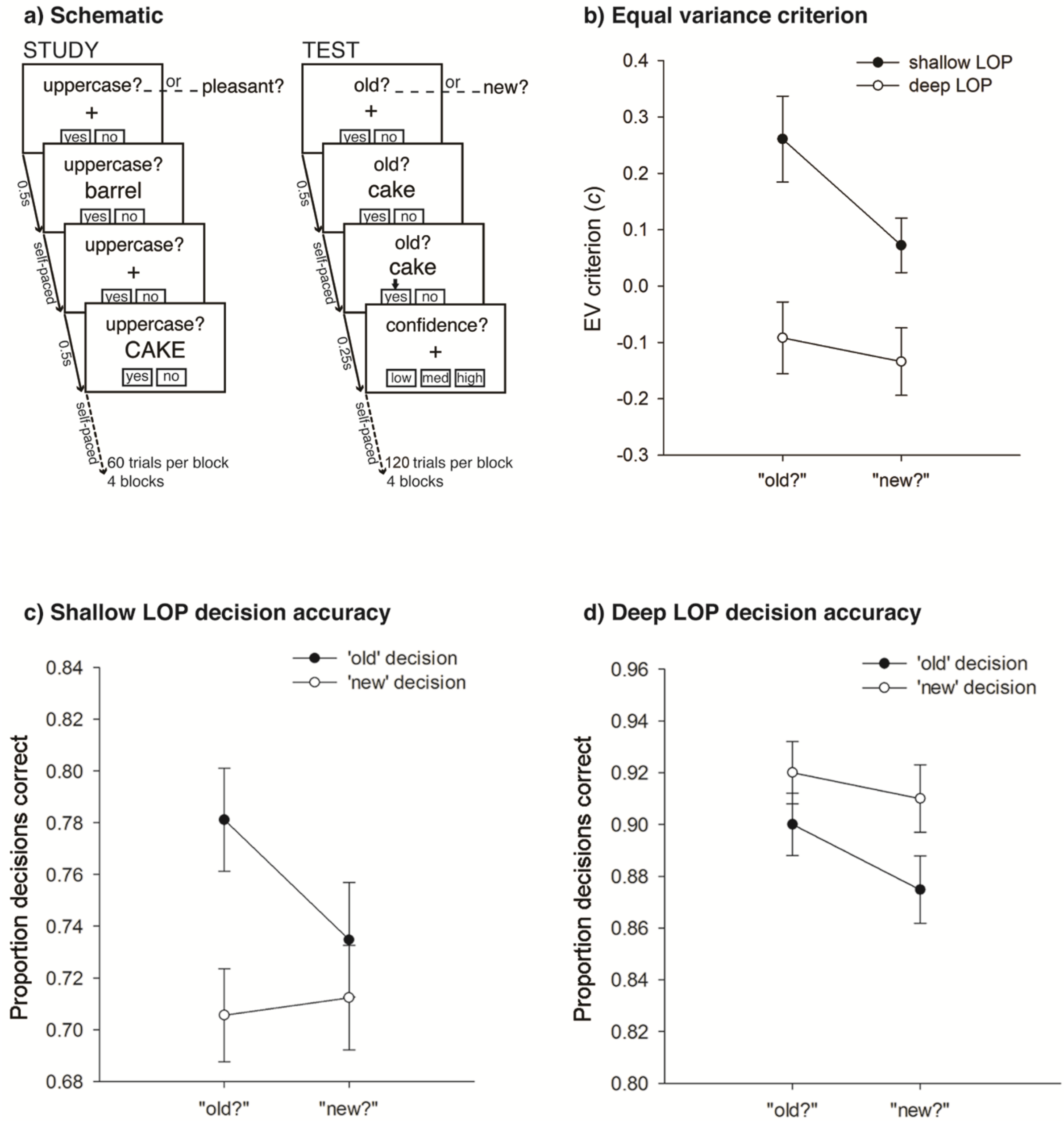


Figure 2.2. Design and results of Experiment 2. **a)** Design schematic showing study and test phase manipulations. **b)** Question emphasis effects on equal variance criterion (EV *c*), with separate lines denoting levels-of-processing (LOP) achieved at study. The remaining graphs show effects of question emphasis on decision accuracy measures in **c)** the shallow LOP condition and **d)** the deep LOP condition, with separate lines representing decision types. Error bars represent standard error of the mean.

2.4. Introduction to Experiment 3

The observed question format effects on recognition performance have thus far been suggestive of a goal-driven strategic bias, wherein the test question implicitly emphasises a particular memory decision as a higher-order goal and enforces more cautious endorsement of that decision. However, in both presented experiments, the varying question emphases were confounded by static “yes” and “no” response options. The question format effect may hence alternatively reflect a tendency to respond “no” irrespective of implicitly emphasised decision goals.

To adjudicate between these interpretations, we independently manipulated question format as in Experiment 2, and introduced an orthogonal response format manipulation (see Figure 2.3a.) Response format was manipulated by providing only one physical, visible response option at test, with the alternative response indicated by withholding from using that response option. This is similar to laboratory-based go/no go procedures (e.g. Brutkowski, 1964) and applied eye-witness identification protocols where participants are given the option of not-responding in perpetrator-free line-ups (e.g. Weber & Perfect, 2011). Across study-test blocks, the single response option was assigned either a "yes" or "no" response, thereby requiring participants to integrate both the question and the response formats to infer emphasised decisions. Varying these two test environment factors independently allowed us to test whether criterion and decision accuracy effects shift according to the decision emphasised as a goal by the combination of the question and response formats, or whether these follow the decision which maps onto a "no" response irrespective of emphasis (see Figure 2.3b for a detailed explanation of these predictions).

EXPERIMENT 3

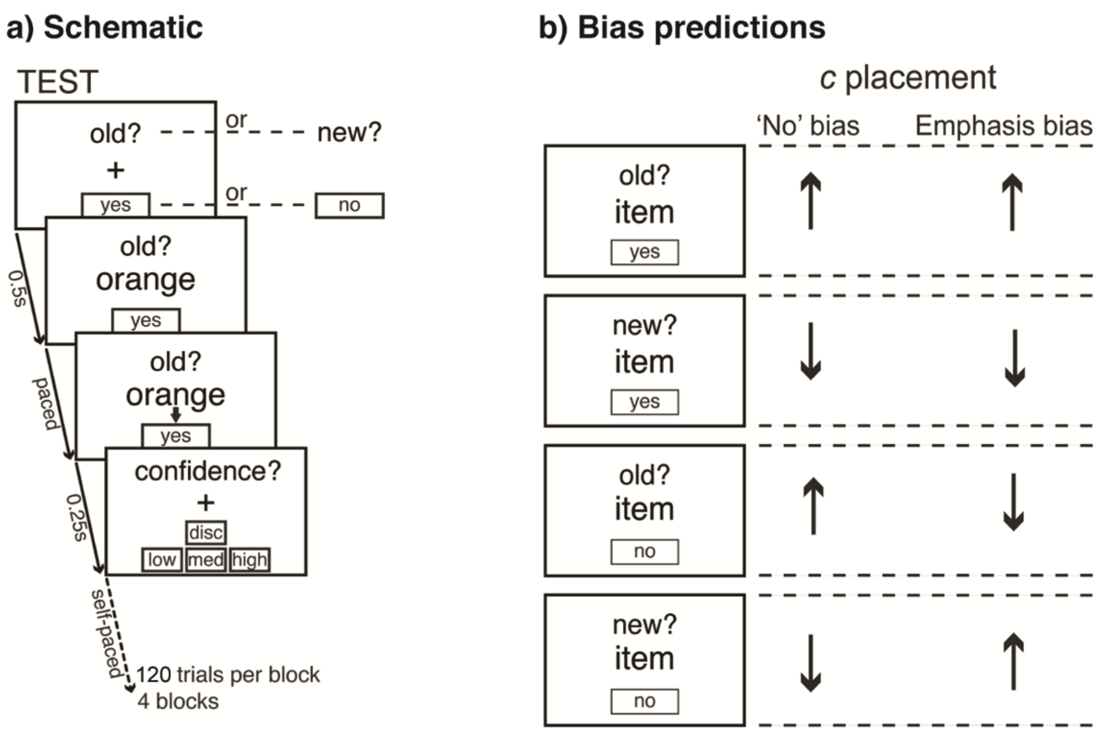


Figure 2.3. Experiment 3 design. **a)** Design schematic showing blockwise test format manipulations (study phase comprised the same case judgement task from Experiment 2). **b)** Criterion placement (*c*) as generated by the competing bias predictions for Experiment 3. Arrows denote criterion shifts, with increasing *c* associated with decreased “old” responding (conservative bias) and decreasing *c* associated with increased “old” responding (liberal bias). According to the “no” attraction prediction, participants consistently prefer responding “no”, and hence *c* should decrease when an “old” decision maps to the “no” response option (“new?”-“yes” and “new?”-“no” test blocks), and increase when it maps to the ‘yes’ response option (“old?”-“yes” and “old?”-“no” test blocks). According to the goal emphasis prediction, the *combination* of the question and response formats serves to emphasise one class of decision, and participants decrease endorsement of that decision. Hence, *c* should increase when “old” decisions are emphasised (“old?”-“yes” and “new?”-“no” test blocks) and decrease when “new” decisions are emphasised (“new?”-“yes” and “old?”-“no” test blocks).

2.4.1. Method

2.4.1.1. Participants. The sample comprised 29 native English speakers who reached the minimum performance threshold of $d' > 0.1$ (21 female; mean age = 20.6; age range = 18 to 25) from a total of 35 completing the experiment (4 were excluded for failing to understand task instructions, 2 for poor task performance). Informed consent and compensation procedures were identical to those in the previous two experiments.

2.4.1.2. Stimuli. Word stimuli were randomly drawn from the same pool used in Experiments 1 and 2. Each of five study-test blocks utilised 120 words (60 *old* and 60 *new* at each test phase) and the experiment was presented using PCs running MATLAB and Psychophysics Toolbox (as before).

2.4.1.3. Procedure. Following on-screen instructions and a practice phase, participants completed five study-test blocks (see Figure 2.3.a). All study phases employed a self-paced case judgement task presented and responded to in a manner identical to Experiment 2, and were immediately followed by old/new recognition test phases comprising 60 *old* words and 60 *new* words. The first test phase was used to calibrate the paced response window used in subsequent test phases, and employed the composite “old/new?” question format presented in Experiment 1 followed by a three-point confidence rating (with both trial responses being self-paced). The response window for subsequent test phases was set, on a participant-by-participant basis, as the 80th percentile of the slowest category of recognition decisions broken down by confidence (typically “low” confidence decisions). This liberal response window ensured that participants had sufficient time to make paced recognition decisions in subsequent test phases (M = 3.84 s, SD = 1.69).

The next four test phases comprised blocked combinations of question format ("old?" or "new?") and response format manipulations ("yes" keypress or "no" keypress; see Figure 2.3.a). Question format was manipulated across blocks as described for Experiment 2. Response format was manipulated across blocks by allocating either "yes" or "no" to a single keyboard response. Participants endorsed the allocated response with a keypress or endorsed the opposing response by withholding the keypress for the duration of the paced trial (similar to a go/no-go task, Brutkowski, 1964). To prevent preferential keypress responding to hasten completion of the task, the second stage of responding was initiated after the full response window had elapsed, irrespective of whether a keypress was made or withheld. Subsequently, participants received the prompt "confidence?" and provided a self-paced confidence rating of "low", "medium" or "high". Participants could alternatively make a "discard" response if they wanted their previous recognition decision to be ignored. This option was intended for when participants had failed to render a response due to an attentional lapse and prevented such trials from being coded as deliberately withheld responses. A 0.5 s fixation cross preceded each recognition assessment and a 0.25 s fixation cross preceded each confidence assessment. Hence, across the four study-test blocks, participants completed two test phases with an old decision emphasis ("old?"-"yes" and "new?"-"no"), and two test phases with a new decision emphasis ("old?"-"no" and "new?"-"yes"). The order of the four tests was pseudorandomised to minimise the combined switching of question and response formats across study-test blocks, leading to four orders in total.

2.4.2. Results and Discussion

2.4.2.1. Equal variance sensitivity and bias. A 2(response format: "yes" keypress or "no" keypress) x 2(question format: "old?" or "new?") repeated measures ANOVA on equal variance estimates of *d′* revealed no significant main or interaction effects, all *Fs* < 1. Sensitivity was therefore unaffected by the test format manipulations.

The same 2(response format) x 2(question format) ANOVA on EV *c* found no main effects of response or question format, $F(1,28) = 3.73$, $p = .064$, $\eta_p^2 = .12$ and $F < 1$ respectively (see Figure 2.4a). Crucially, the interaction of response and question format was significant, $F(1,28) = 8.50$, $p = .007$, $\eta_p^2 = .23$ (see Table 2.1. for means). Planned pairwise t-tests revealed significant increases in *c* (reflecting a reduced likelihood of responding "old") when old was emphasised by the combination of response and question formats. For the "yes" response format, *c* was placed higher for "old?" than "new?" questions, $t(28) = 2.41$, $p = .023$, $d = 0.45$. For the "no" response format, *c* was lower for "old?" than "new?" questions, $t(28) = 2.77$, $p = .010$, $d = 0.52$. Note that in the "no" response format condition, old emphasis is imparted by a "new?" question and new emphasis is imparted by an "old?" question (see Figure 2.3b for further details). These findings therefore replicate the biases observed in Experiments 1 and 2, and suggest that the effect is indeed driven by which decision is emphasised as a goal, and not by a general tendency to respond "no".

2.4.2.2. Unequal variance sensitivity and bias. Test format effects were also examined with UEV estimates of performance. A 2(response format) x 2(question format) repeated measures ANOVA on UEV *d′* revealed no significant main or interaction effects, all *Fs* < 1. The same 2 x 2 ANOVA for UEV c_{rel} found no main effects of response format or question format, both *Fs* < 1. As above, the response format x question format interaction was significant, $F(1,28) = 4.57$, p = 0.041, $\eta_p^2 = 0.140$. Planned pairwise comparisons revealed that within the "yes" response format condition, old question format c_{rel} estimates were significantly higher than new question format c_{rel} estimates, $t(28) = 2.65$, $p = .013$, $d = 0.49$. Within the "no" response format condition, c_{rel} estimates did not significantly differ across question format conditions, $t(28) = 1.35$, $p = .187$, $d = 0.25$ (see Table 2.1.). These findings coalesce with the equal variance analyses in suggesting that the disconfirmatory question bias is indeed driven by goal emphasis.

2.4.2.3. Decision accuracy. The accuracy of "old" and "new" decisions was also analysed, separately for each response format condition. A 2(decision type) x 2(question format) repeated measures ANOVA for the "yes" response condition yielded no main effects of decision type or question format, $F(1,28) = 1.98$, $p = .171$, $\eta_p^2 = .066$ and $F < 1$ respectively (see Figure 2.4b; see Table 2.1. for means). Importantly, the decision type x question format interaction was significant, in a direction such that "old?" improved the accuracy of "old" decisions and "new?" improved the accuracy of "new" decisions, $F(1,28) = 14.19$, $p = .001$, $\eta_p^2 = .336$. Pairwise comparisons revealed a numerical difference approaching significance in "old" decision accuracy across question conditions, $t(28) = 2.01$, $p = .054$, $d = 0.37$, with a nonsignificant difference direction obtained for "new" decision accuracy, $t(28) = 0.48$, $p = .637$, $d = 0.09$. The same 2 x 2 ANOVA for the "no" response condition yielded no significant main effects of decision type or question format, both *F*s < 1 (see Figure 2.4c). Crucially, the decision type x question format interaction was significant, with "new?" improving the accuracy of "old" decisions and "old?" improving the accuracy of "new" decisions, $F(1,28) = 14.19$, $p = .001$, $\eta_p^2 = .336$ (as mentioned before, this bias is in a goal-driven direction, given the inverse emphasis imparted by questions in the “no” response condition; see Figure 2.3c). Pairwise comparisons revealed a trend approaching significance for new decision accuracy across question conditions, $t(28) = 2.04$, $p = .051$, $d = 0.39$, with a nonsignificant effect obtained for old decision accuracy, $t(28) = 1.12$, $p = .271$, $d = 0.21$. The decision accuracy findings from Experiments 1 and 2 were hence replicated in both response format conditions - question emphasis improved the accuracy of endorsing the emphasised decision.

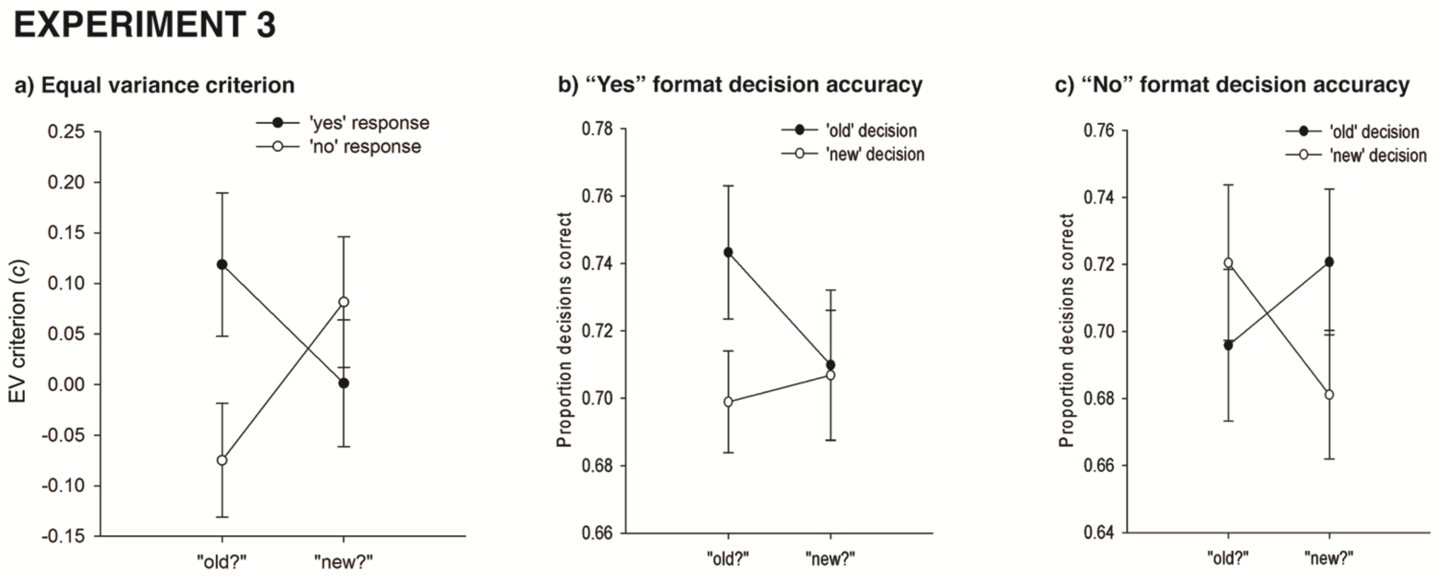


Figure 2.4. Experiment 3 results. **a)** Question format effects on equal variance criterion (EV *c*), with separate lines denoting response format conditions. The remaining graphs show effects of question emphasis on decision accuracy measures in **b)** the "yes" format condition and **c)** the "no" format condition, with separate lines representing decision types. Error bars represent standard error of the mean.

2.5. Overall Discussion

The experiments presented in this chapter provide evidence of a hitherto overlooked role of test question format as a source of goal information in memory evaluation. The behavioural trends of Experiment 1 suggested that test questions adopted as standard in recognition research are capable of influencing memory decision outcomes. Experiment 2 followed up these preliminary findings with systematic manipulation of memory strength, and recovered heightened test question effects under conditions of low overall memory strength. Experiment 3 tested two competing explanations for the question effects, and found evidence in favour of a goal emphasis mechanism rather than a habitual preference for responding “no”. In all three experiments, we replicated the effect of question emphasis shifting SDT criterion (as estimated from two prominent variants of the recognition SDT model) to instil a disconfirmatory bias that reduced the likelihood of endorsing the emphasised decision. Our analyses involving the more unorthodox decision accuracy measures also recovered a consistent effect, wherein question emphasis improved the accuracy of endorsements of the emphasised decision. Considered with the criterion bias effects, the pattern of results suggests that question emphasis improves the accuracy of making emphasised responses at the expense of their frequency, consistent with the controlled engagement of strategic caution. These results are now discussed with reference to previous findings in the fields of memory and decision-making.

Prior studies investigating effects of test format in standard single item recognition have typically focused on measures of source memory or subjective experiences of memory (Bastin & Van der Linden, 2003; Dodson & Johnson, 1993; Hicks & Marsh, 1999; Khoe, Kroll, Yonelinas, Dobbins, & Knight, 2000; Marsh & Hicks, 1998). For example, Hicks and Marsh (1999) reported an increase in "old" responding in a recognition test phase comprising three response options with two "old" subcategories ("remember *old*"/"know *old*"/"*new*") compared to a test phase with two equally weighted response options ("*old*"/"*new*", followed by "remember"/"know"; though cf. Bruno & Rutherford, 2010). Here we present evidence of a test format bias in the primary measure of recognition performance, which manifests in the absence of unequally weighted response options. Nevertheless, our findings accord with previously reported test format biases in highlighting the salient influence of aspects of the test environment on the outcomes of memory evaluation (Johnson et al., 1993).

Further, the observed disconfirmatory bias was in the opposite direction to previously reported test format biases, such as the acquiescence bias in questionnaire responding ('yea-responding'; Podsakoff, et al., 2003), the Loftus framing effect in eyewitness testimony (Loftus, 1996) and decision cueing in recognition research (e.g. O'Connor, et al., 2010; Rhodes & Jacoby, 2007). We suggest that our emphasis-driven bias differs from these other biases in that it establishes a decision goal against which diagnostic evidence is evaluated, rather than contributing to the evidence itself. This aspect serves to distinguish the present question effects from the classic Loftus framing effect, in which questions that framed details of potentially encoded contexts ("leading questions") led to a confirmatory response tendency that enhanced fallacious endorsement of contexts as having been previously experienced (Loftus, 1996). Our questions lacked any such allusion to explicit aspects of encoding, which would putatively impact upon memory evaluation by directly interfering with the assessment of memory evidence i.e. via the retrieval control processes outlined in the previous chapter (see Section 1.3.). The question bias is similarly differentiated from the cueing bias, which despite also acting on the same broad category of decision control processes (operating independent of memory strength) as the emphasis bias, nevertheless serves to instil cued expectations that serve as an additional diagnostic evidence base for responding. Rather, our question manipulations impacted via a different strategy, by implicitly emphasising a particular memory decision as a higher order goal, the endorsement of which demanded a higher *overall* level of diagnostic evidence relative to the non-emphasised decision. Our findings therefore inform a tentative distinction between three major strategic biases wrought by different aspects of the retrieval environment: decision cueing (confirmatory), evidence framing (confirmatory) and goal emphasis (disconfirmatory). Chapter 4 examines the interaction of cueing and goal emphasis manipulations with a view to directly interrogating the above bias trichotomy.

The present effects hence reveal a strategic influence on memory evaluation that is potentially consistent with findings in the rewarded decision-making literature, particularly mixed incentive research. In mixed incentive research, monetising one of two available response options leads to a reluctance to endorse the monetised decision (Newman et al., 1985). A similar behavioural trend indicative of this disconfirmatory bias was reported in a previous recognition experiment that differentially provided mixed incentives for “old” or “new” decisions (Han, Huettel, Raposo, Adcock & Dobbins 2010; see also Section 1.5.). The present findings are notable for inducing comparable caution without the provision of incentives, and demonstrate that even in the absence of any explicit mention of reward, participants display a goal-directed bias that is driven by implicit information gleaned from the testing environment. Chapter 4 directly tests the correspondence of the present question emphasis effects with more explicit emphasis mechanisms engaged by mixed incentives.

We also found evidence that the test question bias, whilst decreasing the frequency of the emphasised decision, improved the accuracy of these decisions when they were made. As outlined in the previous chapter, manipulations enacted solely at retrieval have little or no impact on old/new discrimination sensitivity, which is largely determined by encoding processes (Craik & Lockhart, 1972; Stretch & Wixted, 1998; albeit with some exceptions; Dobbins & McCarthy, 2008; Marsh & Hicks, 1998). Indeed, across all three experiments our retrieval manipulations had little effect on measures of *d′* sensitivity that binned responses by 'objective' item categories and collapsed across old and new decisions. However, the observed effects on decision accuracy measures old_{corr} and new_{corr} (which binned responses by 'subjective' decision categories) highlight the efficacy of these more unorthodox analyses in elucidating a strategic effect which would otherwise be somewhat obscured by a reliance on SDT analyses alone. Analogous influences of the test environment on strategic use of retrieved memory evidence have been highlighted previously by models of memory control (Koriat & Goldsmith, 1996). In the present context, it furthers our suggestion of a goal-directed mechanism underlying the question format effects, which prompted participants to adjust their criterion towards a more rigorous evaluation of diagnostic evidence promoting the emphasised decision. The question therefore served to instil the goal of getting a greater proportion of the emphasised decisions correct, even if that entailed making fewer endorsements of that decision.

Overall, the present findings allude to an implicit method by which the test environment heightens memory control under variations in goal state – an important yet neglected environmental precipitant of memory control identified in the previous chapter (see Section 1.1. and 1.2.). This provides preliminary evidence for the previously speculated propensity for participants in a memory test to infer higher-order goals from uncontrolled environmental aspects (Johnson et al., 1993; Koriat & Goldsmith, 1996). Further, the results also validate SDT criterion as a behavioural index of memory control processes, albeit whilst also demonstrating the virtue of conducting complementary analyses of other behavioural measures to enable stronger inferences (indeed, this complementary analytic technique will be maintained in subsequent chapters). The real-world implications of these laboratory findings are made apparent by revisiting the previous eyewitness test scenario in which heightened goal emphasis acted to reduce the likelihood of identifying a suspect in a lineup as the perpetrator of a crime. The present findings raise the additional possibility that cross-examiners might be capable of biasing eyewitness reports merely by selectively mentioning either “old” or “new” decisions in their manner of questioning (especially if the strength of the eyewitness’ memory of the crime itself is weak). The real-world significance of these goal emphasis effects inspired further investigation in ensuing experimental chapters.

The most immediate avenue of investigation followed from the literature review of the previous chapter, which highlighted a neglect in determining whether environmentally provoked strategies observed in episodic memory evaluation also manifest in other cognitive domains (see Section 1.4. and Section 1.6.). This represents an important step in elucidating the cognitive underpinnings of behaviourally observable memory strategies, and specifically whether such strategies are rooted in control processes confined to the episodic domain (“domain-specific” control) or those that are capable of modulating behaviour in a number of different domains (“cross-domain” control). Hence, the next experimental chapter investigated the emergence of the presently established question emphasis bias across episodic and non-episodic domains.

<u>Chapter 3:</u> Question format as an implicit source of goal emphasis across cognitive domains

3.1. Introduction to Experiment 4

The previous experimental chapter established question format as an implicit source of goal emphasis in the test environment, which acts to regulate the controlled evaluation of episodic memories. A related question – one that was raised in the literature review of Chapter 1 (see Section 1.6.) and which follows on from the findings of Chapter 2 – is whether question emphasis impacts on controlled processing in other cognitive domains. Hence, the aim of Chapter 3 was to instantiate similar variations of question format across a number of single item discrimination tasks spanning different cognitive domains, as performed by the same sample of subjects. Evidence of comparable behavioural effects across these different discrimination tasks would suggest that the controlled processes engaged by question emphasis manipulations are not specific to the domain of episodic memory. This would also serve to address the longstanding yet rarely empirically tested question as to the potential cross-domain nature of behavioural correlates of memory control (Fleming & Dolan, 2012; Jacoby, Kelley & McElree, 1999; White & Poldrack, 2014; Whittlesea, 2004). Of these correlates, this chapter will focus on both “primary” measures of performance (directly reflecting yes/no decision outcomes) and “secondary” measures (reflecting self-rated confidence in these decisions).

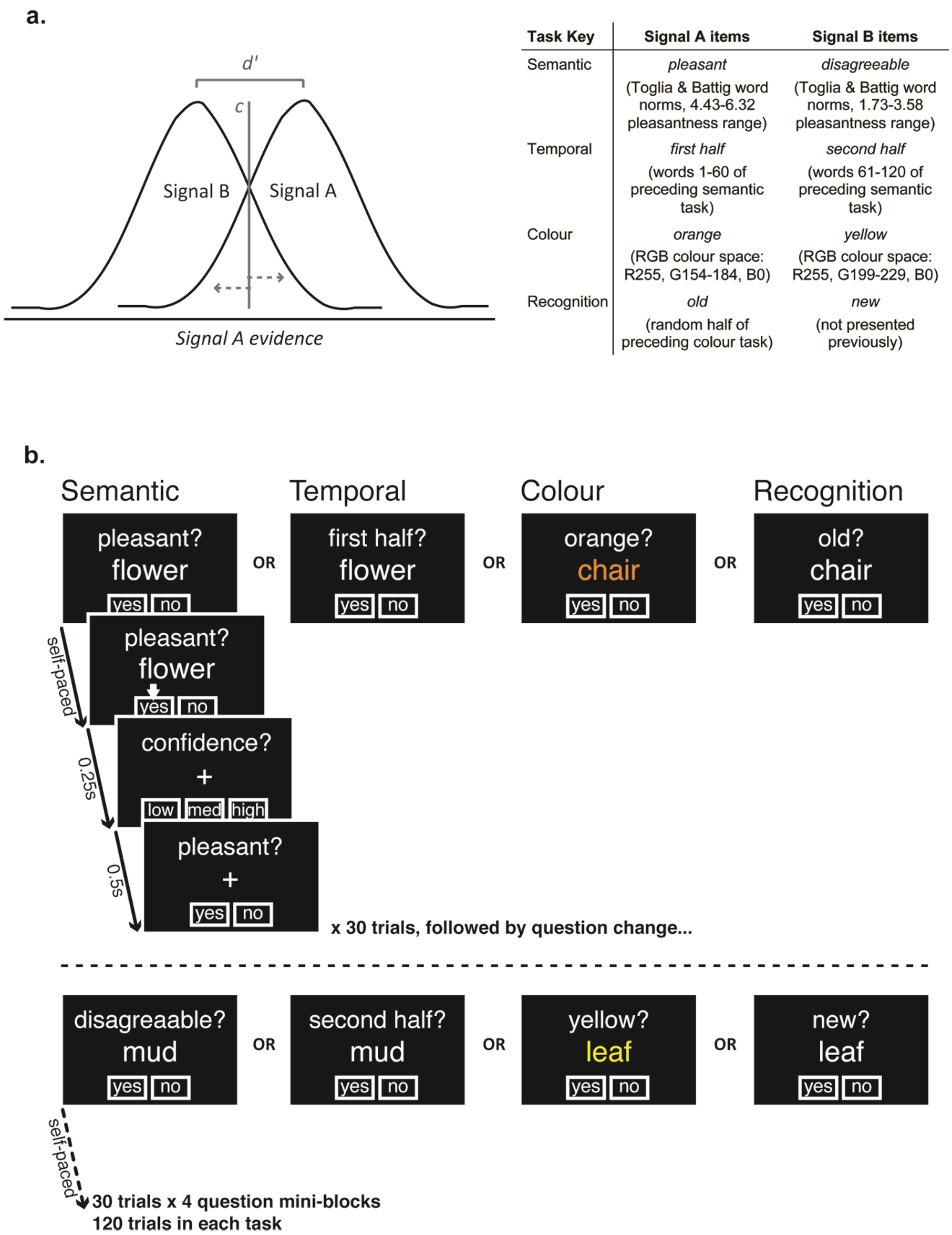
The main focus of the primary response analyses was in examining the effects of question format in instilling decision strategies, as characterised by the criterion bias parameter of the signal detection theory model (SDT; Green & Swets, 1966; Macmillan & Creelman, 2005). This follows from the widespread application of SDT in modelling episodic evaluation in yes/no recognition tasks (as outlined in Chapter 1, Section 1.4.) and more fine-grained aspects of source memory (e.g. the Remember/Know procedure; Donaldson, 1996), as well as in discrimination tasks beyond the episodic domain, such as visual perception (Tanner & Swets, 1954), auditory perception (Paul & Sutton, 1972) and semantic word recognition (Townsend & Ashby, 1982). Analyses of response criterion were of particular interest, given the question emphasis effects on criterion described in Chapter 2 and prior allusion to criterion as an overarching index of strategic biases often left unconstrained in laboratory psychology experiments, such as the perceived sequential dependencies of randomly presented stimuli and the inferred goals of the experimenter (Swets, 2014). However, direct empirical scrutiny of criterion as a cross-domain strategic index has been lacking, both in terms of the active manipulation of environmental sources of strategic bias, as well as in assessing the comparability of strategic influences across different tasks in the same sample of participants.

The present experiment aimed to rectify this empirical dearth, by probing the correspondence of the strategic question effect established in the previous chapter across four tasks encompassing processing in different cognitive domains. These four decision-making tasks shared the same basic decision format of discriminating between two item categories from which individually presented word stimuli were sampled. The tasks differed according to the cognitive domain under evaluation when making discrimination decisions, and spanned two non-episodic domains –

colour perception and semantic judgement – as well as two sub-domains within episodic memory – episodic recognition and episodic temporal judgement. Question emphasis was manipulated in analogous fashion across all tasks, with the wording of the question variably emphasising one of the two available decision categories (as in the selective question emphasis manipulation employed in Chapter 2, see Section 2.2.). Hence, across all tasks, participants' discrimination decisions were assumed to be subject to an assessment of the strength of available discrimination evidence (indexed by d' sensitivity) and the adjustment of a response criterion (c) in accordance with varying strategic biases. Whether criterion was influenced by question emphasis in a similar fashion to that evidenced in the recognition experiment of the previous chapter was of particular interest within this SDT framework (see Figure 3.1a).

As an extension to the analytic approach of the previous chapter, the present experiment also examined effects of question emphasis on behavioural measures other than SDT criterion. This encompassed two further primary response measures – decision accuracy (the accuracy of endorsements of each category of discrimination decision when made) and decision RT (the time taken to endorse each decision category) – as well as a secondary measure of proportion confidence (the proportion of endorsements of each decision category made at each level of a three-point confidence scale). As in the previous chapter, decision accuracy was calculated according to decision status rather than "objective" item status (the latter being used in calculating the SDT measures; see Chapter 2 Section 2.2.1.4 for further details), and was expected to increase under concordant question emphasis. Decision RT complemented the other response-based measures with insight into how the time taken to make discrimination decisions varied according to question emphasis. Prior research involving incentivized manipulations has reported quicker reaction times when making decisions that were congruent with immediately emphasised goals, both in episodic (Han, Huettel, Raposo, Adcock & Dobbins, 2010) and non-episodic decision-making tasks (Small et al., 2005). Hence, question emphasis was expected to serve as an implicit form of these goal-driven effects and induce similar reductions in RT for emphasised decisions.

Proportion confidence served as an important index of evidence strength, which was shown in Experiment 2 of the previous chapter (Section 2.3.) to modulate the strategic influence of question emphasis. Direct manipulation of evidence strength in each of the four presented tasks would have led to a prohibitively long overall run time. For example, manipulating memory strength in the episodic memory task would require the presentation of separate study phases (as in Experiment 2, Section 2.3.), and calibrating evidence strength for the colour discrimination task would have involved time-consuming psychophysical methods applied to each individual participant. Self-rated confidence therefore served as a less precise but more practicable alternative in inferring levels of evidence strength. To clarify, decisions made with high confidence were assumed to entail high evidence strength, whereas those made with low confidence were assumed to reflect low evidence strength, as in previous studies of episodic memory (Henson, Rugg, Shallice & Dolan, 2000; Fleck, Daselaar, Dobbins & Cabeza, 2006) and visual discrimination (Fleming, Huijgen & Dolan, 2012). Following from the findings of Experiment 2, a related assumption was that strategic processes would be heightened under low evidence strength and associated decision uncertainty, and it was therefore expected that question emphasis effects would primarily be driven by a reduction in the proportion of emphasised decisions *specifically* made with “low” confidence (and hence characterised by greater uncertainty). The results from analyses of these behavioural measures provide partial support for question emphasis effects being underpinned by cross-domain control processes, but also serve to highlight the intricacies in elucidating behavioural effects of implicit goal emphasis manipulations.



Task Key	Signal A items	Signal B items
Semantic	*pleasant* (Toglia & Battig word norms, 4.43-6.32 pleasantness range)	*disagreeable* (Toglia & Battig word norms, 1.73-3.58 pleasantness range)
Temporal	*first half* (words 1-60 of preceding semantic task)	*second half* (words 61-120 of preceding semantic task)
Colour	*orange* (RGB colour space: R255, G154-184, B0)	*yellow* (RGB colour space: R255, G199-229, B0)
Recognition	*old* (random half of preceding colour task)	*new* (not presented previously)

Figure 3.1. Model definition and experiment design. **a.** Operationalization of the four discrimination tasks according to the equal variance signal detection model. Across task domains, participants were assumed to be discriminating between Signal A and Signal B, as determined by their sensitivity (*d'*) to the strength of evidence in favour of Signal A and their criterion placement (*c*) along the Signal A evidence axis (as demonstrated by the left panel distributions). Positive criterion shifts therefore reflect a reduced likelihood of endorsing a

“Signal A” decision (i.e. “Decision A”) and relatedly an increased likelihood of endorsing a “signal B” decision (i.e. “Decision B”), whereas negative criterion shifts reflect the inverse bias. The assignment of item categories in each task to Signal A (correct identification of which leads to a ‘Hit’) and Signal B status (correct identification of which leads to a ‘Correct Rejection or ‘CR’) is illustrated in the right panel table. Summaries of the sampling methodologies used to create stimuli in each task are also provided in parentheses (see the companion Section 3.2.2. for further details). **b.** Design schematic illustrating the common yes/no discrimination format maintained across the question emphasis manipulations instantiated in the four task domains. Each task involved 120 trials in which question format was varied between selectively emphasising the two available response options (i.e. 60 trials “Decision A” question emphasis and 60 trials “Decision B” emphasis).

3.2. Method

3.2.1. Participants. Thirty-two participants (21 female, mean age = 21.2, age range = 19-25 years) were recruited for this experiment, all of whom had normal colour vision (confirmed by Ishihara colour test; Ishihara, 1917) and reported English as their first language. Only participants that met the minimum performance threshold of $d' > .1$ (maintained from the previous chapter) were included in the "primary" response analyses specific for each task. This led to the exclusion of four participants from the episodic recognition task analyses (n = 28, 18 female, mean age = 21.1, age range = 19-25) and eleven participants from the episodic temporal judgement task analyses (n = 21, 14 female, mean age = 21.1, age range = 19-25). Additional participants were excluded from the "secondary" proportion confidence analyses in each task if they failed to make a decision at that confidence level across all emphasis conditions. This led to the exclusion of three participants from the low confidence proportion analyses for the colour task (n = 29, 18 female, mean age = 21.3, age range = 19-25), one participant from the medium confidence proportion analyses for the semantic task (n = 31, 20 female, mean age = 21.1, age range = 19-25) and one participant from the low confidence proportion analyses for the semantic task (n = 31, 20 female, mean age = 21.2, age range = 19-25). Informed consent for all participants was obtained in accordance with the University Teaching and Research Ethics Committee at the University of St Andrews. Participants were compensated for their time at a rate of £5/hour.

3.2.2. Stimuli. Stimulus creation methodologies for each task are summarised in Figure 3.1a. For each participant, a different set of words was randomly sampled from a pool of 2584 common nouns taken from the Toglia and Battig semantic word norms (Toglia & Battig, 1978). Although these words served as stimuli for all tasks, the Toglia and Battig database was used primarily for its appropriateness to the *semantic task*, given that it contains words rated for pleasantness on a 7-point scale in a large sample (with low ratings for unpleasant words and high ratings for pleasant ones). These ratings were used to divide words in the Toglia & Battig database into *pleasant* and *disagreeable* semantic item categories, which were separated from each other by 1 standard deviation in their pleasantness ratings to enforce a relatively robust pleasant/disagreeable semantic status divider, and ensure that words of ambiguous pleasantness were excluded from either category. These two semantic item categories formed the "objective" basis for participants' pleasant/unpleasant discrimination decisions in the semantic task, and were hence used when estimating sensitivity and criterion bias parameters in the SDT model (*pleasant* items = semantic Signal A; *disagreeable* items = semantic Signal B; see Figure 3.1a).

Each resultant semantic category spanned a pleasantness rating range encompassing both strongly diagnostic and weakly diagnostic semantic items (*disagreeable* category range = 1.73-3.58 pleasantness rating, *pleasant* category range = 4.43-6.32 pleasantness rating), and was randomly sampled from at regular intervals across this range to create task lists. Words from the low semantic diagnosticity range were sampled in an attempt to equate the overall levels of evidence strength in the semantic task with that in the two episodic memory tasks (i.e. the recognition and temporal tasks). Similar efforts were made for the colour task (outlined below), as both these non-episodic tasks involved processing "in the present" of immediately accessible item evidence, and would hence likely be characterised by greater evidence strength compared to the analysis of the more volatile memory traces in the two episodic tasks. Each semantic task wordlist hence comprised 120 words, of which 60 had *pleasant* meanings and 60 had *disagreeable* meanings, with words from each category assigned equally to each question emphasis condition.

For the *colour task*, participants had to decide if word stimuli were presented in either *orange* or *yellow* coloured font (ascribed *Signal A* and *Signal B* statuses respectively in the adopted SDT framework; see Figure 3.1a). These specific colours were chosen as they span a relatively small portion of the visible colour spectrum – one that does not fall under any perceptually distinct colour category as determined by the sensitivities of retinal cone photoreceptor cells (i.e. short, medium and long wavelength cones, broadly sensitive to blue, green and red colours respectively; Wright & Pitt, 1934). Choosing the perceptually ambiguous orange-yellow spectrum therefore avoided ceiling levels of colour evidence strength and discrimination sensitivity, which would reduce scope for the influence of higher-order strategies (including the question emphasis bias) on decision-making performance (Kahneman, Slovic & Tversky, 1982; see also the findings of Experiment 2, Section 2.3.). All colour hues were sampled from the sRGB colour space, which comprises three adjustable scales for levels of red, green and blue colour intensities (‘R’, ‘G’ and ‘B’ respectively) that all range from 0-255 in value. Three orange and three yellow hues were created from the available RGB orange-yellow spectrum by varying the G intensity value whilst keeping R and B values constant (R for all colours = 255, B for all colours = 0, G sampling range = 128-255). This broad range enabled creation of ‘strong’, ‘medium’ and ‘weak’ evidence diagnosticity sub-types for each colour category. An orange/yellow category divider was instantiated by spacing the “weak orange” and “weak yellow” hues equally on opposing sides of the G value that was “objectively” of equal orange/yellow intensity (i.e. the median of the G sampling range 191.5). The inclusion of weakly diagnostic colour hues situated close to the orange-yellow category divider was undertaken to equate levels of evidence uncertainty in the colour decision-making task with that in the two episodic tasks.

The six resultant hues had the following sRGB values, with associated International Commission on Illumination (CIE) Yxy values provided in brackets: strong orange = 255, 154, 0 (44.03, 0.51, 0.43); medium orange = 255, 169, 0 (49.60, 0.49, 0.45); weak orange = 255, 184, 0 (56.17, 0.47, 0.46); weak yellow = 255, 199, 0 (63.33, 0.46, 0.47); medium yellow = 255, 214, 0 (71.34, 0.44, 0.49); strong yellow = 255, 229, 0 (80.33, 0.43, 0.50). These six coloured fonts were randomly assigned to words sampled from the Toglia and Battig word norms (1978), leading to the presentation of 60 *orange* and 60 *yellow* words in each participants’ 120 task wordlist, with each category comprising 20 of each colour diagnosticity sub-type. To minimise confounding influences on colour perception, stimuli were presented against an achromatic background on a monitor positioned in a dark room without ambient sources of illumination. All participants viewed stimuli in the colour task (and all other task phases) on CRT monitors that were calibrated for colour display using the ColorCAL MKII Colorimeter (Cambridge Research Systems).

The semantic task served as the study phase for the ensuing episodic temporal judgement task, and the colour task served as study for the ensuing episodic recognition task. All 120 words appearing in the semantic task were presented in a randomised order for the *temporal task*, with 60 words having been previously presented in the *first half* of the semantic task (assuming temporal task *Signal A* status in the SDT framework, see Figure 3.1a) and 60 in the *second half* (*Signal B*). Each of these two temporal item categories was further divided into three subcategories based on the precise order of presentation in the semantic task and hence the assumed evidence strength (“strong”, “medium” or “weak”) when discriminating the temporal status of these items. To clarify, the “strong” subcategory of *first half* items comprised the first 20 words presented in the semantic task, the “medium” *first half* subcategory comprised the moderately diagnostic 21-40 presented words, and the “weak” subcategory comprised the final 41-60 *first half* words that were presented close to the onset of the *second half* items (and hence more likely to be confused with them). Equivalent stratification was employed for *second half* subcategories e.g. the “strong” category of *second half* items comprised words ordered from 101-120 in the semantic task wordlist. As with the stratified sampling strategies employed in the colour and semantic tasks, these subcategories were allocated equally to task lists in the temporal question conditions.

The 120 words used in the *recognition task* comprised a random selection of 60 *old* words (recognition *Signal A;* see Figure 3.1a) that were presented in the preceding colour task and 60 *new* words (recognition *Signal B*) not previously seen that were sampled afresh from the Toglia and Battig word database. The experiment was presented and responses recorded using PCs running MATLAB (The MathWorks Inc., Natick, MA, 2000) and Psychophysics Toolbox (Brainard, 1997).

3.2.3. Procedure. After the presentation of onscreen instructions, participants worked through a practice phase composed of brief versions of all four tasks. Each task in the main phase involved the presentation of 120 words broken into 4 mini-blocks of 30 words each, within which the format of the test question was varied to emphasise one of the two decision categories available for that task (i.e. semantic task questions: “pleasant?”, “disagreeable?”; colour task: “orange?”, “yellow?”; recognition task: “old?”, “new?”; temporal task: “first half?”, “second half?”; see

Figure 3.1b). Participants were alerted to the change in question via an instruction screen presented for 5s prior to the appearance of the first word in each question mini-block (e.g. “Next question: pleasant?”). Across all tasks, the appearance of to-be-judged words on each trial was preceded by a 0.5s fixation cross. Participants submitted their self-paced task decisions using the number keys, where 1 denoted a “yes” response and 2 a “no” response. After a 0.25s delay, participants supplemented their yes/no decision with a self-paced rating of decision confidence using a 3-point scale (1 = “low”, 2 = “medium” and 3 = “high” confidence). The semantic task was always succeeded by the temporal task, and the colour task was always succeeded by the recognition task. The ordering of these task pairings was counterbalanced so that half the participants performed the semantic-temporal block first, and the other half performed the colour-recognition block first. Question emphasis order was partially counterbalanced within these task orders, with half the participants in each task order exposed to the following first questions in each respective task phase (with the questions alternating thereafter for the remaining 3 mini-blocks): “orange?-old?-pleasant?-first half?” questions first; while the other half were exposed to the inverse sequence first: “yellow?-new?-disagreeable?-second half?”

3.2.4. Calculation. Analyses were conducted on sensitivity (*d′*) and criterion bias (*c*) parameters estimated from the equal variance signal detection model (Green & Swets, 1966; Macmillan & Creelman, 2005), following a correction for errorless responding (Snodgrass & Corwin, 1988; see Section 2.2.1.4 for further details). Across the four discrimination tasks, item categories were assigned *Signal A”*(the correct identification of which led to a “hit”) and *Signal B* status (correct identification of which led to a “correct rejection” or CR) in accordance with the “discrimination” rather than "detection" operationalization of the SDT model (Benjamin, Diaz & Wee, 2009; Tanner, 1956; see Figure 3.1a). Participants were hence assumed to be discriminating between *Signal A* and *Signal B* item categories cross the four tasks, as informed by their estimated *d′* sensitivity (the distance between the two item distributions) and criterion *c* parameters (the threshold amount of evidence required to decide that *Signal A* was present). This common formalisation of *Signal A* and *Signal B* item categories in each task is extended to describe cross-task analyses involving decision categories (i.e. endorsements of *Signal A* are termed "Decision A" and endorsements of *Signal B* are termed "Decision B") and the cross-task question emphasis conditions (i.e. questions that impart emphasis to "Decision A" and "Decision B" in each task).

Cross-domain question emphasis effects were also interrogated by analyses of other behavioural measures. The analyses of old and new decision accuracy conducted for the recognition task in Chapter 2 (old_{corr} and new_{corr} respectively; see Section 2.2.1.4.) were extended to encompass decision categories across all four tasks, in the form of the $DecisionA_{corr}$ and $DecisionB_{corr}$ parameters. To clarify with reference to the semantic task, the proportion of correct "pleasant" decisions out of all "pleasant" decisions made by each individual subject was indexed by semantic $DecisionA_{corr}$ (given that *pleasant* is assigned *Signal A* item status in that task; see Figure 3.1a), whereas the proportion of correct "disagreeable" decisions out of all "disagreeable" decisions made was indexed by $DecisionB_{corr}$ (as *disagreeable* is assigned *Signal B* status). Decision RT was calculated as the time taken to render each category of decision available in each task (irrespective of the correctness), in the form of the $DecisionA_{RT}$ and $DecisionB_{RT}$ parameters. The proportion confidence analyses were conducted on calculations of the proportion of high, medium and low confidence decisions made for each category of decision in each task (Decision A confidence measures: $HighA_{prop}$, $MedA_{prop}$, $LowA_{prop}$; Decision B confidence measures: $HighB_{prop}$, $MedB_{prop}$, $LowB_{prop}$). For example, the number of "pleasant" decisions rendered with high confidence divided by the *total* number of "pleasant" decisions made was indexed by $HighA_{prop}$. Comparisons of the above behavioural measures across the different question emphasis conditions enabled a thorough examination of the cross-domain properties of any observed question emphasis effects.

3.3. Results

3.3.1. Overall task performance. Differences in overall performance across the four tasks were first quantified by analyses of *d′* sensitivity in the complete sample[4] (i.e. without performance-based exclusions), which collapsed across question emphasis conditions in each individual task. A one-way repeated measures ANOVA of task format (semantic, colour, temporal and recognition) on overall *d′* yielded a significant main effect, $F(3,93) = 263.11$, $p < .001$, $\eta_p^2 = .90$. Pairwise t-tests confirmed that sensitivity in the colour task ($M = 2.92$, $SD = 0.46$) was significantly greater than sensitivity in the semantic task ($M = 2.11$, $SD = 0.38$), $t(31) = 8.65$, $p < .001$, $d = 1.53$, which was in turn greater than sensitivity in the recognition task ($M = 0.82$, $SD = 0.37$), $t(31) = 12.95$, $p < .001$, $d = 2.29$, which was in turn greater than sensitivity in the temporal task ($M = 0.57$, $SD = 0.41$), $t(31) = 3.57$, $p = .001$, $d = 0.63$.

[4] A similar pattern of differences in overall task performance was observed for *d′* sensitivity analyses that excluded sub-threshold performers in each task.

The differences in overall performance across the four tasks imply a considerable degree of variation in the levels of underlying evidence strength, which has been previously highlighted as a salient impediment to recovering cross-domain control effects in the field of metacognition (Fleming & Dolan, 2012). This variability in evidence strength might have had the collateral effect of instilling strategies other than the question emphasis strategy of primary interest (see the Discussion Section 3.4. for further details). The overall sensitivity differences are likely to account for the lack of consistent cross-domain question emphasis effects on “primary” behavioural measures centring on the basic yes/no response selection (detailed in ensuing sections).

3.3.2. Signal detection sensitivity and bias. Means and standard deviations for analyses of SDT measures across the question emphasis conditions are presented in Table 3.1. The effects of question emphasis on discrimination sensitivity were first analysed via paired-samples *t*-test conducted for each task separately. For the semantic task, d' was significantly lower for “disagreeable?” compared to “pleasant?” questions, $t(31) = 3.68$, $p = .001$, $d = 0.65$. There was also a significant effect of question emphasis on sensitivity in the temporal judgement task, such that d' was lower for “second half?” than “first half?” questions, $t(20) = 2.21$, $p = .039$, $d = 0.49$. Question emphasis did not significantly influence d' in the colour and recognition tasks, $t(31) = 1.36$, $p = .183$, $d = 0.25$, and $t(27) = 1.06$, $p = .301$, $d = 0.20$ respectively. The reduced sensitivity for the “disagreeable?” and “second half?” questions is potentially due to the greater difficulty entailed in fluently integrating these questions into ongoing decision-making. For “disagreeable?” questions in the semantic task, this processing difficulty might result from its lower frequency of usage in the English language and associated greater lexical difficulty compared to the alternate “pleasant?” question. For “second half?” questions in the temporal task, the processing difficulty might arise from the dissonance between its featuring the term “second” whilst requiring participants to press the “1” number key to endorse the question (unlike the alternate “first half?” question, for which the numerical descriptor worded in the question is congruent with its method of endorsement via the “1” key). These suggested difficulties in processing the questions themselves would have served to reduce overall sensitivity both by using up pre-decision cognitive resources otherwise directed towards the assessment of evidence strength[5], as well as by increasing the post-decision likelihood of pressing the wrong response buttons.

Admittedly, this interpretation of the question effects on semantic and temporal discrimination sensitivity is speculative, however the discussion of these findings is limited as such given that the primary aim was to interrogate question effects on criterion bias. Nevertheless, the ensuing criterion analyses are complemented by analyses of a measure that takes into account the above variation in sensitivity – c' (Macmillan & Creelman, 2005). This is simply criterion divided by d', which serves as a scaling operation that reduces the potential for differences in sensitivity across question emphasis conditions to exert confounding influences on bias.

The analyses of unscaled criterion bias (c) were also conducted via separate paired samples t-tests for each individual task (see Table 3.1. for means; see Figure 3.2.). For the semantic task, c was significantly higher for “pleasant?” compared to “disagreeable?” questions, $t(31) = 6.08$, $p < .001$, $d = 1.09$. The disconfirmatory or

[5] The proposed disruption of evidence-strength processing is comparable to the interference of source monitoring processes following divided attention and response speeding manipulations enacted at recognition test phases, as described in Chapter 1 section 1.3. (Johnson, Kounios & Reeder, 1994; Jacoby, Toth & Yonelinas, 1993).

"counter-emphasis" direction of this criterion shift parallels that observed in the previous chapter, in reflecting the reduced likelihood of endorsing the semantic decision emphasised by the question (i.e. "pleasant?" reduces the likelihood of responding "pleasant"). However, the difference in *c* across question emphasis conditions in the recognition task was non-significant, $t(27) = 0.39$, $p = .697$, $d = 0.07$. Despite the counter-emphasis direction of this shift in recognition criterion, the failure for the shift to reach significance was contrary to the findings of the previous chapter (in which significant question emphasis effects on recognition criterion were recovered in two separate samples) and an independent study involving a large online sample (Mill & O'Connor, 2014). Further, *c* shifted in the opposing "pro-emphasis" direction in the colour and temporal tasks (i.e. reflecting an increased likelihood of endorsing emphasised decisions) albeit to non-significant degrees, $t(31) = 1.29$, $p = .206$, $d = 0.23$, and $t(20) = 1.54$, $p = .138$, $d = 0.10$ respectively (see Figure 3.2.).

Given the previously described effects of question emphasis on d' sensitivity in the semantic and temporal tasks, the *c* analyses were complemented by analysis of the sensitivity-scaled bias parameter c'. The pattern of bias results involving c' was the same as that for unscaled *c*, with a significant counter-emphasis bias shift observed between question conditions in the semantic task and a non-significant shift in the same direction observed for the recognition task, $t(31) = 6.42$, $p < .001$, $d = 1.16$, and $t(27) = 0.70$, $p = .489$, $d = 0.14$ respectively (see Table 1 for means). Sensitivity-scaled bias did not differ between question conditions in the colour and temporal tasks, $t(31) = 0.77$, $p = .446$, $d = 0.13$, and $t(20) = 0.458$, $p = .652$, $d = 0.10$ respectively. Collectively, analyses involving both *c* and sensitivity-scaled c' reveal that the only reliable effect of question emphasis in shifting response criterion was evidenced in the semantic task. Hence, the overall pattern across the four tasks does not provide compelling evidence in support of cross-domain effects of question emphasis being registered in SDT estimates of bias.

Table 3.1. Design key and descriptive statistics for cross-task analyses of equal variance signal detection theory (SDT) parameters, including sensitivity (SDT *d′*), unscaled bias (SDT *c*) and sensitivity-scaled bias (SDT *c′*).

Task		**Semantic**		**Colour**		**Temporal**		**Recognition**	
Question		pleasant?	disagreeable?	orange?	yellow?	first half?	second half?	old?	new?
Emphasis		Decision A	Decision B	Decision A	Decision B	Decision A	Decision B	Decision A	Decision B
SDT *d′*	*M*	2.29	1.92	2.85	2.99	0.84	0.68	0.92	0.82
	SD	0.47	0.49	0.49	0.59	0.38	0.33	0.49	0.38
SDT *c*	*M*	-0.05	-0.41	-0.30	-0.22	0.01	0.11	0.12	0.09
	SD	0.35	0.29	0.41	0.40	0.30	0.21	0.38	0.32
SDT *c′*	*M*	-0.02	-0.22	-0.11	-0.09	0.12	0.19	0.19	0.08
	SD	0.16	0.20	0.15	0.16	0.66	0.66	0.82	0.55

Note: *M*, mean; *SD*, standard deviation.

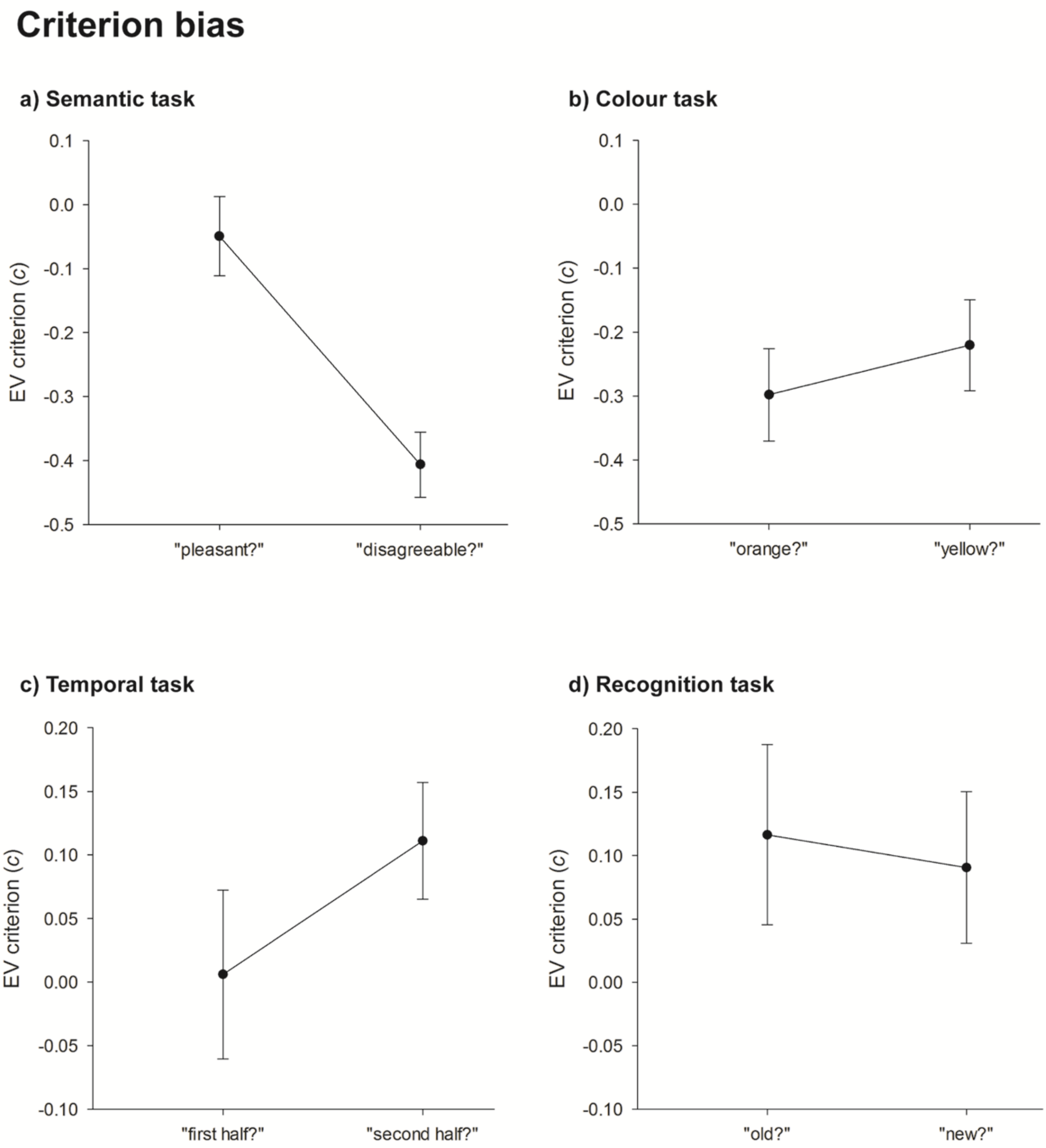


Figure 3.2. Question emphasis effects on equal variance (EV) estimates of criterion bias (*c*) across the four presented tasks: **a.** semantic task, **b.** colour task, **c.** temporal task and **d.** recognition task. In each panel, graphs depict shifts in criterion bias according to the decision emphasised by the question, such that increasing *c* values reflect a *reduced* likelihood of endorsing the “Signal A” decision (i.e. “pleasant”, “orange”, “first half” and “old” decisions respectively; see Figure 3.1. for further detail on this formalization). Error bars represent standard error of the mean.

3.3.3. Decision accuracy. Question emphasis effects were also examined on the accuracy of endorsements of the two decision categories available in each task (i.e. $DecisionA_{corr}$ and $DecisionB_{corr}$, see Calculation Section 3.2.4). Means and standard deviations for decision accuracy analyses are provided in Table 3.2. A 2(decision type: Decision A and Decision B) x 2(question emphasis: “Decision A?” or “Decision B?”) repeated measures ANOVA was conducted on decision accuracy for each separate discrimination task (see Figure 3.3.). For the semantic task, there was a main effect of decision type, such that “disagreeable” decisions (M = .89, SD = .06) were more accurately made overall than “pleasant” decisions (M = .82, SD = .06), $F(1,31) = 23.39$, $p < .001$, $\eta_p^2 = .43$. A main effect of decision type was also observed in the colour task, such that “yellow” decisions (M = .96, SD = .05) were more accurately made than “orange” decisions (M = .89, SD = .06), $F(1,31) = 16.13$, $p < .001$, $\eta_p^2 = .34$. Considered with the overall biases revealed by the SDT analyses in the colour and semantic tasks, reflective of a reduced tendency to endorse “disagreeable” and “yellow” decisions (i.e. absolute criterion in relation to statistically optimal placement at 0 was negative across question emphasis conditions in both tasks; see Table 3.1.), these main effects of decision type suggest that participants were more strategically cautious in endorsing “disagreeable” and “yellow” decisions than their respective decision alternatives (regardless of question emphasis). There were no main effects of decision type in the temporal task and the recognition task, $F(1,20) = 1.38$, $p = .254$, $\eta_p^2 = .06$, and $F(1,27) = 3.92$, $p = .058$, $\eta_p^2 = .13$ respectively. The variable effects of decision type on decision-making across the four tasks once again highlight the performance differences arising from the variable evidence domains under evaluation in each task.

The decision accuracy ANOVAs also revealed main effects of question emphasis in the semantic and temporal tasks, $F(1,31) = 17.82$, $p < .001$, $\eta_p^2 = .37$, and $F(1,20) = 5.09$, $p = .036$, $\eta_p^2 = .20$ respectively. Overall accuracy was worse following "disagreeable?" ($M = .83$, $SD = .06$) than "pleasant?" questions ($M = .88$, $SD = .05$), and for "second half?" ($M = .64$, $SD = .06$) than "first half?" questions ($M = .67$, $SD = .07$). There were no main effects of question emphasis on decision accuracy in the colour and recognition tasks, $F(1,31) = 1.75$, $p = .196$, $\eta_p^2 = .05$, and $F(1,27) = 1.00$, $p = .325$, $\eta_p^2 = .04$ respectively. The question effects in the semantic and temporal tasks parallel the earlier SDT sensitivity effects in suggesting that "disagreeable?" and "second half?" questions led to worse overall discrimination performance. As before, these deleterious effects on performance can be speculated to arise from the greater difficulty in integrating these question types into ongoing decision-making.

The interaction of decision type and question emphasis was of primary interest in determining whether the hypothesised increase in decision accuracy under concordant question emphasis was borne out by the data. The 2x2 ANOVAs revealed a significant decision type x question emphasis interaction for the semantic task, $F(1,31) = 40.36$, $p < .001$, $\eta_p^2 = .57$. In support of the observed counter-emphasis bias on semantic criterion, planned pairwise comparisons revealed that the accuracy of "pleasant" decisions was significantly improved for "pleasant?" compared to "disagreeable?" questions, with a non-significant shift in a counter-emphasis direction observed for "disagreeable" decisions, $t(31) = 8.08$, $p < .001$, $d = 1.43$, and $t(31) = 0.60$, $p = .554$, $d = 0.10$ respectively (see Table 3.2.; see Figure 3.3The decision type x question emphasis interactions were non-significant in the colour, recognition tasks and temporal tasks, $F(1,31) < 1$, $F(1,27) < 1$ and $F(1,20) = 3.34$, $p = .083$, $\eta_p^2 = .14$ respectively.

Hence, the sole goal emphasis-consistent effect of question format on decision accuracy was evidenced in the semantic task, with question emphasis increasing the accuracy of emphasised semantic decisions. Considered with the earlier SDT analyses, which demonstrated a counter-emphasis criterion bias exclusively in this task, the pattern of findings suggests that the influence of question format in instilling strategic caution towards particular decision categories was successfully isolated in the semantic task by the same measures used in Chapter 2. However, the fact that the remaining tasks failed to recover consistent strategic question effects on SDT bias and decision accuracy highlights the potential contamination of these measures by other factors, such as the observed difficulty in integrating particular questions and the adoption of alternative evidence-based strategies in different tasks. The subsequent analyses of decision RT and proportion confidence were therefore employed to attempt to further interrogate the question emphasis effects of primary interest.

Table 3.2. Design key and descriptive statistics for cross-task analyses of decision status measures, including the accuracy of endorsing available decision categories (DecisionA$_{corr}$ and DecisionB$_{corr}$) and the reaction time for endorsing these categories (DecisionA$_{RT}$ and DecisionB$_{RT}$).

Task		Semantic		Colour		Temporal		Recognition	
Question		pleasant?	disagreeable?	orange?	yellow?	first half?	second half?	old?	new?
Emphasis		Decision A	Decision B	Decision A	Decision B	Decision A	Decision B	Decision A	Decision B
DecisionA$_{corr}$	*M*	.87	.76	.88	.90	.67	.65	.69	.67
	SD	.07	.06	.08	.09	.08	.08	.11	.08
DecisionB$_{corr}$	*M*	.88	.89	.96	.96	.67	.62	.66	.65
	SD	.07	.08	.06	.05	.08	.05	.08	.08
DecisionA$_{RT}$	*M*	1.70	2.07	1.37	1.43	1.85	2.18	1.87	2.07
	SD	0.54	0.69	0.52	0.48	0.44	0.58	0.70	0.85
DecisionB$_{RT}$	*M*	1.90	1.95	1.57	1.29	1.79	1.94	1.88	1.91
	SD	0.61	0.72	0.65	0.44	0.49	0.55	0.64	0.58

Note: *M*, mean; *SD*, standard deviation; RT measures are in seconds.

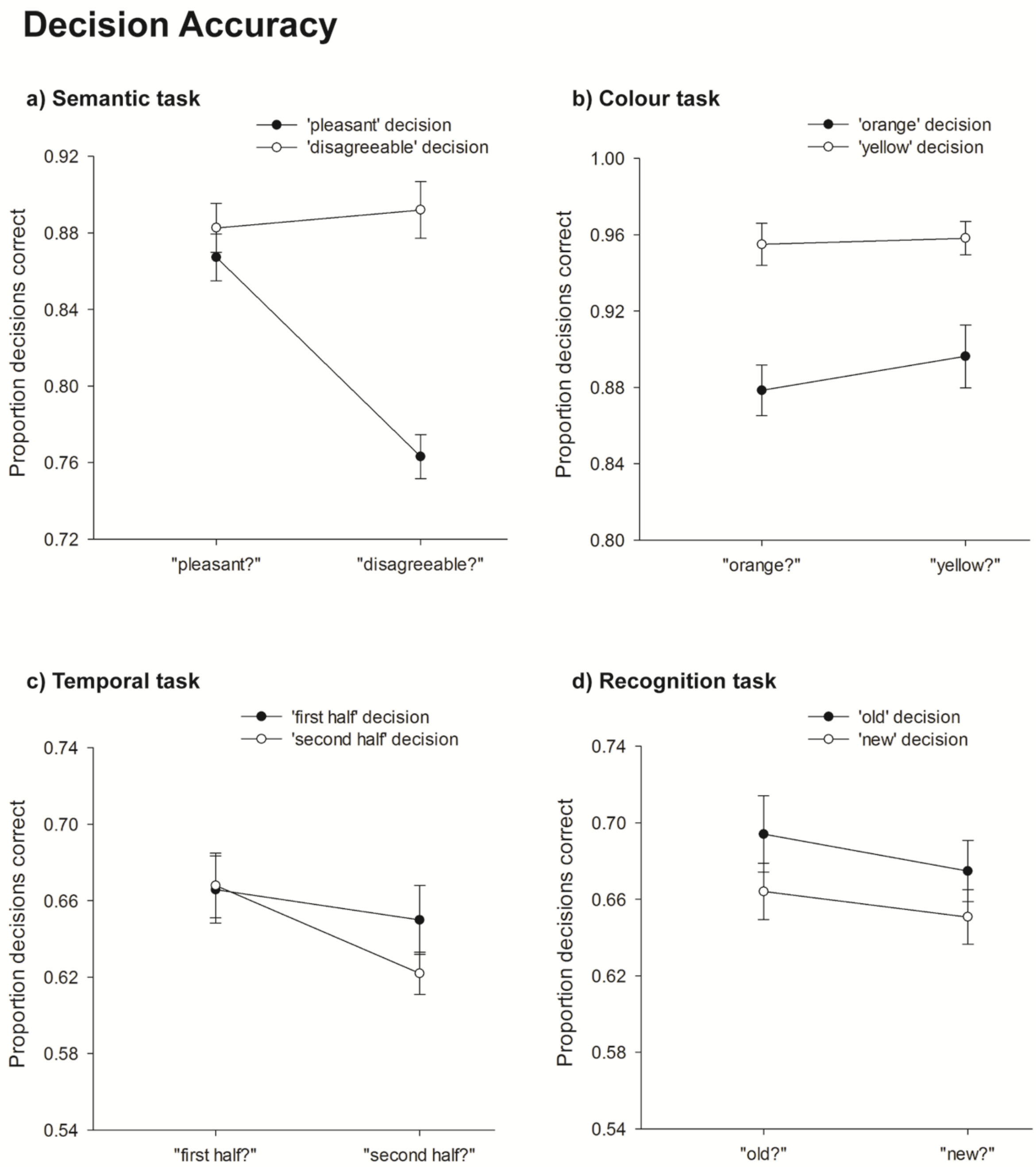


Figure 3.3. Question emphasis effects on the accuracy of endorsements of decision categories available across tasks: **a.** semantic task, **b.** colour task, **c.** temporal task and **d.** recognition task. Error bars represent standard error of the mean.

3.3.4. Decision RT. Effects of question emphasis were examined on the time taken to endorse each of the two decision categories available in the four tasks (i.e. $DecisionA_{RT}$ and $DecisionB_{RT}$, see Calculation Section 2.4 above), via 2(decision type: Decision A and Decision B) x 2(question emphasis: "Decision A?" or "Decision B?") repeated measures ANOVAs conducted for each task separately. Means and standard deviations for the decision RT analyses are provided in Table 3.2. The main effect of decision type was significant in the temporal task, such that "first half" decisions (M = 2.01s, SD = 0.48) were made slower overall than "second half" decisions (M = 1.87s, SD = 0.48), $F(1,20) = 8.72$, $p = .008$, $\eta_p^2 = .30$. Inspection of the decision RT means in the temporal task (see Table 3.2.; see Figure 3.4.) suggests that the decision type effect is driven by the slower reaction time for "first half" decisions made under "second half?" questions, compared to other condition cells. This decision type main effect was specific to the temporal task, with non-significant effects observed in the semantic, colour and recognition tasks, $F(1,31) = 1.08$, $p = .307$, $\eta_p^2 = .03$, $F(1,20) < 1$ and $F(1,27) = 1.28$, $p = .267$, $\eta_p^2 = .05$ respectively. As before, this finding can be linked with the dissonance between decision categories and available response options (i.e. the need to press the "2" key to submit a "first half" decision).

Main effects of question emphasis on decision RT were observed in the semantic, temporal and colour tasks, $F(1,31) = 18.29$, $p < .001$, $\eta_p^2 = .37$, $F(1,20) = 12.92$, $p = .002$, $\eta_p^2 = .39$, and $F(1,31) = 6.30$, $p = .018$, $\eta_p^2 = .17$ respectively. The main effect of question emphasis was non-significant in the recognition task, $F(1,27) = 3.38$, $p = .077$, $\eta_p^2 = .11$. The direction of the main effects in the semantic and temporal tasks complements the earlier decision accuracy analyses in highlighting the greater processing difficulties associated with "disagreeable?" and "second half?" questions than their task counterpart (linked with lexical difficulty and dissonant response options respectively). Hence, “disagreeable?” questions (M = 2.01, SD = 0.69) were responded to slower than “pleasant?” questions (M = 1.80, SD = 0.56), and “second half?” questions (M = 2.06, SD = 0.53) were responded to slower than “first half?” questions (M = 1.82, SD = 0.44). The question main effect in the colour task was driven by the slower reaction time for “orange?” questions (M = 1.47, SD = 0.55) compared to “yellow?” ones (M = 1.36, SD = 0.43). Given the absence of complementary effects of question emphasis on decision accuracy in the colour and recognition tasks, interpretation of this latter main effect remains speculative.

As with the earlier decision accuracy analyses, the interaction of decision type and question emphasis was the primary index of the effects on decision RT expected if the test question served to impart goal emphasis to specific decision categories (i.e. a reduction in the time take to endorse the decision category emphasised by the question). The decision type x question emphasis interaction was significant in the semantic and colour tasks, $F(1,31) = 23.28$, $p < .001$, $\eta_p^2 = .43$, and $F(1,31) = 15.85$, $p < .001$, $\eta_p^2 = .34$ respectively. The colour task interaction was in the anticipated goal emphasis-driven direction, with pairwise t-tests revealing a significant reduction in the time taken to make “yellow” decisions for “yellow?” compared to “orange?” questions, and a non-significant reduction in the time taken to make “orange” decisions for “orange?” compared to “yellow?” questions, $t(31) = 4.55$, $p < .001$, d = 0.99, and $t(31) = 1.02$, $p = .314$, d = 0.18 respectively (see Table 3.2.; see Figure 3.4.). However, the direction of the semantic task interaction was not consistent with the goal emphasis prediction, and merely extended the prior main effect of question emphasis in increasing RT for “disagreeable?” compared to “pleasant?” questions. Pairwise t-tests demonstrated that the greater RT under “disagreeable?” questions was significant for “pleasant” but not “disagreeable” decision types, $t(31) = 6.25$, $p < .001$, $d = 1.22$, and $t(31) = 0.92$, $p = .367$, $d = 0.17$ respectively. The decision type x question emphasis interactions were non-significant and non-supportive of emphasis predictions in the temporal and recognition tasks, $F(1,20) = 2.51$, $p = .129$, $\eta_p^2 = .11$, and $F(1,27) = 2.46$, $p = .129$, $\eta_p^2 = .08$ respectively.

Hence, the only goal emphasis-consistent effect recovered in the decision RT analyses was the significant decision type x question emphasis crossover interaction in the colour task, reflecting the reduced time taken to make emphasised colour decisions. Those remaining decision RT effects that reached significance bore the influence of extraneous behavioural factors, such as the previously described difficulty in integrating particular question types in the semantic and temporal tasks. Overall, the decision RT analyses parallel the earlier SDT and decision accuracy analyses in revealing the general absence of cross-domain effects of question emphasis. A likely cause of these modest effects is the fact that all behavioural measures presented thus far collapse across the variable evidence strength signals in each task, which were shown to be present to a significant degree by the overall differences in task sensitivities reported in Section 3.3.1. The ensuing proportion confidence analyses therefore aimed to counteract these variations in task evidence signals by segregating decisions in each task according to self-rated confidence and inferred levels of evidence uncertainty.

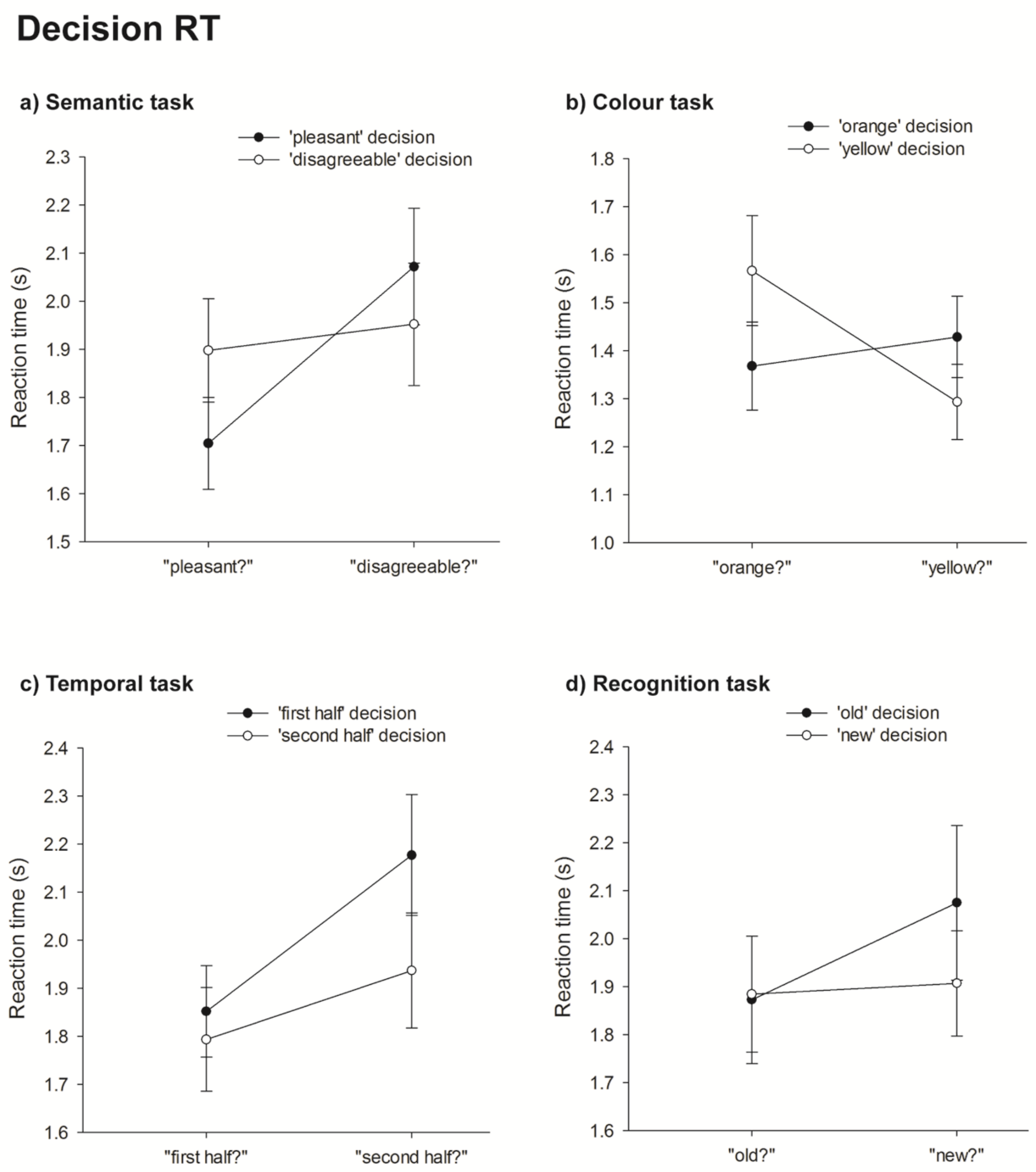


Figure 3.4. Question emphasis effects on the reaction time (RT) taken to endorse decision categories across tasks: **a.** semantic task, **b.** colour task, **c.** temporal task and **d.** recognition task. RT is provided in seconds (s). Error bars represent standard error of the mean.

3.3.5. Proportion confidence. Across the four tasks, the proportion of high, medium and low confidence decisions made for each decision category was calculated (i.e. Decision A confidence measures: $HighA_{prop}$, $MedA_{prop}$, $LowA_{prop}$; Decision B confidence measures: $HighB_{prop}$, $MedB_{prop}$, $LowB_{prop}$). To reiterate, confidence was assumed to serve as an index of evidence strength, such that decisions rendered with low confidence were assumed to reflect heightened uncertainty. Following on from the counter-emphasis bias described in the previous chapter (i.e. the reduced likelihood of endorsing decisions emphasised by the question; see Section 2.3.), it was expected that question emphasis would serve to reduce the proportion of those emphasised decisions specifically given low confidence ratings (and associated with greater uncertainty). Hence, a series of 2(decision type: Decision A and Decision B) x 2(question emphasis: “Decision A?” or “Decision B?”) repeated measures ANOVAs were conducted on the proportion of high, medium and low confidence decisions, separately for each of the four tasks. Means and standard deviations for these proportion confidence analyses are provided in Table 3.3.

For the analyses of the proportion of high confidence decisions ($HighA_{prop}$ and $HighB_{prop}$), the main effect of decision type was significant in the temporal and recognition tasks, $F(1,20) = 4.88$, $p = .039$, $\eta_p^2 = .20$, and $F(1,27) = 5.22$, $p = .030$, $\eta_p^2 = .16$ respectively. The proportion of high confidence ratings was higher for “second half” decisions ($M = .46$, $SD = .22$) compared to “first half” decisions ($M = .40$, $SD = .22$), and for “old” decisions ($M = .38$, $SD = .18$) compared to “new” decisions ($M = .31$, $SD = .18$). The main effects of decision type did not generalize across the two tasks, with non-significant effects observed in the semantic and colour tasks, both $Fs(1,31) < 1$. The main effects of question emphasis were non-significant in the semantic, colour, temporal and recognition tasks, $F(1,31) < 1$,

F(1,31) < 1, F(1,20) < 1, and F(1,27) = 1.21, p = .281, η_p^2 = .04 respectively. Crucially, the decision type x question emphasis interactions were also non-significant across the four tasks, F(1,31) = 2.06, p = .162, η_p^2 = .06, F(1,31) < 1, F(1,20) < 1, and F(1,27) = 2.91, p = .100, η_p^2 = .10 respectively. As before, interactions of decision type and question emphasis served as the primary index of counter-emphasis effects in the proportion confidence analyses. Hence, the complete absence of such interactions for the analyses of the high confidence decision proportions across tasks is consistent with the inferred high evidence strength associated with these decisions. High evidence strength ensures sufficiently accurate responding that precludes the need to engage controlled strategies, such as heightening strategic caution towards the rendering of emphasised decisions.

A similar pattern of results was observed for the analyses of the proportion of medium confidence decisions ($MedA_{prop}$ and $MedB_{prop}$). The 2(decision type) x 2(question emphasis) ANOVAs revealed non-significant main effects of decision type across the semantic, colour, temporal and recognition tasks, F(1,30) = 2.74, p = .108, η_p^2 = .08, F(1,31) < 1, F(1,20) < 1, and F(1,27) < 1 respectively. The main effects of question emphasis across these four tasks were also non-significant, F(1,30) < 1, F(1,31) < 1, F(1,20) < 1, and F(1,27) = 3.53, p = .071, η_p^2 = .12. Finally, the interaction of decision type and question emphasis was non-significant in the semantic, colour and recognition tasks, F(1,30) < 1, F(1,31) < 1 and F(1,27) < 1 respectively. A significant interaction was observed in the temporal task, although in a converse direction to that expected of the counter-emphasis bias i.e. question emphasis increased the proportion of emphasised decision endorsements at medium confidence (see Table 3.3. for means), F(1,20) = 23.20, p < .001, η_p^2 = .54. Pairwise t-tests revealed that this accuracy improvement under emphasis was significant for both “first half” and “second half” decisions, $t(20) = 4.43$, $p < .001$, $d = 0.98$, and $t(20) = 4.46$, $p < .001$, $d = 1.01$ respectively. Overall, the proportion of medium confidence decisions did not display consistent counter-emphasis shifts across tasks. Rather, the converse “pro-emphasis” shift evidenced in the temporal task highlights the virtue of segregating performance by confidence level and inferred evidence strength, as converse behavioural trends might be present at different confidence levels.

Unlike the analyses of high and medium proportion confidence measures (and those involving the earlier “primary” response measures), the analyses of the proportion of low confidence decisions yielded a reliable cross-domain effect in the anticipated goal emphasis direction. The 2(decision type) x 2(question emphasis) ANOVAs revealed main effects of decision type in the semantic, temporal and recognition tasks, $F(1,30) = 9.42$, $p = .005$, $\eta_p^2 = .24$, $F(1,20) = 5.34$, $p = .032$, $\eta_p^2 = .21$, and $F(1,27) = 6.98$, $p = .014$, $\eta_p^2 = .21$ respectively. Across these tasks, a greater proportion of low confidence ratings were given for “disagreeable” decisions ($M = .13$, $SD = .11$) compared to “pleasant” decisions ($M = .09$, $SD = .09$), as well as for “first half” ($M = .23$, $SD = .15$) compared to “second half” decisions ($M = .18$, $SD = .13$), and for “new” ($M = .31$, $SD = .22$) compared to “old” decisions ($M = .24$, $SD = .17$). The main effect of decision type was non-significant in the colour task, $F(1,28) < 1$. These decision type main effects are putatively driven by domain-specific characteristics of the evidence strength under evaluation in each task. Further, the main effects of question emphasis were non-significant in the semantic, colour, temporal and recognition tasks, $F(1,30) < 1$, $F(1,28) < 1$, $F(1,20) = 1.06$, $p = .317$, $\eta_p^2 = .05$, $F(1,27) = 1.48$, $p = .234$, $\eta_p^2 = .05$ respectively.

Crucially, the decision type by question emphasis interactions were significant in the semantic, temporal and recognition tasks, $F(1,30) = 13.76$, $p = .001$, $\eta_p^2 = .31$, $F(1,20) = 17.50$, $p < .001$, $\eta_p^2 = .47$, and $F(1,27) = 14.12$, $p = .001$, $\eta_p^2 = .34$ respectively. The interactions were in the same direction across these tasks, reflecting "counter-emphasis" shifts in response proportion i.e. a reduced proportion of low confidence decisions made under question emphasis (see Figure 3.5.). Although the interaction was clearly non-significant in the colour task, the direction was nevertheless in the same counter-emphasis direction as the other three tasks, $F(1,28) = 1.48$, $p = ..233$, $\eta_p^2 = .05$. Pairwise t-tests demonstrated significant counter-emphasis low confidence shifts for "disagreeable" but not "pleasant" decisions in the semantic task, $t(30) = 2.60$, $p = .014$, $d = 0.48$, and $t(30) = 1.99$, $p = .055$, $d = 0.36$ respectively; for both "first half" and "second half" decisions in the temporal task, $t(20) = 4.68$, $p < .001$, $d = 1.23$, and $t(20) = 3.00$, $p = .007$, $d = 0.70$ respectively; for "new" but not "old" decisions in the recognition task, $t(27) = 3.17$, $p = .004$, $d = 0.60$, and $t(20) = 1.85$, $p = .075$, $d = 0.35$ respectively; and neither for "orange" nor "yellow" decisions in the colour task, $t(28) = 0.75$, $p = .459$, $d = 0.15$, and $t(28) = 1.34$, $p =$ .190, $d = 0.26$ respectively. The low confidence proportion results therefore bear out the prediction that strategic effects of question emphasis would manifest across cognitive domains *exclusively* in the presence of low evidence strength and heightened levels of uncertainty.

Table 3.3. Design key and descriptive statistics for cross-task analyses of proportion of endorsements of available decision categories (Decision A and Decision B) made at particular confidence levels, comprising the proportion of decisions made with high ($HighA_{prop}$ and $HighB_{prop}$), medium ($MedA_{prop}$ and $MedB_{prop}$) and low confidence levels ($LowA_{prop}$ and $LowB_{prop}$).

Task		**Semantic**		**Colour**		**Temporal**		**Recognition**	
Question		pleasant?	disagreeable?	orange?	yellow?	first half?	second half?	old?	new?
Emphasis		Decision A	Decision B	Decision A	Decision B	Decision A	Decision B	Decision A	Decision B
$HighA_{prop}$	*M*	.63	.60	.77	.75	.40	.40	.40	.35
	SD	.18	.19	.18	.17	.22	.25	.20	.18
$HighB_{prop}$	*M*	.60	.63	.75	.76	.47	.45	.30	.32
	SD	.20	.19	.16	.19	.25	.21	.19	.19
$MedA_{prop}$	*M*	.31	.31	.17	.18	.43	.31	.37	.39
	SD	.12	.13	.13	.12	.15	.12	.15	.14
$MedB_{prop}$	*M*	.27	.28	.17	.18	.30	.42	.35	.41
	SD	.13	.15	.11	.16	.13	.16	.11	.14
$LowA_{prop}$	*M*	.08	.10	.07	.07	.16	.29	.22	.26
	SD	.11	.10	.08	.10	.12	.20	.18	.17
$LowB_{prop}$	*M*	.15	.11	.09	.07	.23	.13	.35	.27
	SD	.13	.11	.09	.07	.18	.12	.23	.22

Note: *M*, mean; *SD*, standard deviation.

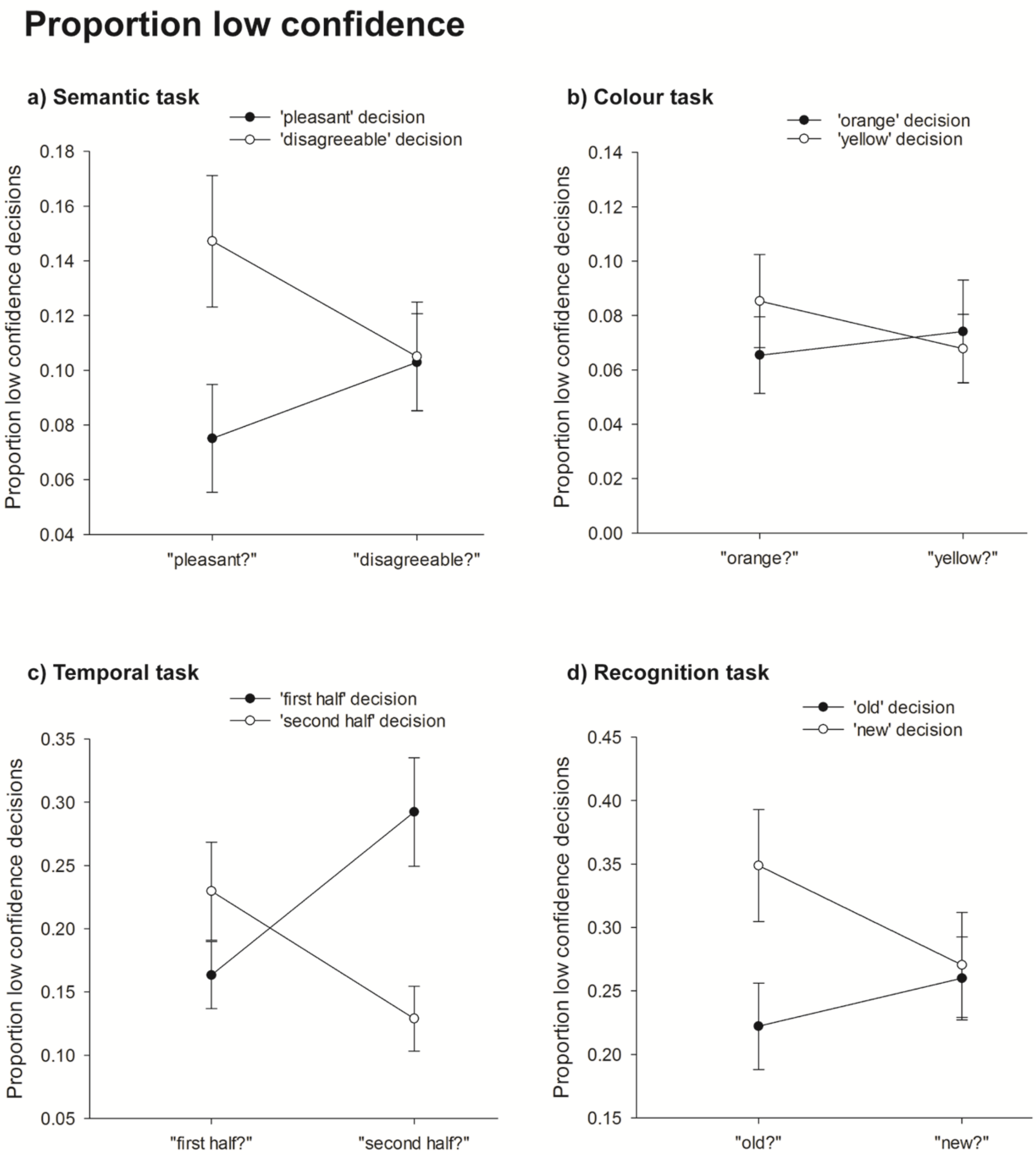


Figure 3.5. Question emphasis effects on the proportion of low confidence decisions made for decision categories across tasks: **a.** semantic task, **b.** colour task, **c.** temporal task and **d.** recognition task. Error bars represent standard error of the mean.

3.4. Discussion

The present experiment probed the strategic influence of implicit goal information imparted by test question format on evaluative performance across different domains. This followed from the findings of the recognition experiments conducted in Chapter 2, in which question format was found to impart goal emphasis to particular decision categories and instil strategies that directly influenced memory decision outcomes. The present experiment interrogated whether the question emphasis effect was rooted in cross-domain cognitive control mechanisms via presentation of four item discrimination tasks performed by the same sample of subjects: the semantic judgement task, the colour perception task, the episodic temporal judgement task and the episodic recognition task. The results revealed variable influences of question emphasis manipulations across behavioural measures, with a general lack of evidence in support of cross-domain emphasis effects emerging in analyses of “primary” measures of yes/no decision-making, including SDT estimates of criterion bias, as well as the accuracy and time taken to endorse task decision categories. However, consistent cross-domain emphasis effects were observed in the final set of “secondary” proportion confidence analyses, which segregated the proportion of endorsements of decision categories in each task according to self-rated confidence, and demonstrated reliable effects of question emphasis in shifting the proportion of low confidence decisions across the four tasks. The findings serve to highlight the intricacies in elucidating control processes that exert “general” influences on cognitive processing across different domains.

The establishment of the question emphasis effect in the previous chapter centred on analyses of SDT criterion and decision accuracy measures of recognition performance. Specifically, question emphasis shifted SDT criterion so as to reduce the likelihood of endorsing emphasised decisions, whilst improving the accuracy of emphasised endorsements when made. These effects of question emphasis on recognition behaviour were demonstrated to be reliable via replication of these concomitant effects on criterion and decision accuracy in two separate experiments presented in Chapter 2, as well as in an independent online study recruiting a large sample (Mill & O'Connor, 2014). However, in the present experiment, the only reliable effect of question emphasis on these two behavioural measures was exclusive to the semantic task. As with the findings of the previous chapter, question emphasis shifted SDT criterion to reduce the frequency of endorsing emphasised semantic decisions, whilst concurrently increasing the accuracy of these emphasised endorsements. Hence, the SDT criterion and decision accuracy effects in the semantic task suggest that question format might serve as a hitherto unaccounted for source of strategic bias on "primary" yes/no response measures in laboratory tests of semantic memory. Response biases operating on semantic memory have been alluded to previously, albeit as nested within recognition tasks that engage aspects of semantic processing, such as the proposed criterion-gating of semantic elaboration of retrieved content in the remember/know procedure (Donaldson, 1996) and in the recognition of emotional stimuli (Kapucu, Rotello, Ready & Seidl, 2008; Talmi & Moscovitch, 2004). The findings of the present experiment suggest that tasks directly assessing semantic memory might also be subject to goal-driven strategic biases.

However, rather than describing sources of bias specific to semantic evaluation, the primary aim of the present experiment was to elucidate strategic biases that manifest across different cognitive domains. Indeed, the question emphasis effects on SDT criterion and decision accuracy in the semantic task did not generalize to the other discrimination tasks. Of particular note is the failure to recover question effects on criterion and decision accuracy in the current recognition task, which was of a broadly similar format to the recognition tasks of the previous chapter that yielded reliable question effects on these measures. One potential factor accounting for the replication failure is the reduced power afforded by the present recognition task, in that only 60 trials of each question emphasis condition were presented to participants, whereas 120 trials of each question condition were presented in the Chapter 2 experiment. Hence, it is possible that the present recognition criterion shift, which operated in the expected counter-emphasis direction, would have reached conventional significance if more trials were provided.

Further, the ostensibly minor differences in format between the present recognition task and those of Chapter 2 might nevertheless have introduced significant behavioural differences, as supported by prior reports of subtle aspects of task format influencing measures of recognition performance (Hicks & Marsh, 1999; Marsh & Hicks, 1998). The implied difficulties in isolating specific strategic influences on recognition performance are substantiated by prior failure to recover reliable biases following more explicit test environment manipulations than the present subtle variations of question format. For instance, both extreme variations in the base rate proportion of *old* to *new* items (Herron, Quayle & Rugg, 2003; Cox & Dobbins, 2011) and cueing of likely evidence strength on a trial-by-trial basis (Morell, Gaitan & Wixted, 2002; Stretch & Wixted, 1998b) have been observed to yield surprisingly modest effects on response criterion. An important factor contributing to these mixed findings is participants' willingness to adopt controlled strategies (Koriat & Goldsmith, 1996) and the related ease of integrating strategies into ongoing evaluative processing (Benjamin & Bawa, 2004; Rhodes & Jacoby, 2007; Stretch & Wixted, 1998b). Hence, the present partial intermixing of the question format manipulations (i.e. question emphasis alternated after 30 trials within each task's 120 test list) might have contributed to the weak effects on criterion and decision accuracy measures, by increasing the cognitive effort entailed in integrating this source of goal information into the ongoing analysis of memory strength. This is contrasted with the fully blocked manipulations of question format in the Chapter 2 experiments, which putatively affords easier goal emphasis integration.

A major cause of the difficulty in isolating question emphasis effects in criterion and decision accuracy measures is the observed variance in overall sensitivity across tasks. These cross-task sensitivity differences putatively arise from differences in the underlying evidence strength signals under evaluation in each task domain (Green & Swets, 1966). Variability in evidence strength has been highlighted as a primary impediment in eliciting stable cross-task measures of performance monitoring in the field of metacognition (Lau, 2010; Fleming & Dolan, 2012). Evidence strength variability is also capable of impacting more directly on "primary" measures of performance, given that evaluative outcomes are widely assumed to emerge from the interaction of evidence strength assessment and higher-order strategic processes, as postulated in research in episodic (Benjamin, 2007) and non-episodic domains (Cabeza, Ciaramelli, Olson & Moscovitch, 2008; Corbetta & Shulman, 2002). Hence, failure to adequately equate the evidence strength signals might have encouraged differences in the relative contributions of evidence and strategic processes to evaluative processing in different task domains. For instance, high levels of evidence strength in the colour task (for which overall sensitivity was at ceiling levels) might have served to reduce the general need to engage strategic biases to optimise performance, as is suggested by the broad absence of strategic effects of question format in this task.

Variation in evidence strength across the four tasks might also have enhanced strategic influences other than the question emphasis bias, such as the observed strategic caution towards particular decision categories *within* tasks (as demonstrated by the patterns of negative absolute criterion placement in the semantic and colour tasks, and complementary effects of decision type on decision accuracy and RT). Other extraneous strategic influences might have arisen from the online monitoring of the different performance levels *across* the different tasks. For instance, given the greater difficulty of the episodic recognition and temporal tasks compared to the two non-episodic tasks (as revealed by the overall task sensitivities), participants might have attempted to counteract the subjective experience of weaker episodic evidence by adopting confirmatory biases acting in converse directions to the question emphasis bias. Similar biases resulting from evidence strength monitoring have been reported in prior recognition research (Hirshman, 1995; see Chapter 1, Section 1.4.). Overall, both the SDT criterion and decision accuracy measures suffered from collapsing across the observed variability in overall levels of evidence strength across tasks. The related variability introduced in the strategic processes engaged during task performance therefore made it harder to isolate the question emphasis bias of primary interest.

Indeed, reliable effects of question emphasis emerged in the proportion confidence analyses that aimed to account for the cross-task variability in evidence strength. Implicit in the proposed dichotomy between evidence strength and strategic processes is the dynamic up-regulation of strategic processes when evidence strength is weakly diagnostic and leads to the experience of uncertainty. Uncertainty drives a heightened reliance on strategic sources of diagnostic information to reach adaptive decision outcomes, as documented in diverse forms of decision-making (Kahneman, Slovic & Tversky, 1982) and perception (Whiteley & Sahani, 2008). Hence, stronger question emphasis effects were expected in the present experiment for the proportion confidence analyses, which segregated decision proportions according to levels of confidence and inferred levels of evidence uncertainty. Decisions rendered with low confidence were assumed to reflect high levels of uncertainty, as predicated on similar assumptions in functional neuroimaging studies of decision-making in episodic (Henson et al., 2000; Fleck et al., 2006) and visual domains (Fleming et al., 2012). The results demonstrated reliable cross-domain reductions in the endorsements of decisions emphasised by the question, *specifically* at low levels of confidence and heightened uncertainty. Indeed, the only proportion low confidence shift that failed to reach significance was observed in the colour task, in which evidence strength was at ceiling levels and the scope for the question emphasis bias to impact on behaviour was reduced.

Further, the present strategic confidence shifts acted in the same “counter-emphasis” direction as the question biases described in the previous chapter i.e. in reducing the likelihood of endorsing emphasised decisions. The fact that these strategic shifts were confined to low confidence decisions (and, in the case of the temporal task, manifested despite presence of a converse “pro-emphasis” criterion bias) corroborates allusion in the previous chapter to the mediating influence of evidence strength uncertainty on environmental strategies. To clarify, stronger question emphasis effects were obtained in Experiment 2 (Sections 2.2. and 2.3.) when episodic evidence was systematically weakened by encoding tasks that yielded shallow levels-of-processing (LOP; Craik & Lockhart, 1972), with progressively weaker question emphasis effects elicited following intermediate and deep LOP encoding tasks.

The selective impact of strategic question effects on the low confidence portion of the evidence spectrum also clarifies the relative involvement of the ‘retrieval control’ (or ‘evidence control’, to use a less domain-specific term) and ‘decision control’ sub-processes defined in Chapter 1 (see Section 1.3.). If question emphasis shifts were underpinned by evidence control mechanisms, selective impacts on the proportion of high confidence decisions would be expected, resulting from heightened attempts to recover stronger evidence for emphasised decisions. In this scenario, question emphasis would increase the proportion of emphasised decisions registered at high confidence, consistent with prior suggestion of heightened evidence control leading to heightened recollective source monitoring in episodic evaluation (given that recollection is typically associated with high confidence responding; Jacoby, Kelley & McElree, 1999; Jacoby, Toth & Yonelinas, 1993). Rather, the observed results bear out the selective impact of question emphasis effects on low confidence decisions, consistent with the recruitment of decision control mechanisms in the face of uncertainty. The findings serve to highlight that environmental manipulations primarily impact on evaluative performance via the up-regulation of decision control mechanisms, thereby validating the preferential study of this sub-category of cognitive control during the course of my PhD.

The present question emphasis effects also highlight the potential role for decision confidence as a behavioural index of cross-domain processing. This complements prior report of cross-modality metacognitive processes being elicited in measures of response confidence across different perceptual tasks (Song et al., 2011; see Section 1.6.). The cross-domain consistency of the present confidence effects is especially noteworthy when considering the large differences in overall sensitivity and evidence strength across tasks. The earlier discussion illustrated the propensity for evidence strength variability to impede the isolation of higher-order processes active in different cognitive domains – a problem that is typically controlled by psychophysical methods of performance equation for perceptual domains (e.g. Fleming, Weil, Nagy, Dolan & Rees, 2010) and by levels-of-processing manipulations for episodic memory domains (as in the previous chapter, Craik & Lockhart, 1972). The present findings illustrate the utility of response confidence as an inferential alternative to these more time-consuming methods of evidence strength control (especially as overall run time is a key concern in experiments involving a number of task domains). A confidence-based segregation of evidence strength might be especially useful when trying to access a large sample, in which case the psychophysical calibration of each individual's available evidence strength becomes particularly impracticable. These concerns might become increasingly pertinent with the advent of online methods of psychological experimentation (Birnbaum, 2000).

In conclusion, whilst the findings of the present experiment provide evidence suggesting that question emphasis effects are indeed underpinned by cross-domain strategic control processes, they nevertheless also serve to emphasise the difficulty in isolating these control processes in overt behaviour. The variable evidence strength signals across different task domains impaired the recovery of cross-task question emphasis in "primary" measures of performance, such as SDT criterion and decision accuracy. Stronger effects in these primary measures might be expected following more explicit manipulations of strategic control. This was demonstrated by a recent study that explicitly manipulated the base rate proportion of presented items in perceptual and episodic discrimination tasks, and observed comparable shifts in diffusion model estimates of response bias across both task domains (White & Poldrack, 2014; see Section 1.6.). Hence, the next experimental chapter implemented stronger manipulations of goal emphasis (in the form of monetary incentives) with the aim of recovering more pronounced cross-domain strategic effects in primary measures of evaluative performance. These manipulations were applied to the semantic and episodic tasks that have thus far recovered the strongest strategic biases across the four presented experiments. The reduction in presented task domains raised the additional benefits of increasing the trial counts for conditions of interest in each task (and associated power), as well as reducing the scope for extraneous strategic effects driven by the monitoring of relative performance across task domains (as discussed previously). This enabled a more targeted interrogation of cognitive control mechanisms capable of influencing behaviour across different evaluative domains.

Chapter 4: The interaction of explicit goal emphasis and cued expectation in cross-domain control

4.1. Introduction to Experiment 5

The previous chapter investigated the capacity for an implicit feature of the test environment – goal emphasis imparted by the test question – to heighten strategic control across different cognitive domains. The experiment yielded only partial evidence of cross-domain question format manipulations leading to common strategic effects, which were confined to “secondary” measures of decision confidence. The present chapter sought clearer evidence of cross-domain effects on “primary” response measures (directly related to yes/no decision performance) via more explicit manipulation of the test environment. Goal emphasis was therefore varied by the provision of mixed monetary incentives (Han, Huettel, Raposo, Adcock & Dobbins 2010; Newman, Widom & Nathan, 1985) and expectations were varied by the provision of textual cues (O’Connor, Han & Dobbins, 2010; Posner, Snyder & Davidson, 1980; Squires, Squires & Hillyard, 1975). These two features of the test environment were manipulated in identical fashion across an episodic recognition task and a semantic judgement task, both comprising the same yes/no discrimination format as the tasks described in the previous chapter (see Section 3.2.3.). The experiment afforded a direct behavioural interrogation of cross-domain control mechanisms previously speculated in the literature (Cabeza, Ciaramelli, Olson & Moscovitch, 2008; Corbetta & Shulman, 2002; Fleming & Dolan, 2012).

Manipulation of goal emphasis was effectuated by the mixed incentives procedure, wherein both a monetary reward *and* punishment are selectively associated with one of the two decision categories available in each task domain (i.e. for “old” or “new” decisions in the recognition task, and for “pleasant” or “unpleasant[6]” decisions in the semantic task). As highlighted in the introduction to Chapter 2 (Section 2.1.), findings in the human rewarded decision-making literature suggest that mixed incentives selectively emphasise a particular decision category as a higher-order goal, leading to a disconfirmatory or “counter-emphasis” bias that reduces the likelihood of endorsing the emphasised decision (Newman et al., 1985; Hartung, Milich, Lynam & Martin, 2002). This reward-related bias differs from more widely studied effects of general presence or absence of performance incentives (applied to both available decision categories) that serve to heighten overall strategic caution (Duffy, 1962; Knutson, Adams, Fong & Hommer, 2001), as well as of “reward only” incentives applied to particular decision categories, which encourage biases in converse confirmatory directions (i.e. increasing the likelihood of endorsing the decision selectively associated with the reward; Swets, Tanner & Birdsall, 1961; Healy & Kubovy, 1978). Mixed incentives therefore permit empirical scrutiny of a goal-driven constraint on evaluative performance that has been thus far neglected in the human rewarded decision-making literature.

The neglect in researching effects of mixed incentives also extends to the field of episodic memory, despite demonstration in the introductory chapter of how goal emphasis mechanisms might serve to heighten strategic caution over memory

[6] Note that in the present experiment, the term “unpleasant” replaces the term “disagreeable”, which was used in Chapter 3 to describe the equivalent semantic item and decision categories. This follows from allusion in the previous chapter to fundamental difficulties in lexically processing “disagreeable?” question formats, which led to reduced sensitivity and decision accuracy, as well as slower reaction time. Adoption of the alternative “unpleasant” descriptor is motivated by its comparatively greater frequency of occurrence in the English language and hence likely greater ease of integration into ongoing evaluative processing.

evaluation in the real world (in the eyewitness test scenario, Section 1.1.; see Figure 1.2.). Indeed, a behavioural pattern reflective of a “counter-emphasis” strategy was observed in a prior functional neuroimaging study of recognition memory, in which mixed incentives were selectively applied to “old” or “new” decision categories (Han et al., 2010). However, this study failed to link the implied strategic effects to formalized measures of bias, such as signal detection theory estimates of response criterion (SDT *c*; MacMillan & Creelman, 2005), and instead found an effect on secondary measures of response confidence (specifically, an increase in confidence for decisions associated with mixed incentives). Indeed, the application of mixed incentives in the wider rewarded decision-making literature has also suffered from a lack of formalization, with bias effects predominantly inferred from shifts in raw response proportions (Newman et al., 1985; Hartung et al., 2002), that conflate sensitivity and bias (MacMillan & Creelman, 2005). It was therefore a primary aim of the present chapter to assess whether mixed incentives led to analogous “counter-emphasis” shifts in criterion bias (and other primary response measures) as those reported in the recognition experiments of Chapter 2 (Experiments 1-3). Furthermore, the inclusion of an emphasis-neutral condition (the “No token” condition), in which no monetary incentives were provided for either decision category, enabled examination of whether particular decision categories in each task were driving goal emphasis effects. Another focal aim, given the mixed findings of Chapter 3, was to determine whether more explicit manipulation of goal emphasis information led to common strategic effects across the semantic and episodic task domains.

The present manipulation of cued expectation was similarly motivated by a desire to probe the cross-domain underpinnings of expectation-related strategies. Whilst strategic expectation effects have been more widely studied than those associated with goal emphasis, research into expectation has nonetheless been conducted within largely separated fields focusing on isolated cognitive domains. This is true for expectation manipulations enacted both prior to the onset of a task block (i.e. by explicating the “base rate” proportion of items presented from available categories; Swets et al., 1961) and cued on a trial-by-trial basis within the block itself (Posner et al., 1980). Despite the failure to interrogate cross-domain expectation effects, expectation-driven strategies have often been indexed by formalized SDT criterion estimates, which reveal biases in confirmatory “pro-expectation” directions (increasing the endorsement of expected decisions) across a number of isolated task domains, including visual perception (Swets et al., 1961), spatial attention (Posner et al., 1980), motor response selection (Kawashima, Roland & O’Sullivan, 1995), and identification of both visual and auditory items in the oddball paradigm (Polich, 2007; Squires et al., 1975). Similar pro-expectation biases have been documented in the domain of episodic memory, following both blockwise base rate (Ratcliff, Gronlund & Sheu, 1992) and trial-by-trial cueing manipulations (Rhodes & Jacoby, 2007; O’Connor et al., 2010; see Chapter 1, Section 1.4. for further details). An additional behavioural marker of the pro-expectation bias is the reliable reduction in reaction time for decisions that are made congruent to prior expectations, as observed in tasks spanning perceptual (Posner et al., 1980), motor (Abrams & Jonides, 1988) and episodic domains (O’Connor et al., 2010).

The reliability of the above behavioural effects has led to the proposal of a dedicated neural system that enforces “top-down” expectation biases on the assessment of evidence accumulated in different sensory and cognitive domains (Cabeza, Ciaramelli, Olson & Moscovitch, 2008; Corbetta & Shulman, 2002). However, direct evidence of the cross-domain function of this proposed system has been lacking, given the failure to probe the correspondence of expectation-driven biases across different cognitive domains within the same sample of subjects. The design of the present experiment – in which the same sample is exposed to episodic and semantic discrimination tasks – enabled the interrogation of such cross-domain effects.

A final aim of the present study was to examine how cued expectations and goal emphasis interact during cognitive evaluation. The interaction of these two environmental factors has not been systematically studied, despite both factors often exerting concomitant strategic influences on evaluation in the real world. To demonstrate by revisiting the eyewitness recognition scenario (Section 1.1.; Figure 1.2.), the eyewitness' evaluation of the suspect line-up might be coloured both by the greater emphasis imparted to "old" decisions (due to the greater perceived consequences of making incorrect "old" decisions, as explained in Chapter 2, Section 2.1.), as well as an expectation of encountering suspects with specific characteristics at the scene of the crime, leading to confirmatory pro-expectation biases towards members of the lineup that share such characteristics. Hence, both goal emphasis and cued expectation are capable of influencing real-world evaluative outcomes, yet these environmental factors have rarely been studied in conjunction in the laboratory. To rectify this neglect, the present experiment independently manipulated cued expectation and goal emphasis during the same task phase, enabling the examination of whether the combined presence of these environmental factors failed to interfere with their isolated strategic effects (i.e. leading to recovery of both a "counter-emphasis" bias and a "pro-expectation" bias), or whether these individual strategies were altered. Testing these competing predictions across both presented task domains enabled further insight into the cross-domain underpinnings of the expectation by emphasis interaction.

To summarise, three aspects of cross-domain control were under examination: the isolated effects of goal emphasis in selectively heightening caution towards emphasised decision categories, the isolated effects of cued expectation in increasing the tendency to endorse expected decision categories, and the combined effects of emphasis and expectation on strategic processing. These effects were examined on “primary” yes/no response measures analysed in previous chapters i.e. SDT criterion, as well as the accuracy and time taken to endorse decision categories (decision accuracy and decision RT). A final series of correlational analyses aimed to further interrogate the cross-domain properties of the emphasis- and expectation-driven effects on SDT criterion, as well as the relationship between elicited cross-domain task performance and individual differences in psychometric measures of personality. The findings bear out the prediction of stronger cross-domain behavioural effects following more explicit manipulations of strategic control, and provide novel insights into the methods by which different higher-order strategies interact.

4.2. Method

4.2.1. Participants. The sample comprised 29 self-reported native English speakers (20 female, mean age = 21.41, age range = 18-50) out of a total number of 32 participants who completed the experiment (all of whom reached the minimum performance threshold of $d' > .1$). Three participants were excluded for adopting the extreme strategy of responding only to make emphasised decisions associated with the potential to earn incentives. Not only did this strategy complicate the analysis by introducing empty cells for measures of non-emphasised decision performance, but it also served as a notable difference in the evaluative processes engaged by these participants compared to the remaining sample, given that all participants were instructed to factor in evidence strength, cues and the available incentives when making their decisions. The extreme incentive strategy implied an unwillingness to effectively analyse evidence strength or cues, and hence evidence of its adoption in any of the presented experimental phases led to the complete exclusion of three participants. Informed consent for all participants was obtained in accordance with the University Teaching and Research Ethics Committee at the University of St Andrews. All participants were compensated for their time at a rate of £5/hour, in addition to the monetary value of the tokens earned during the course of the experiment (see Section 4.2.3 below for further details).

4.2.2. Stimuli. As in the previous chapter, word stimuli for all task phases were sampled from the Toglia and Battig semantic word norms (Toglia & Battig, 1978), which contains 2584 common nouns rated for pleasantness in a large sample (with a higher rating denoting greater pleasantness). These ratings were used to create *pleasant* and *unpleasant* item categories in the *semantic task* (spanning pleasantness rating ranges of 4.43-6.32 and 1.73-3.58 respectively). As before, these semantic item categories were separated from each other by 1 standard deviation in pleasantness ratings, so as to enforce a relatively robust semantic item boundary and ensure that words of ambiguous pleasantness were excluded from either category. The wordlist for the single uncued semantic task comprised 192 words, of which 96 had *pleasant* and 96 had *unpleasant* meanings, whereas wordlists created for the three cued versions of the semantic task each comprised 180 words, of which 90 had *pleasant* and 90 had *unpleasant* meanings. Across both uncued and cued task phases, words of each semantic item category were sampled from at regular pleasantness-rating intervals to encompass a range of high to low evidence strength (which was shown in the previous chapter to modulate cross-domain goal emphasis effects, see Section 3.3.).

For the *recognition task* phases, a randomly selected half of the preceding semantic task wordlist was carried over to form words in the studied *old* recognition item category. Hence, the single cued recognition task phase constituted 192 words, of which 96 were *old* and 96 were *new* (i.e. not seen before in the experiment), whereas wordlists created for the three cued recognition task phases each constituted 180 words, of which 90 were *old* and 90 were *new*. In creating these recognition task wordlists, potential confounding influences of the preceding semantic phase were counteracted by ensuring that *old* words across recognition conditions of interest comprised an equal proportion of semantic status tems (*pleasant/unpleasant*) and prior semantic incentives (emphasised/non-emphasised). The experiment was presented and responses recorded using PCs running MATLAB (MathWorks Inc., Natick, MA, 2000) and Psychophysics Toolbox (Brainard, 1997).

4.2.3. Procedure. All experimental phases were preceded by onscreen instructions and brief practice versions of the presented tasks. The main phases involved variations of two discrimination tasks – the semantic task and the recognition task – which shared the same underlying discrimination format (i.e. requiring a yes/no discrimination decision between *Signal A* and *Signal B* item categories; see Figure 3.1a), but differed according to the cognitive domain under evaluation (semantic memory or episodic recognition memory). In the *semantic task,* participants discriminated between words based on whether their meaning could be deemed *pleasant* (*Signal A*) or *unpleasant* (*Signal B*). In the recognition task, participants decided if words were *old* (seen in the previous semantic phase; *Signal A*) or *new* (not seen before in the experiment; *Signal B*). Across conditions of interest in both tasks, participants submitted their discrimination decisions via the number keys to which the assignment of decision categories remained fixed (semantic task: 1 = "pleasant", 2 = "unpleasant"; recognition task: 1 = "old"; 2 = "new"). No question prompt was provided at any stage, in an attempt to minimize the implicit goal emphasis effects of question format detailed in Chapters 2 and 3. Further, response confidence was not measured given both the need to minimise the overall run time in light of the addition of cued task phases, and the anticipation that the present explicit goal emphasis manipulations would lead to effects on primary yes/no response measures (unlike the implicit effects of the previous chapter).

Each participant's session began with a single "uncued" semantic-recognition task block, in which mixed incentives were provided to manipulate goal emphasis in isolation (i.e. in the absence of cues; see Figure 3.1b). Incentives took the form of performance-related "tokens" that were converted to money at the end of the experiment. Participants were not informed of the total number of tokens available during the experiment nor the precise value of each token (i.e. 488 tokens each worth £0.02), merely that acquisition of tokens would lead to a maximum monetary bonus of £10 on top of the hourly compensation rate (rounded up to the nearest £0.50 denomination). The mixed incentive procedure involved the selective association of both a token reward for a correct response (+1 token) and a token punishment for an incorrect response (-1 token) for one of the two decision categories available in the task being performed. To demonstrate for the semantic task, goal emphasis was imparted to "pleasant" decisions in the "pleasant token" condition by adding one token for each correct "pleasant" decision and subtracting one token for each incorrect “pleasant” decision, whilst providing no tokens for “unpleasant” decisions irrespective of accuracy. This payoff scheme was reversed in the “unpleasant token” incentive condition, such that correct/incorrect “unpleasant” decisions led to the gain/loss of tokens, whereas “pleasant” decisions did not. A “No token” condition was also presented as a baseline measure of performance in the absence of goal emphasis, such that performance-related tokens in this condition were entirely absent from either semantic decision category. Analogous mixed incentives were provided in the recognition task token conditions (“old token”, “new token” condition and “No token” conditions).

The variation in token emphasis was enacted via a similar partially blocked scheme to that used in Experiment 4 (Chapter 3, Section 3.2.3). Hence, uncued wordlists for semantic and recognition tasks each comprised 192 words, which were presented in six mini-blocks of 32 trials, comprising two repetitions of the three token conditions. Participants were initially notified of the ensuing token emphasis mini-block by an instruction screen appearing for 5s. Feedback was also provided on a trial-by-trial basis whenever emphasised decisions were made (and hence whenever a token was lost or gained; see Figure 3.1b), so as to heighten awareness of the incentives manipulation and thereby increase its strategic influence on performance (following from prior research e.g. Estes & Maddox, 1995; see Section 1.4.). To prevent inequalities in response pacing between token emphasised and non-emphasised decisions, participants had to wait the same duration after making their yes/no decision irrespective of whether token-related feedback was provided or not.

The uncued phase was followed by three “cued” semantic-recognition blocks, in which participants’ expectations were manipulated in addition to token emphasis (see Figure 3.1b). Emphasis was manipulated in identical fashion to the “uncued” phase, with the gain/loss of monetary tokens selectively associated with endorsements of one of the two available decision categories, as varied in a partially blocked fashion across each 180-word task phase (i.e. 6 mini-blocks of 30 trials, comprising 2 repetitions of the 3 token emphasis mini-block conditions). Participants were also provided with textual cues that appeared before items and remained onscreen for the remainder of the trial (semantic task cues: “likely pleasant”, “likely unpleasant”; recognition task cues: “likely old”, “likely new”). These cues were valid in predicting the status of ensuing test items on approximately 70% of trials and invalid on the remaining 30% (i.e. setting up ~70% encounters with “expected” items and ~30% “unexpected” items). Participants were informed of the validity of the cues and encouraged to utilise them when making their discrimination decisions (in addition to the varying token emphases). The cue and emphasis manipulations were fully orthogonalized, such that each token emphasis condition had the same number of validly and invalidly cued trials, comprising an equal number of words from each item category.

The uncued phase preceded the cued phase for all participants. Further, the token emphasis mini-blocks within each task were presented in an order that equated fatigue effects for each emphasis condition across the entire participant run, based on an ordinal ranking of the anticipated extent of fatigue effects related to presentation order (i.e. first mini-block = low fatigue, mini-block 6 = high fatigue). Three of these fatigue-equated emphasis mini-block orders were created, to which participants were randomly assigned.

After performing both uncued and cued task phases, participants finally completed a short battery of psychometric questionnaires. This included paper versions of the Behavioural Inhibition/Behavioural Activation (BIS/BAS; Carver & White, 1994), Big 5

(John, Donahue & Kentle, 1991) and Brief Sensation-Seeking (Hoyle, Stephenson, Palmgreen, Lorch & Donohew, 2002) questionnaires. Once participants filled out the psychometric battery, they were monetarily reimbursed based on their overall run time and the converted value of their token amount, and allowed to leave.

4.2.4. Calculation. Analyses in the uncued and cued phases involved the same measures as those featured in Chapters 2 (Section 2.2.1.4.) and 3 (Section 3.2.4.). This included estimates of sensitivity (*d'*) and criterion bias (*c*) parameters from the equal variance signal detection model (Green & Swets, 1966; Macmillan & Creelman, 2005), corrected for errorless responding (Snodgrass & Corwin, 1988). As in Chapter 3, item categories in each task domain were assigned *Signal A* and *Signal B* status (see Figure 3.1a), which participants discriminated between based on their sensitivity and bias. To reiterate, an increasingly positive criterion denotes a reduced likelihood of endorsing the "Signal A" response option. This nomenclature was extended to describe analyses calculated according to decision status across the two tasks i.e. "Decision A" (used to describe "pleasant" and "old" decisions in the respective tasks) and "Decision B" categories ("unpleasant" and "new"). Decision accuracy was therefore calculated as the proportion of correct endorsements out of all endorsements made for each decision category ($\text{DecisionA}_{\text{corr}}$ and $\text{DecisionB}_{\text{corr}}$; see Section 2.2.1.4.). Decision RT was calculated as the time taken to endorse each decision category ($\text{DecisionA}_{\text{RT}}$ and $\text{DecisionB}_{\text{RT}}$; see Section 3.2.4.). These measures were compared across the different token emphasis conditions, which were also characterised within the same common discrimination framework: "Token A" (i.e. "pleasant token emphasis" and "old token emphasis" conditions in the semantic and recognition tasks respectively), "Token B" ("unpleasant token emphasis" and "new token emphasis") and "No token" conditions. Cue type conditions in the cued phase were also defined via the same discrimination framework (i.e. as “likely A” and “likely B” cue conditions).

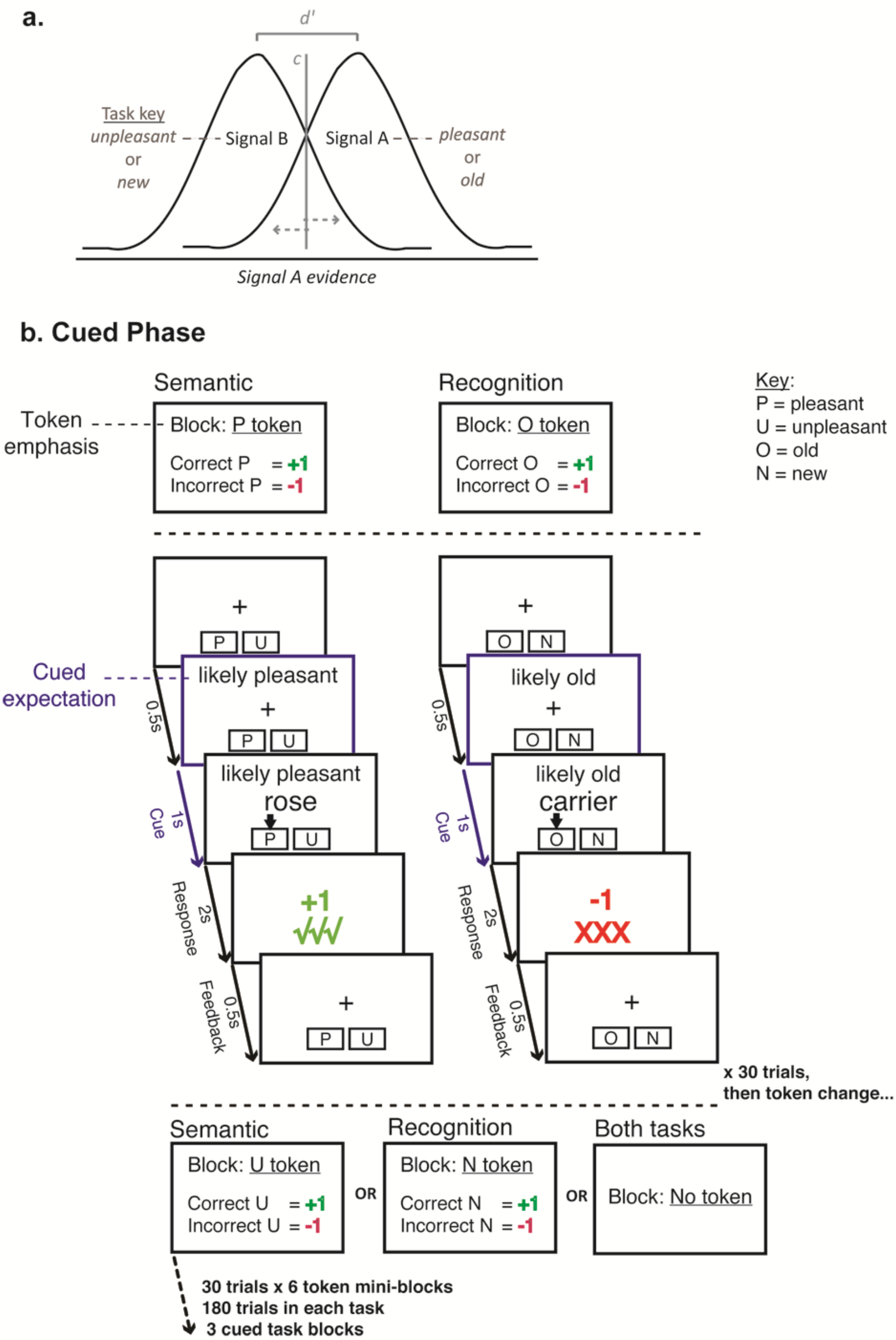


Figure 4.1. Model definition and cued phase design schematic. **a.** Common formalization of the evaluative process for the semantic and recognition tasks, with participants assumed to be discriminating between *Signal A* and *Signal B* on the basis of sensitivity (*d'*) and criterion bias (*c*). Positive criterion shifts reflect a reduced likelihood of endorsing a “Signal A”

decision (i.e. responding “Decision A”) and relatedly an increased likelihood of endorsing a “Signal B” decision (i.e. “Decision B”), whereas negative criterion shifts reflect the inverse bias. **b.** Design schematic demonstrating the common yes/no discrimination format for the cued semantic and recognition tasks. Although only the cued phase is displayed, the uncued phase was identical in format and pacing except for the absence of textual cues, which onset during the trial screen outlined in blue during the cued phase. Token emphasis was imparted by notification screens that appeared for 5s at the start of each 30-trial mini-block and explicated the decision category associated with mixed incentives i.e. +1 token for each correct emphasised decision (as demonstrated here in the example semantic task trial), -1 for incorrect emphasised decisions (as demonstrated in the example recognition task trial) and 0 tokens for correct/incorrect non-emphasised decisions (refer to the key in the top right of the panel for further clarification). A “No token” condition was also presented in both tasks, wherein neither category of decision was associated with the gain/loss of tokens irrespective of accuracy. Feedback was provided whenever emphasised decisions were made, and was replaced by a fixation cross of the same duration when non-emphasised decisions were made. Cues in the cued phase were of the form “likely A” (i.e. “likely pleasant” and “likely old” in the semantic and recognition tasks respectively) and “likely B” (“likely unpleasant” and “likely new”), and were valid in predicting the status of the ensuing item on ~70% of trials.

4.3. Results

4.3.1. Overall task performance

Differences in overall performance were first assessed with analyses of overall *d′* sensitivity that collapsed across the token emphasis and cued expectation conditions. A 2(task phase: uncued or cued) x 2(task domain: semantic or recognition) repeated measures ANOVA conducted on overall *d′* yielded a significant main effect of task domain but not phase, $F(1,28) = 16.93$, $p < .001$, $\eta_p^2 = .38$ and $F(1,28) = 1.57$, $p = .221$, $\eta_p^2 = .05$ respectively. The interaction of phase and domain was not significant, $F(1,28) = 2.57$, $p = .120$, $\eta_p^2 = .08$. The main effect of domain paralleled a similar finding in the previous chapter (Section 3.3.1.), such that overall sensitivity was greater in the semantic ($M = 2.46$, $SD = 0.25$) compared to the recognition task ($M = 2.16$, $SD = 0.39$). This difference in overall sensitivity raises the potential for variation in evidence strength in each task, which was highlighted in the previous chapter as a salient contributor to the broad lack of cross-domain effects of implicit goal emphasis manipulations on primary measures of performance (Section 3.4.). However, the present more explicit environmental manipulations of goal emphasis (via monetary token incentives) and cued expectation were expected to overcome these overall task sensitivity issues (as detailed in ensuing sections).

4.3.2. Effects of explicit token emphasis in isolation (uncued phase analyses)

4.3.2.1. SDT sensitivity and bias. Means and standard deviations for all analyses of the uncued phase are presented in Table 4.1. The effects of token emphasis on discrimination sensitivity were first analysed in one-way repeated-measures ANOVAs with emphasis as a factor (“Token A”, “Token B” or “No token”), conducted separately for the semantic and recognition tasks. The main effect of token emphasis on *d′* was significant in the recognition but not the semantic task, $F(2,56) = 6.65$, $p = .003$, $\eta_p^2 = .19$ and $F(2,56) = 1.58$, $p = .214$, $\eta_p^2 = .05$ respectively. Pairwise t-tests confirmed that the cue effect on recognition sensitivity was due to the significantly greater sensitivity observed under "old token" emphasis compared to both "new token" and "No token" conditions, $t(28) = 3.80$, $p < .001$, $d = 0.71$, and $t(28) = 2.77$, $p = .010$, $d = 0.52$ respectively (the difference between the "new" and No token conditions was non-significant, $t < 1$). The improvement in recognition sensitivity under "old" emphasis parallels a similar finding in Experiment 2 (Section 2.3.2.1.), wherein implicit "old" emphasis had the same positive effect on sensitivity. The cause of this sensitivity effect is unclear, although prior recognition studies documenting heightened source monitoring following test format manipulations designed to aid the recovery of retrieved content offer one potential mechanism (Johnson, Hashtroudi & Lindsay, 1993; Marsh & Hicks, 1998; see Chapter 7 for greater discussion of this effect, Section 7.2.).

The primary SDT analyses of criterion bias also involved one-way token emphasis ANOVAS conducted separately for each task domain. A significant main effect of token emphasis was observed in both the semantic and recognition tasks, $F(2,56) = 7.53$, $p = .001$, $\eta_p^2 = .21$ and $F(2,56) = 7.84$, $p = .001$, $\eta_p^2 = .22$ respectively. As expected, the direction of these main effects reflected bias shifts that reduced the likelihood of endorsing the decision emphasised by available token incentives (see Table 4.1; see Figure 4.2a and 4.2b). Planned pairwise t-tests conducted for the semantic task demonstrated that criterion in the "pleasant" token condition was significantly greater than criterion in both the No token and "unpleasant" token conditions, $t(28) = 2.38$, $p = .024$, $d = 0.45$ and $t(28) = 3.32$, $p = .003$, $d = 0.62$ respectively (the difference between the "unpleasant" and No token conditions was non-significant, $t(28) = 1.70$, $p = .101$, $d = 0.32$ respectively). Pairwise t-tests in the recognition task revealed that criterion in the “old” token condition was significantly greater than criterion in both the No token and “new” token conditions, $t(28) = 4.46$, $p < .001$, $d = 0.83$ and $t(28) = 2.92$, $p = .007$, $d = 0.55$ respectively (the difference between “new” and No token conditions was non-significant, $t < 1$). These pairwise comparisons suggest that cross-domain effects of emphasis on criterion were primarily driven by positive bias shifts in the respective Token A emphasis conditions.

The overall pattern of SDT results suggest that the present explicit goal emphasis manipulations led to similar counter-emphasis biases as those observed in the experiments of Chapter 2 (in which goal emphasis was implicitly manipulated). Further, the counter-emphasis bias was reliable across both presented task domains, in contrast to the findings of Chapter 3 in which cross-domain effects of implicit goal emphasis manipulations were not registered in SDT criterion.

Table 4.1. Design key and descriptive statistics for cross-task analyses of the uncued phase, including SDT sensitivity (*d′*) and criterion bias (*c*), decision accuracy ($DecisionA_{corr}$, $DecisionB_{corr}$) and decision RT measures ($DecisionA_{RT}$, $DecisionB_{RT}$).

Task		**Semantic**			**Recognition**		
Token Emphasis		pleasant Decision A	unpleasant Decision B	No token No token	old Decision A	new Decision B	No token No token
SDT *d′*	*M*	2.60	2.46	2.38	2.49	2.10	2.13
	SD	0.54	0.40	0.51	0.49	0.52	0.55
SDT *c*	*M*	-0.03	-0.30	-0.21	0.14	-0.08	-0.08
	SD	0.31	0.27	0.31	0.31	0.25	0.30
$DecisionA_{corr}$	*M*	.90	.85	.85	.91	.84	.84
	SD	.05	.06	.08	.07	.09	.07
$DecisionB_{corr}$	*M*	.90	.93	.92	.87	.87	.87
	SD	.07	.05	.05	.06	.08	.06
$DecisionA_{RT}$	*M*	0.88	0.95	0.94	0.87	1.01	1.01
	SD	0.13	0.11	0.13	0.09	0.11	0.11
$DecisionB_{RT}$	*M*	0.91	0.88	0.93	0.93	0.96	1.06
	SD	0.13	0.12	0.14	0.12	0.12	0.12

Note: *M*, mean; *SD*, standard deviation.

Criterion bias

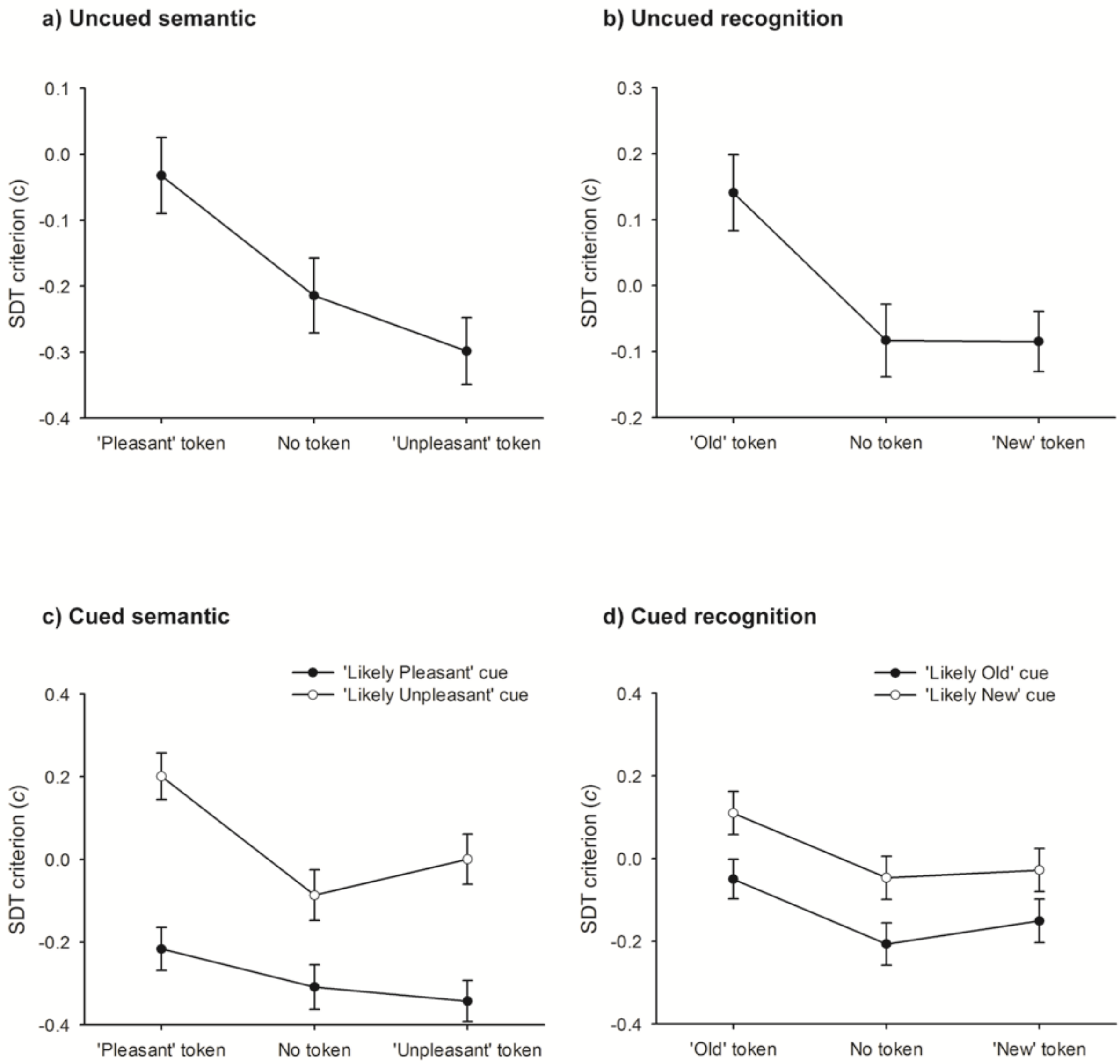


Figure 4.2. SDT criterion bias across token emphasis conditions in all experimental task phases: **a)** uncued semantic, **b)** uncued recognition, **c)** cued semantic and **d)** cued recognition phases. Note that token conditions on the x axis were also formalized according to the common discrimination format outlined in the Calculation section (4.2.4.) as follows: “Token A” emphasis (“pleasant” and “old” token conditions) and “Token B” emphasis (“unpleasant” and “new” token conditions). Separate lines in the cued plots represent different cue type conditions. Error bars represent standard error of the mean.

4.3.2.2. Decision accuracy. The effect of token emphasis on the accuracy of endorsing emphasised decisions in the uncued phase was interrogated via 2(decision type: Decision A or Decision B) x 3(token emphasis: Token A, Token B or No token) repeated-measures ANOVAs conducted separately for each task. The main effect of decision type was significant for the semantic but not the recognition task, $F(1,28) = 24.06$, $p < .001$, $\eta_p^2 = .46$ and $F < 1$ respectively. The decision type effect in the semantic task arose from the greater overall accuracy of endorsing "unpleasant" ($M = .92$, $SD = .03$) compared to "pleasant" decisions ($M = .87$, $SD = .05$), which considered with the pattern of absolute criterion placement in this task (i.e. negative criterion values across all emphasis conditions, indicating a reduced overall likelihood of making "unpleasant" decisions; see Table 4.1.), suggests that participants were more strategically cautious in making "unpleasant" compared to "pleasant" decisions. A similar decision type effect was observed in Chapter 3 (Section 3.3.3.) and is likely to reflect processes specific to the semantic task, given that the main effect did not generalize to the recognition domain.

Further, the main effect of token emphasis was significant in the recognition but not the semantic task, $F(2,56) = 5.34$, $p = .008$, $\eta_p^2 = .16$ and $F(2,56) = 1.96$, $p = .150$, $\eta_p^2 = .07$ respectively. Planned comparisons of the token condition means confirmed that the token main effect in the recognition task reflected the greater overall decision accuracy under "old" token emphasis ($M = .89$, $SD = .05$) compared to both the No token ($M = .85$, $SD = .05$) and "new" token ($M = .86$, $SD = .07$) conditions, $F(1,28) = 14.03$, $p = .001$, $\eta_p^2 = .33$ and $F(1,28) = 6.46$, $p = .017$, $\eta_p^2 = .19$ respectively (there was no difference between the "new" and no token conditions, $F < 1$). This finding complements the earlier improvement in recognition sensitivity under “old” emphasis (Section 4.3.2.1.) in suggesting an overarching performance benefit under “old” emphasis.

The interaction of decision type and token emphasis was the primary index of the anticipated effects of explicit token emphasis in improving the accuracy of endorsing emphasised decisions. This interaction was significant for both the semantic and recognition tasks, $F(2,56) = 9.54$, $p < .001$, $\eta_p^2 = .25$ and $F(2,56) = 9.57$, $p < .001$, $\eta_p^2 = .26$ respectively. However, the direction of the interactions in both tasks suggests a more complex pattern of token emphasis effects than anticipated (see Table 4.1.; see Figure 4.3a and 4.3b). Planned pairwise t-tests in the semantic task revealed that the accuracy of “pleasant” decisions was significantly greater under “pleasant” token emphasis than No token or “unpleasant” token emphasis, $t(28) = 3.57$, $p = .001$, $d = 0.69$ and $t(28) = 4.12$, $p < .001$, $d = 0.78$ respectively (there was no difference in “pleasant” decision accuracy between “unpleasant” and No token conditions, $t(28) = 0.10$, $p = .920$, $d = 0.02$). However, the accuracy of “unpleasant” decisions under “unpleasant” token emphasis, despite being numerically greater, was not significantly different from the No token and “new” token conditions, $t(28) = 1.62$, $p = .117$, $d = 0.30$ and $t(28) = 1.77$, $p = .087$, $d = 0.33$ respectively (there was also no significant difference between “pleasant” and No token conditions, $t(28) = 0.66$, $p = .511$, $d = 0.12$). A similar pattern was recovered in the recognition task, in which pairwise t-tests revealed that “old” decision accuracy was significantly greater under “old” token emphasis than both No token or “new” token emphasis, $t(28) = 5.19$, $p < .001$, $d = 0.96$ and $t(28) = 4.06$, $p < .001$, $d = 0.77$ respectively (there was no difference in “old” decision accuracy between “new” and No token conditions, $p = .767$). The accuracy of “new” decisions did not reliably shift according to token emphasis, with “new” decision accuracy being numerically lowest in the “new” token condition compared to the No token and “old” token conditions, although not significantly so, both *t*s < 1 (there was no difference in “old” decision accuracy between “new” and No token conditions, $t < 1$). Hence, the results of the pairwise comparisons in each task suggests that decision accuracy effects were confined to an improvement in Decision A accuracy under concordant Token A emphasis.

The overall pattern of decision accuracy analyses demonstrate that explicit emphasis served to heighten strategic caution, albeit primarily towards how Decision A categories were evaluated. Considered with the SDT bias results, the decision accuracy effects suggest that the reduced likelihood of endorsing “old” and “pleasant” decisions under concordant emphasis was undertaken to maximise their accuracy. The persistence of token emphasis effects across the presented semantic and episodic task domains validates the prediction of stronger cross-domain influences following more explicit manipulation of goal emphasis.

Uncued Decision Analyses

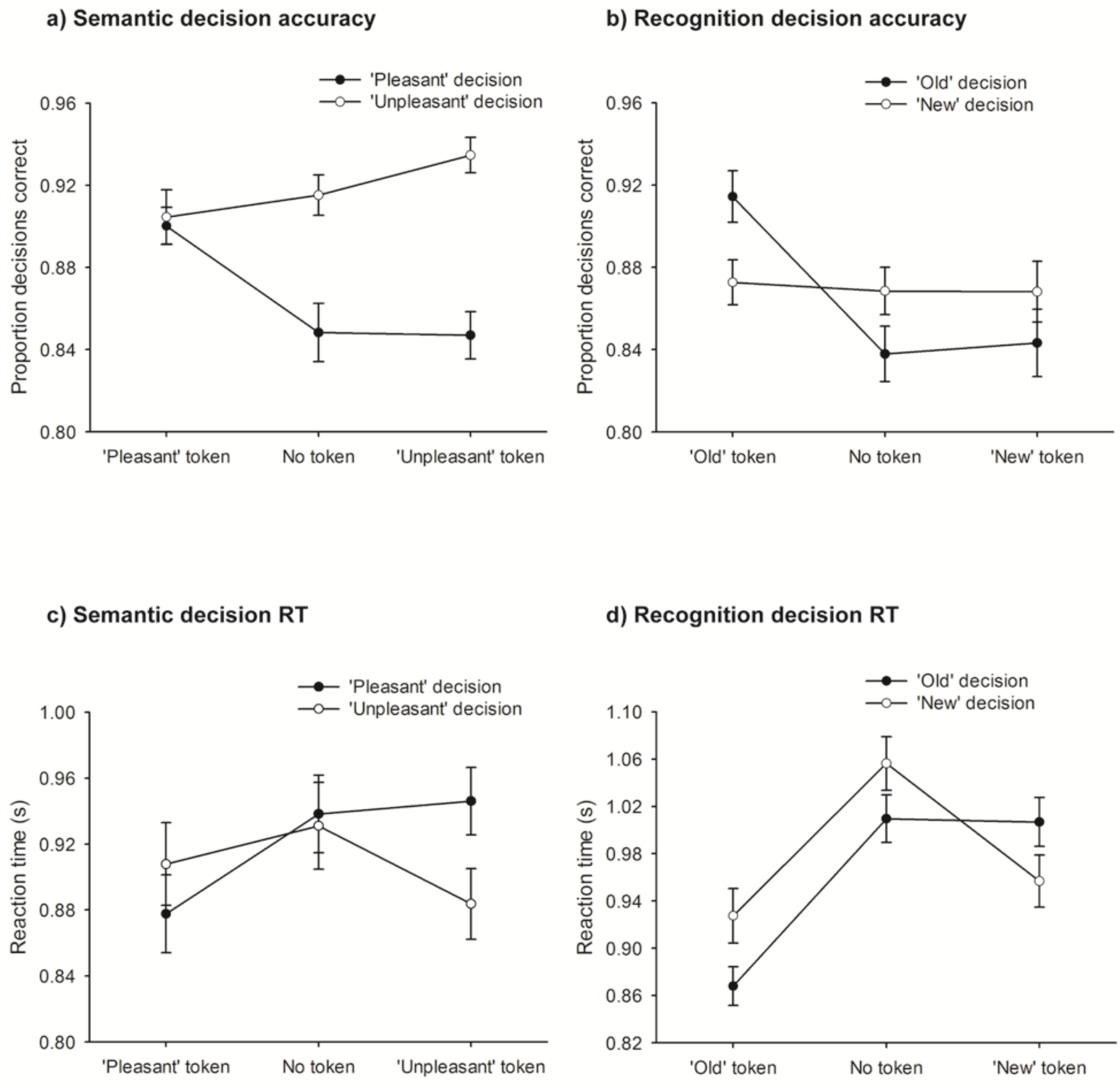


Figure 4.3. Effects of token emphasis on decision accuracy and decision RT in the uncued phase. Decision accuracy plots are presented for **a)** semantic and **b)** recognition tasks. Decision RT plots (in seconds, s) are presented for **c)** semantic and **d)** recognition tasks. Separate lines denote decision categories available in each domain. Error bars represent standard error of the mean.

4.3.2.3. Decision RT. The effects of token emphasis on the time taken to make emphasised decisions was also analysed via 2(decision type) x 3(token emphasis) repeated-measures ANOVAs conducted on decision RT. The main effect of decision type was non-significant in both the semantic and recognition tasks, $F(1,28) = 1.44$, $p = .240$, $\eta_p^2 = .05$ and $F(1,28) = 3.60$, $p = .068$, $\eta_p^2 = .11$ respectively. However, significant main effects of emphasis were observed on decision RT in both tasks, $F(2,56) = 5.25$, $p = .008$, $\eta_p^2 = .16$ and $F(2,56) = 45.18$, $p < .001$, $\eta_p^2 = .62$ respectively. These token main effects were suggestive of an overall reduction in decision RT specifically under Token A emphasis conditions in each task. Comparisons of token condition means in the semantic task revealed a significant reduction in decision RT in the "pleasant" token condition ($M = 0.89$, $SD = 0.12$) relative to the No token ($M = 0.94$, $SD = 0.12$) but not the "unpleasant" token condition ($M = 0.92$, $SD = 0.11$), $F(1,28) = 10.40$, $p = .003$, $\eta_p^2 = .27$ and $F(1,28) = 2.38$, $p = .134$, $\eta_p^2 = .08$ respectively (the difference between "unpleasant" and No token decision RT was also non-significant, $p = .092$). Comparison of condition means in the recognition task demonstrated a significant reduction in decision RT under "old" token emphasis ($M = 0.90$, $SD = 0.10$) relative to both No token ($M = 1.03$, $SD = 0.11$) and "new" token conditions ($M = 0.98$, $SD = 0.11$), $F(1,28) = 64.01$, $p < .001$, $\eta_p^2 = .70$ and $F(1,28) = 45.27$, $p < .001$, $\eta_p^2 = .62$ respectively (the difference between "new" and No token decision RT was also significant, $F(1,28) = 14.71$, $p = .001$, $\eta_p^2 = .34$).

As before, the interaction of decision type and token emphasis was the primary index of the hypothesised effects of goal emphasis on decision RT. The interaction was significant in both the semantic and recognition tasks, $F(2,56) = 11.79$, $p < .001$, $\eta_p^2 = .30$ and $F(2,56) = 18.57$, $p < .001$, $\eta_p^2 = .40$ respectively. As with the earlier decision accuracy analyses, the interactions in both tasks reflected a more complex pattern than the anticipated reduction in Decision A and Decision B under concordant token emphasis (see Table 4.1.; see Figure 4.3c and 4.3d). Planned pairwise t-tests conducted across token emphasis conditions in the semantic task revealed that “pleasant” decision RT was significantly reduced under “pleasant” emphasis compared to both No token and “unpleasant” emphasis, $t(28) = 3.72$, $p = .001$, $d = 0.69$ and $t(28) = 4.05$, $p < .001$, $d = 0.76$ respectively (there was no difference between “unpleasant” and No token conditions, $p = .573$). Decision RT for “unpleasant” decisions was significantly reduced in the “unpleasant” token compared to the No token but not the “pleasant” token condition, $t(28) = 2.97$, $p = .006$, $d = 0.60$ and $t(28) = 1.47$, $p = .153$, $d = 0.30$ respectively (there was no difference between “pleasant” and No token conditions, $t(28) = 1.36$, $p = .185$, $d = 0.25$). For the recognition task, pairwise t-tests revealed a significant reduction in the time taken to make “old” decisions in the “old” token condition compared to both the No token and “new” token conditions, $t(28) = 6.79$, $p < .001$, $d = 1.27$ and $t(28) = 8.57$, $p < .001$, $d = 1.64$ respectively (there was no difference between “new” and No token conditions, $p = .863$). Further, whilst “new” decision RT was significantly reduced in the “new” token emphasis relative to the No token condition, it was also significantly reduced in the “old” token compared to the No token condition and to a numerically larger degree, $t(28) = 6.10$, $p < .001$, $d = 1.13$ and $(28) = 7.17$, $p < .001$, $d = 1.33$ respectively (the reduction in “new” decision RT in the “old” compared to the “new” token condition was non-significant, $t(28) = 1.76$, $p = .090$, $d = 0.33$).

Overall, the decision RT analyses complement the earlier decision accuracy analyses in demonstrating that the only reliable cross-domain effect of token emphasis was that imparted by the Token A condition to Decision A categories.

Token A emphasis served to reduce the time taken to endorse Decision A categories, consistent with the greater goal emphasis imparted to this decision category.

4.3.2.4. Cross-domain emphasis correlations. To provide further validation of the cross-domain persistence of the above token emphasis effects, a correlation analysis was conducted for token-driven shifts in SDT criterion bias in the semantic and recognition tasks. Emphasis criterion shift was calculated as criterion in the Token A emphasis condition minus criterion in the Token B emphasis condition, such that positive criterion shift values indicated the presence of a counter-emphasis shift in bias. The results revealed a significant correlation between emphasis criterion shifts in each task (emphasis c shift: semantic M = 0.27, SD = 0.43; recognition M = 0.23, SD = 0.42), thereby providing direct evidence (at the participant level) in support of a cross-domain control mechanism underlying the strategic effects of token emphasis on evaluative performance, r (27) = .44 (p = .017).

4.3.3. Interaction of emphasis and expectation (cued phase analyses)

4.3.3.1. Cued SDT sensitivity and bias. Means and standard deviations for all cued phase analyses are provided in Table 4.2. The effects of cues and token emphasis on SDT sensitivity were first investigated via 2(cue type: “likely A”, “likely B”) x 3(token emphasis: Token A, Token B, No token) repeated-measures ANOVAs conducted separately for each task. The main effect of cue type on *d′* was non-significant in both the semantic and recognition tasks, $F < 1$ and $F(1,28) = 1.08$, $p = .308$, $\eta_p^2 = .04$ respectively. The main effect of token emphasis was significant in both tasks, $F(2,56) = 3.79$, $p < .029$, $\eta_p^2 = .12$ and $F(2,56) = 16.52$, $p < .001$, $\eta_p^2 = .37$ respectively. Finally, the interaction of cue type and token emphasis was non-significant for both tasks, both *Fs* < 1. Token main effects in both tasks reflected the greater sensitivity elicited by the presence of some form of token emphasis. This was confirmed by comparison of the token condition means, which in the semantic task yielded significantly lower sensitivity in the No token ($M = 2.27$, $SD = 0.44$) compared to both the “pleasant” ($M = 2.45$, $SD = 0.36$) and “unpleasant” token conditions ($M = 2.49$, $SD = 0.30$), $F(1,28) = 4.30$, $p = .047$, $\eta_p^2 = .13$ and $F(1,28) = 6.44$, $p = .017$, $\eta_p^2 = .19$ respectively (the difference between “pleasant” and “unpleasant” token sensitivity was not significant, $F < 1$). Comparison of token condition means in the recognition task illustrated a significant reduction in sensitivity in the No token ($M = 1.88$, $SD = 0.52$) relative to both the “old” token ($M = 2.24$, $SD = 0.54$) and “new” token conditions ($M = 2.03$, $SD = 0.54$), $F(1,28) = 31.83$, $p < .001$, $\eta_p^2 = .53$ and $F(1,28) = 6.44$, $p = .017$, $\eta_p^2 = .19$ respectively. Recognition sensitivity was also significantly greater in the “old” compared to the “new” token condition, $F(1,28) = 10.14$, $p = .004$, $\eta_p^2 = .27$. Hence, the observed main effects of token emphasis on sensitivity broadly reflect the greater level of task engagement engendered by the opportunity to receive a monetary reward in the Token A and Token B task conditions. The additional observation of improved recognition sensitivity under "old" emphasis is comparable to effects reported in the uncued phase, and provides preliminary evidence of the recovery of analogous token emphasis effects in the presence and absence of cues.

The focal analysis of cue and token emphasis effects on SDT criterion was also conducted via 2(cue type) x 3(token emphasis) repeated-measures ANOVAs for *c* in each separate task (see Table 4.2.; see Figure 4.2c and 4.2d). The main effect of cue type was significant in the semantic and recognition tasks, $F(1,28) = 41.26$, $p < .001$, $\eta_p^2 = .60$ and $F(1,28) = 48.43$, $p < .001$, $\eta_p^2 = .63$ respectively. As expected, cue types shifted criterion in "pro-expectation" directions so as to increase the likelihood of endorsing the cued decision in both task domains (semantic task *c*: "likely pleasant" $M = -0.29$, $SD = 0.16$; "likely unpleasant" $M = 0.04$, $SD = 0.25$; recognition task *c*: "likely old" $M = -0.14$, $SD = 0.21$; "likely new" $M = 0.01$, $SD = 0.22$).

The main effect of token emphasis was also significant in the semantic and recognition tasks, $F(2,56) = 6.02$, $p = .004$, $\eta_p^2 = .18$ and $F(2,56) = 6.13$, $p = .004$, $\eta_p^2 = .18$ respectively. These token main effects reflected a goal-driven reduction in the likelihood of endorsing the emphasised decision. Planned comparisons of token condition means in the semantic task confirmed that criterion was significantly more positive in the "pleasant" emphasis ($M = -0.01$, $SD = 0.26$) relative to the "unpleasant" ($M = -0.17$, $SD = 0.25$) and No token emphasis conditions ($M = -0.20$, $SD = 0.22$), $F(1,28) = 5.25$, $p = .030$, $\eta_p^2 = .16$ and $F(1,28) = 11.9$, $p = .002$, $\eta_p^2 = .30$ respectively (the difference between "unpleasant" and No token criterion was not significant, $F < 1$). Comparisons in the recognition task also demonstrated a significantly more positive criterion in the “old” emphasis (*M* = 0.03, *SD* = 0.25) compared to the “new” (*M* = -0.09, *SD* = 0.24) and No token emphasis conditions (*M* = -0.13, *SD* = 0.25), $F(1,28) = 4.58$, $p = .041$, $\eta_p^2 = .14$ and $F(1,28) = 14.19$, $p = .001$, $\eta_p^2 = .34$ respectively (the difference between “new” and No token criterion was not significant, $F < 1$). The interaction of cue type and token emphasis was significant in the semantic but not the recognition task, $F(2,56) = 3.64$, $p = .033$, $\eta_p^2 = .12$ and $F(2,56) < 1$ respectively. Post-hoc t-tests demonstrated that the semantic task interaction was driven by stronger emphasis effects on criterion in the “Likely Unpleasant” compared to the “Likely Pleasant” cue condition (see Figure 4.2c)[7].

Hence, SDT analyses of the cued phase recovered virtually identical effects of explicit emphasis on cross-domain bias to those reported in the uncued phase, namely a bias shift that reduced the likelihood of endorsing decisions emphasised by the available incentives. The “counter-emphasis” direction of these token biases acted independently and in opposite directions to the numerically larger “pro-expectation” biases instilled by presented cue types. The findings therefore provide clear evidence that independent and opposing strategic effects of emphasis and expectation are recovered even when both environmental factors are conjunctly manipulated.

[7] Pairwise comparisons for token emphasis effects on semantic criterion were conducted separately for the “Likely Pleasant” cue condition (“pleasant” > No token, $t(28) = 1.26$, $p = .219$, $d = 0.23$; “pleasant” > “unpleasant” token, $t(28) = 1.55$, $p = .132$, $d = 0.29$; No token > “unpleasant”, $p = .605$) and the “Likely Unpleasant” condition (“pleasant” > No token, $t(28) = 4.73$, $p < .001$, $d = 0.88$; “pleasant” > “unpleasant” token, $t(28) = 2.56$, $p = .016$, $d = 0.48$; “unpleasant” > No token, $t(28) = 1.58$, $p = .125$, $d = 0.29$).

Table 4.2. Design key and descriptive statistics for the cued analyses, including SDT sensitivity (*d′*) and bias (*c*), decision accuracy (DecisionA$_{corr}$, DecisionB$_{corr}$) and decision RT measures (DecisionA$_{RT}$, DecisionB$_{RT}$). Cue types are described using task-specific and cross-task descriptors (parenthetic "Cue A" and "Cue B"; see Figure 4.1.). Token emphasis conditions are presented using task-specific descriptors only: P = "pleasant" emphasis, U = "unpleasant" emphasis, O = "old" emphasis and N = "new" emphasis.

Task		**Semantic**						**Recognition**					
Cue type		Likely Pleasant (Cue A)			Likely Unpleasant (Cue B)			Likely Old (Cue A)			Likely New (Cue B)		
Token emphasis		P	U	no	P	U	no	O	N	no	O	N	no
SDT *d′*	*M*	2.45	2.50	2.29	2.44	2.48	2.24	2.22	1.97	1.88	2.27	2.08	1.88
	SD	0.42	0.38	0.46	0.41	0.47	0.57	0.70	0.55	0.53	0.55	0.61	0.59
SDT *c*	*M*	-0.22	-0.34	-0.31	0.20	0.00	-0.09	-0.05	-0.15	-0.21	0.11	-0.03	-0.05
	SD	0.28	0.27	0.29	0.30	0.33	0.33	0.26	0.28	0.28	0.28	0.28	0.28
DecisionA$_{corr}$	*M*	.92	.91	.90	.85	.81	.75	.92	.89	.88	.80	.73	.70
	SD	.03	.04	.04	.07	.10	.11	.06	.06	.06	.12	.12	.13
DecisionB$_{corr}$	*M*	.85	.89	.84	.92	.94	.93	.77	.75	.75	.92	.92	.91
	SD	.09	.06	.09	.04	.04	.05	.11	.12	.12	.04	.05	.06
DecisionA$_{RT}$	*M*	0.85	0.91	0.94	0.92	0.99	1.04	0.86	0.96	1.00	0.93	1.00	1.05
	SD	0.10	0.10	0.11	0.13	0.12	0.11	0.09	0.10	0.11	0.11	0.10	0.11
DecisionB$_{RT}$	*M*	0.95	0.94	1.00	0.89	0.87	0.96	0.98	1.01	1.11	0.95	0.96	1.06
	SD	0.11	0.11	0.13	0.11	0.11	0.11	0.12	0.11	0.11	0.11	0.11	0.11

Note: *M*, mean; *SD*, standard deviation.

4.3.3.2. Cued decision accuracy. The effects of cued expectation and token emphasis on decision accuracy were also interrogated via 2 (cue type) x 3 (token emphasis) repeated-measures ANOVAs, which were conducted separately for decision categories within each task domain for reasons of clarity (i.e. $DecisionA_{corr}$ and $DecisionB_{corr}$; see Table 4.2.; see Figure 4.4.). For the semantic task, significant main effects of cue type were observed for both “pleasant” and “unpleasant” decisions, $F(1,28) = 96.45$, $p < .001$, $\eta_p^2 = .78$ and $F(1,28) = 57.85$, $p < .001$, $\eta_p^2 = .67$ respectively. Both cue type effects reflected increased decision accuracy under concordant cueing, such that the accuracy of “pleasant” decisions was improved under “Likely Pleasant” ($M = .91$, $SD = .03$) compared to “Likely Unpleasant” cue conditions ($M = .80$, $SD = .07$), whereas “unpleasant” decision accuracy was improved under “Likely Unpleasant” ($M = .93$, $SD = .03$) compared to “Likely Pleasant” cue conditions ($M = .86$, $SD = .05$). Significant main effects of cue type were also observed on the accuracy of “old” and “new” decisions in the recognition task, $F(1,28) = 151.74$, $p < .001$, $\eta_p^2 = .84$ and $F(1,28) = 169.80$, $p < .001$, $\eta_p^2 = .86$ respectively. Cue type effects once again reflected an accuracy improvement under concordant cueing, with the accuracy of “old” decisions being improved under “Likely Old” ($M = .90$, $SD = .05$) compared to “Likely New” ($M = .74$, $SD = .11$) cue conditions, whereas “new” decision accuracy showed the inverse pattern (“Likely New” $M = .92$, $SD = .04$; “Likely Old” $M = .76$, $SD = .10$). These results provide evidence that participants were utilising the cues strategically to improve the accuracy of their evaluations.

The 2 x 3 ANOVA for the semantic task recovered a significant main effect of token emphasis on “pleasant” decision accuracy but not “unpleasant” decision accuracy, $F(2,56) = 11.40$, $p < .001$, $\eta_p^2 = .29$ and $F(2,56) = 2.44$, $p = .097$, $\eta_p^2 = .08$

respectively. Emphasis effects on decision accuracy in the cued semantic task therefore paralleled equivalent analyses of the uncued phase, in illustrating a more complex pattern of emphasis-driven adjustments to task performance (see Figure 4.4a and 4.4b). Comparisons of the condition means demonstrated that the effect of token emphasis on “pleasant” decision accuracy was driven by a significant accuracy improvement in the “pleasant” (M = .89, SD = .04) compared to the No token (M = .83, SD = .06) but not “unpleasant” emphasis condition (M = .86, SD = .06), $F(1, 28) = 25.05$, $p < .001$, $\eta_p^2 = .47$ and $F(1,28) = 3.36$, $p = .078$, $\eta_p^2 = .11$ respectively (the difference between “unpleasant” and No token conditions was also significant ($F(1, 28) = 8.56$, $p = .007$, $\eta_p^2 = .23$). Further, the cue by emphasis interaction was significant for “pleasant” but not “unpleasant” decision accuracy, $F(2,56) = 10.19$, $p < .001$, $\eta_p^2 = .27$ and $F(2,56) = 1.51$, $p = .230$, $\eta_p^2 = .05$ respectively. Post-hoc t-tests demonstrated that the “pleasant” decision accuracy interaction emerged from the stronger emphasis effects in the “Likely Unpleasant” compared to the “Likely Pleasant” cue condition (in which performance levels were at ceiling; see Figure 4.4a)[8].

A similar pattern was observed in the recognition task, with the 2 x 3 ANOVA yielding a significant main effect of token emphasis on “old” but not “new” decision accuracy, $F(2,56) = 19.78$, $p < .001$, $\eta_p^2 = .41$ and $F(2,56) = 1.27$, $p = .289$, $\eta_p^2 = .04$ respectively. Comparisons of the condition means illustrated that the effect of emphasis on “old” decision accuracy was driven by a significant accuracy improvement in the “old” (M = .86, SD = .09) compared to both the No token (M =

[8] Pairwise comparisons for emphasis effects on “pleasant” decision accuracy were conducted separately for the “Likely Pleasant” cue condition (“pleasant” > No token, t(28) = 2.39, p = .024, d = 0.45; “pleasant” > “unpleasant” token, t(28) = 1.00, p = .324, d = 0.19; “unpleasant” > No token, t(28) = 1.13, p = .268, d = 0.21) and the “Likely Unpleasant” condition (“pleasant” > No token, t(28) = 5.17, p < .001, d = 1.10; “pleasant” > “unpleasant” token, t(28) = 1.97, p = .059, d = 0.37; “unpleasant” > No token, , t(28) = 3.02, p = .005, d = 0.42).

.79, *SD* = .09) and "new" emphasis conditions (*M* = .81, *SD* = .09), $F(1, 28) = 44.71$, $p < .001$, $\eta_p^2 = .62$ and $F(1, 28) = 11.70$, $p = .002$, $\eta_p^2 = .30$ respectively (the difference between "new" and No token conditions was also significant, $F(1, 28) = 6.25$, $p = .019$, $\eta_p^2 = .18$). Further, the cue type by token emphasis interaction was significant for "old" but not "new" decisions, $F(2,56) = 7.35$, $p = .001$, $\eta_p^2 = .21$ and $F< 1$ respectively. As with the semantic decision accuracy analyses, post-hoc t-tests revealed that the interaction for "old" decision accuracy resulted from stronger token emphasis effects in the "Likely New" compared to the "Likely Old" cue condition (in which "old" decision accuracy levels were at ceiling; see Figure 4.4c)[9].

Overall, the decision accuracy analyses of the cued phases suggest that the only reliable improvement in emphasised decision accuracy occurred as a result of Token A emphasis imparted to Decision A categories in each task. The unilateral nature of these effects provides a direct parallel to the earlier uncued decision accuracy effects, and therefore serves to once again highlight that equivalent cross-domain goal emphasis effects were recovered in the presence of cued expectations.

[9] Pairwise comparisons for emphasis effects on "old" decision accuracy were conducted separately for the "Likely Old" cue condition ("old" > No token, t(28) = 3.55, p = .001, d = 0.66; "old" > "new" token, t(28) = 2.17, p = .038, d = 0.41; "new" > No token, t(28) = 1.24, p = .225, d = 0.23) and the "Likely New" condition ("old" > No token, t(28) = 5.92, p < .001, d = 1.11; "old" > "new" token, t(28) = 3.50, p = .002, d = 0.65; "new" > No token, t(28) = 2.28, p = .030, d = 0.43).

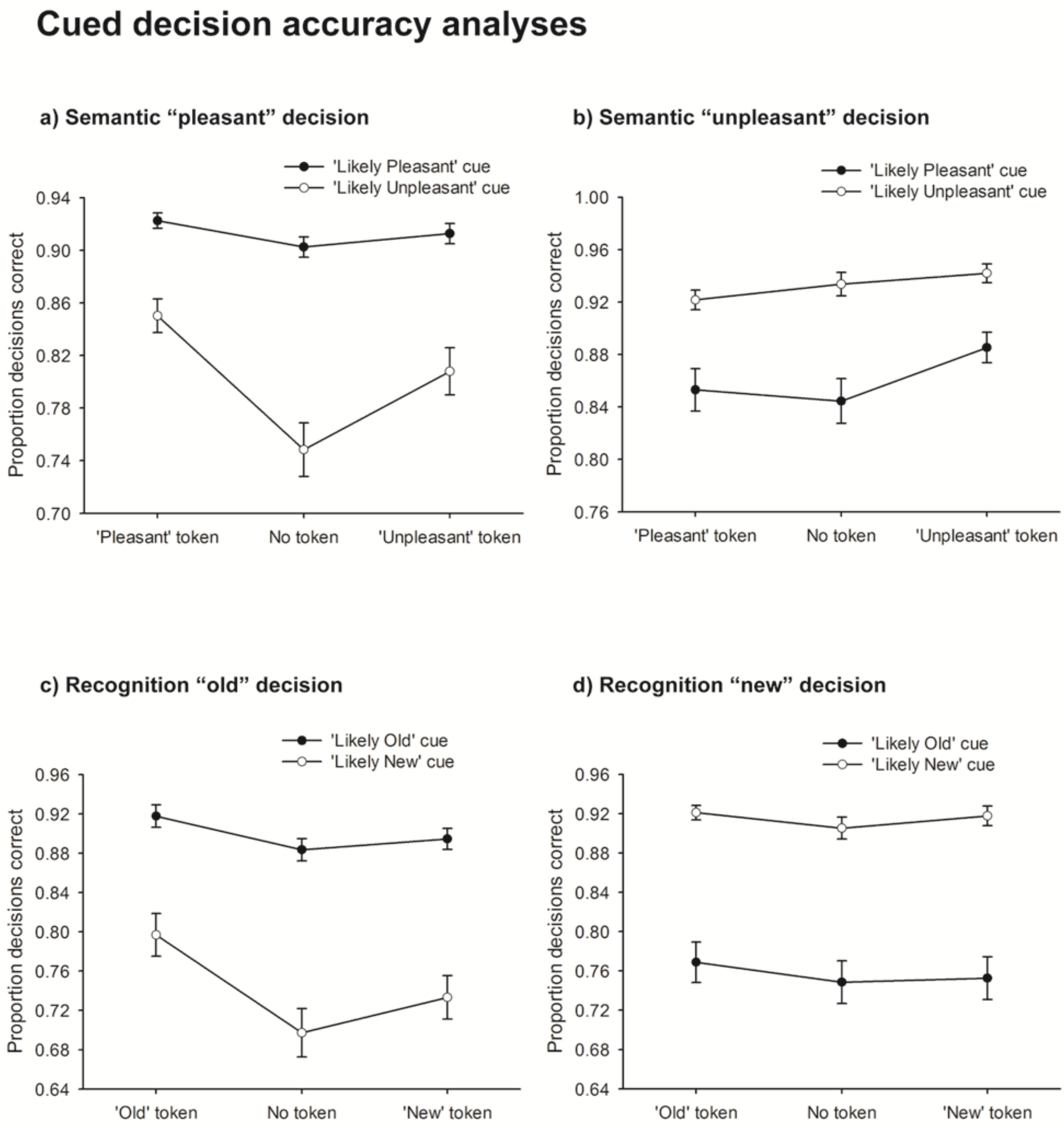


Figure 4.4. Effects of cues and token emphasis on decision accuracy in the cued phase. Top panel depicts decision accuracy effects in the semantic task, with separate plots for **a)** “pleasant” and **b)** “unpleasant” decision accuracy analyses. Bottom panel depicts decision accuracy effects in the recognition task, with separate plots for **c)** “old” and **d)** “new” decision accuracy analyses. Separate lines in each plot denote cue type conditions for the relevant task. Error bars represent standard error of the mean.

4.3.3.3. Cued decision RT. 2(cue type) x 3(token emphasis) repeated-measures ANOVAs were also conducted on the time taken to endorse decision categories in each separate task (i.e. $DecisionA_{RT}$ and $DecisionB_{RT}$ respectively; see Table 4.2.; see Figure 4.5.). For the semantic task, significant main effects of cue type were observed for both “pleasant” and “unpleasant” decision RT, $F(1,28) = 87.97$, $p < .001$, $\eta_p^2 = .76$ and $F(1,28) = 26.68$, $p < .001$, $\eta_p^2 = .49$ respectively. Both cue type effects signified a reduction in the time taken to make cued decisions, such that “pleasant” decision RT was reduced in the “Likely Pleasant” ($M = 0.90$, $SD = 0.10$) compared to “Likely Unpleasant” cue condition ($M = 0.99$, $SD = 0.11$), whereas “unpleasant” decision RT showed the inverse pattern (“Likely Unpleasant” $M = 0.91$, $SD = 0.10$; “Likely Pleasant” $M = 0.96$, $SD = 0.10$). Comparable main effects of cue type were also observed for “old” and “new” decision RT in the recognition task, $F(1,28) = 70.98$, $p < .001$, $\eta_p^2 = .72$ and $F(1,28) = 17.86$, $p < .001$, $\eta_p^2 = .39$ respectively. Recognition cue type effects once again reflected a reduction in decision RT under concordant cueing, with “old” decision RT being reduced in the “Likely Old” ($M = 0.94$, $SD = 0.09$) compared to the “Likely New” cue condition ($M = 1.00$, $SD = 0.09$), and “new” decision RT showing the inverse pattern (“Likely New” $M = 0.99$, $SD = 0.10$; “Likely Old” $M = 1.03$, $SD = 0.10$). These reaction time effects complement the earlier cued decision accuracy analyses in highlight the common strategic influence of cued expectations in both task domains.

The main effect of token emphasis in the semantic task was significant for both “pleasant” and “unpleasant” decision RT, $F(2,56) = 35.82$, $p < .001$, $\eta_p^2 = .56$ and $F(2,56) = 16.76$, $p < .001$, $\eta_p^2 = .37$ respectively. As in the uncued decision RT analyses, token main effects on cued decision RT were more complex than the anticipated reduction in the time taken to endorse emphasised decisions (see Figure

4.5a and 4.5b). Comparisons of the condition means demonstrated a reduction in "pleasant" decision RT in the "pleasant" token (M = 0.89, SD = 0.11) compared to both the "unpleasant" (M = 0.95, SD = 0.11) and No token emphasis conditions (M = 0.99, SD = 0.11), $F(1,28) = 39.05$, $p < .001$, $\eta_p^2 = .58$ and $F(1,28) = 55.97$, $p < .001$, $\eta_p^2 = .67$ respectively ("pleasant" decision RT was also significantly reduced in the "unpleasant" compared to the No token condition, $F(1,28) = 9.12$, $p = .005$, $\eta_p^2 = .25$). "Unpleasant" decision RT was significantly reduced in the "unpleasant" token (M = 0.90, SD = 0.11) compared to the No token (M = 0.98, SD = 0.11) but not the "pleasant" token condition (M = 0.92, SD = 0.10), $F(1,28) = 32.38$, $p < .001$, $\eta_p^2 = .54$ and $F(1,28) = 1.42$, $p = .243$, $\eta_p^2 = .05$ respectively (the reduction in "unpleasant" decision RT for "pleasant" compared to No token emphasis was also significant, $F(1,28) = 19.08$, $p < .001$, $\eta_p^2 = .41$). Finally, the cue type by token emphasis interaction was non-significant for both "pleasant" and "unpleasant" decision RT, $F(2,56) = 1.95$, $p = .151$, $\eta_p^2 = .07$ and $F(2,56) = 2.43$, $p = .097$, $\eta_p^2 = .08$ respectively.

A similar pattern was observed in the recognition task, with the 2 x 3 ANOVA yielding main effects of token emphasis on both "old" and "new" decision RT, $F(2,56) = 47.04$, $p < .001$, $\eta_p^2 = .63$ and $F(2,56) = 42.52$, $p < .001$, $\eta_p^2 = .60$ respectively. Token effects on RT for both recognition decision categories were seemingly driven by an overarching reduction in decision RT under "old" token emphasis (irrespective of decision type; see Figure 5c and 5d). Comparisons of the token condition means revealed that "old" decision RT was significantly reduced under "old" token (M = 0.89, SD = 0.10) compared to "new" (M = 0.98, SD = 0.10) and No token emphasis (M = 1.03, SD = 0.11), $F(1,28) = 37.46$, $p < .001$, $\eta_p^2 = .57$ and $F(1,28) = 73.07$, $p < .001$, $\eta_p^2 = .72$ respectively (the reduction in "old" decision RT for "new" compared to

No token emphasis was also significant, $F(1,28) = 15.21$, $p = .001$, $\eta_p^2 = .35$). Whilst "new" decision RT was significantly reduced in the "new" token ($M = 0.99$, $SD = 0.10$) compared to the No token condition ($M = 1.09$, $SD = 0.10$), it was also reduced in the "old" token condition ($M = 0.97$, $SD = 0.11$) and to a numerically larger degree, $F(1,28) = 41.99$, $p < .001$, $\eta_p^2 = .60$ and $F(1,28) = 65.35$, $p < .001$, $\eta_p^2 = .70$ respectively ("new" decision RT did not differ significantly between the "old" and "new" token conditions, $F(1,28) = 2.79$, $p = .106$, $\eta_p^2 = .09$). Finally, the cue type by token emphasis interaction was non-significant for both "old" and "new" decision RT, $F(2,56) = 1.91$, $p = .157$, $\eta_p^2 = .06$ and $F < 1$ respectively.

Overall, the cued decision RT analyses parallel the equivalent uncued analyses in suggesting that the only reliable cross-domain effect of token emphasis on decision RT emerged from Token A emphasis reducing the time taken to endorse Decision A categories in both task domains. RT effects for the alternate Decision B categories broadly reflected a decrease in reaction time whenever the opportunity to win monetary tokens was present, with equivalent reduction in Decision B_{RT} in the Token A and Token B conditions compared to the No token condition. Although the pattern of results is more complex than the predicted crossover reductions in decision RT under congruent decision emphasis, the fact that this pattern was recovered in both task domains, as well as across uncued and cued phases, highlights their reliability.

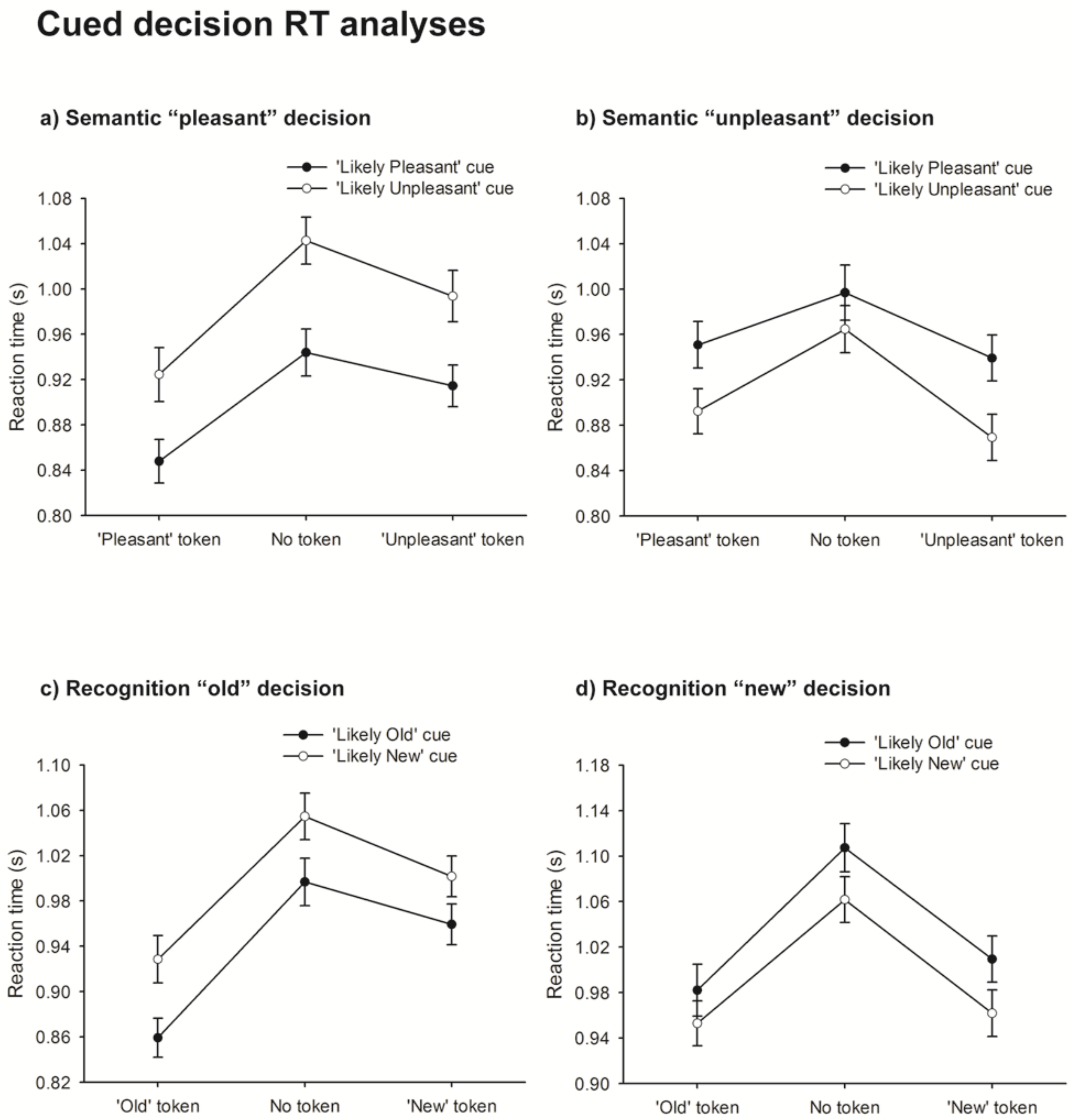


Figure 4.5. Effects of cues and token emphasis on decision RT in the cued phase. Top panel depicts decision RT effects in the semantic task, with separate plots for **a)** “pleasant” and **b)** “unpleasant” decision RT analyses. Bottom panel depicts decision RT effects in the recognition task, with separate plots for **c)** “old” and **d)** “new” decision RT analyses. Separate lines in each plot denote cue type conditions for the relevant task. Error bars represent standard error of the mean.

4.3.3.4. Cross-domain cue and emphasis correlations. A series of correlation analyses were conducted to directly test the cross-domain persistence of cued phase effects at the participant level. These analyses centred on two difference measures calculated for SDT criterion values in each task domain: “cue criterion shift” and “emphasis criterion shift”. Cue criterion shift was an index of the tendency for participants to adjust their bias in accordance with environmentally-cued expectations, and was calculated as criterion in the “Likely B” condition minus criterion in the “Likely A” condition (such that positive cue criterion shift values indicated an increased tendency to adhere to cued expectations). Emphasis criterion shift served as an index of the tendency to adjust bias in accordance with environmentally-emphasised goals, and was calculated as criterion in the “Token A” condition minus criterion in the “Token B” condition (with positive emphasis criterion shift values reflecting an increased adherence to emphasised goals, as in the uncued cross-task analyses).

The results demonstrated a non-significant positive correlation between cue criterion shifts in the semantic (M = 0.32, SD = 0.28) and recognition tasks (M = 0.14, SD = 0.12), r (27) = .24 (p = .239). However, a significant positive relationship was observed between emphasis criterion shifts in the cued semantic (M = 0.15, SD = 0.39) and recognition tasks (M = 0.12, SD = 0.30), r (27) = .60 (p = .001). To probe the correspondence of this positive token relationship with the analogous positive token relationship described in the uncued cross-task analyses (see Section 3.1.4.), emphasis criterion shifts for tasks in both phases were also correlated. The emphasis criterion shift in the uncued semantic task was significantly positively correlated with that observed in the cued recognition task and non-significantly (albeit marginally so) positively correlated with the emphasis shift in the cued semantic task, r (27) = .40 (p = .031) and r (27) = .36 (p = .055) respectively. The emphasis criterion shift in the uncued recognition task was significantly positively correlated with emphasis criterion shifts in both the cued semantic and cued recognition tasks, r (27) = .48 (p = .009) and r (27) = .61 (p < .001) respectively.

Hence, whilst a stable relationship between expectation-driven bias shifts across different task domains was not observed, reliable positive relationships between emphasis-driven bias shifts were recovered in both cued and uncued experimental phases. The results of the cross-task analyses suggest that the tendency to adjust performance in light of varying goal emphases in one task domain correlates with the tendency to effectuate similar adjustments in other task domains, and this positive relationship persists in the presence and absence of cued expectations. The findings therefore serve to corroborate, at the participant level, the assertion that goal-related information is processed by a unified strategic control system, capable of influencing evaluative performance in different cognitive domains.

4.3.4. Individual differences in emphasis and expectation biases.

A final series of correlation analyses were conducted to assess the relationship between task-evoked tendencies to adopt controlled evaluative strategies and psychometric measures of personality. These measures were taken from the BIS/BAS (Carver & White, 1994), Big 5 (John et al., 1991) and Brief Sensation-Seeking scales (B-SS; Hoyle et al., 2002), and were correlated with emphasis criterion shifts in the uncued and cued phase, as well as the cue criterion shift in the cued phase (as described in the previous section). No significant correlations were found between uncued emphasis criterion shifts (in either task domain) and any of the elicited psychometric measures. Psychometric associations with strategic shifts in the cued phase were also sparse, with only a positive relationship between emphasis criterion shift in the semantic task and the BAS fun-seeking component of the BIS/BAS (r (27) = .42, p = .023), as well as a positive relationship between cue criterion shift in the semantic task and the boredom susceptibility component of the B-SS observed (r (27) = .47, p = .009). The association of BAS fun-seeking and an increased propensity to be biased by token emphasis is plausible given that this component attempts to capture the willingness to engage in novel and rewarding experiences, of which maximising the number of tokens earned in the present task is likely a constituent (Carver & White, 1994). The relationship between boredom susceptibility and an increased tendency to be biased by cued expectation is also somewhat plausible, given that this component attempts to capture individual differences in the tendency to experience boredom (Hoyle et al., 2002). The experience of boredom is a likely consequence of the lengthy and repetitive nature of the presented task, which might sequentially drive a greater reliance on the static cue validities as a basis for responding, rather than the variable evidence strength associated with each test item.

However, the fact that none of the above relationships generalized to corresponding criterion measures in the recognition task (or in the case of the emphasis criterion shifts, even to corresponding measures in uncued task phases) casts doubt over their reliability. As such, the overall results of the psychometric analyses do not reveal any stable relationships between personality traits and the cross-domain control processes under focal scrutiny.

4.4. Discussion

The present experiment interrogated the influence of two major environmental strategies (goal emphasis and cued expectation) on evaluative processing in two different cognitive domains (semantic memory and episodic recognition memory). Goal emphasis was manipulated via the application of mixed token incentives with one of two task decision categories, whereas cued expectation was manipulated by the provision of trial-by-trial textual cues that were mostly valid in predicting the status of ensuing task items. Analyses of the uncued phase, in which token incentives were manipulated in isolation, recovered reliable shifts in signal detection estimates of bias acting in "counter-emphasis" directions, such that the likelihood of endorsing the decision category emphasised by the available incentives was reduced. The strategic underpinnings of these bias effects were highlighted by complementary findings of greater accuracy and reduced reaction time for decision categories endorsed under concordant emphasis. The results also reveal particularly pronounced emphasis-driven adjustments in behaviour for "Decision A" categories under concordant "Token A" emphasis.

Analyses of the cued phase, in which both explicit token incentives and cued expectations were manipulated, successfully recovered anticipated bias shifts in "pro-expectation" directions, such that the likelihood of endorsing the cued decision category was increased. Notably, the cued phase also recovered identical effects of token incentives to those reported in the uncued phase, namely bias shifts that reduced the likelihood of endorsing emphasised decisions, as complemented by analogous effects on decision accuracy and decision RT. The cued analyses therefore successfully isolated independent strategic influences of expectation and goal emphasis on behavioural measures of task performance. Finally, strategic effects on all elicited measures analysed in both uncued and cued task phases were observed to generalise across the presented task domains. These cross-domain effects were further corroborated by a series of cross-task correlations, which recovered stable positive relationships between goal emphasis-driven bias shifts (but not expectation-driven bias shifts) across task domains. The findings provide behavioural evidence in support of a cross-domain cognitive control system.

The present goal emphasis manipulations yielded findings comparable to prior research involving mixed incentives, in which counter-emphasis shifts were observable in raw forms of both primary and secondary response measures (Han et al., 2010; Hartung et al., 2002; Newman et al., 1985). However, the present experiment is the first to explicitly link mixed incentives with strategic adjustments to a formalized measure of bias, namely response criterion as estimated from the signal detection model. As such, the counter-emphasis direction of the observed bias also serves to differentiate the strategic influence of goal emphasis from that related to more widely studied “reward only” incentive manipulations, which yield bias in a converse confirmatory direction (Healy & Kubovy, 1978; Van Zandt, 2000). The strategic underpinnings of the counter-emphasis bias were further elucidated by complementary effects on decision accuracy, suggesting that participants reduced endorsements of the emphasised decision so as to maximise their accuracy. The bias and decision accuracy results parallel the findings of the recognition experiments of Chapter 2, and strengthen interpretation of the question format effects as resulting from implicit forms of the same goal-driven mechanism engaged by mixed incentives. The decision RT analyses provided additional insight into how the counter-emphasis bias actively benefitted task performance, by demonstrating quicker reaction times when decision categories were endorsed under emphasis (similar to prior reports, e.g. Han et al., 2010). Collectively, the findings of Chapter 2 and the present chapter serve to elucidate goal emphasis as an overlooked source of strategic control that is reliably indexed by estimates of criterion bias.

The present study also revealed asymmetries in how goal emphasis impacted on the evaluation of specific decision categories. This added insight was afforded by the inclusion of the baseline “No token” emphasis condition, in which mixed incentives were entirely absent and performance was consequently assumed to be relatively emphasis-neutral. In both presented task domains, deviations in performance from the baseline No token condition primarily emerged for “Decision A” categories under concordant emphasis (i.e. in the Token A condition), with weaker emphasis-driven adjustments observed for “Decision B” categories under emphasis (i.e. in the Token B condition). Specifically, emphasis shifted criterion bias most prominently in the Token A condition (i.e. in a positive direction) and affected decision category measures in a similar unilateral fashion (i.e. the only reliable improvement in decision accuracy and decision RT occurred for endorsements of “Decision A” in the Token A condition).

The observed “Decision A” asymmetry is somewhat unexpected, although the fact that it manifested across both task domains implies the involvement of a cross-domain rather than domain-specific factor. One such factor is the ordering of the presented response format, which was held constant across task domains and phases as “Decision A” assigned to the “yes” response key and “Decision B” assigned to the “no” number key. Goal emphasis imparted by the Token A emphasis condition might therefore have combined with emphasis imparted by the fixed order of the response format might to produce stronger overall emphasis effects on

“Decision A” performance. Although speculative, this “response emphasis” interpretation of the observed “Decision A” asymmetry is plausible, given both evidence of response format emphasis provided by Experiment 3 (Chapter 2, Section 2.4.2.) and previously reported biases resulting from similarly subtle aspects of test response formats (Lindsay & Johnson, 1989; Marsh & Hicks, 1998). The General Discussion chapter examines this “Decision A” preferential emphasis finding in greater detail (Chapter 7, Section 7.3.).

The present experiment also interrogated the interaction of conjunct manipulations of goal emphasis and cued expectation. Pronounced strategic effects of cued expectation were recovered, comprising “pro-expectation” shifts in criterion bias, which served to increase endorsement of the cued decision, as complemented by greater accuracy and reduced RT for endorsements of the cued decision. These cue-specific effects are consistent with prior research examining expectancy modulations in a number of isolated cognitive domains (O’Connor et al., 2010; Posner et al., 1980; Squires et al., 1975), yet are demonstrated in the present study to manifest equivalently for the same sample of participants across different cognitive domains. More notably, analyses of the cued phase also recovered analogous strategic effects of goal emphasis as those evidenced in the uncued phase. This finding is most clearly highlighted by criterion bias effects in the cued phase, which shifted in independent and opposing directions in accordance with the confirmatory “pro-expectation” and disconfirmatory “counter-emphasis” biases instilled by the respective manipulations. The results therefore provide compelling evidence for the functional role of response criterion as a behavioural index of multiple sources of strategic bias, as has been conjectured previously in the fields of psychophysics (Swets, 2014) and recognition memory (Benjamin, 2007).

A notable feature of the observed strategic effects of both goal emphasis and cued expectation manipulations is their persistence across semantic and recognition task domains. This cross-domain persistence is particularly notable given the reported differences in overall sensitivity across the two task domains (see Section 4.3.1.), which has previously been highlighted as an impediment to the recovery of cross-domain strategic processes (Fleming & Dolan, 2012). Indeed, variation in overall sensitivity across the tasks presented in Experiment 4 (Chapter 3) putatively served as a major cause of the broad lack of cross-domain effects following the implicit emphasis manipulations instantiated. The present findings reveal that more explicit manipulations of the test environment are capable of overcoming underlying variation in task domain sensitivities to successfully recruit cross-domain control processes. The results also extend a recent demonstration of cross-domain effects on expectation on bias parameters of the drift diffusion model (White & Poldrack, 2014) with evidence that cross-domain influences of *both* expectation and goal emphasis are indexed by signal detection estimates of response bias. Criterion bias might therefore serve as an alternative behavioural index of cross-domain control – one that is less computationally intensive than its diffusion model analogue.

The cross-domain properties of the observed strategic effects were further examined via correlations of emphasis- and expectation-driven bias shifts elicited in each task domain. The results revealed a strong positive relationship between the tendency to adopt the counter-emphasis bias in the semantic and recognition task domains. Counter-emphasis bias shifts were also positively correlated across uncued and cued versions of each task domain, thereby highlighting that the integration of goal emphasis into ongoing evaluative processing conforms to a stable individual trait. However, correlations of the pro-expectation bias revealed only modest positive relationships across semantic and recognition task domains. Similarly weak relationships also emerged from the correlations of counter-emphasis and pro-expectation bias shifts with psychometric measures of personality. Further research involving a larger sample and psychometric battery will therefore be necessary to test the reliability of the present individual difference relationships.

Overall, the present experimental chapter provides behavioural evidence in support of a cross-domain cognitive control system that has been widely conjectured yet rarely empirically validated (Cabeza et al., 2008; Corbetta & Shulman, 2002; Fleming & Dolan, 2012). The findings from the first three experimental chapters collectively emphasise the need for explicit manipulation of environmental factors in successfully isolating cross-domain control processes. These insights were incorporated in the next chapter, in which I interrogated the neural underpinnings of the behaviourally evidenced cross-domain control system via the application of functional neuroimaging methods.

Chapter 5: A multi-modal neuroimaging approach to the study of cross-task control

5.1. Introduction to Experiment 6

The previous chapters aimed to provide behavioural evidence of a unified cognitive system that is engaged whenever evaluation of a stimulus requires greater control, irrespective of the specific domain in which that controlled evaluation takes place. With a particular focus on episodic memory, the conducted experiments sought to isolate processes that likely perform a “general” cross-domain control function from those that are specific to the episodic domain. The findings presented thus far suggest that signal detection estimates of response criterion (as supplemented by measures of decision accuracy and RT) might serve as behavioural correlates of cross-domain control processes, recruited both when stimulus evaluation is conducted under explicated goals or reward, as well as in the presence of expectations cued by the immediate environment. However, these common behavioural effects might nevertheless arise from distinct neural computations, and direct functional neuroimaging investigation is required to elucidate the emergence of cross-domain control function in the brain. The present chapter targeted this aim, with a related focus on clarifying the precise functional properties of brain activations identified in previous studies of episodic memory, spanning EEG and fMRI neuroimaging modalities.

In research involving the EEG modality, prior studies have examined event-related potentials (ERPs) associated with the “old/new” effect – voltage fluctuations emerging at frontal and parietal electrode sites that distinguish correct “old” decisions

(“Hits”) from correct “new” ones (“CRs”; Sanquist, Rohrbaugh, Syndulko & Lindsley, 1980; Warren, 1980). Subsequent research identified two distinct sub-components of the old/new effect – an earlier negative deflection with a peak at around 400 ms and a frontal-parietal scalp maximum (N400), and a later positive deflection with a peak at approximately 600 ms (overlapping into an ensuing slow-wave component) and a central-parietal maximum (P600, also termed the “late positive” component or LPC; Rugg, Mark, Schloerscheidt, Birch & Allan, 1998). Recognition-specific interpretations have related these two components with the dual processes of familiarity and recollection respectively, on the basis of selective effects on these ERPs following manipulations designed to tap one or the other process (Rugg & Curran, 2007; Rugg et al., 1998; see Section 1.5. for further details).

However, this recognition-specific account has been challenged by findings linking the constituent recognition ERP components with more general “decision” factors. The proposed link between the recognition LPC and recollection has been subject to particular scrutiny. For example, Karayanidis and colleagues (Karayanidis, Andrews, Ward & McConaghy, 1991) conducted a recognition task in which the interval between the encoding and retrieval episodes was varied, under the assumption that shorter intervals would lead to stronger encoding. The latency of the N400 component was reduced for the shorter intervals, consistent with a direct relationship between this component and retrieval processes influenced by the interval manipulation. Conversely, the morphology of the LPC was unaffected by the shorter interval, which merely led to the quicker onset of the LPC as a corollary of the reduced latency of the earlier retrieval-related N400.

Further evidence of a “post-retrieval” role for the LPC was provided by Finnigan and colleagues (Finnigan, Humphreys, Dennis & Geffen, 2002), in an experiment manipulating the memory strength of encoded items via study repetition (i.e. “strong” *old* items presented 3 times at study, “weak” *old* items presented once). Whilst the N400 was found to track the memory strength manipulation in a linear fashion (with component positivity being greatest for “strong” Hits and lowest for CRs), no such clear recognition-specific relationship was observed for the LPC. Rather, LPC amplitude was found to increase for correct endorsements of both “old” and “new” decision categories, prompting the authors to re-characterise the LPC as an index of controlled processes engaged post-retrieval to maximise accuracy. This view is supported by prior evidence of a link between LPC positivity and greater old/new decision confidence (Rubin, van Petten, Glisky & Newberg, 1999). The overall pattern of findings therefore suggests a more “general” control function for the late positive component of the old/new ERP effect – one that potentially extends beyond the recognition domain.

Indirect support for a control interpretation of the recognition LPC comes from ERP research in a different cognitive domain, centring on a morphologically and topographically similar component - the syntactic P600 elicited in studies of semantic memory (Coulson, King & Kutas, 1998; Osterhout & Holcomb, 1992). Early studies reported a positive voltage deflection, peaking at around 600 ms and emerging at central-parietal scalp locations, that was enhanced following the detection of syntax violations in grammatically incorrect sentences in comparison to grammatically correct controls. A number of syntactic violations have been found to modulate this LPC component, including garden path sentences (e.g. “The broker persuaded *to sell the stock[10]”; Osterhout & Holcomb, 1992) and subject-verb disagreement (e.g. “The plane took *we to paradise and back’; Hagoort, Brown & Groothusen, 1993),

[10] The asterisks in these sentence examples denote the “critical words” that generate two forms of syntactic violation.

which encouraged a syntax-specific interpretation of the LPC as a dedicated sentence parsing mechanism (Osterhout & Holcomb, 1992).

However, as with the recognition LPC, challenges to the syntactic LPC's proposed domain specificity have also emerged. Coulson and colleagues (Coulson, King & Kutas, 1998) highlighted the potential for grammatically incorrect sentences to serve as a syntactic form of response conflict, given that words appearing to generate the syntactic anomalies (e.g. "*to" in the above garden path example) stand in violation to expectations of grammatical correctness instilled by the sentence beginnings. According to this view, the enhanced positivity of the syntactic LPC instead reflects the heightened engagement of control to resolve conflict, in analogous fashion to ERP control signatures evoked by the detection of infrequent/unexpected targets in classic "oddball" paradigms (e.g. Squires, Squires & Hillyard, 1975). Rather than arguing for the syntactic LPC as a domain-specific marker of syntax processing, this control interpretation brings the LPC component within the purview of the broader P3 complex (specifically the P3b and ensuing slow wave component), which exhibits a highly similar morphology and central-parietal scalp distribution (Polich, 2007; Pritchard, 1981; Ruchkin & Sutton, 1983).

Coulson et al. provided direct evidence of the affiliation between the syntactic LPC and the P3 complex in an experiment manipulating both the grammatical correctness of sentences (grammatically correct vs. incorrect) and the expectedness of encountering either sentence type (expected vs unexpected; see also Hahne & Friederici, 1999). For the latter expectedness manipulation, base rates were varied blockwise such that grammatically correct and grammatically incorrect sentences were differentially associated with either 80% or 20% frequencies of occurrence. The authors observed increased LPC positivity effects for both grammaticality (incorrect >

correct) and expectedness (unexpected/infrequent > expected/frequent). Further, these two effects interacted such that the unexpectedness effect was greater for the grammatically incorrect compared to correct sentences, consistent with the authors' proposal of a unified neural generator accounting for both effects.

A similar control interpretation can be applied to the recognition LPC, given the increased goal salience typically inculcated towards "old" decisions by recognition task formats, which might similarly lead to a heightening of control reflected in the Hits > CRs late positivity enhancement (Neville, Kutas, Chesney & Schmidt, 1986; see Chapters 2 and 3 for supportive behavioural evidence of such "old" emphasis effects). As a parallel to Coulson et al.'s study involving the syntactic LPC, Herron and colleagues (Herron, Quayle & Rugg, 2003) also instantiated orthogonal manipulations of recognition-specific (Hit vs. CR items) and domain-general conditions (unexpected vs. expected items), and observed a modulation of the slow-wave extent of the LPC old/new effect by decision unexpectedness (i.e. unexpected > expected for both Hits and CRs). Hence, the presented findings highlight the sensitivity of late positive ERP components elicited in isolated recognition and semantic task domains to factors typically associated with the P3 complex. Considered with the overlapping morphological and topographical features of all three ERPs, this raises the parsimonious potential for the LPC to serve as a unified index of heightened cognitive control demands across diverse task domains. This possibility is directly interrogated in the present experimental chapter.

The process of re-characterising neural signatures initially interpreted as domain-specific has also extended to the fMRI literature. In fMRI studies of recognition memory, reports of old/new effects on the BOLD response implicated a network of brain regions that reliably elevated for Hits compared to CRs, encompassing regions of the prefrontal cortex, including medial and lateral PFC, the cingulate cortex, including anterior and rostral cingulate regions, the parietal cortex, including the inferior and superior parietal lobules, and sub-cortical regions, including the dorsal striatum (de Zubicaray, McMahon, Eastburn, Finnigan & Humphreys, 2005; Kim, 2013; Konishi, Wheeler, Donaldson & Buckner, 2000; Spaniol et al., 2009). As with the chronology of research perspectives on the equivalent ERP effect, the functional significance of the fMRI old/new effect was originally interpreted as signifying the successful recovery of memory content specific to *old* items (termed the “retrieval success” effect; de Zubicaray et al., 2005; Konishi et al. 2000; Spaniol et al., 2009), followed by later re-interpretations in light of the observed sensitivity of a large number of “retrieval success” regions to aspects of memory control (Dobbins et al., 2002; Kim, 2013).

One of the clearest control-driven re-characterisations of fMRI retrieval success effects was provided by the likelihood cueing paradigm described in previous chapters (see Section 1.5. and Section 4.2.3.), wherein activation in a number of brain regions previously linked with retrieval success was shown to be preferentially sensitive to detecting the mismatch between invalidly cued expectations and the actual *old/new* recognition status of presented test items (O’Connor, Han & Dobbins, 2010). Subsequent research has elucidated more specified sub-functions for constituents of the resultant “invalid cueing network”, including the differential sensitivity to unexpected “old” and unexpected “new” decisions respectively (Jaeger, Konkel & Dobbins, 2013), and a proposed segregation into temporally “early” conflict detection and “later” control allocation sub-networks (Mill & O’Connor, under review).

However, it is thus far unclear as to whether the control functions performed by regions of the fMRI invalid cueing effect are specific to the episodic memory domain or more general in character. Indeed, targeted attempts to segregate domain-specific from cross-domain control activations have been broadly lacking. Rather, the extant literature has primarily focused on task manipulations and analytic techniques that isolate control sub-processes within the isolated domains in which they are elicited. These approaches have nonetheless ascribed cross-domain or (more generally) "cross-task" control properties to brain regions recurrently observed in episodic memory tasks, including lateral and medial PFC regions, anterior and rostral cingulate regions, and regions of the left parietal lobule (Barredo, Oztekin & Badre, 2013; Dobbins, Foley, Schacter & Wagner, 2002; Dobbins & Wagner, 2005). A similar frontal-parietal network has been implicated in cross-task control processes recruited during evaluation of semantic memory (Badre, Poldrack, Paré-Blagoev, Insler & Wagner, 2005; Grossman et al., 2006; Wagner, Maril, Bjork & Schacter, 2001). However, without exposing the *same sample* of participants to tasks spanning *different* cognitive domains, unequivocal demonstration of a unified control system is complicated by the presence of likely between-subject and between-task confounds.

Further, those studies that have exposed the same sample to tasks in different domains have relied on relatively coarse measures of cognitive control. For instance, Nyberg and colleagues (2003) conducted PET imaging as participants performed a range of tasks spanning episodic, semantic and working memory domains, and examined conjunctions of baseline contrasts (e.g. episodic task > fixation) across all three task domains to identify likely cross-task control regions. Beyond limitations associated with the use of the PET modality (which suffers from lower temporal and spatial resolution than fMRI), the simplistic baseline approach is likely to recover a large number of general task-positive activations and is therefore less sensitive to recovering activations specifically involved in the heightening of control. In a more controlled fMRI study reported by Fleck and colleagues (Fleck, Daselaar, Dobbins & Cabeza, 2006), participants performed a standard episodic recognition task and a visual discrimination task (requiring participants to discriminate which of two coloured onscreen segments was larger in size), with the authors localizing cross-task control activations on the basis of the conjunction of regions showing greater activation for "low" compared to "high" confidence decisions. Employing the low > high confidence contrast in this manner serves as a *post-hoc* method of identifying brain regions linked in a relatively broad sense with some form of conflict and associated control, in comparison to the more systematic manipulation of control demands on a trial-by-trial basis afforded by the likelihood cueing paradigm (O'Connor et al., 2010).

My aims for the present experimental chapter were hence threefold. Firstly, to probe the cross-task properties of the late positive ERP previously elicited in isolated episodic recognition and semantic memory domains. Secondly, to segregate brain regions that display fMRI BOLD activations consistent with a recognition-specific function from those with a function that extends beyond the recognition domain. A final aim was to integrate findings from both the EEG/ERP and fMRI modalities to provide robust multi-modal neuroimaging support for the existence of a unified cross-domain control system. These aims were targeted in an experiment that simultaneously acquired EEG and fMRI data whilst participants performed a novel "match/mismatch" decision-making task designed to systematically manipulate control demands across episodic and semantic memory domains. The task format was held constant across these two domains, with participants judging whether the status of visually presented item stimuli "matched" or "mismatched" with cue screens presented on the same trial (with similar cross-domain investigations described in Chapters 3 and 4). Following from related findings from the likelihood cueing paradigm and studies of response conflict in the cognitive control literature, it was assumed that control demands were heightened in “mismatch” compared to “match” trials, given the need for additional controlled processing to resolve the conflict between the cue and item. In addition, the present experiment examined the generality of elicited control activations at a further level of specification via orthogonal manipulation of the order in which cues and items appeared within the course of a given trial. Neural signatures that index the presence of cue-item conflict (and associated control demands) should do so irrespective of not only the domain in which the item is evaluated but also the order in which the item is presented in relation to the cue. Match/mismatch effects were therefore examined both for when cues preceded items (the “Cue First” condition) as well as when items preceded cues (the “Item First” condition). The adopted task format and multi-modal imaging approach enabled a detailed elucidation of the temporal and spatial properties of putative “cross-task” control activations.

5.2. Method

5.2.1. Participants. The sample comprised 22 right-handed, native English speaking participants (12 female; age mean = 23.45, range = 19-32 years) out of a total number of 24 who completed the experiment. One participant was excluded for failing to respond during the entirety of the Item First condition, and another participant was excluded due to excessive artefacts contaminating their EEG data, which persisted after offline pre-processing. Informed consent procedures were approved by the University Teaching and Research Ethics Committee at the

University of St Andrews, and the East Scotland Research Ethics Committee, Ninewells Hospital and Medical School.

5.2.2. Stimuli. As in the previous two chapters, word stimuli for the *semantic task* were sampled from *pleasant* and *unpleasant* item categories created from the Toglia and Battig semantic word norms (Toglia & Battig, 1978). These item categories were separated by 1.5 standard deviations in their pleasantness ratings so as to exclude words of ambiguous pleasantness proximal to the neutral rating (i.e. “4” on the 7-point pleasantness scale) and were randomly sampled from at regular intervals in collating task wordlists for each participant (sampled *pleasant* word range: 4.64-6.32; *unpleasant* word range: 1.73-3.36). The two cued semantic task wordlists each comprised 128 words (64 *pleasant*, 64 *unpleasant*) and a single uncued semantic task wordlist comprised 80 words (40 *pleasant,* 40 *unpleasant*).

Recognition task wordlists were created by randomly selecting half of the words presented in the preceding semantic phase to form the studied *old* item category. This randomisation was stratified such that *old* and *new* recognition item categories were equated for the number of *pleasant* and *unpleasant* words used to constitute them. Hence, 128 words were presented in each cued recognition phase (64 studied *old*, 64 unstudied *new*) whereas 80 words were presented in the uncued phase (40 *old*, 40 *new*). The experiment was presented and responses recorded using a presentation PC running E-prime (Schneider, Eschman & Zuccolotto, 2002) that was synchronized with the EEG-fMRI apparatus.

5.2.3. Procedure. Brief practice versions of all task phases were presented during EEG cap preparation, after which participants were led to the scanner to undergo a sequence of structural scans followed by a 4-minute resting-state scan (fcMRI).

Participants then performed variations of the match/mismatch task in separate blocks spanning semantic and recognition domains (such that semantic blocks served as "study" phases for ensuing recognition blocks; see Figure 5.1.). Task format was identical across both domains, with participants deciding whether textual cue screens highlighting a specific item category (semantic cues: "pleasant" or "unpleasant"; recognition cues: "old" or "new") matched or mismatched with the actual status of word items presented in the same trial (semantic item categories: *pleasant* or *unpleasant*; recognition item categories: *old* or *new*).

The ordering of cues and items was also varied blockwise, such that cues preceded items in the "Cue First" semantic-recognition task phase, and items preceded cues in the "Item First" semantic-recognition task phase. To clarify, match/mismatch decisions in the Cue First phase therefore emerged from 1) processing the cue screen, 2) discriminating the status of the subsequently presented item (as determined by the task domain in effect for that block), and 3) assessing whether the cue and the discriminated item matched or mismatched. The match/mismatch decision process for the Item First condition was identical except for the reversal of steps 1) and 2). The critical stage 3 in which the match/mismatch between cues and items is actively evaluated is hereafter referred to as the "active decision" period.

Participants submitted their match/mismatch decisions combined with a high/low confidence rating via a single button-box response, submitted using the left hand (1 = "high confidence match", 2 = "low confidence match", 3 = "low confidence mismatch", 4 = "high confidence mismatch"). Participants also performed a single "uncued" semantic-recognition task phase, in which discrimination decisions were made for items in the absence of cues, with response format adjusted accordingly (semantic task format: 1-4 = "high confidence pleasant"-"high confidence unpleasant”; recognition task format: 1-4 = “high confidence old” – “high confidence new”). A response window of 2s was enforced across all cued and uncued task phases.

Design schematic

Domain key	Cue type	Item status
Semantic	“pleasant”/“unpleasant”	*pleasant/unpleasant*
Recognition	“old”/“new”	*old/new*

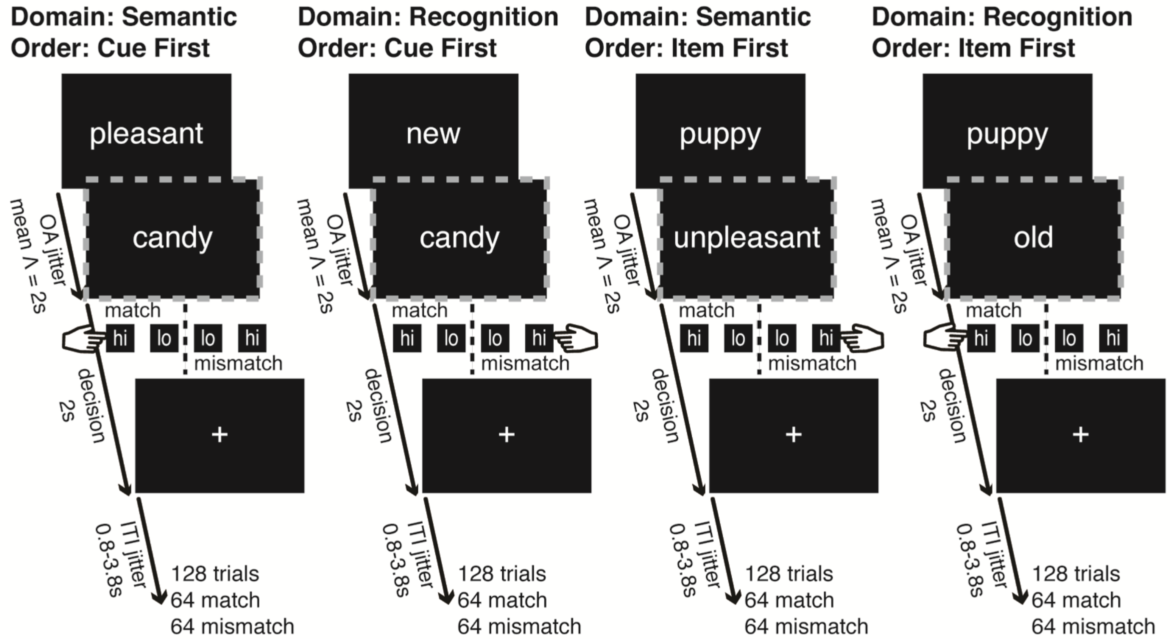


Figure 5.1. Design schematic for cued task trials. Participants decided if cue and item screens matched or mismatched across four task blocks, crossing manipulations of the cognitive domain in which items were discriminated (Semantic or Recognition) and the order in which cues and items were presented (Cue First or Item First). Example trials are depicted for all four blocked conditions, with the key in the upper panel defining the possible values of cue and item screens in each task domain, and the correct decision in each trial denoted by position of the hand icon. Onset of the “active decision” period in each cue-item order condition is highlighted by the dashed grey screen border. ‘OA’ = onset asynchrony; ‘λ’ = mean of Poisson jitter distribution; ‘ITI’ = inter-trial interval; ‘hi’ = high confidence response option; ‘lo’ = low confidence response option.

To clarify, the present match/mismatch cueing procedure differed from the previously described likelihood cueing paradigm (O'Connor et al., 2010; see Section 4.2.3. and Section 5.1.) in that cues across task phases were not explicitly predictive i.e. all cues had a 50% chance of matching items. This format led to the heightening of control demands in 50% of available trials, compared to the lower 20-25% proportion of "invalid cue" trials typically presented in the likelihood cueing paradigm (at the commonly adopted 70-75% cue validity level) in which control demands are also heightened. Increasing the number of heightened control trials was desirable given that the blockwise variations of cue-item ordering and task domain, as well as the inclusion of additional resting and uncued task phases, served to increase the overall experimental run-time. This precluded the alternative of repeating blocks of the likelihood cueing paradigm to raise invalid cue trial counts to a comparable level. Rather, the greater number of trials per block afforded by the match/mismatch procedure ensured adequate statistical power for the EEG and fMRI analyses, whilst permitting additional task manipulations and analyses to also be interrogated. Given that cues in the present study were not valid, and hence did not set up expectations that could be used strategically to improve task performance, participants might have been discouraged from attending to them entirely. However, the match/mismatch response format counteracted this possibility, as participants' final responses had to directly reflect the match or mismatch between cues and items, rather than merely the status of the items.

To improve design efficiency in both cued phases, the onset asynchrony (OA) between cues and items was jittered on an interval range sampled from a zero-truncated Poisson distribution, centred on a mean λ of 2s with 1s units of increment

(Hagberg, Zito, Patria & Sanes, 2001)[11]. One set of cue-item jitter values was sampled and randomly allocated to trials in each task phase, so as to ensure that the overall run time was the same across task phases for all recruited participants. The inter-trial interval (ITI) was also jittered via randomly interspersed blank fixation periods of a fixed 3.8s interval, with 26 blank fixation trials presented in each cued task phase (16 blank fixations were presented in the uncued phase). Design efficiency was further augmented by subjecting participants' task wordlists to a genetic algorithm that determined the optimal ordering of task conditions (Wager & Nichols, 2003).

The uncued task phase always concluded participants' scanned sessions, with the ordering of the preceding Cue First and Item First conditions counterbalanced. Analyses described in this experimental chapter are restricted to the cued task phases, with scrutiny of the uncued task phase and the resting-state fcMRI phase planned for future investigations.

5.2.4. EEG acquisition and pre-processing. EEG data were acquired with an MR compatible system manufactured by Brain Products (Gilching, Germany), comprising 64 sintered Ag/AgCl scalp electrodes (see Appendix B for the electrode map of this EEG system) sampled at 5000 Hz, with a 250 Hz low pass filter (10 second time constant, 0.5 microVolt resolution). An additional ECG electrode was placed on the lower back to monitor participants' electrocardiogram. The impedance between all EEG electrodes and participants' scalps was kept below 20 kΩ. Data were acquired with the FCz electrode as reference. EEG recordings made inside the MRI scanner

[11] Sampling cue-item OA values from the Poisson distribution served to minimize the presentation of long-duration jitters in comparison to rectangular and normal distributions sampling methods, whilst still maintaining equivalent efficiency to these alternative methods (Hagberg et al., 2001). This was advantageous given the need to minimise the overall runtime.

room with the MR-compatible EEG cap and amplifier were relayed via fibre optic cable to a dedicated PC placed outside the scanner room.

Offline pre-processing of the EEG signal was undertaken to remove two primary sources of artefact introduced by recording in the MR environment – the gradient artefact and the pulse artefact (see Figure 5.2. for a schematic representation of EEG pre-processing steps conducted for an example participant). Gradient artefacts arise from electromotive forces produced by the rapid switching between gradient coils during fMRI image acquisition (Allen, Josephs & Turner, 2000). Removal of gradient artefacts was facilitated by the high dynamic range of the MR-compatible EEG amplifier and a hardware device that synchronised EEG acquisition with the MRI system's gradient clock (Syncbox, manufactured by Brain Products). These hardware features respectively enabled the adequate recording of the large gradient contaminants of the EEG signal (without signal saturation) and the writing of precise timepoints at which each fMRI volume was acquired. The recorded gradient artefacts were subsequently removed by average artefact subtraction (AAS; Allen et al., 2000) as implemented in Brain Vision Analyzer 2 (Brain Products), using a baseline-corrected sliding average of the gradient artefact waveform computed over 20 consecutive fMRI volumes. Dummy scans at the beginning and end of each task phase were used to improve stability of the averaged artefact template. Residual gradient artefacts were removed by low-pass filtering at a 40 Hz cut-off frequency.

Subsequent EEG pre-processing steps focused on removal of pulse artefacts. These arise from complex effects of cardiac activity on EEG recorded in the MR environment, encompassing pulse-related head movement, pulse-related electrode displacement and pulsatile movement of blood itself (Debener, Mullinger, Niazy & Bowtell, 2008). The appearance of the pulse artefact roughly corresponds to the R-peak of the cardiac cycle, which was precisely recorded by the MR-compatible ECG electrode placed on participants' lower backs. The R-peaks were used in the offline removal of pulse artefacts via the optimal basis set method (OBS; Niazy et al., 2005), as implemented in EEGLAB 12.0 (Delorme & Makeig, 2004) running the fMRIB plug-in (Niazy et al., 2005). Residual pulse artefacts in the OBS-corrected EEG data were identified by visual inspection of the results of an ensuing temporal independent components analysis (ICA; Benar et al., 2003; Debener et al., 2008), which was conducted for each individual participant in Brain Vision Analyzer 2 using the Infomax algorithm. This process led to the removal of pulse artefact components from the back-projected EEG data, as well as other characteristic artefact components such as eye movements and eye blinks (Jung et al., 2000). It is worthwhile highlighting that jittering of both the cue-probe OA and ITI served to reduce the correlation between gradient/pulse artefacts and the task-evoked EEG signal, thereby improving the efficacy of the above offline pre-processing steps. The gradient- and pulse-corrected EEG data were finally re-referenced to a common average and band-pass filtered (0.05-30 Hz).

Event-related potentials (ERPs) were then calculated from the raw electrode channels for the conditions of interest, all of which were time-locked to the "active decision" period wherein both cues and items had been presented and participants could evaluate the match/mismatch between them. To clarify, the "active decision" period in the Cue First condition is marked by appearance of the item, whereas for the Item First condition it is appearance of the cue (see Figure 5.1.), and therefore active decision ERPs were extracted over 2800 ms post-item and post-cue epochs in each respective cue-item order condition. Active decision ERPs were baseline corrected to the 200 ms period immediately preceding these epochs and averaged over separate match and mismatch conditions in each cued task phase (for correct responses only). Statistical analyses focused on the average amplitude in central and parietal channels of *a priori* interest, as computed over ERP onset periods identified in active decision periods in each task phase (see Results, Section 5.3.2. for further details).

5.2.5. Group ICA of EEG data. Analysis of the raw electrode ERPs was supplemented by a group ICA approach, as implemented in the EEGIFT toolbox (Eichele, Rachakonda, Brakedal, Eikeland & Calhoun, 2011) running in EEGLAB. This method seeks to identify independent components common to all sampled participants by applying ICA to all aggregated observations at the group level. Group ICA enables more effective signal decomposition than ICA applied to individual participant datasets, given that the stochastic nature of ICA (i.e. different iterations will generate different components, arranged in a different order and with different scaling) makes it difficult to identify common components across participants (Eichele et al., 2011). Further, the reduced signal to noise ratio of participant-level ICA data complicates identification of the task-related components themselves, which is of added concern given the significant MR-induced artefacts in simultaneously acquired EEG (Wirsich et al., 2014). Given that group ICA uses “model-free” methods to isolate task-related components maximally free from artefact sources, these analyses were aimed at validating the topography and temporal onset of the cross-domain task-related ERPs identified from the raw channels.

Two separate group ICAs were conducted for the Cue First and Item First conditions respectively, with semantic and recognition domain phases entered as separate sessions within these cue-item order conditions. Each participant’s pre-processed dataset was segmented across the same “active decision” period used in the raw electrode ERP analyses, for all trials presented in each task phase. Unlike in the raw electrode ERP analyses, participants’ data was not re-referenced to common average so as to reduce linear dependencies in the data (which would reduce the algorithm’s efficiency in isolating independent components). Datasets were then baseline corrected (over the 200 ms period preceding the decision period) and smoothed using a sliding 3-trial average, followed by a 2-step reduction to 10 principle components. The principle components were concatenated across participants and the resulting group dataset underwent Infomax temporal ICA, recovering 10 independent components trained at a final convergence rate of 10^{-6}. The stability of these components was validated by the ICASSO method (Himberg, Hyvarinen & Esposito, 2004) over 5 repeated ICA iterations with random initial starting points. All resultant group ICA components were found to be stable, in terms of both the precision of their clustering across the 5 ICASSO runs and their aggregated quality index (I_q). Components were back-projected onto individual participants’ data and scaled as standardized z-scores. Task-related components were finally identified by examination of the grand average component topographies and waveform timecourses.

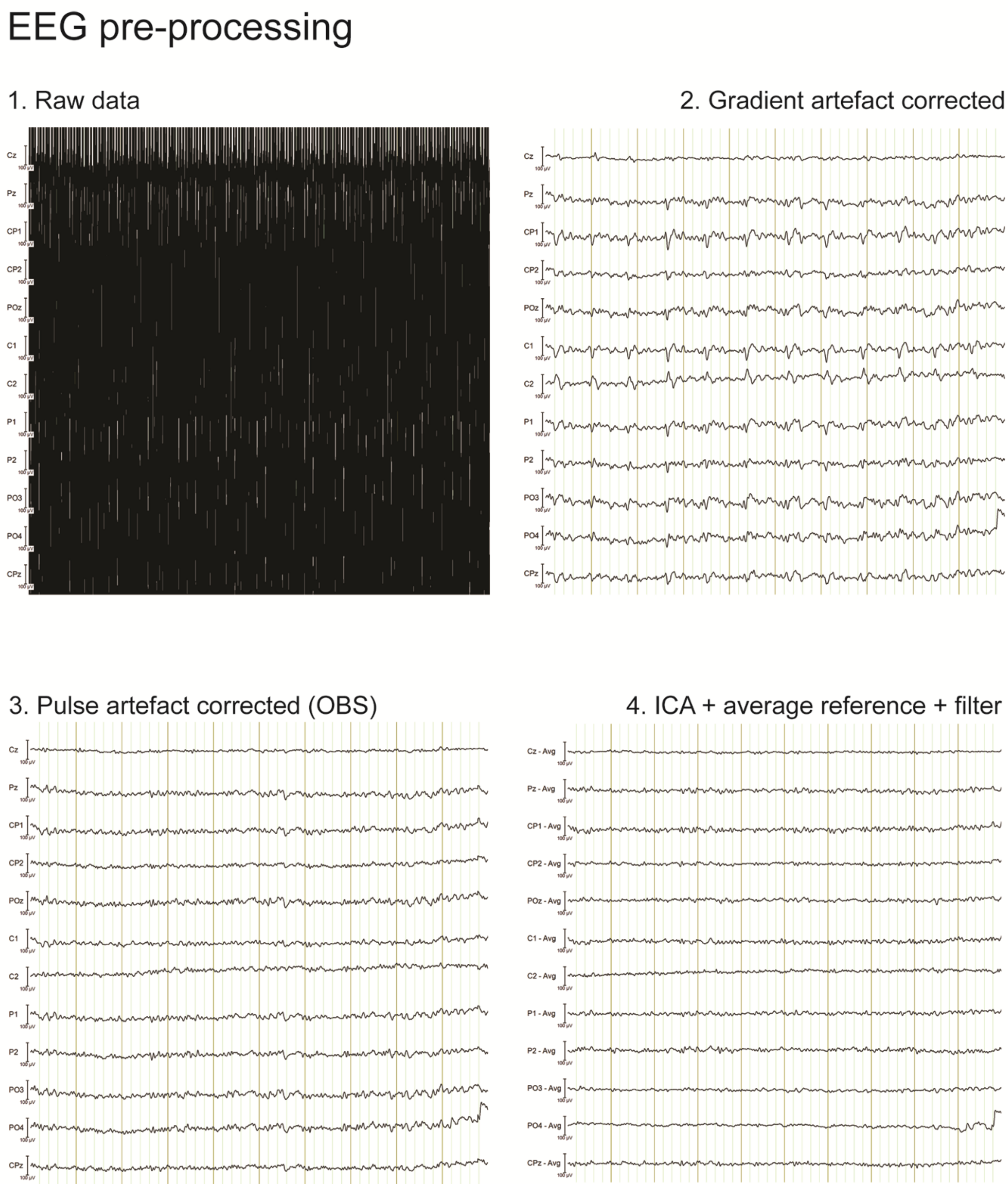


Figure 5.2. EEG pre-processing steps conducted for an example participant. Data are shown for central and parietal electrodes of interest over a 100 μV amplitude range for a 10 second period. Raw data (**1.** in the figure) was firstly corrected for fMRI gradient artefacts by average artefact subtraction (**2.**), followed by removal of pulse artefacts by optimal basis sets (**3.**). The resultant data underwent temporal ICA to remove residual pulse and other sources of artefact, prior to re-referencing to common average and band-pass filtering (0.05-30 Hz; **4.**).

5.2.6. fMRI acquisition and analysis. fMRI images were acquired via a 3T Siemens Trio whole-body MRI scanner (Siemens Medical Solutions, Erlangen, Germany) using a standard 12 channel receive-only head coil. Task phase functional scans were acquired using a descending interleaved echo-planar pulse sequence (TR = 2000ms, TE = 30ms, 90 degree flip angle, 33 axial slices parallel to the AC-PC plane, with 3.5 x 3.5 x 4mm voxel resolution and no inter-slice gap). The 4-minute resting-state fcMRI scan was conducted with identical acquisition parameters. Head motion was minimized using foam padding. All fMRI data were processed in SPM8 (Wellcome Department of Imaging Neuroscience, London). Slice timing was corrected by temporal resampling, followed by rigid body head motion correction. Scanned volumes were then spatially normalized to the canonical echo-planar template, resampled to 3mm isotropic voxels and spatially smoothed using a Gaussian kernel of 6mm full width half maximum.

The standalone fMRI analyses (detailed in Section 5.3.3.) employed an event-related design based on assumptions of the general linear model (GLM; Friston et al., 1995). For each participant, a single GLM convolved a series of event onset vectors (spanning all cued task phases) with the canonical hemodynamic response function to extract voxel-wise BOLD amplitude estimates for conditions of interest. Event regressors were specified for correct “match” and “mismatch” decisions in each cued task phase, separately for item categories in each task domain. fMRI amplitudes were estimated for the same “active decision” period used for the extraction of ERPs (i.e. when cue-item match/mismatch was being interrogated; see Section 5.2.4. above), which was modelled as a 2000 ms epoch. Incorrect match/mismatch decisions in each task phase were modelled as regressors of no interest. The GLM also included two regressors for “pre-decision” events in each task phase (i.e. all cues in the Cue First condition and all items in the Item First condition), irrespective of the correctness and match/mismatch status of the trial.

Group level effects were assessed by treating participants as a random effect and brain volumes as a temporally correlated timeseries (Worsley & Friston, 1995). Linear contrast images of the mismatch > match effect were specified at the participant level and subjected to one-sample t-tests at the group level, against the null hypothesis of a zero contrast effect at each voxel. Analyses of mismatch effects in the Cue First recognition condition adopted a standard threshold of $p < .001$ (uncorrected for multiple comparisons, 5 contiguous voxels). Conjunction analyses investigating mismatch effects across cue-item order and task domain conditions were thresholded at a combined threshold of $p < .001$ uncorrected, 5 contig. vox. (i.e. at individual condition thresholds of $p < .1$), consistent with prior fMRI studies employing a conjunction approach in the study of cognitive control (Badre et al., 2005; Fleck et al., 2006).

5.2.7. EEG-fMRI integration. The final analysis stage integrated the simultaneously acquired data from both neuroimaging modalities via an EEG-informed fMRI approach (Debener et al., 2005; see Section 5.3.4.). A new fMRI GLM was constructed with trial-by-trial amplitude averages of the late positive potential identified in the ERP analysis entered as parametric modulators of the BOLD response (Buchel et al., 1998). At the participant level, event regressors were specified for “correct” and “incorrect” decisions for all four cued phases (trials without a response were modelled as regressors of no interest), modelled as 2000 ms epochs from “active decision” onset (as in the standalone fMRI analyses). Trialwise averages of the late positive ERP were entered as 1^{st}-order linear modulators of the BOLD response during “correct” and “incorrect” decisions in each task. Linear contrast images were specified for ERP-modulated “correct” decision events collapsed across all 4 cued phases. Group effects were interrogated by one sample t-tests conducted on these modulated contrast images, so as to identify spatially precise brain regions in which fMRI BOLD changes were reliably predicted by the temporally precise late positive ERP component, across all cue-item order and task domain conditions. These analyses adopted a relatively liberal threshold of $p < .05$ (uncorrected, 5 contiguous voxels), which were nevertheless inclusively masked with the results of the standalone fMRI analyses conducted at more typical thresholds (i.e. $p < .001$ uncorrected, 5 contig. vox.; see Section 5.3.4. for further details).

5.3. Results

5.3.1. Behavioural validation of mismatch control.

The success of the match/mismatch paradigm in systematically manipulating cognitive control demands was interrogated in behavioural analyses of the reaction time (RT) and response confidence associated with correct decisions. A 2 (cue-item order: Cue First or Item First) x 2 (task domain: recognition or semantic) x 2 (match status: match or mismatch) repeated-measures ANOVA was conducted on RT (see Figure 5.3a). A significant main effect of order was observed, such that RT was slower in the Cue First (M = 1368.08 ms, SD = 156.21) compared to the Item First condition (M = 1035.97 ms, SD = 156.59), $F(1,21) = 180.34$, $p < .001$, $\eta_p^2 = .90$. The order effect on RT likely resulted from the greater time required to process the status of items compared to cues, and the increased time afforded to evaluate the former more resource-intensive event in the Item First condition. In this condition, items appeared in the initial “pre-decision” period for an average duration of ~2000 ms (depending on the Poisson-sampled jitter) and were amenable to further evaluation in the ensuing 2000 ms “active decision” period. In comparison, the Cue First condition required participants to conduct the item status evaluation entirely in the later active decision period. Indeed, this interpretation of the order effect on RT is corroborated by the differing temporal onsets of the late positive ERP component described in Section 5.3.2. A significant order by task domain interaction was also observed, such that RT was slower for the recognition compared to the semantic domain in the Cue First condition (recognition RT: M = 1393.76 ms, SD = 155.63; semantic RT: M = 1342.41 ms, SD = 180.49), whereas this pattern was reversed in the Word First condition (recognition RT: M = 1008.67 ms, SD = 158.77; semantic RT: M = 1063.28 ms, SD = 168.57), $F(1,21) = 22.39$, $p < .001$, $\eta_p^2 = .52$.

Crucially, there was also a significant main effect of match status, such that RT was slower in the mismatch (*M* = 1245.34 ms, *SD* = 145.03) compared to the match condition (*M* = 1158.71 ms, *SD* = 151.17), $F(1,21) = 48.89$, $p < .001$, $\eta_p^2 = .70$. All other main effects and interactions were non-significant (all *F*s < 1, except order by match interaction, $F(1,21) = 1.09$, $p = .307$, $\eta_p^2 = .05$). The effect of match status on RT highlights that mismatch decisions – irrespective of cue-item order or task domain – demonstrated a slowing of reaction time that is a characteristic behavioural signifier of heightened control demands.

The same 2 (order) x 2 (domain) x 2 (match status) ANOVA was conducted for response confidence (see Figure 5.3b). The main effect of match status was again significant, reflecting an overall reduction in confidence for mismatch (*M* = 1.80, *SD* = 0.15) compared to match decisions (*M* = 1.85, *SD* = 0.12), $F(1,21) = 15.58$, $p = .001$, $\eta_p^2 = .43$. All other main and interaction effects were non-significant (main effect of order and order by domain interaction, both *F*s < 1; main effect of domain, $F(1,21) = 1.57$, $p = .224$, $\eta_p^2 = .07$; order by match interaction, $F(1,21) = 2.83$, $p = .107$, $\eta_p^2 = .12$; domain by match interaction, $F(1,21) = 1.35$, $p = .258$, $\eta_p^2 = .06$; three-way interaction, $F(1,21) = 2.67$, $p = .117$, $\eta_p^2 = .11$). As with the earlier RT analyses, the reduced confidence observed for mismatch compared to match decisions serves as another characteristic behavioural signifier of the greater control required to make the former class of decision. The overall pattern of behavioural effects is consistent with heightened effortful processing undertaken to resolve the conflict associated with cue-item mismatch, irrespective of the cue-item order and task domain in which this mismatch occurs. The behavioural findings therefore validated the control assumptions of the ensuing multi-modal neuroimaging analyses.

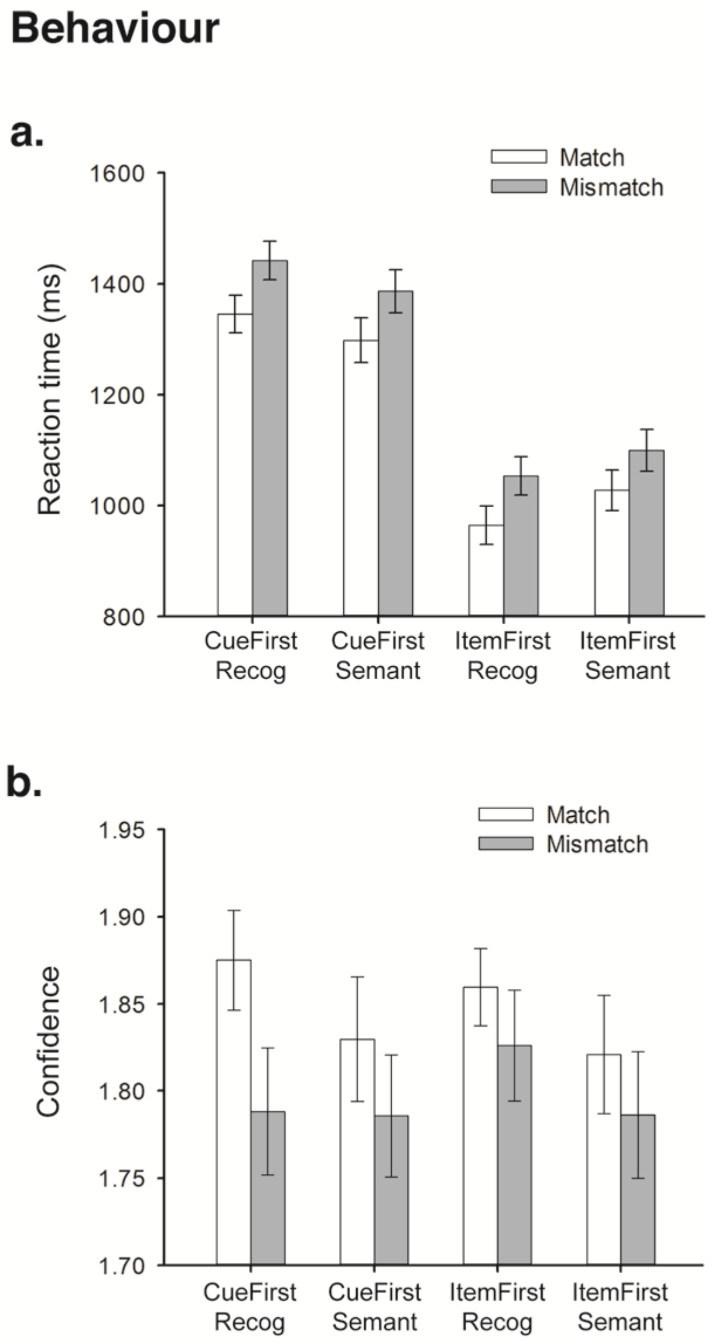


Figure 5.3. Behavioural effects of match/mismatch control manipulation on **a.** reaction time (RT) and **b.** response confidence across task phases. Error bars represent standard error of the mean. “Recog” = recognition task domain, “Semant” = semantic domain.

5.3.2. Standalone EEG analyses recover late positive ERP.

The standalone EEG analyses focused on raw channel ERPs extracted over the "active decision" period within each trial, separately for correct "match" and "mismatch" decisions. Figure 5.4. illustrates the common "late positive" component (LPC) recovered across the presented task phases. Scalp topographies of the "mismatch minus match" difference wave revealed a voltage positivity evoked over central and parietal electrode sites across all task phases, which emerged over the period of putative assessment of cue-item match/mismatch for a duration of approximately 600 ms. This temporal profile was confirmed by examination of the grand average late positive waveforms, which were pooled across central-parietal midline electrodes (CPz, Pz and POz; see Figure 5.4.). The onset of the late positive waveform corroborated the earlier main effect of cue-item order on RT, such that the ERP onset earlier in the Item First (~400 ms) compared to the Cue First condition (~600 ms). As described in the behavioural validation results section (see 5.3.1.), the differing onsets likely reflect the quicker completion of item evidence assessment, and hence quicker engagement of subsequent cue-item match/mismatch evaluation, in the Item First condition.

The difference wave topographies and grand average waveforms also highlighted the sensitivity of the LPC to the match/mismatch manipulation, in the form of an increased amplitude positivity for mismatch compared to match decisions. This difference was interrogated in formal statistical analyses, wherein the amplitude of the pooled late positive ERP was averaged over the relevant onset period in each cue-item order condition (i.e. Cue First ERP average: 600-1200 ms; Item First ERP average: 400-1000 ms), and subjected to a 2 (cue-item order: Cue First or Item First) x 2 (task domain: recognition or semantic) x 2 (match status: match or mismatch) repeated-measures ANOVA. The main effect of match status was significant, reflecting the increased LPC amplitude for mismatch (*M* = 0.76, *SD* = 0.89) compared to match decisions (*M* = 0.28, *SD* = 0.79), $F(1,21) = 11.60$, $p = .003$, $\eta_p^2 = .36$. All other main and interaction effects were non-significant, all *F*s < 1[12]. The standalone EEG analyses therefore recovered a late positive ERP component that onset within a sub-epoch of the "active decision" period (one in which controlled cue-item match/mismatch assessment was most likely to occur) and which showed an increased voltage potential for mismatch compared to match decisions across all presented task phases.

Supplementary group ICA analyses were also conducted to validate the topographical location and temporal morphology of the LPC component identified in the raw channel analyses. All task-related components identified by the separate group ICAs applied to the Cue First and Item First conditions are presented in Figure 5.5. All other group ICA components not displayed pertained to various artefactual sources, such as head motion or residual pulse artefact. The group ICA results demonstrate that task-related components identified by the "model-free" group ICA (i.e. without prioritisation of particular electrode locations) were evoked at similar central-parietal scalp locations and with comparable grand average component waveforms to the raw channel ERPs (with peak ~600 ms for the Cue First condition and ~400 ms for the Item First condition).

[12] Separate 2 (cue-item order) x 2 (task domain) x 2 (match status) ANOVAs conducted for the individual CPz, Pz and POz electrode channels that were pooled in the main analyses also yielded significant main effects of match status, in the absence of any other significant main or interaction effects, $F(1,21) = 7.26$, $p = .014$, $\eta_p^2 = .26$ (CPz mismatch: *M* = 0.68, *SD* = 1.54; match: *M* = 0.06, *SD* = 1.19), $F(1,21) = 10.66$, $p = .004$, $\eta_p^2 = .34$ (Pz mismatch: *M* = 0.73, *SD* = 1.23; match: *M* = 0.39, *SD* = 0.91), and $F(1,21) = 5.01$, $p = .036$, $\eta_p^2 = .19$ (POz mismatch: *M* = 0.73, *SD* = 1.50; match: *M* = 0.39, *SD* = 1.50) respectively.

Given that the group ICAs were trained on each subject's complete dataset, the resulting task-related components were unlikely to reliably differentiate between subtler aspects of responding, such as accuracy and match/mismatch status. Hence, the trial-by-trial averages of the raw late positive ERP (displaying clear mismatch > match effects across all task phases) were used in the integrated EEG-fMRI analyses**,** rather than the trialwise participant-level reconstruction of the group ICAs.

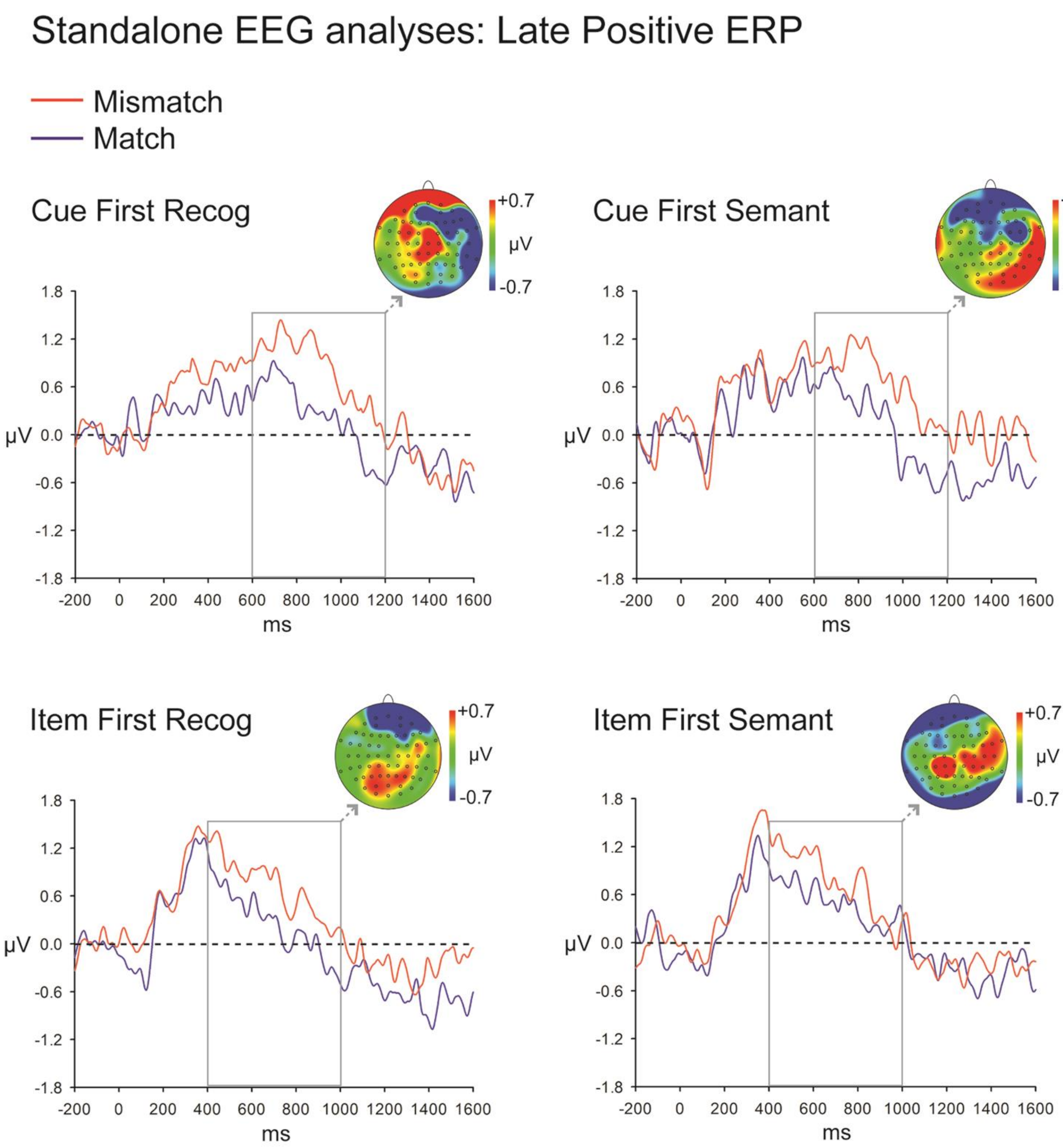


Figure 5.4. Standalone EEG analyses identify late positive ERP across all task phases, spanning cue-item order (Cue First, Item First) and task domain conditions ('Recog' = recognition; 'Semant' = semantic). Grand average waveforms are shown for pooled central-parietal midline electrodes (CPz, Pz and POz), with red waves denoting averaged 'mismatch' decisions and blue waves denoting 'match' decisions. Grey rectangles illustrate the 600 ms period over which ERP amplitudes were averaged for statistical comparisons. Scalp topographies are presented for the difference wave (mismatch – match) calculated over the 600 ms ERP onset period relevant for each task phase.

Standalone EEG analyses: Group ICA

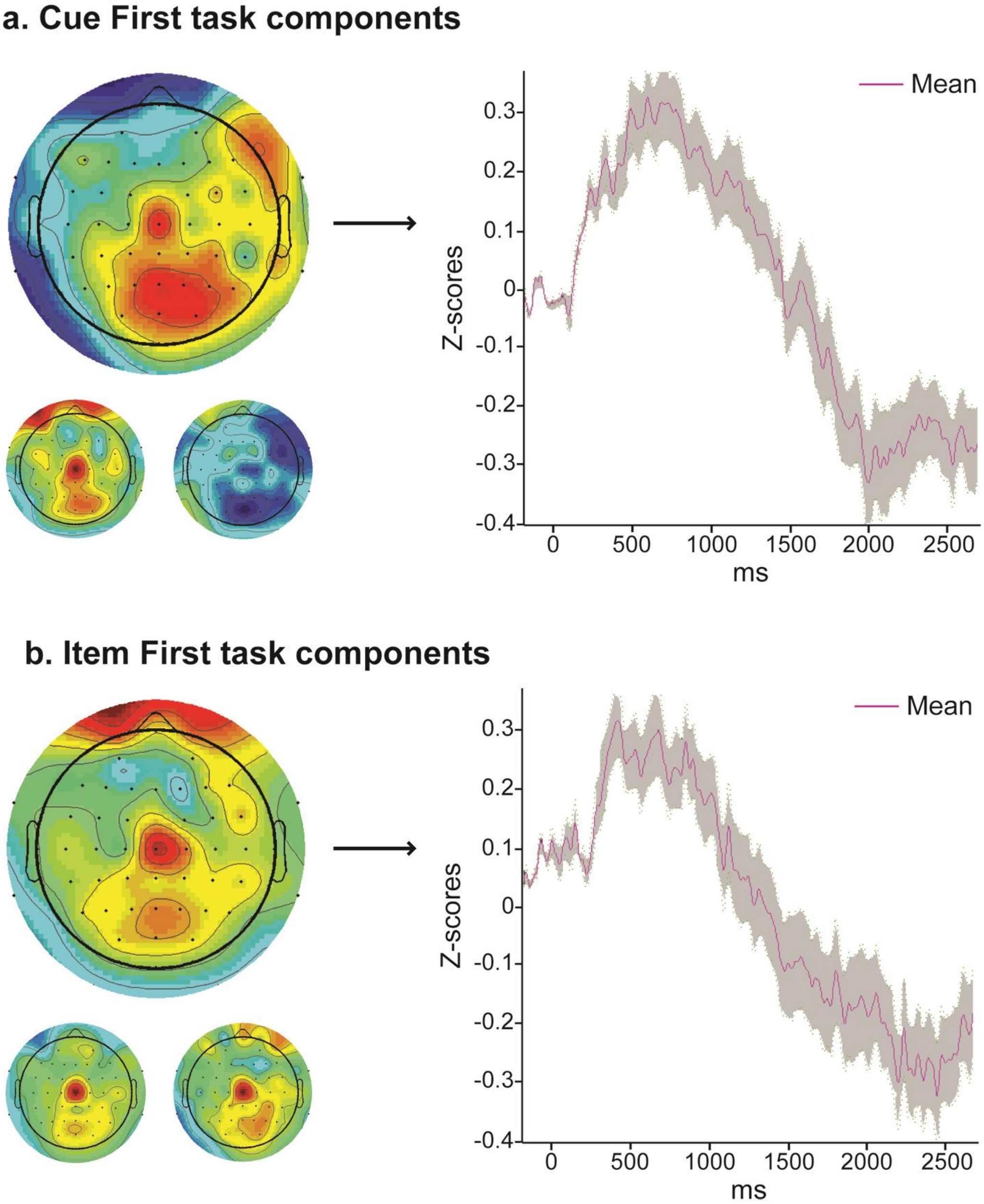


Figure 5.5. Supplementary group ICA of EEG data recovers central-parietal task components, in separate analyses for **a.** the Cue First condition and **b.** the Item First condition. Left panels present scalp topographies for all task-related independent components identified by the Group ICA (scaled in arbitrary units). Right panels present grand average waveforms for selected task components (scaled in z-scores), with standard error of the mean shaded in grey for each data point.

5.3.3. Standalone fMRI analyses recover memory control network.

An initial standalone fMRI analysis aimed to identify brain regions linked broadly with heightened memory control, which were subsequently probed for cross-task function in the masking procedures described in the next section. The recognition task in the Cue First order condition was the focus of this preliminary analysis, given that its format is most similar to the likelihood cueing paradigm (with cues guiding the evaluation of ensuing recognition items in both tasks, O'Connor et al., 2010), and therefore enabled comparison of fMRI control correlates emerging from the present mismatch control manipulation with those associated with the reliably replicated invalid cueing effect. Indeed, the mismatch control contrast for the Cue First recognition phase (mismatch > match decisions, $p < .001$ uncorrected, 5 contiguous voxels; see Figure 5.6a, see Table 5.1.) yielded a distributed network similar in extent and threshold to that observed in previous studies employing the invalid cuing contrast (invalid > validly cued items, O'Connor et al., 2010). This "recognition mismatch network" prominently encompassed frontal regions, including bilateral inferior frontal gyri (IFG, ~BA 45/47), left insula (~BA 47/48), left premotor cortex (BA 6) and right motor cortex (BA 4). Additional mismatch-sensitive clusters were located in the rostral cingulate zone (RCZ, BA 32; just inferior to the pre-SMA) and a large swath of the left parietal lobe, centring on the intraparietal sulcus (IPS, BA 7/40). Greater activation for mismatch compared to match Cue First recognition decisions was also observed in bilateral occipital cortices (~ BA 17/18/19), left precuneus (~ BA7), regions of the left dorsal striatum (putamen extending into caudate), right temporal cortex (BA 20/21) and right cerebellum. The inverse match > mismatch contrast yielded no suprathreshold clusters at the same threshold. Considered with the behavioural analyses described earlier, the standalone fMRI analyses of the Cue

First recognition phase suggest that the heightened control demands associated with correctly identifying cue-item mismatch engaged a number of brain regions that have been reliably observed in previous studies of recognition memory control.

Further standalone fMRI analyses sought to segregate the recognition mismatch network into regions that likely serve a cognitive control function specific to the Cue First recognition task, and those that serve a more “general” cross-task function (irrespective of the cue-item order and task domain in which control is heightened). A conjunction analysis was conducted to identify regions in which fMRI BOLD amplitude increased for the mismatch > match contrast across all four presented task phases (combined $p < .001$ uncorrected, 5 contiguous voxels). Figure 5.6b shows the result of inclusively masking this cross-task conjunction with the initial Cue First recognition mismatch contrast (see Table 5.1. also). Recognition mismatch regions with likely cross-task control properties were largely left-lateralised, with prominent segregation observed in the RCZ and left IPS clusters. The cross-task fMRI conjunction is examined further in the ensuing section as part of a comprehensive examination of the precise control functions subserved by regions of the recognition mismatch network, which encompasses integration with the late positive ERP component described previously.

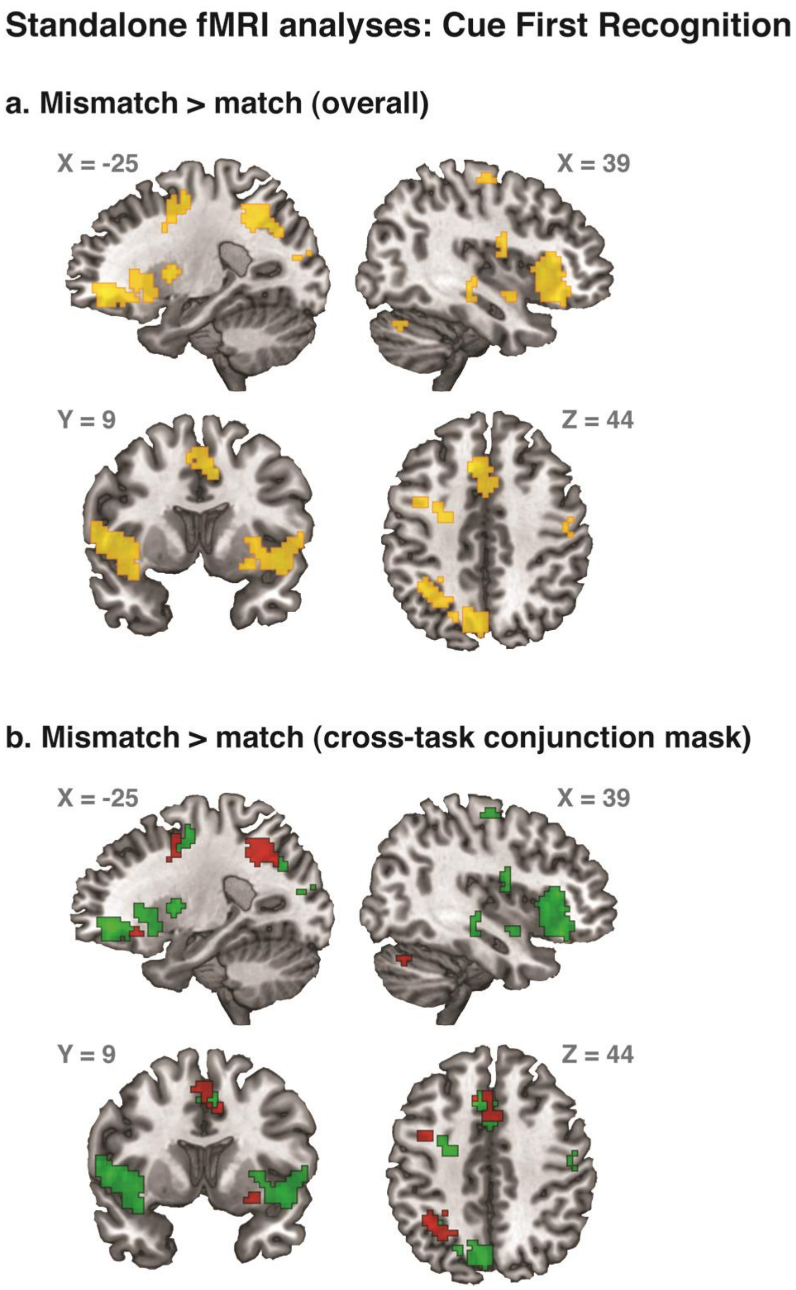


Figure 5.6. Standalone fMRI analyses recovering recognition mismatch network. **a.** All regions demonstrating significant activation in the mismatch control contrast of the Cue First recognition task (mismatch > match decision trials; $p < .001$, uncorrected, 5 contiguous voxels). **b.** The same mismatch contrast in a. masked by the results of a conjunction analysis of mismatch > match effects across all four task phases (combined $p < .001$, 5 contig. vox.). Regions active in both the Cue First recognition mismatch contrast and the cross-task conjunction are shown in red, whereas regions solely active in the Cue First recognition contrast are shown in green. X, Y and Z coordinates are provided in MNI space.

5.3.4. Integrated EEG-fMRI isolates reliable cross-task control regions.

The approach adopted in the integration of the EEG and fMRI data followed from the observation that whilst the late positive ERP reliably increased for mismatch compared to “match” decisions, it nevertheless demonstrated a voltage potential increase relative to baseline for both decision categories (see Section 5.3.2; see Figure 5.4.). The control process underpinned by the late positive component is hence likely engaged for both decision categories, albeit to an enhanced degree when making mismatch decisions. fMRI BOLD activation correlated with the ERP would be at a risk of cancelling out if both decision types were modelled separately due to shared variance between the two regressors. A simplified model was therefore adopted, wherein “correct” decisions irrespective of match/mismatch status were parametrically modulated by trialwise estimates of the late positive ERP (averaged over the 600 ms sub-epoch used in the standalone ERP analyses). The resultant activation map yielded brain regions in which BOLD amplitude was predicted by variations in the late positive ERP (thresholded at $p < .05$ uncorrected, 5 contig. vox.). The ERP modulation map and the earlier fMRI cross-task conjunction map (combined $p < .001$ uncorrected, 5 contig. vox.) were then utilised in a series of masking procedures centring on the core recognition mismatch control contrast (Cue First recognition mismatch > match, $p < .001$ uncorrected, 5 contig. vox.), so as to elucidate memory control regions with reliable cross-task control functions.

To validate the above approach in isolating cross-task control activations, it is worthwhile highlighting that inclusively masking the recognition mismatch > match contrast with the negative ERP modulation map yielded no suprathreshold clusters. This demonstrates that parametric relationships between the late positive ERP and the fMRI mismatch control effects were exclusive to the positive modulation. Further, inclusively masking the same positive ERP modulation map with the inverse Cue First recognition match > mismatch contrast (itself emerging at a lower threshold of $p < .05$ uncorrected, 5 contig. vox.) also yielded no suprathreshold clusters. This latter finding confirms that the relationship between the late positive ERP and the standalone fMRI analyses was preferentially sensitive to the fMRI mismatch > match contrast that identified brain regions underlying aspects of cognitive control.

The results of the masking procedures are rendered in Figure 5.7. and indexed in Table 5.1. Panel a. of the figure depicts mismatch-sensitive regions that were exclusive to the Cue First recognition phase, and not active in either the fMRI cross-task conjunction or the ERP parametric modulation. These "recognition-specific" regions included the entirety of the occipital mismatch cluster, including bilateral lingual gyrus (~ BA 18/19) and left precuneus (BA 7), as well as the left IFG (BA 45/47), left insula (BA 47/48), right temporal pole (BA 20/21) and right primary motor cortex (BA 4). The control function of these regions is likely confined to the detection of cue-item mismatch *specifically* when items are evaluated in the recognition memory domain.

Panel 7b depicts recognition mismatch regions that demonstrated masked activation patterns consistent with some form of general cross-task function. Regions overlaid in dark blue are those that were active in both the recognition mismatch contrast and the fMRI cross-task conjunction, but that did not correlate with the late positive ERP. This pattern of overlap was observed in right cerebellum, left IFG (~ BA 47) and left premotor cortex (BA 6). These regions potentially perform a cross-task control function, albeit one that does not directly relate to the late positive ERP evoked over the 600 ms sub-epoch within the active decision period. The light blue overlay highlights recognition mismatch regions that positively correlated with the late positive ERP (at a combined threshold of $p < .00005$, 5 contig. vox.), but which were not active in the cross-task fMRI conjunction. Regions displaying this particular masking pattern included right IFG (~ BA 45), left insula (~ BA 47), left premotor (BA 6) and left dorsal striatum. These regions also potentially perform a cross-task function, although this function is likely not primarily sensitive to processes that differentiate mismatch from match decisions.

Finally, regions overlaid in red in Figure 7b are those that were active in the recognition mismatch contrast, the fMRI cross-task conjunction and the positive ERP modulation. This confluence of activation masks emerged in two clusters localized in the RCZ (BA 32) and the IPS (~ BA 7/40) respectively. These "cross-task" clusters were reliably engaged when control demands were heightened in recognition and semantic domains, irrespective of whether cues or items were presented first. Further, the observed association with the ERP modulation links the spatially precise fMRI localization of these cross-task regions with a precise temporal engagement at the 600Ms period in which the late positive ERP was evoked (with onset 600Ms for the Cue First phase, and 400 Ms for the Item First phase). It is worthwhile highlighting that the larger regions in which these cross-task clusters were located were in fact split by their affiliation to the various masking procedures. The cross-task RCZ cluster was flanked by a posterior ERP modulation-inclusive cluster (shown in light blue) and an anterior fMRI conjunction-inclusive cluster (shown in dark blue), whereas the cross-task IPS cluster was solely flanked by an anterior fMRI conjunction-inclusive cluster. These activation patterns allude to the potential sensitivity of even broader regions of the rostral cingulate and left parietal lobe to aspects of cross-task control.

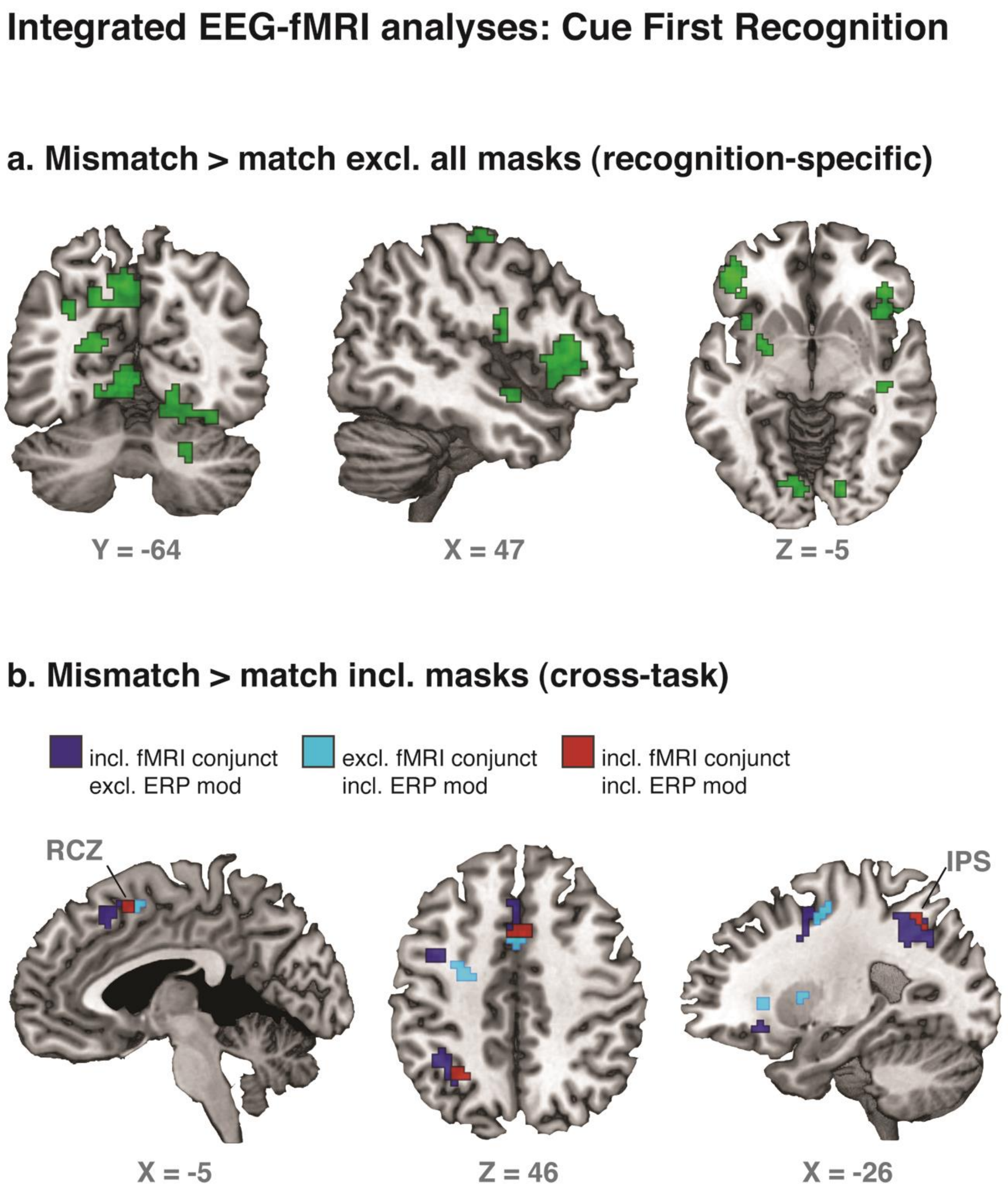


Figure 5.7. Integrated EEG-fMRI analyses segregate recognition-specific and cross-task control regions. Activation maps show the results of masking the recognition mismatch network emerging from the Cue First mismatch > match contrast (p < .001 uncorrected, 5 contig. vox.) with the fMRI cross-task conjunction (mismatch > match across all 4 presented phases, combined p < .001 uncorrected, 5 contig. vox.; “fMRI conjunct”) and the late positive ERP parametric modulation (positive modulation, p < .05 uncorrected, 5 contig. vox.; “ERP mod”). **a.** Regions solely active in the Cue First recognition mismatch > match contrast (i.e. non-overlapping with both fMRI conjunction and ERP modulation maps) are displayed in

green. **b.** Regions active in both the recognition mismatch contrast and various combinations of the fMRI conjunction and ERP modulation maps, with regions overlapping solely with the fMRI conjunction shown in navy blue, those overlapping solely with the ERP modulation shown in light blue, and those inclusive to both masks shown in red. X, Y and Z coordinates are provided in MNI space. ‘Excl.’, exclusive mask; ‘incl.’, inclusive mask; ‘RCZ’, rostral cingulate zone; ‘IPS’, intraparietal sulcus.

Table 5.1. Regions active in the fMRI mismatch > match contrast for the Cue First recognition phase ($p < .001$, 5 contiguous voxels), broken down by their association with masking procedures involving the fMRI cross-task mismatch > match conjunction ("fMRI conjunct"; combined $p < .001$, 5 contig. vox.) and the positive EEG-fMRI parametric modulation ("ERP mod"; $p < .05$, 5 contig. vox.).

Contrast	Region	Lat.	BA	*x*	*y*	*z*	Vox.	*Z* score
Excluding fMRI conjunct and ERP mod								
	Occipital							
	Calcarine/Lingual	L	17/18	-3	-82	4	57	4.38
	Precuneus	L	7	-3	-73	43	83	4.09
	Lingual	R	18/19	18	-82	-8	46	3.76
	Superior Occipital	L	19	-18	-82	19	19	3.71
	Middle Occipital	L	19	-33	-67	37	7	3.53
	Cingulate							
	RCZ/ACC	L	32	-9	20	43	19	4.31
	Frontal							
	IFG	L	47	-48	44	-8	78	4.26
	IFG	R	45/47	48	17	4	144	4.09
	Insula	L	48	-36	14	-5	13	3.45
	IFG	L	45	-54	17	2	19	3.43
	Insula	L	47	-33	20	-2	5	3.26
	Primary motor	R	4	48	-16	61	6	3.18
	Limbic							
	Putamen	L	-	-30	2	4	14	4.10
	Cerebellum							
	Ansiform	R	-	33	-55	-35	7	3.68
	Ansiform	R	-	24	-67	-32	7	3.49
	Temporal							
	TC	R	20	42	-25	-11	13	3.53
	Superior TC	R	21	48	-1	-11	6	3.44
	Parietal							
	IPS	L	7/40	-30	-52	40	8	3.51
	Postcentral gyrus	R	43	66	-4	19	10	3.49
Including fMRI conjunct, excluding ERP mod								
	Cerebellum							
	Ansiform	R	-	33	-58	-35	22	3.99
	Parietal							

IPS	L	7/40	-30	-61	37	59	3.96
Cingulate							
RCZ/ACC	L	32	-9	23	43	16	3.85
Frontal							
IFG	L	47	-27	26	-11	7	3.67
Premotor	L	6	-36	2	49	18	3.62
Including ERP mod, excluding fMRI conjunct							
Frontal							
IFG	R	45	45	17	4	36	4.24
Premotor	L	6	-24	-7	52	15	4.05
Insula	L	47	-30	23	4	5	3.38
Limbic							
Putamen/Caudate	L	-	-24	-1	7	17	4.13
Cingulate							
RCZ	-	32	0	8	52	6	3.47
Including fMRI conjunct and ERP mod							
Cingulate							
RCZ	L	32	3	11	52	11	3.81
Parietal							
IPS	L	7/40	-27	-55	49	7	3.24

Note: x, y, and z coordinates refer to cluster maxima. Lat., Laterality; BA, approximate Brodmann's area; Vox., number of significant voxels; RCZ, rostral cingulate zone; IFG, inferior frontal gyrus; TC, temporal cortex; IPS, intraparietal sulcus.

5.4. Discussion

The present multimodal neuroimaging study focused on a research question that is often conjectured but rarely empirically tested – is the engagement of cognitive control across diverse task contexts mediated by common neural substrates? The adopted design exposed the same sample of participants to cued tasks that differed according to both the domain under evaluation (episodic or semantic) and the ordering of task events (Cue First or Item First), but which were identical in terms of sensorimotor input and the precise form of control manipulation. Behavioural analyses revealed a slowing of RT and reduction in response confidence when cues and items mismatched across all task phases, consistent with the hypothesised increase in control demands for the former category. Standalone EEG analyses revealed a late positive ERP component (LPC) that manifested across all task phases at similar central-parietal scalp locations and onset at the period of putative controlled cue-item match/mismatch assessment. The evoked LPC also reliably indexed the control manipulation in the form of an increased positive slow-wave component for mismatch decisions. The ERP findings were used to clarify the properties of a core “mismatch control network” identified in standalone fMRI analyses of the Cue First recognition phase, which was comparable in extent and threshold to previous recognition control effects. Masking procedures involving this mismatch control network, and combinations of the conjunction of fMRI mismatch effects across all four task phases, and the positive parametric modulation of the fMRI response by trialwise LPC amplitudes, collectively segregated brain regions underpinning cross-task control from those associated with recognition-specific control. The findings raise important refinements to prior interpretations of neural responses elicited in isolated episodic and semantic memory domains, as well as in the wider field of cognitive control.

The results speak directly to prior equivocation in ERP studies of episodic and semantic memory. Research in these isolated domains has documented ERP components with highly similar topographical and morphological properties to the presently observed LPC, which have nevertheless been linked with disparate domain-specific functions. LPC-like components elicited in recognition tasks have been characterised as markers of the recollection of episodic details (Rugg & Curran, 2007; Rugg et al., 1998), whereas LPC-like components elicited in semantic tasks have been associated with syntactic parsing mechanisms (Hagoort, Brown & Groothusen, 1993; Osterhout & Holcomb, 1992). These domain-specific interpretations have been challenged by reports of the sensitivity of both recognition and syntactic LPCs to the detection of expectancy violations (Coulson et al., 1998; Herron et al., 2003). The resolution of conflict associated with unexpected events has long been considered a core cognitive control process, and one that is typically associated with the P3 complex (Squires et al., 1975). Indeed, the P3b and ensuing slow-wave sub-components of the P3 complex also share topographical and morphological features with the previous and presently observed LPCs, raising the possibility that each ERP indexes a common process (as was noted by Coulson et al., 1998).

The results suggest that this common process is likely a “general” feature of cognitive control, as indicated by the modulation of the LPC in response to the conflict generated by cue-item mismatch across different task phases. The findings therefore highlight the virtue of going beyond the traditional paradigms in which P3-like components are elicited (e.g. the oddball paradigm and the go/no-go paradigm) so as to more effectively characterise their oft-conjectured involvement in domain-general processes (Polich, 2007; Pritchard, 1981). The overall pattern of ERP effects favours a functional link between the LPC and cognitive control processes recruited in the resolution of conflict, irrespective of the specific task context in which this conflict arises. This interpretation further implies that the recognition LPC, syntactic LPC and the P3 LPC are all electrophysiological markers of a common cross-task control process.

The findings suggest similar refinements to previously reported fMRI effects. Mere identification of the mismatch control network in the Cue First recognition phase adds to growing evidence that a number of regions previously linked with episodic “retrieval success” effects are in fact sensitive to aspects of memory control (Dobbins et al., 2002; Han, Huettel, Raposo, Adcock & Dobbins, 2010; O’Connor et al., 2010). The ensuing masking procedures enabled the specification of more precise sub-functions for regions within this core memory control network. Regions associated with a “recognition-specific” role demonstrated an increased BOLD response for mismatch compared to match decisions, which did not generalise to the other task phases and did not correlate with the control processes captured by the LPC amplitudes entered as parametric modulators. This exclusive activation pattern likely reflects an involvement in controlled episodic retrieval processes, engaged in an attempt to confirm the recognition status of items that appeared to generate mismatch conflict. Indeed, a number of brain regions identified as “recognition-specific” in the present analyses have been reliably linked with controlled retrieval and source monitoring processes in previous studies, including the precuneus (Dobbins et al., 2002; Lundstrom, Ingvar & Petersson, 2005), regions of the ventrolateral PFC (IFG and insula; Barredo et al., 2013; Dobbins et al., 2002;

Dobbins & Wagner, 2005) and temporal cortex (Dobbins et al., 2002; Dobbins & Wagner, 2005). The fact that these retrieval-related processes did not generalise to the other task phases serves to validate the proposed dichotomy between "retrieval control" and "decision control" outlined in the introductory chapter (see Section 1.3.).

fMRI control activations that were not confined to the Cue First recognition task were identified by the various inclusive masks. Recognition mismatch control regions that were also active in the fMRI cross-task mismatch conjunction, but that did not correlate with the LPC parameters, were likely involved in cross-task control processes that were not primarily engaged over the onset period of the LPC effect. Such processes might contribute to inhibition of the cued response, which putatively would begin at the appearance of the item at the beginning of the "active decision" period, preceding the period of cue-item match/mismatch assessment linked with the LPC. This speculative interpretation finds some support from prior studies linking response inhibition with both cerebellar (e.g. Bunge, Dudukovic, Thomason, Vaidya & Gabrieli, 2002a) and premotor regions (e.g. Ridderinkhof, Wery, van den Wildenberg, Segalowitz & Carter, 2004) that emerged in this particular masking scheme.

Interpreting the functional role of regions observed in the other "partial" inclusive masking pattern, identifying recognition mismatch regions that overlapped with the LPC modulation map but not the fMRI mismatch conjunction, is more complicated. These regions were putatively driven by a process indexed by the LPC, albeit one that does not serve as a *primary* contributor to the ERP's observed sensitivity to cross-task mismatch control effects. As noted in the results Section 5.3.4., the ERP analyses revealed that while LPC amplitude was greater for mismatch compared to match decisions, the component still visibly deviated from baseline for match decisions, suggesting some sensitivity to general task-positive processes common to both decision categories. Considering that prior research involving intracranial recordings has sourced P3-like components with a large network of underlying neural generators, exhibiting a substantial degree of functional heterogeneity among them (Friedman, Cycowicz & Gaeta, 2001), it is likely that processes other than those *directly* driving the mismatch > match effects are also reflected to some degree in trialwise LPC amplitudes. One such process might be the likelihood of reward attainment, which would explain the emergence in this LPC-inclusive mask of a dorsal striatal region reliably associated with reward in prior research (e.g. Knutson, Fong, Adams, Varner & Hommer, 2001). This view might also explain the selective involvement of the striatum in the Cue First recognition phase, given prior reports linking activity in this region with processes underlying the increased salience imparted to “old” decisions by typical recognition task formats (Han et al., 2010).

Contrasting the nuances of the above “partial” masking patterns, recognition mismatch regions displaying the most reliable cross-task control properties were identified in fairly unambiguous fashion – namely as those regions present in all three activation maps. The confluence of masks recovered fMRI activations in two regions – the rostral cingulate zone (RCZ) and the left intraparietal sulcus (IPS) – that elevated for mismatch compared to match decisions across all task domain and cue-item order conditions, and which also correlated with the LPC component shown to be sensitive to cross-task control effects in of itself. Relating the engagement of the RCZ and IPS regions with the precise temporal scale of the correlated LPC effect suggests that the integrated EEG-fMRI cross-task activation emerged between 400-600 ms following onset of the active decision period and lasted for a duration of approximately 600 ms. This temporal profile suggests that the cross-task control processes mediated by the RCZ and IPS were putatively engaged after initial cue- or item-specific processing had been completed (depending on the order condition), with an emergence over the period in which cue-item match/mismatch was being actively evaluated prior to commission of a response. The clear elucidation of a cross-task function for these two regions should clarify the nature of their recurrent involvement in prior studies of recognition memory control (Barredo et al., 2013; Dobbins et al., 2002; Dobbins & Wagner, 2005; O'Connor et al., 2010).

Despite the conjunct elicitation of RCZ and IPS regions in the present cross-task analyses, prior research suggests these two regions might perform subtly distinct sub-functions within the broader service of cognitive control. For instance, the reliable involvement of the RCZ in resolving various forms of response conflict (Botvinick et al., 2001; Kerns et al., 2004) has been more specifically linked with the process of selecting a response amongst competing alternatives (Rushworth, Walton, Kennerley & Bannerman, 2004; Mueller, Brass, Waszak & Prinz, 2007). This response selection role is supported by the anatomical connectivity of RCZ with downstream pre-SMA and primary motor cortices, as well as a direct efferent to the spinal cord, all of which serve as possible pathways via which the actions selected by the RCZ might be translated into overt motor responses (Ball et al., 1999; Dum & Strick, 1991).

In comparison, the role of the IPS in mediating cross-task control might involve the complementary representation and maintenance of available response options in working memory. Support for this representational function comes from neuropsychological studies in which patients with IPS lesions were found to exhibit pronounced working memory deficits in tasks involving spatial memory (Gillebert et al., 2011) and arithmetic operations (Rivera, Reiss, Eckert & Menon, 2005). Indeed, a representational role for the IPS and adjacent left parietal lobe regions has also been repeatedly conjectured in functional neuroimaging studies of recognition control (Barredo et al., 2013; Dobbins et al., 2002; Dobbins & Wagner, 2005). In a recent variant of the likelihood cueing paradigm, Jaeger and colleagues (2013) highlighted the potential cross-task character of the IPS representation function, observing activation in this region for the detection of invalidly cued recognition items irrespective of the actual *old/new* status of the item (contrasting proximal left parietal regions which showed item-specific invalid cueing effects). The notion that the IPS represents and maintains information that is ultimately selected by the RCZ is supported by the findings of Bunge and colleagues (2002a), who successfully dissociated working memory operations linked to the IPS from response selection processes associated with the RCZ in the Eriksen flanker task (Eriksen & Eriksen, 1974). Future research will undoubtedly be required to interrogate this proposed differentiation between maintenance and selection functions of the IPS and RCZ cross-task regions respectively.

In conclusion, the presented multi-modal neuroimaging experiment enabled the precise functional specification of memory control correlates with reliable cross-domain properties, both in time (via association with an ERP that onset between 400-600 ms during the active decision period for a duration of ~ 600 ms) and space (localized with fMRI to rostral cingulate and intraparietal sulcus regions). The findings highlight the virtue of employing simultaneous EEG-fMRI to recover a spatially and temporally rich dataset capable of providing novel insights into the study of episodic and non-episodic control processes. Indeed, this rich dataset will be explored in future analyses seeking to elucidate the large-scale network connectivity dynamics underpinning the observed regional effects, as well as aim to examine the thus far overlooked uncued phase (see Chapter 7, Section 7.4. for further details.). Connectivity methods might be especially useful in clarifying the functional significance of the overlap amongst the various masked effects visible at cortical sites proximal to both the cross-task RCZ and IPS clusters. Nevertheless, the described analyses have provided an insight into the emergence of memory control processes with cross-domain properties in cortical brain regions. The final experimental chapter employed a different functional neuroimaging modality – high resolution infra-red pupillometry – to complement the emergent understanding of cortical dynamics, with insight into the involvement of neuromodulatory brainstem systems that interact with cortical regions in the service of adaptive control.

<u>Chapter 6:</u> Probing the autonomic correlates of memory control using pupillometry

6.1. Introduction to Experiment 7

The previous chapter examined the cortical dynamics underpinning the allocation of cognitive control across diverse task contexts, encompassing different domains of evaluation and different orderings of control-inducing events. The simultaneous EEG-fMRI method revealed the temporal signature of this cross-task control function – linked with onset of a late positive ERP component – which was spatially sourced to regions of the cingulate cortex (rostral cingulate zone, RCZ) and left parietal lobe (intraparietal sulcus, IPS). The findings clarified the functional significance of electrophysiological and hemodynamic signatures of episodic memory control, which have been subject to considerable speculation in prior research (Barredo, Oztekin & Badre, 2013; Dobbins, Foley, Schacter & Wagner, 2002; Finnigan, Humphreys, Dennis & Geffen, 2002).

The aim of the present chapter was to complement this emerging cortical control system with insight into the autonomic systems with which it putatively interacts to facilitate adaptive responses to environmentally induced conflict. Burgeoning research has linked such on-task autonomic changes with neuromodulatory activity of the brainstem locus coeruleus-norepinephrine system (LC-NE; Aston-Jones & Cohen, 2005), as evidenced by animal research documenting the increased activation of individual LC neurons to motivationally salient events (Aston-Jones & Bloom, 1981) and the concomitant release of NE in distributed cortical sites via efferent projections (Abercrombie, Keller & Zigmond, 1988). Whilst these autonomic changes have traditionally been interpreted as signifying general physiological arousal (Berridge & Waterhouse, 2003), more recent accounts suggest a role for the LC-NE system in providing a “boosting” function more intimately tied with cognitive processing (Servan-Schreiber, Printz & Cohen, 1990). In this view, the LC-NE system increases the responsivity or “gain” of cortical control networks in light of environmental factors, such as changes to goal state and heightened response conflict (Aston-Jones & Cohen, 2005; Nieuwenhuis, de Geus & Aston-Jones, 2010). Increasing evidence therefore suggests that the LC-NE neuromodulatory system serves a seminal role in complementing cortical processing to regulate adaptive behaviour. However, the precise nature of this adaptive cortical-autonomic system interaction and the mechanisms of its engagement across different cognitive domains are in need of further interrogation.

In investigating the role of the LC-NE system in episodic memory control (the cognitive domain of primary interest in this thesis), the present experimental chapter led to my use of another functional neuroimaging modality - high resolution infra-red eye-tracking. This followed from considerable evidence suggesting that pupil diameter serves as a physiological index of LC-NE activity (Aston-Jones & Cohen, 2005; Nieuwenhuis et al., 2010). For example, single-cell recordings in primates have highlighted the strong, positive relationship between LC neuronal discharge and increases in pupil diameter (Rajkowski, Rubiak, Aston-Jones, 1993). Furthermore, human pupillometry research spanning almost fifty years has established increased pupillary dilation (PD) as a correlate of heightened cognitive processing in diverse task domains, including mental arithmetic (Hess & Polt, 1964), working memory (Kahneman & Beatty, 1966), perceptual decision-making (Kahneman & Beatty, 1967), semantic decision-making (Ahern & Beatty, 1981) and economic decision-making (Fiedler & Glockner, 2012). Interpretations of the PD response have traditionally focused on theories of processing load, with pupil diameter proposed as an index of the amount of effort or “cognitive load” allocated in ongoing processing (Beatty, 1982; Kahneman, 1973).

Similar load theories have been extended to emerging findings from pupillometry research into episodic recognition memory, which report an increased PD for the presentation of studied *old* compared to unstudied *new* items at test (Gardner, Mo & Borrego, 1974; Goldinger & Papesh, 2012; Heaver & Hutton, 2011; Naber, Frässle, Rutishauser & Einhäuser, 2013; Otero, Weekes & Hutton, 2011; Papesh, Goldinger & Hout, 2012; Vo et al., 2008). Termed the “pupil *old/new* effect”, the increase in recognition PD has been interpreted as a neurophysiological correlate of the heightened cognitive load associated with processing retrieved memory content (Goldinger & Papesh, 2012; Papesh et al., 2012; Vo et al., 2008). The content-related account of pupil *old/new* effects raises clear parallels with analogous “retrieval success” interpretations of *old/new* effects in other functional neuroimaging modalities, as described in the previous chapter (Konishi, Wheeler Donaldson & Buckner, 2000; Rugg & Curran, 2007). To reiterate, this view attributes the heightened neural response for the correct identification of *old* items (“Hits”) compared to *new* items (“CRs”) to the successful recovery of memory content, but has been challenged by studies controlling for strategic control processes that overtly contribute to evaluative outcomes. Of these processes, systematic manipulation of participants’ expectations of encountering *old* or *new* test items has been shown to both modulate event-related potential signatures of retrieval success (Herron, Quayle & Rugg, 2003) and fMRI BOLD activation in brain regions previously associated with retrieval success (Herron, Henson & Rugg, 2004; Jaeger, Konkel & Dobbins, 2013; O'Connor, Han & Dobbins, 2010).

The potential for memory expectation to exert similar influences on the pupil *old/new* effect has thus far been overlooked, despite the documented sensitivity of the PD response to expectancy violations in non-episodic task domains. The link between PD and expectation follows from its early identification as part of the group of neurophysiological changes associated with orienting to change in the environment (Sokolov, 1963). Increased PD has also been reliably observed in the detection of infrequent (i.e. unexpected) "target" stimuli embedded in a stream of frequent "lures" in the oddball paradigm (Friedman, Hakerem, Sutton & Fleiss, 1973; Van Olst, Heemstra & Kortenaar, 1979). More recent research has implicated PD as a specific index of "surprise" resulting from the countermanding of prior expectations (Preuschoff, 't Hart & Enhauser, 2011; De Gee, Knapen & Donner, 2014). Hence, unconstrained expectations operating in prior studies of recognition memory might have served as covert contributors to the pupil *old/new* effect, in which case the observed recognition PD would serve not as a signifier of retrieval success *per se*, but rather as an index of controlled processing undertaken when mnemonic item evidence conflicts with prior expectations.

The present experiment sought to directly interrogate the influence of memory expectation on the pupil *old/new* effect, as systematically manipulated by the likelihood cueing paradigm described in previous chapters (e.g. O'Connor et al., 2010). Participants underwent high-resolution infra-red eye-tracking whilst making recognition decisions for *old* and *new* test items that were either expected or unexpected based on the validity of anticipatory cues (e.g. expected *old* trial: "likely old" cue followed by an *old* item; unexpected *old* trial: "likely new" cue followed by an *old* item). Based on the findings summarised above, we anticipated an increased PD following encounters with unexpected recognition test items i.e. those instances when participants countermanded invalid cues. We also interrogated more specific PD effects; namely, whether pupillary responses were exclusive to unexpected novelty or unexpected familiarity, and whether they tracked response accuracy under these situations. This focus followed from a recent fMRI study (Jaeger et al., 2013) which split the invalid cueing effect into three functionally distinct sub-networks, respectively comprising regions that activated for unexpected familiarity, unexpected novelty, and unexpected events irrespective of memory status.

In this endeavour, analyses of the mean PD amplitude were complemented by multi-level modelling (MLM) techniques that sought to disentangle the contributions of different cognitive processes to changes in the trial-by-trial PD. Similar attempts to separate the cognitive constituents of the PD response have emerged recently in research into predictive rule-learning (Nassar et al., 2012) and visual perception (de Gee et al., 2014). MLM provides a particularly effective method of separating components of the PD response, given that it explicitly models participant-level variation (i.e. as a random effect), thereby enabling more efficient partitioning of variance that aids segregation of task-evoked cognitive processes (Hayes, 2006). Overall, the findings provide novel evidence in support of link between pupil dilation and two functionally dissociable memory control sub-processes, namely unexpected familiarity and decision uncertainty.

6.2. Method

6.2.1. Participants. The sample comprised 34 participants (22 female; mean age = 24.8 years, range = 18-37 years) out of a total number of 40 who completed the experiment[13]. Three participants were excluded for artifacts contaminating more than 20% of their eye-tracking data, 2 for falling asleep during the experiment, and 1 due to a computer malfunction All had normal or corrected-to-normal vision and abstained from caffeine in the hour immediately preceding their participation (as caffeine intake leads to tonic pupil dilation; Tryon, 1975). Informed consent was obtained in compliance with Washington University's human subjects guidelines, with all participants compensated at a rate of $10/hour.

6.2.2. Stimuli. Word stimuli were sampled from a departmental pool of 1216 (Kucera-Francis corpus frequency = 8.85), yielding five, 360 item wordlists to which participants were randomly assigned. Word presentation order within these five wordlists was randomised across participants. Word length was controlled by excluding words less than four letters and greater than 10 letters in length, which served to minimise luminance differences across presented words (which could introduce confounding variation on pupil dilation). Each participant completed three study-test blocks, with 60 words presented at study and 120 words at test (the latter constituting 60 studied *old* words and 60 unstudied *new* words).

6.2.3. Procedure. The experiment was conducted on a standard PC running E-Prime software (Psychology Software Tools, Pittsburgh, PA, USA) that synchronized with both the participant display PC and another PC running the eye-tracker

[13] I conducted the present experiment during a 2 month research placement at Washington University in St Louis, under the supervision of Professor Ian Dobbins. The placement was funded by a SINAPSE Early Career Researcher Exchange Award.

recording software. Following the full setup of the eye-tracker (detailed below), participants were presented with onscreen instructions and a brief practice phase. A self-paced syllable counting task constituted the study phase, with response options ‘1’, ‘2’ and ‘3+’, to facilitate incidental encoding of subsequent *old* items (see Figure 6.1. for a design schematic). The test phase followed immediately, with participants deciding if words appearing onscreen were studied (*old*) or unstudied (*new*) for a randomized list containing equal proportions of *old* and *new* probe items. Each test trial commenced with anticipatory cues as to the likely old/new status of ensuing probes. These took the form of “Likely Old” (LO) and “Likely New” (LN) cue types, which were each valid on approximately 70% of trials. Participants were informed of the cues’ validity and encouraged to use them to guide their recognition decisions. On a subset of trials, participants were provided non-informative cues of the form “Likely ???” (“uncued” trials, UC) that assessed behavioural and eye-tracking responses in the absence of cued information. All cues appeared in isolation for 2 s at the start of each trial, followed by appearance of both the test item and a response prompt for 3.5 s, and finally a 2 s fixation cross period (See Figure 6.1.). Hence, each test phase comprised a total of 120 cued test trials, of which 70 were validly cued, 30 invalidly cued and 20 uncued. Participants completed three study-test block runs in total.

6.2.4. Eye-tracking data collection. Pupillometry data were recorded using an Eyelink 1000 infrared eye-tracker (SR Research Ltd., Mississauga, ON, Canada) running Eyelink software (v 4.48), sampling at 1000 Hz and at spatial resolution <0.01° root mean square. The eye-tracker was placed centrally under the presentation monitor at a fixed distance 60 cm from each participant. Participants were seated on a comfortable chair that included an adjustable headrest to minimize head movement. The experiment was displayed in a room maintained at a constant low level of ambient illumination. Calibration and validation of gaze direction was conducted at the start of each participant's session. Pupil data was acquired at study and test across all three blocks. The data was pre-processed using in-house software written in Java (Oracle Corporation, Redwood Shores, CA, USA). The pupil data were corrected for blinks and other sources of signal dropout by linear interpolation, and were subsequently downsampled to 20 Hz. As mentioned, participants with signal dropout in excess of 20% were excluded from the analysis. The pupil data was extracted as the percent signal change from a baseline mean of the 200 ms fixation period preceding the onset of each trial. Separate analyses were conducted for the mean trial-locked (i.e. locked to cue onset) and mean response-locked pupil dilation, as well as a linear decomposition of the dilation response during the probe epoch that informed the final multi-level model analyses (see Results section for further details).

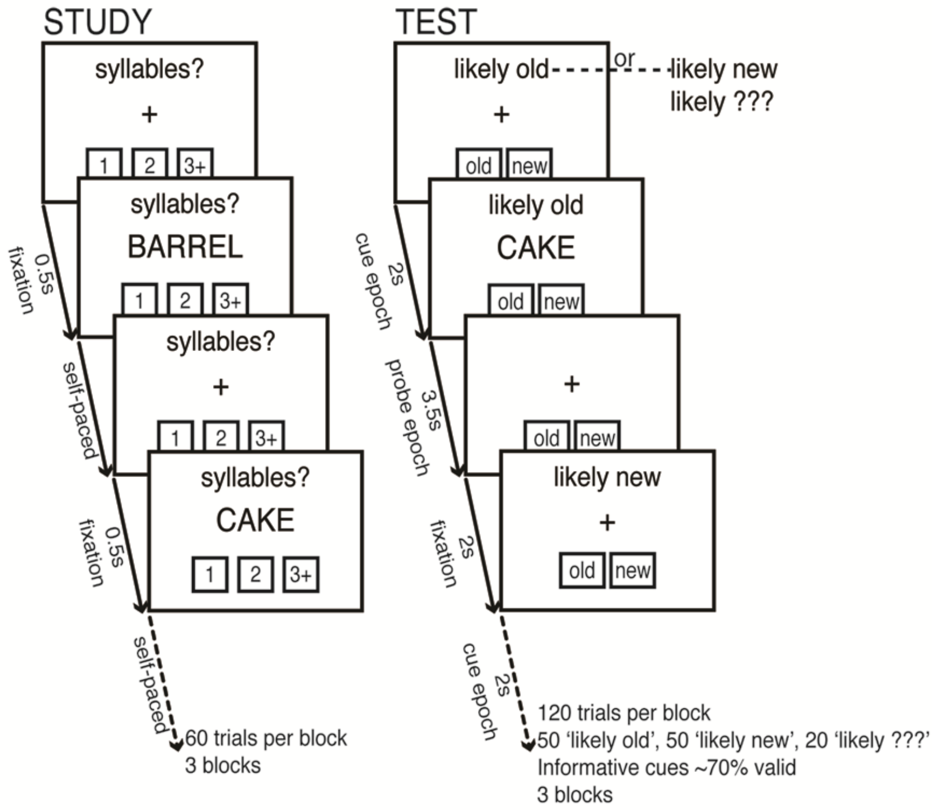


Figure 6.1. Design schematic for the presently employed likelihood cueing paradigm. Study comprised a syllable counting task for 60 presented words. At test, participants decided if presented words were studied *old* or unstudied *new,* for a wordlist comprising 60 *old* and 60 *new* words. Each test trial was preceded by appearance of anticipatory cues of the form “Likely Old” and “Likely New”, which were valid in predicting the status of ensuing test items on ~70% of trials. Uninformative cues of the type “Likely ???” were also presented on a subset of trials in the “uncued” condition. Trial timings are in provided in seconds (s).

6.3. Results

6.3.1. Behavioural validation of likelihood cueing manipulation. The cueing manipulation was validated by behavioural analyses of sensitivity (*d′*) and response criterion (*c*) estimated from the signal detection model[14] (Green & Swets, 1966; Macmillan & Creelman, 2005). A one-way repeated-measures ANOVA was firstly conducted on *d′* with cue type as the factor ("likely old", LO; "likely new", LN; "likely ???", UC), yielding a non-significant main effect, $F < 1$. Participants' sensitivity in discriminating between *old* and *new* probe items was not influenced by the cueing manipulation (LO cue d' $M = 1.48$, $SD = 0.39$; LN cue d' $M = 1.49$, $SD = 0.44$; UC cue d' $M = 1.50$, $SD = 0.57$). The same one-way ANOVA on *c* did yield a significant main effect of cue type in the anticipated direction, such that criterion placement was most liberal in the LO cue condition ($M = -0.25$, $SD = 0.32$), most conservative in the LN cue condition ($M = 0.27$, $SD = 0.28$), and relatively neutral in the uncued condition ($M = 0.01$, $SD = 0.30$), $F(2,64)=37.25$, $p < .001$, $\eta_p^2 = .54$. Criterion placement was confirmed to significantly differ between these individual cue types by Tukey's Highest Significant Difference (HSD), all *ps* < .001. As with the variant of the likelihood cueing paradigm employed in Chapter 4 (see Section 4.3.3.1.) and prior research involving the same paradigm (e.g. O'Connor et al., 2010), the cueing manipulation shifted criterion bias so as to increase confirmation of the cued decision, without impacting on sensitivity.

[14] One further subject was removed from the *d′* and *c* analyses for not committing any incorrect "old" decisions (i.e. false alarms) in one of the cueing conditions, prohibiting the calculation of these measures for the cued phase. This participant was included in supporting analyses of the uncued trials as sufficient data was available. Whereas the protocol for dealing with such instances of missing data in previous experimental chapters has been to apply the Snodgrass & Corwin (1988) correction for errorless responding, this method was not favoured by the lab where I was placed at Washington University, and I hence did not apply it in this chapter.

6.3.2. Mean trial-locked pupil dilation following informative cues. Pupillometry analyses were first conducted on the pupil dilation (PD) response time-locked to the onset of each trial (i.e. appearance of the cue) and averaged for the informative cue conditions (i.e. "Likely Old", LO, and "Likely New", LN). To probe the selectivity of the PD response to processes directly contributing to correct decisions (as integrally proposed by "retrieval success" accounts), these trial-locked mean PD responses were calculated separately for correct and incorrect trials. Figure 2 plots the results of these analyses, with bootstrapped 95% confidence intervals for each timepoint in the mean PD plots shaded in grey. These intervals were calculated for each timepoint of the PD response via a resampling by replacement procedure implemented in the R statistical program (R Project for Statistical Computing, http://www.r-project.org/), wherein each PD timepoint was reconstituted over 34 random samples taken from participants' mean PD timecourses. The final confidence intervals were calculated from the standard deviation across 1000 iterations of this resampling by replacement procedure.

The resultant plots in Figure 6.2. clearly demonstrate the lack of any difference in the PD response between "old" and "new" decisions in the "Likely Old" cue condition that led subjects to expect *old* items, irrespective of decision correctness (Figure 6.2., left columns). In contrast, a pronounced difference was observed in the "Likely New" cueing condition that instilled expectations of encountering *new* items, with an increased PD response for "old" compared to "new" decisions (Figure 6.2., right columns). This elevated PD response onset several hundred milliseconds after appearance of the *old* item and lasted for the remainder of the trial. Furthermore, the dilation occurred for correct decisions (correct "old" i.e. "hits" > correct "new" i.e.

“correct rejections”, CRs) as well as incorrect decisions (incorrect “old” i.e. “false alarms”, FAs > incorrect “new” i.e. “misses”).

This visible difference under the LN cue was substantiated by submitting the mean PD response during the probe epoch (2000-5500 ms) to a 2 (cue type: LO or LN) x 2 (decision: “old” or “new”) ANOVA, conducted separately for correct and incorrect decisions. The ANOVA for correct decisions (Figure 2, upper panel) revealed significant main effects of decision and cue type, which were tempered by a significant decision by cue interaction, $F(1,33) = 9.64$, $p = .004$, $\eta_p^2 =.23$, $F(1,33) = 4.83$, $p = .035$, $\eta_p^2 =.13$, and $F(1,33) = 7.23$, $p = .011$, $\eta_p^2 =.18$ respectively. Consistent with Figure 6.2., Tukey’s HSD confirmed that the pattern of ANOVA results reflected a significant increase in PD for Hits compared to CRs *only* under the LN cue ($p =.002$), with no reliable difference observed for the LO cue (p ~ 1). The same ANOVA conducted for incorrect responses (Figure 6.2., lower panel) yielded only an interaction of decision by cue type, $F(1,32) = 4.61$, $p = .039$, $\eta_p^2 = .13$. The interaction was in the same direction as that observed for correct decisions, such that an elevated PD response for FAs compared to misses verged on significance under the LN cue ($p = .055$), but was clearly non-significant under the LO cue ($p > .95$, via Tukey’s HSD).

Overall, the trial-locked mean PD analyses reveal the high selectivity of the recognition dilation response, which differentiates “old” from “new” decisions *only* when familiarity is unexpectedly encountered in the environment (i.e. in the LN cue condition). The high dependency of the PD response on this specific form of expectation, and its emergence for both correct and incorrect decisions, both stand in opposition to prior reports implicating pupil dilation as a general marker of the successful retrieval of *old* item content.

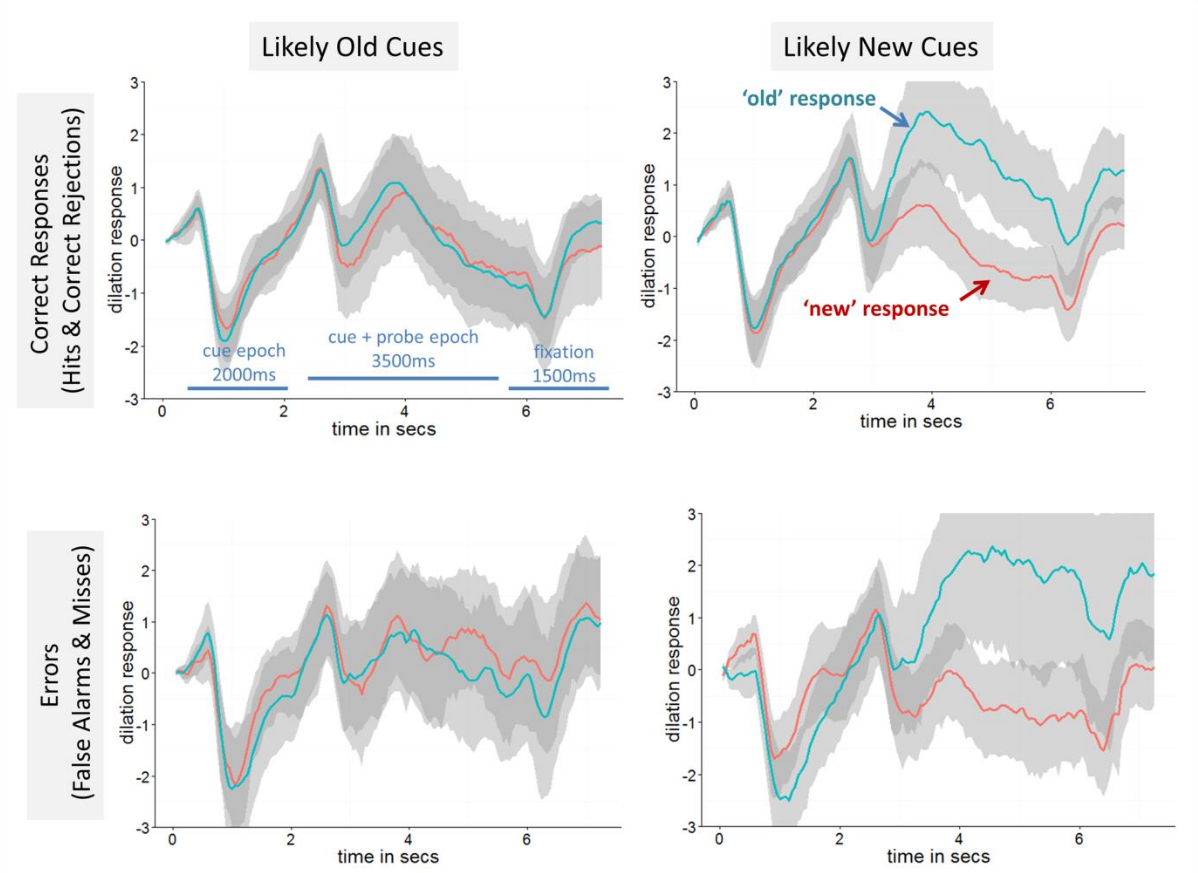


Figure 6.2. Trial-locked pupil dilation (PD) across cued trial types (left column = "Likely Old" cue, right column = "Likely New") and decision correctness (upper panel = correct responses, lower panel = errors). For all plots, blue lines denote the PD response during "old" decisions and red lines indicate PD during "new" decisions. The upper left panel also depicts the temporal onset of events within each trial (cue, probe and fixation epochs). Light grey areas denote the bootstrapped 95% confidence interval of the mean response across the 34 subjects, established at each sampled time point using 1000 replications.

6.3.3. Mean trial-locked pupil dilation following uninformative cues. Analyses of the trial-locked mean PD response were also conducted for the uncued condition (i.e. “Likely ???”, UC), wherein participants made recognition decisions in the absence of explicitly cued expectations. Figure 6.3. shows the PD response for uncued “old” and “new” decisions, averaged separately for correct (upper panel) and incorrect trials (lower panel). The bootstrapped 95% confidence intervals demonstrate that the uncued plots are noisier than the cued trial plots, and this is particularly apparent in the early period of incorrect uncued trials. Nevertheless, the upper panel depicting correct responses suggests an elevated PD response for hits compared to CRs during the probe epoch, although comparison of the mean response during this period did not yield significance, $t(33) = 1.37$, $p = .18$, $d = 0.24$. Consistent with the lower panel of Figure 6.3., the PD response for incorrect “old” and “new” decisions also did not significantly differ, $t < 1$.

One explanation of the lack of reliable differentiation between “old” and “new” decisions in the uncued condition is that these trials were overall lower in number than the informative cue conditions, and therefore subject to greater noise contamination. However, this view is challenged by the fact that trial counts were on average even lower for incorrect responses under the LN cue (mean number of incorrect LN cue trials = 30.97; mean number of correct UC trials = 43.47), yet a clear PD response difference was observed in this condition, $t(33)= 2.62$, $p = .013$, $d = 0.50$ (FAs > misses; see Figure 6.2., lower right). Hence, the lack of a reliable PD difference in the uncued condition might instead arise from a general weakening of the dilation response due to the lack of experimental control over participants’ expectations. In support of this view, the previous analyses of the informative cue conditions demonstrated that the PD response is minimal when observers are cued to expect *old* items and instilled with a confirmatory “old” bias (under LN cues, see Figure 6.2.). Similar expectations and resulting confirmatory “old” strategic biases are likely to have operated in unconstrained fashion the uncued phase, and therefore contributed to the general diminishment of the averaged PD response for this condition. These analyses therefore provide further support for the critical dependency of the recognition dilation response on encounters with unexpected familiarity.

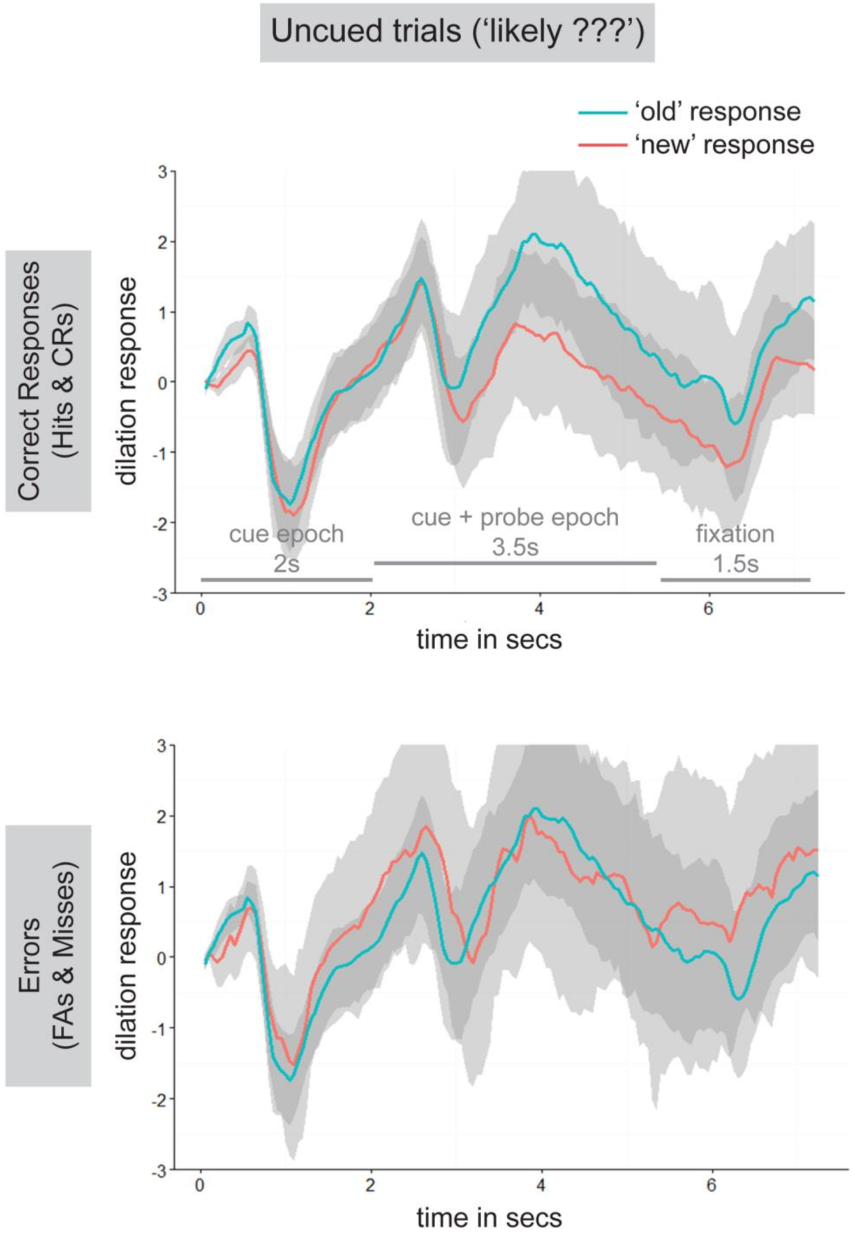


Figure 6.3. Trial-locked pupil dilation across uncued trial types for correct decisions (upper panel) and errors (lower panel). Blue lines denote PD response during "old" decisions and red lines represent the response for "new" decisions. The upper panel also depicts the temporal onset of events within each trial (cue, probe and fixation epochs). Light grey areas denote the bootstrapped 95% confidence interval of each PD response across the 34 subjects, established at each time point using 1000 replications. 'CRs', correct rejections; 'FAs', false alarms.

6.3.4. Mean response-locked pupil dilation following informative cues. The ensuing analyses focused on pupil dilation during the LN cue condition so as to further elucidate the temporal emergence of the unexpected familiarity effect. The mean response-locked PD response was calculated for the LN cue condition, spanning the 1500 ms period either side of participants' overt response. Figure 6.4. plots the results of these response-locked analyses, separately for correct (upper panel) and incorrect decisions (lower panel). The bootstrapped 95% confidence intervals in the figure indicate that the elevated PD response to unexpected "old" decisions onsets comfortably prior to commission of the actual response.

To confirm this impression, correct and incorrect decision trials were collapsed (given that prior analyses showed unexpected familiarity effects persisted regardless of decision correctness), and a series of paired samples t-tests were conducted to compare "old" and "new" decision mean response-locked PDs at each timepoint (at 50 ms sampling intervals). Type I error was controlled by adjusting p-values using the false discovery procedure (FDR; Benjamini & Hochberg, 1995). Figure 6.5. plots the results of these FDR analyses, which reveal that the earliest reliable difference in the PD response emerged 450 ms before the participants' button-press response. These analyses suggest that the unexpected familiarity PD response is closely tied to cognitive processes directly involved in the active decision-making process, which *precedes* commission of the overt response.

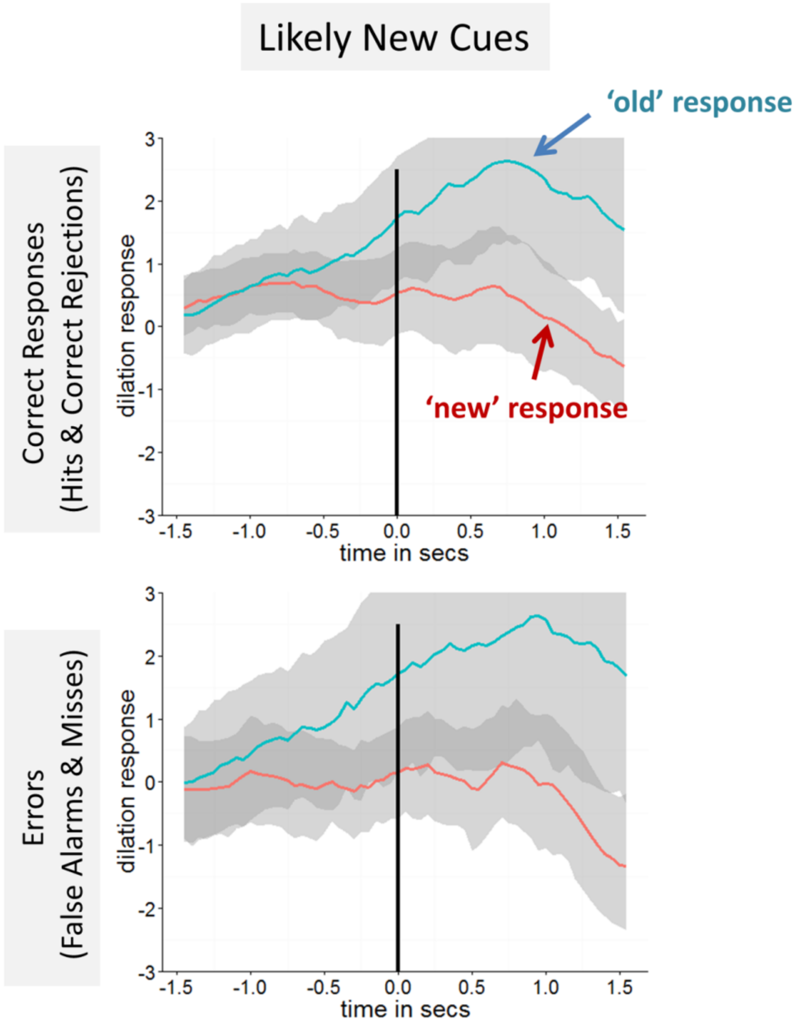


Figure 6.4. Response-locked PD response under the "Likely New" cue for correct decisions (upper panel) and errors (lower panel). Blue lines represent the PD response during "old" decisions whereas red lines indicate PD during "new" decisions. The vertical line marks the response of the participants at 0 seconds. Light grey areas denote the bootstrapped 95% confidence interval of each response across the 34 subjects, established at each time point using 1000 replications.

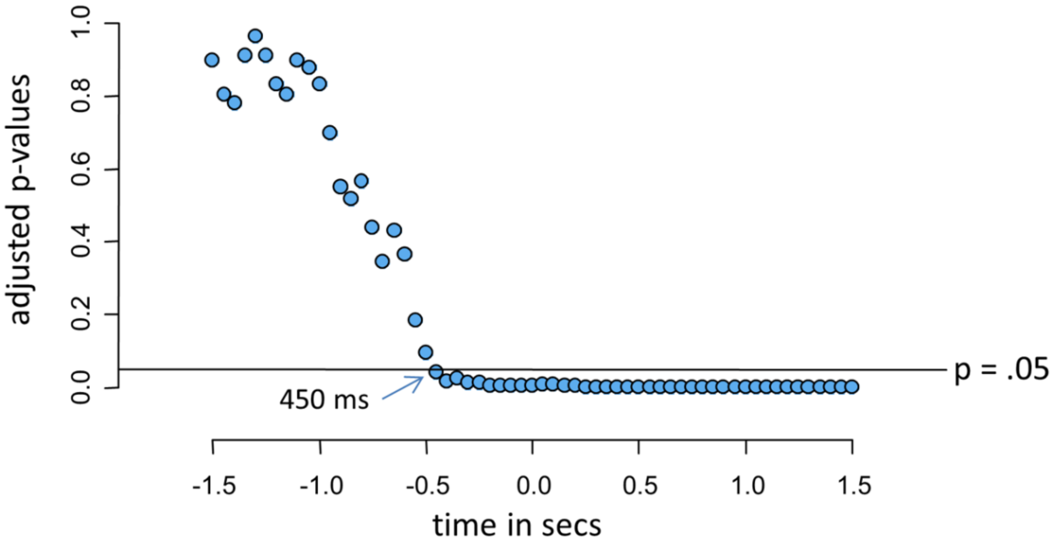


Figure 6.5. False Discovery Rate (FDR) adjusted p-values of the paired samples t-tests comparing response-locked pupil dilation for “old” versus “new” decisions under the “Likely New” cue condition (collapsed across correctness; sampled at 50 ms intervals). Participants’ responses onset at 0 seconds on the x axis and the horizontal line denotes the adopted significance threshold on the y axis of p = .05.

6.3.5. Multi-level modelling of trial-by-trial pupil dilation. The mean PD response analyses presented thus far suggest that pupil dilation during the evaluation of episodic memory is critically dependent on the experience of unexpected familiarity. This PD effect is observable irrespective of whether the experience of unexpected familiarity is accurate or not, as it occurs for both hits and FAs made under the LN cue. However, inspection of the trial-locked PD response plots in Figure 6.2. reveals a subtle differentiation between the PD morphologies for these categories of "old" decision. This is more clearly demonstrated in Figure 6.6., which plots the mean PD response during the LN cue for correct and incorrect "old" decisions time-locked to appearance of the probe. Although both hits and FAs displayed an increased amplitude that onsets around 1750 ms after appearance of the probe, the PD response for FAs returned to baseline at a slower rate than that for hits.

This variation in the morphology of the trailing PD slope might be related to the significant reduction in reaction time observed for hits (M = 1443.93 ms, SD = 237.74) compared to FAs (M = 1616.45 ms, SD = 291.75), $t(33) = 3.19$, p =.003, d = 0.55. Indeed, FAs are typically associated with slower reaction times and lower response confidence (e.g. Ratcliff & Murdock, 1976) – both characteristic behavioural signifiers of heightened response conflict and associated decision uncertainty (see also Chapter 5, Section 5.3.1.). The visibly greater positivity of the trailing PD slope for FAs compared to hits might therefore relate to the greater uncertainty associated with the former decision category. In this view, the more sustained positivity of the false alarm PD slope reflects a slower, more uncertain decision process, whereas the quicker return to baseline of the hit PD (i.e. with a more negative slope) reflects a quicker, less uncertain process. The variation in PD morphology between correct and incorrect "old" decisions made in the LN cue condition therefore suggests that the early amplitude of the dilation response and its later trailing slope might index different cognitive processes – namely unexpected familiarity and decision uncertainty respectively.

To test this process differentiation, a multi-level modelling (MLM) approach was adopted in which the early amplitude and trailing slope components of the PD response were contrasted by their ability to predict behaviour at the trial level. Measures of these PD components were firstly obtained via a simple linear regression fitted to the PD response over a period beginning 1750 ms in the probe epoch and ending at its offset (3500 ms after probe onset). This decomposed the trialwise PD response into an early amplitude estimate (corresponding to the intercept of the regression) and a later slope estimate (the slope of the regression), which were both entered in a series of MLMs predicting three aspects of behaviour – reaction time, accuracy (dummy-coded as 1 = “correct”, 0 = “incorrect”) and decision type (coded as 1 = “old”, 0 = “new”). Figure 6.7. shows the distribution of early amplitude and trailing slope estimates obtained from this linear decomposition – with both distributions spanning a range of positive and negative values obscured by the mean PD plots. The MLM method optimally models participant-level variation as a random effect (with a varying intercept component), and therefore serves as an improved method of isolating cognitive processes indexed by the PD response than multiple regression (in which participant-level variation is imperfectly modelled as a fixed effect). Furthermore, modelling at the single trial level ensured that any observed link between components of the PD response and aspects of behaviour could not be attributed to the averaging process (which is more greatly distorted by outliers and RT variability). The lme4 and lmerTest packages in the R statistical program were used to implement these MLM analyses.

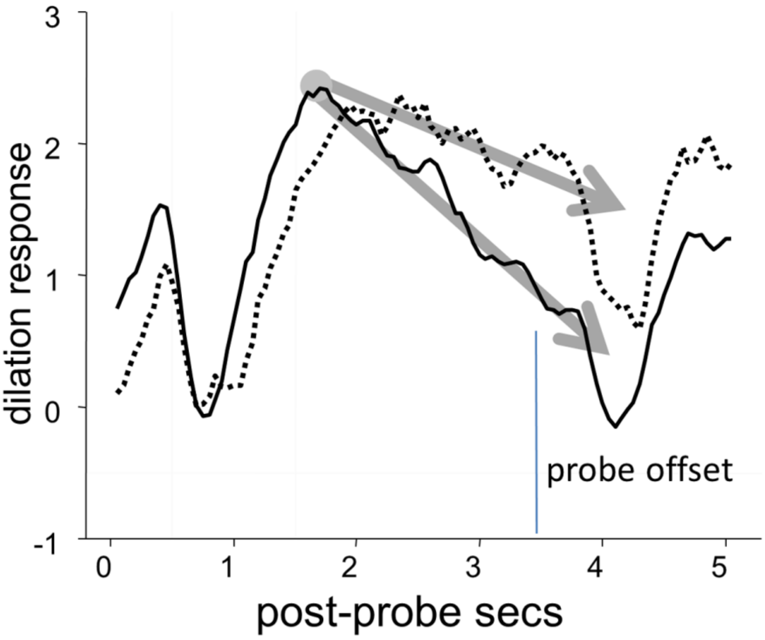


Figure 6.6. Comparison of the PD response for correct (solid line) and incorrect (dashed line) “old” decisions under the Likely New cue condition, locked to the onset of the probe at 0 seconds. The dilation responses appear to differ in the rate at which the early amplitude elevation, triggered by unexpectedly familiar items, returns to baseline. These morphological differences suggest that errors are associated with more positive-going PD slopes, raising the potential for the early amplitude and trailing slope of the dilation response to predict different cognitive processes at the trial level.

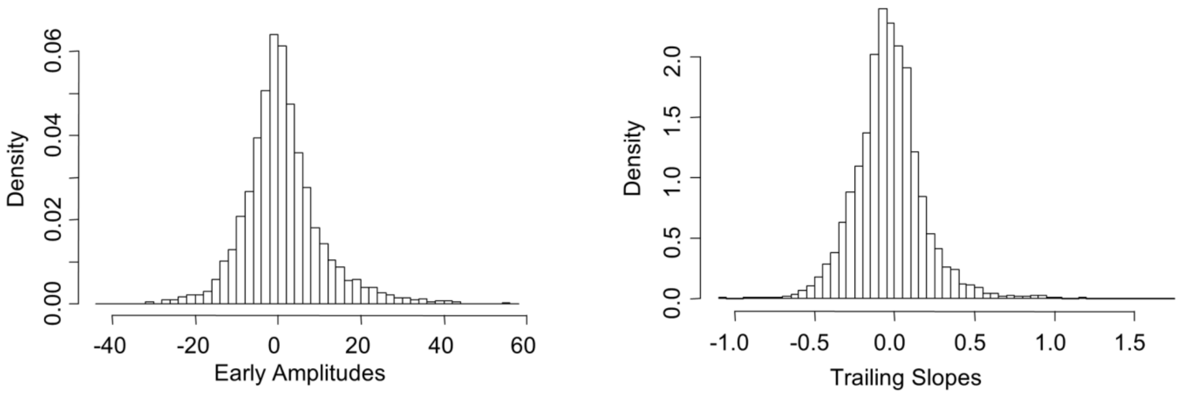


Figure 6.7. Distribution of trial-wise “early amplitude” and “trailing slope” values obtained from simple linear decomposition of the pupillary dilation response during the probe period (1750-3500 ms post-probe onset) into regression intercept and slope components respectively.

Reaction Time MLM

Table 6.1. displays the results of all MLM analyses. The first MLM predicted trialwise reaction time by the early amplitude (“Early Amplitude” in the table) and trailing slope components of the dilation response (“Trailing Slope”), separately for each informative cue condition (“Likely Old” and “Likely New”). Under both cue types, the trailing slope reliably and positively predicted reaction time whilst the early amplitude did not. Consistent with Figures 6.2. and 6.6., this finding confirms that greater positivity of the PD slope reflects an increasingly slow, uncertain decision process. In contrast, no relationship between the RT measure of decision uncertainty and the early amplitude of the PD response was observed.

Accuracy MLM

The next MLM focused on predicting trialwise accuracy, under the hypothesis that the PD slope would reliably anticipate the commission of errors. Indeed, the accuracy MLM converged with the reaction time MLM, with the trailing slope found to reliably and negatively predict accuracy, reflecting a link between greater PD slope positivity and a greater likelihood of incorrect responding for both informative cue types. Conversely, the early amplitude of the PD response did not predict accuracy. The PD components therefore dissociated across two separate behavioural measures of uncertainty, with increasingly positive slopes selectively linked to both slower reaction times and incorrect decisions.

Decision Type MLM

The final MLM included decision type (“old” or “new”) as the outcome variable and directly tested the hypothesis that the early amplitude component of the PD response would exclusively predict the rendering of “old” decisions under the LN cue – consistent with the unexpected familiarity effect recovered in the mean PD response analyses. In support of this prediction, the results in Table 6.1. demonstrate that the early PD amplitude reliably and positively predicted “old” responding *only* in the LN cue condition (in which familiarity was unexpectedly encountered). No relationship between the early amplitude component and decision type was observed in the LO cue condition (in which familiarity was expected). Interestingly, the PD slope component was also reliably predictive of decision type across both cueing conditions although in inverse fashion. Under the LN cue, the MLM estimate was positive (linking increased PD with “old” responding) whereas under the LO cue it was negative (linking increased PD with “new” responding). This finding is consistent with the proposed sensitivity of the PD slope to general decision uncertainty, given that slope positivity increased when making decisions that were in conflict with the available cue (i.e. for “old” decisions under LN cues, and “new” decisions under LO cues), and hence associated with greater uncertainty.

Overall, the MLM analyses confirmed the hypothesis that separate morphological components of the recognition PD response represent different memory control sub-processes. The early amplitude of the probe-provoked PD response was found to exclusively predict the experience of unexpected familiarity (i.e. an increase in early PD amplitude during the LN cue predicted an increased likelihood of responding “old”) and the later trailing slope predicted decision uncertainty (i.e. greater slope positivity predicted slower RT, reduced accuracy and the presence of cue-item conflict).

Table 6.1. Multi-level modelling of behavioral variables by pupil dilation components.

PREDICTING REACTION TIME					
Likely Old Cues	**Estimate**	**SE**	**~df**	**t**	**p-value**
Trailing Slope*	0.730	0.069	4821	10.59	<.001
Early Amplitude	0.001	0.002	4821	0.43	.664
Likely New Cues					
Trailing Slope*	0.712	0.068	3930	10.51	<.001
Early Amplitude	0.001	0.002	3934	0.89	.375
PREDICTING ACCURACY					
Likely Old Cues	**Estimate**	**SE**	**~df**	**t**	**p-value**
Trailing Slope*	-0.080	0.029	4764	-2.77	.006
Early Amplitude	0.000	0.001	4733	0.03	.973
Likely New Cues					
Trailing Slope*	-0.118	0.029	4650	-4.13	<.001
Early Amplitude	-0.001	0.001	4661	-0.87	.387
PREDICTING DECISION TYPE					
Likely Old Cues	**Estimate**	**SE**	**~df**	**t**	**p-value**
Trailing Slope*	-0.085	0.032	4820	-2.61	.009
Early Amplitude	0.000	0.001	4819	-0.23	.817
Likely New Cues					
Trailing Slope*	0.174	0.032	4769	5.35	<.001
Early Amplitude*	0.005	0.001	4775	6.78	<.001

Note: Displayed are the results of trialwise multi-level modelling analyses involving the linearly decomposed ‘early amplitude’ and later ‘trailing slope’ components of the pupil dilation response (from 1750-3500 ms post probe-onset). Asterisks denote multi-level model estimates that are significantly different from 0 via one-sample t-tests; ‘SE’, standard error of the model estimate; ‘~df’, approximate degrees of freedom.

6.4. Discussion

The present experiment probed the relationship between memory control processes and autonomic nervous system activity, as indexed by changes in pupil dilation. The findings complement the insight into cortical memory control systems provided in the previous chapter, whilst also refining understanding of the general cognitive underpinnings of the pupil dilation response itself. The results of this first attempt at explicitly manipulating mnemonic expectation in a pupillometry study of recognition memory pose another challenge to the “retrieval success” interpretation of old/new functional neuroimaging effects. Rather than directly signalling the successful recovery of episodic content for *old* items, the observed pattern of trial-locked PD response instead favoured a highly specific role in signalling unexpected familiarity (with dilation amplitude increasing exclusively for “old” decisions made under “Likely New” cues). The subsequent multi-level modelling of the trialwise PD response revealed that the averaged dilation response in fact reflects the contribution of two dissociable memory control sub-processes, each associated with a distinct feature of PD morphology; namely, unexpected familiarity linked to an early amplitude component, and general decision uncertainty linked to a later trailing slope component. The findings raise a number of implications for prevailing theories of adaptive memory function and its autonomic correlates.

The prevailing view is that the recognition pupil dilation response is critically linked to the heightened cognitive load associated with processing successfully retrieved episodic content (Goldinger & Papesh, 2012; Papesh et al., 2012; Vo et al., 2008). The present experiment suggests that this basic interpretation is untenable, given that the pupil old/new effect was fully eliminated when observers were cued to expect familiar materials (i.e. under the LO cue). Rather, the amplitude of the mean trial-locked PD response only reliably increased when familiarity was unexpectedly experienced in the test environment (i.e. under the LN cue; see Section 6.3.2. and Figure 2). Additionally, this PD response was evoked for both correct and incorrect unexpected “old” decisions, providing further evidence against a role in representing processes that directly contribute to successful or accurate recognition decision-making. To the extent that mnemonic expectations were not under experimenter control in previous recognition pupillometry studies, and have been shown to vary with aspects of standard recognition testing environments (Hicks & Marsh, 1999; see also Chapters 2, 3 and 4) and as a function of individual differences (Aminoff et al., 2012), the preferential sensitivity of the PD response to unexpected familiarity has been hitherto unnoticed. The findings therefore coalesce with research challenging the retrieval success interpretation of recognition old/new effects spanning a number of functional neuroimaging modalities (Herron et al., 2003; Herron et al., 2004; Jaeger et al., 2013; O’Connor et al., 2010), including the multi-modal neuroimaging experiment described in the previous chapter.

The highly specific nature of the observed expectancy violation effect also suggests a number of refinements to conceptions of the general cognitive underpinnings of the PD response. For instance, previous assertion that the PD response primarily reflects “post-decision” consolidation processes (e.g. Einhauser, Koch & Carter, 2010) cannot be considered to apply here, as the increased dilation to unexpected familiarity began comfortably prior to commission of the physical response (as revealed by the response-locked analyses in Section 6.3.4). The results also show that changes in pupil dilation cannot *solely* reflect task effort or difficulty (e.g. Beatty, 1982), given that the MLMs revealed the insensitivity of the early amplitude component to decision uncertainty, as reflected in slower reaction times and more error-prone decision-making (see Table 1). Rather, this early amplitude component was functionally dissociated from the later trailing slope of the PD response, which did show a positive relationship with behavioural markers of uncertainty (i.e. slow responding, erroneous responding, and responding under cue-item conflict).

Furthermore, the specificity of the unexpected familiarity response also casts doubt on prior suggestion that pupil dilation serves as a “general” expectation violation or “surprise” signal (Friedman et al., 1973; Preuschoff et al., 2011; van Olst et al., 1979). The present data instead accords with fMRI evidence of distinct neural signatures underlying different expectancy violation sub-processes (Jaeger et al., 2013; O’Connor et al., 2010). In support of this, a recent fMRI study by Jaeger and colleagues (2013) isolated three distinct cortical networks underlying the invalid cueing effect (i.e. evoked when memory decisions are made against cued expectations), two of which mirror the observed early amplitude and trailing slope components of the PD response. With relevance to the former component, the authors identified an “unexpected familiarity” network that activated only when “old” decisions were made under LN cues, comprising prefrontal, left lateral parietal and posterior cingulate regions. These regions have been reliably linked with effortful source monitoring operations (e.g. Dobbins et al., 2002), leading Jaeger et al. to suggest that encounters with unexpectedly familiar items trigger controlled source memory operations that attempt to identify their episodic origin. Whilst the adaptive value of engaging this confirmatory source monitoring process would be high when confronted with an unexpectedly familiar stimulus, no such benefit would be associated with the engagement of similar operations for unexpectedly novel stimuli, given that these present no potential to retrieve further episodic content. The source monitoring interpretation of the unexpected familiarity effect is therefore able to account for the complete insensitivity of the observed PD response to experiences of novelty, whether expected/unexpected or accurate/inaccurate. Although future research will be necessary to test this source monitoring interpretation of the early amplitude response, the present findings nevertheless highlight fundamental differences in how unanticipated familiarity and novelty are oriented towards in the environment.

Regarding the trailing slope component of the PD response, Jaeger and colleagues (2013) also identified a more “general” expectancy violation response, in which BOLD activation increased whenever decisions were made under conflict with available cues, irrespective of old/new status. This network comprised bilateral intraparietal, lateral and medial prefrontal regions, consistent with the literature linking these regions to the engagement of cognitive control under different forms of response conflict (Bunge, Dudukovic, Thomason, Vaidya & Gabrieli, 2002; Kouneiher, Charron & Koechlin, 2009; Miller & Cohen, 2001; see also Section 1.2.). Whilst the early amplitude component is proposed to underlie source monitoring operations aimed at confirming the sense of familiarity, the later slope component might be linked with cognitive control processes performing a similar cross-domain function as scrutinized in Chapters 4 and 5. This function pertains to the controlled evaluation of information in favour of both the expected decision (associated with the anticipatory cue) and the actual strength of the retrieved memory evidence, so as to diagnose the eventual response.

Indeed, the demonstration of a reliable effect of decision uncertainty on the PD response addresses an enduring ambiguity as to whether pupil dilation is sensitive to uncertainty at all. This is as the link between pupil size and uncertainty has thus far not been reliably substantiated, with some prior studies reporting an increase in pupil size under uncertain task conditions (Nassar et al., 2012; Friedman et al., 1973; Satterthwaite et al., 2007) whereas others have failed to elicit any clear relationship with uncertainty (Schacht, Dimigen & Sommer, 2010) or one that is moderated by other cognitive factors (Preuschoff et al., 2011). Indeed, when considering the mean dilation response over the entire probe epoch in the present study (and ignoring its morphological components), the only reliable effect recovered was the PD response for unexpected familiarity. The direction of this mean PD response was also converse to what one would predict based purely on an uncertainty account, given that "new" decisions are typically associated with less confidence and slower reaction times (and hence more putative uncertainty) than "old" decisions (Otero et al., 2011; Ratcliff & Murdock, 1976). Rather, the influence of judgment uncertainty was only successfully isolated in the present study by decomposing the trial-wise dilation response into early amplitude and trailing slope components, and subsequently using MLM analyses to elucidate their separate influences on behaviour. This represents a different approach to elucidating the contribution of decision uncertainty from previous pupillometry studies relying on process dissociations between pre-trial baseline PD and task-evoked PD effects (e.g. Nassar et al., 2012).

Overall, the findings highlight that the PD response is not a "process-pure" measure of cognitive processing, and instead reveal for the first time that different morphological aspects of the same task-evoked PD response are representative of different cognitive processes. Interpreting the precise functions sub-served by the dissociable early amplitude and trailing slope PD components requires further research (see Chapter 7, Section 7.5 for further details.), however previous findings would indicate that the former relates to source monitoring operations whereas the latter relates to evaluative control processes (with potentially cross-domain properties). Both processes are putatively required to generate an adaptive response to conditions of environmental conflict. This interpretation is also parsimonious with the previously outlined dichotomy between domain-specific “retrieval control” processes and cross-domain “decision control” processes (see Chapter 1). Irrespective of the validity of these more precise characterisations, the present data clearly indicate that two aspects of memory control are dissociable by their autonomic nervous system correlates. These dissociable correlates potentially reflect the gain modulation of distinct cortical networks via upstream projections of the LC-NE neuromodulatory system that is strongly associated with the pupil dilation response (as outlined in the introduction; Aston-Jones & Cohen, 2005).

This final experimental chapter serves as the culmination of the interrogation of memory control processes conducted during the course of my PhD, incorporating behavioural and functional neuroimaging experiments spanning episodic and non-episodic task domains. The overarching themes emerging from these empirical studies are summarised and discussed in the final chapter.

Chapter 7: General Discussion

7.1. Overview

The primary aim of this thesis was to investigate the behavioural indices and functional neuroimaging correlates underpinning the regulation of episodic memory control. The preceding experimental chapters were motivated by a desire to go beyond the traditional research focus on characterising the nature of retrieval output, and scrutinize factors that constrain these retrieval processes in light of immediate environmental demands. This endeavour sought to bring the laboratory study of episodic memory more in keeping with real-world evaluative experience, in which the assessment of the strength of retrieved memories interacts with varying strategic influences in the test environment – cued expectations and emphasised goals prominent among them – to determine eventual memory decisions. An auxiliary aim was to segregate control processes that operate specifically in episodic memory ("domain-specific" control) from those that extend beyond this cognitive domain ("cross-domain" or "cross-task" control). The conducted experiments therefore employed a number of behavioural and functional neuroimaging approaches, spanning episodic and non-episodic evaluative domains. The key findings from each empirical chapter are summarised in Table 7.1., and discussed in greater detail in the following sections.

Table 7.1. Summary of experimental chapters presented in this thesis, detailing the implemented methods, evidence domains under evaluation, strategic control manipulations and major findings.

Chapter no.	**Experiment no.**	**Method**	**Domain**	**Manipulation**	**Major findings**
2	1-3	Behavioural	Episodic recognition only	Implicit goal emphasis (“old?” vs “new?”)	Question format instils a “counter-emphasis bias”, reducing endorsements of the decision emphasised by the question
3	4	Behavioural	Episodic recognition, episodic temporal judgement, semantic judgement, colour discrimination	Implicit goal emphasis (e.g. “old?” vs “new?”)	Question format reduces the proportion of emphasised decisions made at low confidence, across different cognitive domains
4	5	Behavioural	Episodic recognition and semantic judgment	Explicit goal emphasis (e.g. “old” token vs “new” token vs No token emphasis) and cued expectation (e.g. “likely old” vs “likely new” cues)	Explicit goal emphasis and cued expectation lead to independent and opposing biases (acting in disconfirmatory “counter-emphasis” and confirmatory “pro-expectation” directions respectively), which manifest across different cognitive domains

5	6	EEG-fMRI	Episodic recognition and semantic judgment	Cue-item match/mismatch (e.g. for “old” cues and “new” cues that matched status of presented items on 50% of trials)	Cue-item conflict, instantiated across episodic and semantic domains, and across “cue first” and “item first” orders, elevates the amplitude of the late positive ERP component (LPC) and BOLD activation in the rostral cingulate zone (RCZ) and intraparietal sulcus (IPS)
6	7	Pupillometry	Episodic recognition only	Cued expectation (“likely old” vs “likely new” vs “likely ???” at ~70% validity)	Control recruited under violations of memory expectation leads to two separable effects on autonomic activity, as reflected in components of the pupil dilation (PD) response – an amplitude elevation associated with the experience of unexpected familiarity, and a slope positivity linked with general decision uncertainty

7.2. Environmental goal emphasis influences overt recognition performance

One of the main experimental aims outlined in the introductory chapter (Section 1.4.) was to examine how different aspects of the test environment influence memory evaluation via the instilment of higher-order strategies. This follows from a traditional neglect in interrogating such strategic effects in the field of episodic memory, which has predominantly focused on characterising retrieval output processes that underlie the assessment of memory strength (viz. the single process vs. dual process debate; Squire, Wixted & Clark, 2007; Yonelinas, 2002). Those studies that have examined strategic influences on memory evaluation have focused almost exclusively on the effects of environmentally cued expectations, which provoke confirmatory biases that increase the likelihood of endorsing the expected decision (Ratcliff, Sheu & Gronlund, 1992; O'Connor, Han & Dobbins, 2010; Rhodes & Jacoby, 2007). As demonstrated in the introductory chapter (see Section 1.1.), cued expectation is but one aspect of the test environment capable of influencing the outcomes of memory evaluation, with environmentally emphasised goals serving as another prominent strategic influence.

It was in rectifying the broad neglect in researching goal emphasis strategies that a subtle aspect of the test environment was empirically scrutinised, namely the format of the question presented to prompt an evaluation of episodic recognition. The experiments of Chapter 2 examined whether goal salience was imparted to available decision categories ("old" or "new") merely by virtue of which of these was emphasised by the wording of the test question. Evidence of this question emphasis effect was provided across three separate experiments, in the form of shifts in SDT criterion that reflected a reduced likelihood of endorsing the decision category emphasised by the question. The direction of criterion bias was converse to expectation-driven biases observed in recognition memory, as well as question format biases in other fields, such as the acquiescence bias in questionnaire responding (“yea-responding”; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) and the Loftus framing effect in eyewitness testimony (Loftus, 1996; see Section 2.5.). These previously reported biases all act in confirmatory directions, and hence the observation of a novel disconfirmatory criterion shift supports a role for SDT criterion as a general index of various environmental strategies (Hirshman, 1995; Benjamin, 2007).

The conducted experiments not only demonstrated the reliability of the question emphasis bias (which was replicated in Experiments 2 and 3, as well as in an independent online study; Mill & O’Connor, 2014, Experiment 1, see Appendix A), but also clarified its core strategic underpinnings. Experiment 2 yielded an enhancement of the question emphasis bias under low levels of memory strength, consistent with the up-regulation of a higher-order strategy in the presence of weakly diagnostic memory evidence (see the ensuing section for further discussion of this effect). Experiment 3 tested two competing interpretations of the precise mechanism underlying the question bias – a habitual (potentially non-strategic) tendency to respond “no”, and an emphasis effect imparted by the combination of question format and available response options – finding support for the latter goal-driven mechanism. Furthermore, the criterion effects were complemented by analyses of decision accuracy, which linked the reduced likelihood of endorsing emphasised decisions with the greater accuracy of these decisions when made (although see the following section for possible asymmetries in how emphasis impacts on particular decision categories). Indeed, both the criterion bias and decision accuracy effects observed following the “implicit” goal emphasis manipulations of Chapter 2 were paralleled by the findings of Experiment 5 (Chapter 4), in which goal emphasis was manipulated by the "explicit" provision of monetary mixed incentives. The overall pattern of findings highlights the salient influence of environmentally emphasised goals in constraining the outcome of memory evaluation – both when such emphasis is imparted by implicit aspects of the test format and by explicit mention of reward. The results also illustrate response bias as an important cognitive factor in its own right, rather than as a nuisance variable obscuring memory strength processes indexed by *d'* sensitivity (as assumed in early retrieval output research, e.g. Blackwell, 1963).

The observed goal-driven shifts in criterion bias therefore serve as a formalized index of the strategic mechanisms alluded to in Koriat & Goldsmith's influential memory control model (1996; see Section 1.3.). This stipulates that the primary method by which cognitive control is heightened over episodic memory is by maximising the quality rather than the quantity of endorsed memories, as evidenced by the previously reported inability to increase the number of memories generated during a recall test under performance incentives (Nilsson, 1987), which contrasts with the ability to regulate the quality of memory decisions in "free" compared to "forced" recall report procedures (Koriat & Goldsmith, 1996). This finding is paralleled by the present effects of implicit and explicit goal emphasis on SDT criterion and decision accuracy, which also reflect a reduction in the frequency of emphasised responding so as to maximise accuracy. Environmental sources of goal emphasis therefore instil a higher-order strategy encouraging more cautious use of available memory strength evidence, which was highlighted in the introductory chapter as a salient form of "decision control" exerted over real-world memory evaluation (see Section 1.3.).

It is worthwhile highlighting that *d′* sensitivity was largely unaffected by both implicit and explicit manipulation of goal emphasis, consistent with its proposed role within the SDT framework in indexing the strength of retrieved memories (Glanzer & Adams, 1985; Lockhart, Craik & Jacoby, 1976; see Section 1.4.). However, the results provide some indication of an improvement in sensitivity under “old” emphasis, both following implicit question emphasis (Section 2.3.2.1.) and explicit mixed incentives manipulations (Section 4.3.2.1.). Complementary decision accuracy results suggested an overall performance improvement under “old” emphasis, raising the potential involvement of source monitoring processes underpinning the controlled search for specific details of an encoding event (Johnson, Hashtroudi & Lindsay, 1993; Jacoby, Kelley & McElree, 1999). Source monitoring processes might conceivably be heightened when the need to diagnose “old” decisions is emphasised by implicit and explicit factors within the immediate environment. Indeed, previous studies have documented the heightening of source monitoring under often subtle manipulations of test format, such as when retrieval is framed as a source memory rather than old/new recognition test (Lindsay & Johnson, 1989), and when the response format contains more “old” than “new” options (Marsh & Hicks, 1998; Dobbins & McCarthy, 2008; see also Section 2.5.). Hence, future research will be required to determine whether source monitoring mechanisms associated with heightening the control over the direct act of retrieval (i.e. “retrieval control”) are indexed by the SDT sensitivity parameter. Another possibility is that the present “old” emphasis effects on sensitivity uncover nuances in memory control that are imperfectly captured by the SDT model. Indeed, applications of the drift diffusion model to recognition memory have operationalized two distinct bias parameters, which have been dissociated by their involvement in source monitoring/retrieval control and strategic/decision control processes respectively (Ratcliff, 1978; White & Poldrack, 2014). It might therefore be of future interest to probe the correspondence of goal emphasis effects on bias parameters in the SDT and drift diffusion models.

Future research might also directly interrogate the application of goal emphasis effects in real-world eyewitness testimony. With particular relevance to the implicit question emphasis effects of Chapter 2, laboratory studies might examine whether asking differently worded questions does bias memory evaluation in more naturalistic renditions of eyewitness testimony (e.g. in virtual reality courtroom paradigms, Bailenson et al., 2008). It would also be worthwhile to probe the involvement of other higher-order strategies documented in the field of cognitive control in episodic memory evaluation, such as sequential dependencies between particular memory events (Mozer, Kinoshita & Shettel, 2007) and the monitoring of recognition performance accuracy (Kerns et al., 2004).

Nevertheless, the present findings highlight an overlooked source of strategic bias embedded in standard laboratory recognition testing environments, which was linked through systematic investigation with goal and reward-related processes involved in memory control. SDT criterion served a crucial role in revealing these behavioural effects (as supplemented by decision accuracy measures), and the potential for criterion to index similar strategic control processes beyond the specific domain of episodic memory is discussed in the ensuing section.

7.3. Response criterion as a behavioural index of cross-domain control

Another aim of the conducted behavioural experiments was to examine whether environmental manipulations that instilled higher-order strategies during episodic evaluation also introduced analogous effects in non-episodic domains. This was part of a wider aim of my PhD to probe the existence of a cognitive system with “cross-domain” properties, capable of integrating strategic influences into the evaluation of different evidence domains (see Section 1.6.). Such a cross-domain system has been implicated in a number of theoretical models, spanning attention (Cabeza, Ciaramelli, Olson & Moscovitch, 2008; Corbetta & Shulman, 2002), metacognition (Fleming & Dolan, 2012) and cognitive control (Botvinick, Braver, Barch, Carter & Cohen, 2001; Ridderinkhof, Ullsperger, Crone & Nieuwenhuis, 2004), although rarely empirically verified. The present behavioural studies therefore examined the potential cross-domain effects of implicit (Experiment 4, see Table 7.1.) and explicit manipulations of goal emphasis (Experiment 5), as well as environmentally cued expectations (Experiment 5). Insight into this cross-domain system would yield reciprocal clarification as to whether control processes underlying strategic effects in episodic memory are specific to this domain or more “general” in their operation, which has been the subject of ongoing debate in the literature (Barredo, Oztekin & Badre, 2013; Fleck, Daselaar, Dobbins & Cabeza, 2006; White & Poldrack, 2014).

However, the question emphasis manipulations of Experiment 4 failed to yield reliable cross-domain effects on “primary” measures of yes/no decision performance (SDT criterion, decision accuracy and a newly introduced measure of decision RT). A likely cause of these broadly weak effects is the “implicit” nature of the question format manipulation, which even in Chapter 2 yielded numerically small (albeit statistically reliable in Experiments 2 and 3) strategic adjustments in behavioural performance. The small nature of the question emphasis effect was likely exacerbated in Experiment 4 by exposing participants to tasks involving varying levels of overall evidence strength, which might have reduced both the scope for strategic influences in cases of high overall sensitivity (as in the colour task) and the potential influence of strategies related to evidence monitoring. The latter evidence-based strategies have been demonstrated to impact on criterion in a previous recognition study (Hirshman, 1995), in which manipulations influencing the perception of “strong” and “weak” encoding events respectively instilled disconfirmatory and confirmatory “old” response biases. Similar biases engaged by assessments of the varying evidence domains in Experiment 4 might have contaminated the question emphasis effects of primary interest.

Indeed, when the variation in evidence strength across tasks was controlled in the “secondary” response confidence analyses of Experiment 4, a reliable cross-domain reduction in “low” confidence decisions made under emphasis emerged. Under the assumption that “low” confidence decisions reflect low levels of evidence strength (following from assumptions in previous functional neuroimaging studies, Fleck et al., 2006; Fleming, Huijgen & Dolan, 2012), these confidence shifts parallel the observed enhancement of the question emphasis bias under shallow levels-of-processing in the Experiment 2 recognition task (see Section 7.2.). Collectively, the findings suggest that both effects might be driven by a cross-domain mechanism sensitive to the up-regulation of higher-order strategies under evidence uncertainty (Benjamin, 2007; Kahneman, Slovic & Tversky, 1982).

In comparison to the modest effects of Experiment 4, the more “explicit” manipulation of goal emphasis in Experiment 5 (via mixed incentives; see Table 7.1.) led to the instilment of a stronger goal emphasis strategy that was reflected in reliable “counter-emphasis” criterion shifts across episodic and semantic domains. Furthermore, this cross-domain goal emphasis effect was isolated in the presence of orthogonal manipulation of cued expectation. Both resultant “counter-emphasis” and “pro-expectation” biases exerted independent and opposing strategic effects on response criterion, highlighting the sensitivity of this behavioural measure to various environmentally provoked strategies. The findings of Experiment 4 and 5 also confirm that only explicit environmental manipulations sufficiently motivate participants to adopt higher-order strategies across different cognitive domains. Participants’ willingness to adopt higher-order strategies has been previously highlighted as a primary mediator of the modest biases observed even after explicit manipulations (e.g. expectation via base rate manipulations, Stretch & Wixted, 1998b). Furthermore, the demonstration of equivalent strategic effects across different task domains supports a role for criterion as a behavioural index of cross-domain control – a function that has been consistently speculated in formalizations of the signal detection framework in memory and perceptual domains (Benjamin, 2007; Swets, 2014), although rarely tested by exposing the same sample of subjects to different evaluative domains.

Whilst SDT criterion was the focal index of cross-domain processes in Experiment 5, supplementary analyses of decision accuracy and decision RT elucidated more nuanced strategic effects. For instance, the pattern of criterion effects in both tasks suggested a deviation in bias from the neutral emphasis condition (i.e. “No token” emphasis) primarily when “Decision A” categories were emphasised by congruent incentives (i.e. “pleasant” semantic decisions and “old” recognition decisions). The decision accuracy and RT results corroborated this finding, by demonstrating that only “Decision A” categories were associated with reliable increases in accuracy and decreases in RT under concordant emphasis, whereas the alternative “unpleasant” and “new” decisions were less reliably affected. Two potential causes can be suggested for these preferential “Decision A” effects. Firstly, the evaluations of “old” and “pleasant” decisions might engage fundamentally different evidence strength processes than their respective “Decision B” alternatives. This interpretation is plausible for the “old” emphasis effects observed in the recognition task, given prior suggestion that a primary method of heightening episodic memory control is by prioritising the recovery of recollected content (which is specific to *old* items; Jacoby et al., 1999; see Section 1.3.). However, no recollection-type process has been linked with the evaluation of “pleasant” semantic decisions, and hence the fact that “Decision A” effects were common to both tasks challenges this evidence-based interpretation.

A second more plausible explanation is linked to the static response format maintained throughout Experiment 5, with “Decision A” assigned to the “yes” response key and “Decision B” assigned to the “no” key. The fixed assignment of “Decision A” categories to “yes” response status raises the potential for additional emphasis imparted by the response format to these decision categories. Hence, the effect of the “Token A” incentives conditions and the static “yes” assignment to “Decision A” might have combined additively to enhance the goal salience ascribed to “Decision A” categories. In support of this “response emphasis” view, Experiment 3 (see Section 2.4.2.) demonstrated that it was the combination of available question and response formats that mediated goal emphasis effects. Furthermore, a recent pupillometry study reported an exclusive increase in pupil dilation (and underlying autonomic nervous system activity) when participants performing a visual discrimination task countermanded a pre-potent bias against responding “yes” (de

Gee, Knapen & Donner, 2014). This “counter-yes” bias acted in a similar disconfirmatory direction as the present “Decision A” emphasis effects and hence might represent the instilment of a common goal-driven strategy (see Section 7.5. for further details). Despite these supportive findings, the response emphasis interpretation would benefit from direct investigation of whether ascribing “yes” status to “Decision B” categories leads to a reversal in preferential token emphasis effects. Nevertheless, the observed decision emphasis asymmetries allude to further strategic influences engaged by subtle aspects of the test environment. The findings also highlight the virtue of examining effects of the same environmental manipulations across different task domains, so as to adjudicate between domain-specific and domain-general characterisations of strategic biases.

The decision accuracy and decision RT analyses of Experiment 5 also illuminate a subtle differentiation in the cross-domain control processes engaged by the counter-emphasis and pro-expectation biases. The pattern of criterion shifts underlying both environmentally provoked strategies suggests a common heightening of control over the decision category associated with a disconfirmatory bias. To clarify, both “old” decisions made under “old” emphasis and “old” decisions made under “likely new” expectations are made under a positive shift in criterion, and hence likely require greater cognitive control to countermand the instilled disconfirmatory “old” response tendencies. However, “old” decisions made under emphasis were associated with greater accuracy and quicker reaction times, whereas “old” decisions made in violation of expectations were associated with reduced accuracy and slower reaction times. Similar response profiles have been reported under analogous manipulation of higher-order goals (Small et al., 2005; Han et al., 2010) and expectations respectively (Corbetta, Kincade, Ollinger, McAvoy & Shulman, 2000; O’Connor et al., 2010). These findings raise the potential for cross-domain effects associated with goal emphasis and cued expectation to be mediated by at least partially distinct cognitive control processes.

In characterising this potential distinction, one possibility is that goal emphasis is mediated by a sustained or “proactive” control process, monitoring for evidence in favour of a rewarded or goal salient decision, whereas expectancy violation is characterised by a phasic or “reactive” control process, engaged under conflict between accumulated evidence and environmentally provoked strategies. This dichotomy would explain the different decision accuracy/RT profiles underlying emphasis- and expectation-driven biases, and is consistent with equivalent “dual process” stipulations of a number of cognitive control frameworks (Aston-Jones & Cohen, 2005; Braver, 2012; Corbetta, Patel & Shulman, 2008). Functional neuroimaging methods would be useful in probing the application of dual process control frameworks to the present cross-domain strategies. Whilst the findings of Experiment 6 provided insight into the neural substrates of expectancy violation-like control engagements (i.e. under cue-item conflict, see the ensuing section for further details), time constraints imposed by my PhD prevented similar functional neuroimaging investigation of goal emphasis processes. Hence, it remains for future research to probe the overlap of neural substrates underlying emphasis- and expectation-driven control, and how these neural patterns might relate to the common and distinct behavioural profiles reported in this section.

To summarise, the findings of Experiments 4 and 5 demonstrate that cross-domain effects of environmentally provoked strategies are isolable in overt behaviour, especially following explicit manipulations. The latter experiment also demonstrated that different strategies associated with goal emphasis and cued expectation respectively lead to independent and opposing effects on SDT criterion, across different cognitive domains. This strengthens the functional role ascribed to criterion as a behavioural index of cross-domain control, although a more complete insight into the engaged strategies was furnished by complementary analyses of decision accuracy and decision RT. Whilst these latter analyses highlight nuances in the precise control processes engaged by the emphasis- and expectation-driven strategies warranting future exploration, these behavioural experiments nonetheless laid important foundations for the functional neuroimaging studies conducted during my PhD. The final two sections discuss the main findings emerging from these studies.

7.4. Multi-modal neuroimaging isolates the cortical correlates of memory control

The investigation of memory control processes conducted in the preceding behavioural experiments was extended in the functional neuroimaging approach of Experiment 6. The primary motivation of this experiment was to elucidate the precise functions of neural activations linked in previous research to broad aspects of episodic memory control. More specifically, the experiment sought to separate the neural substrates of memory control processes that operated exclusively in the episodic domain ("domain-specific" or "recognition-specific" control) from those that performed a more general function in the service of cognitive control ("cross-domain" or "cross-task" control). This followed from findings of the previously described behavioural experiments, which despite demonstrating cross-domain strategic effects isolable in over behaviour, were unable to clarify whether these common behavioural effects arose from common or distinct neural computations. A multi-modal neuroimaging approach was hence adopted wherein EEG and fMRI data were simultaneously acquired as participants performed discrimination tasks, spanning semantic and episodic recognition domains, in which cognitive control was manipulated by the match/mismatch procedure (see Table 7.1.). This multi-modal approach enabled investigation of electrophysiological and hemodynamic correlates that have been debated as to their domain-specificity or generality.

Analyses of the standalone EEG data focused on an ERP component scrutinised in research into both recognition and semantic memory – the late positive potential (LPC). In both these domains, ERP studies have sought to adjudicate between “domain-specific” and “domain-general” interpretations of the LPC, with a domain-specific function of the recognition LPC suggested in signalling the recollection of study details (as a sub-component of the ERP “retrieval success” effect; Rugg & Curran, 2007), whereas the suggested domain-specific function of the semantic LPC involves the signalling of syntactic violations (Hagoort, Brown & Groothusen, 1993; see Section 5.1.). The EEG results of Experiment 6 challenge these domain-specific accounts, by demonstrating the positive modulation of an LPC-like component for “mismatch” compared to “match” decisions across all presented task domains (see Section 5.3.2.; see Figure 5.4.). Considered with prior reports linking both the semantic and episodic LPCs with more general control processes recruited under expectancy violation (Coulson, King & Kutas, 1998; Herron, Quayle & Rugg, 2003), the observed pattern of results favours a unified function for the LPC in the controlled resolution of cue-item conflict, irrespective of the specific context in which it arises.

The results raise the possibility that previous elicitations of the LPC in recognition and semantic domains in fact represented variants of the P3 complex (specifically the P3b), which displays a similar topography and temporal morphology and has long been linked with cross-domain control demands (Polich, 2007; Pritchard, 1981). As a supplementary point, the present findings also suggest that P3-like components are engaged in the absence of expectation manipulations, thereby addressing a previously noted ambiguity as to whether reported control ERPs are critically dependent on expectancy violation events (e.g. Polich, 2007). It is also worth highlighting that the refinement of prevailing understandings of these ERPs might have been critically enabled by deviation from the more standard paradigms used in the study of control, such as the oddball, go/no-go and Stroop tasks. In venturing beyond the short duration, fast presentation format of these tasks, the present match/mismatch paradigm enabled a longer evaluative window that led to the full emergence of cognitive control sub-processes and their underlying ERP components. This is in keeping with a speculated link between the P3/LPC component and relatively “slow-release” neuromodulatory effects of the locus coeruleus-norepinephrine system (LC-NE) on cortical regions (Nieuwenhuis, de Gee & Aston-Jones, 2010), which is discussed in more detail in the next section.

It is notable that the present EEG analyses failed to identify any other ERP components sensitive to the cross-task mismatch manipulation, suggesting that the LPC might be the exclusive electrophysiological index of “general” control processes. This was confirmed by the group ICA analyses, which decomposed the EEG data into maximally independent signal components that were not biased by any a priori focus on particular electrode sites or trial epochs (see Section 5.2.5 and 5.3.2. for further details). The only task-relevant components recovered by this “model-free” method were topographically and morphologically similar to the LPC identified in the electrode ERP analyses. The absence of the frontal-central N400 from both analyses is particularly noteworthy, given that it has been observed as the “early” component of both recognition retrieval success (Rugg & Curran, 2007) and syntactic parsing violation effects (Coulson et al., 1998) generating the respective LPCs (see Section 5.1.). These findings imply a more specific role for the N400, confined either to the semantic or episodic domains, or to the “cue first” or “item first” conflict ordering conditions. A role in domain-specific evidence processing is supported by prior reports linking the recognition N400 with the linear tracking of memory strength (Finnigan, Humphreys, Dennis & Geffen, 2002; Karayanidis, Andrews, Ward & McConaghy, 1991), and the semantic N400 with the detection of ungrammatical sentences (Coulson et al., 1998). In moving beyond the primary emphasis of Experiment 6 in examining cross-task control processes, future analyses in the present EEG dataset might explore the conditions that generate the N400 and other potential domain- or order-specific ERP components.

The standalone fMRI analyses targeted a similar refinement of BOLD activation patterns previously interpreted as domain-specific. These analyses focused on mismatch control effects in the recognition “cue first” phase, which enabled comparison with the analogous “invalid cueing” effect in the more widely studied likelihood cueing paradigm, given that both serve to isolate memory control processes engaged when items appear to conflict with prior cues (Jaeger, Konkel & Dobbins, 2013; O’Connor et al., 2010). The resultant “recognition mismatch” network overlapped appreciably with regions previously linked with the retrieval success effect, including anterior and rostral cingulate regions, medial and lateral PFC regions, left parietal regions and striatal regions (de Zubicaray, McMahon, Eastburn, Finnigan & Humphreys, 2005; Konishi, Wheeler, Donaldson & Buckner, 2000; Spaniol et al., 2009). The present findings therefore coalesce with findings from the likelihood cueing paradigm in highlighting the sensitivity of these retrieval success regions to manipulations of memory control.

With more careful comparison of the present episodic mismatch contrast and a previous version of the invalid cueing contrast (see Figure 7.1.), it is apparent that the extent of the activation is somewhat reduced in the former compared to the latter, especially over medial and dorsolateral PFC, and sub-cortical regions. Indeed, future research might more directly interrogate these differences between mismatch control and invalid cueing effects, as a means of elucidating potential differences in the engagement of control in the presence and absence of expectation. Despite these noted differences, common memory control activations in these two paradigms were located in regions of the cingulate cortex, left parietal cortex, posterior parietal/occipital cortices, temporal cortex, striatum and cerebellum.

fMRI recognition control networks

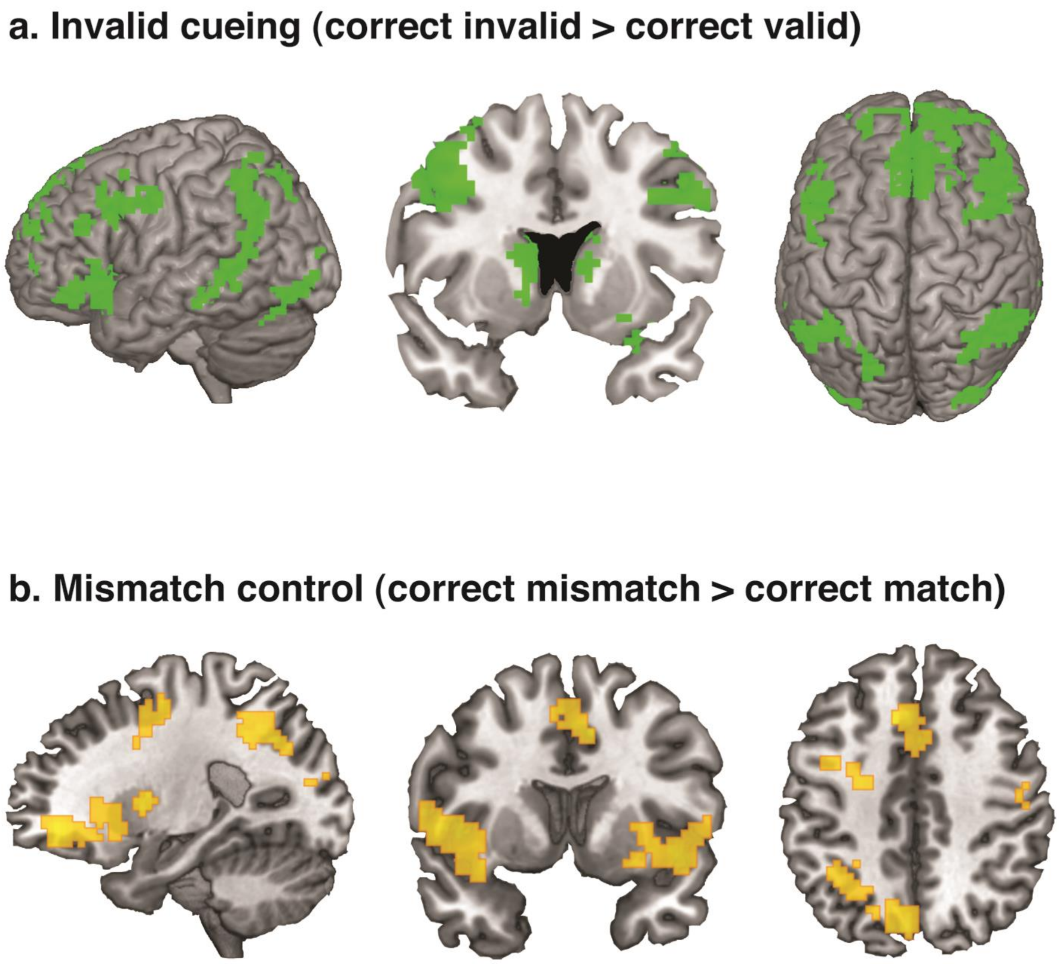


Figure 7.1. Comparison of recognition control networks recovered in different fMRI paradigms. **a.** Activation map for the “invalid cueing” effect (correct invalid cue trials > correct valid cue trials, overlaid in green) observed in the likelihood cueing paradigm, rendered with data provided by Akira O’Connor (from O’Connor et al., 2010). **b.** Activation map for the “mismatch control” effect (correct mismatch > correct match trials, overlaid in yellow) observed in the “cue first” recognition phase of the match/mismatch paradigm (Experiment 6). All activations are overlaid at a threshold of $p < .001$, 5 contiguous voxels.

The recognition control network identified in the standalone fMRI analyses was then parcellated into more precise functional components by a series of masking analyses, involving both the fMRI cross-task mismatch conjunction and the EEG-fMRI parametric modulation analysis (yielding regions that correlated with the cross-task LPC ERP). Recognition control regions that were *exclusive* to both these masks were hence considered as reliable substrates of "recognition-specific" memory control processes (see Figure 5.7a and Table 5.1.). These regions prominently encompassed occipital, ventrolateral PFC and temporal cortex clusters, and were linked with domain-specific source monitoring processes engaged under environmental conflict in an attempt to verify details of prior occurrence (consistent with previous reports; Barredo, Oztekin & Badre, 2013; Dobbins, Foley, Schacter & Wagner, 2002; Dobbins & Wagner, 2005). The imaging analyses of Experiment 6 therefore provide empirical grounds for the distinction between "retrieval control" processes (underpinned by the "recognition-specific" control regions) and "decision control" processes (underpinned by regions active to varying degrees with the two cross-task control masks) raised in the introductory chapter (see Section 1.3.).

Of the recognition mismatch regions that showed some form of inclusivity with the fMRI conjunction and EEG-fMRI modulation maps (see Figure 5.7b), two regions were found to be active in *both* masks and therefore deemed the most reliable cortical substrates of cross-task control – the rostral cingulate zone (RCZ) and the intraparietal sulcus (IPS). Similar "general" control functions have been suggested previously for the identified RCZ and IPS regions, with the RCZ emerging in an fMRI meta-analysis as the region commonly engaged by diverse manipulations of response conflict (Ridderinkhof et al., 2004), and the IPS highlighted as a "general" component of the invalid cueing effect, capable of signalling mnemonic expectancy violation irrespective of the *old/new* status of the violating item (Jaeger et al., 2013). Interestingly, the meta-analysis conducted by Ridderinkhof and colleagues (2004) also identified the N2 ERP component as a potential electrophysiological correlate of fMRI BOLD activation changes in the RCZ. The present integration of the EEG and fMRI data instead suggests that the late positive component of the P3 complex is a more likely ERP correlate of cross-task control.

Although the RCZ and IPS regions were conjunctly identified in the cross-task analyses of Experiment 6, the potential for both to perform subtly different roles within this overarching control involvement persists. One possibility is that the RCZ generates a “top-down” signal that evaluates and selects between representations of available response options in the IPS (as has been suggested previously, Brass, Ullsperger, Knoesche, von Cramon & Phillips, 2005; Bunge, Dudukovic, Thomason, Vaidya & Gabrieli, 2002a). This selection/representation dichotomy is in keeping with prior functional neuroimaging research in humans linking the RCZ with response selection or executive control (Rushworth, Walton, Kennerley & Bannerman, 2004; Mueller, Brass, Waszak & Prinz, 2007) and the IPS with working memory operations (Dobbins & Wagner, 2005; Gillebert et al., 2011; see Section 5.4. for further details). Indeed, a plausible anatomical pathway by which such a signal might be relayed between the RCZ and the IPS via sub-cortical structures is outlined in the next section. Future research is undoubtedly required to elucidate the core function of the RCZ-IPS interaction, and how it relates to another suggested control pathway extending from the mPFC and associated cingulate structures to lateral PFC regions (Kouneiher, Charron & Koechlin, 2009; Ridderinkhof et al., 2004). Exploring the potential for both pathways to mediate different components of the retrieval control/decision control dichotomy (see Section 1.3.) would have clear pertinence in extending the findings of my PhD research, with the RCZ-lateral PFC pathway perhaps involved in the former class of domain-specific evidence operations (such as episodic source monitoring) and the RCZ-IPS pathway involved in the latter domain-general processes that diagnose the eventual response.

Planned future analyses of functional connectivity in the present match/mismatch task might also clarify the specific roles of the RCZ and IPS, as well as the mechanisms of their interaction in the service of cross-task control. This follows from recent evidence suggesting that regional fMRI activations actually reflect on-task changes amongst large-scale functional connectivity networks, linked with flexible and integrative functions within cognitive control (Cole et al., 2013). Indeed, both of the identified cross-task control regions have demonstrated intriguingly ambiguous patterns of functional connectivity in prior research, as mapped using “task-free” resting-state methods (fcMRI). Specifically, the RCZ has been variously identified as part of the frontal-parietal control network (Vincent et al., 2008) and the cingulo-opercular control network (Dosenbach et al., 2006), whereas the IPS has been identified both as part of the dorsal attention network (Corbetta & Shulman, 2002; Fox, Snyder, Zacks & Raichle, 2006) and at the overlap between this network and the frontal-parietal control network (Vincent et al., 2008). The position of these cross-task control regions at the overlap of resting-state networks identified as functionally distinct is in keeping with an integrative function necessary to facilitate the engagement of control across diverse task contexts.

Future analyses might therefore investigate the affiliation of the RCZ and IPS regions to wider functional connectivity networks, and explore their potential operation as cortical “hubs”, in which regional activation coordinates the activity of more specialised functional networks (Cole et al., 2013). Such a hub scheme might mediate both the engagement of mnemonic source monitoring operations (involving lateral PFC, as alluded to earlier in this section), and the generation of the motor response subsequent to its selection by the RCZ-IPS, with both processes facilitating adaptive responses to environmentally induced conflict. Furthermore, devising innovative ways of integrating the EEG data with these fMRI connectivity approaches might yield novel insight into the temporal emergence of the speculated large-scale network dynamics, which would not be feasible if relying on either imaging modality in isolation.

To summarise, the multi-modal imaging approach adopted in Experiment 6 yielded evidence in favour of the cross-domain (and cross-task) control system alluded to in the earlier behavioural experiments (see Section 7.2. and 7.3.) – one that mediates an adaptive response to various forms of conflict arising between item evidence and the environment. The findings demonstrate the virtue in actively manipulating control demands across different task contexts presented to the same sample of subjects, which permits functional inferences that are left ambiguous in studies of isolated task domains. The results also highlight how the present simultaneous EEG-fMRI method overcame limitations specific to each isolated modality, both by effectively improving the poor temporal resolution of the fMRI analyses down to a 400-600ms range (linked with emergence of the LPC ERP), and by overcoming the “inverse problem” of localizing the neural generators of scalp-evoked EEG effects. This insight into the precise functional properties of cortical activations was complemented in Experiment 7 by scrutiny of changes in autonomic nervous system activity that have also been implicated in adaptive control. The major findings from this final experiment, and their relation to the presented EEG-fMRI effects, are discussed in the next section.

7.5. Pupillometry reveals the autonomic correlates of memory control

In investigating the autonomic correlates of memory control, Experiment 7 employed another method of functional neuroimaging – high resolution infra-red pupillometry (see Table 7.1.). This approach follows from single-cell research in animals which established a reliable link between pupil dilation and autonomic nervous system activity mediated by the neuromodulatory locus coeruleus-norepinephrine system (LC-NE, Aston-Jones & Cohen, 2005; see Section 6.1. for further details). The routine involvement of autonomic nervous system changes in cognition has been demonstrated by almost 50 years of human pupillometry research, which has linked increased pupillary dilation (PD) with diverse forms of processing (Beatty, 1982; Kahneman, 1973; Nassar et al., 2012), including episodic recognition memory (Gardner, Mo & Borrego, 1974; Goldinger & Papesh, 2012). As with prevailing interpretations of recognition findings in EEG and fMRI imaging modalities (summarised in the previous section), research into the recognition PD response has also posited pupil dilation as an index of retrieval success, following from the reported increase in dilation for correct “old” compared to “new” decisions (Goldinger & Papesh, 2012; Papesh, Goldinger & Hout, 2012; Vo et al., 2008). It was therefore the primary aim of Experiment 7 to determine whether systematic manipulation of control demands during a recognition pupillometry experiment would lead to a challenge against recognition-specific interpretations of the evoked PD.

Experiment 7 employed a variant of the previously described likelihood cueing paradigm (O’Connor et al., 2010) to probe the sensitivity of the recognition PD response to memory control processes engaged under expectancy violation-induced conflict. Analyses of the mean trial-locked dilation response demonstrated that the amplitude of the recognition PD was exclusively elevated by the experience of unexpected familiarity (i.e. making an “old” decision under a “likely new” cue), with no reliable amplitude elevation observed for any other configuration of decision type or cue type. The amplitude effect was also evoked irrespective of whether the experience of unexpected familiarity was accurate or not. Both the primary sensitivity of the PD response to a specific form of expectancy violation and its complete insensitivity to decision accuracy run contrary to predictions based on retrieval success. Rather, the findings imply that unconstrained expectations operating in previous recognition pupillometry studies may have instilled covert strategic biases that gave rise to old/new effects, consistent with prior research involving the likelihood cueing paradigm (Jaeger et al., 2013; O’Connor et al., 2010) and the behavioural experiments presented in earlier sections (Section 7.2. and 7.3.). The pupillometry findings therefore combine with the findings of Experiment 6 (described in the previous section) in posing a challenge to the retrieval success account, spanning EEG, fMRI and pupillometry neuroimaging modalities.

The highly specific nature of the observed mean PD response also contests prior assertion that pupil dilation signals a general form of expectancy violation or “surprise” (Friedman, Hakerem, Sutton & Fleiss, 1973; Preuschoff, ‘t Hart & Enhauser, 2011; Van Olst, Heemstra & Kortenaar, 1979). Rather, the present findings accord with recent fMRI evidence suggesting that functionally distinct brain networks mediate responses to different forms of mnemonic expectancy violation, respectively encompassing unexpected familiarity, unexpected novelty and more “general” violation responses (Jaeger et al., 2013). The unexpected familiarity network identified by Jaeger and colleagues spanned a number of regions linked previously with aspects of source monitoring, including prefrontal, left lateral parietal and posterior cingulate regions. A similar interpretation can be extended to the present unexpected familiarity effect on PD, such that greater dilation amplitude reflects source monitoring operations engaged to resolve cue-item conflict. A role in source monitoring might explain the observed insensitivity of the PD response to unexpected novelty, given that the engagement of an effortful search for details of prior occurrence would be futile for stimuli perceived as novel. This source monitoring interpretation raises potential overlap with the “recognition-specific” control regions identified in Experiment 6 (see previous section), which were highlighted as neural correlates of similar “retrieval control” processes.

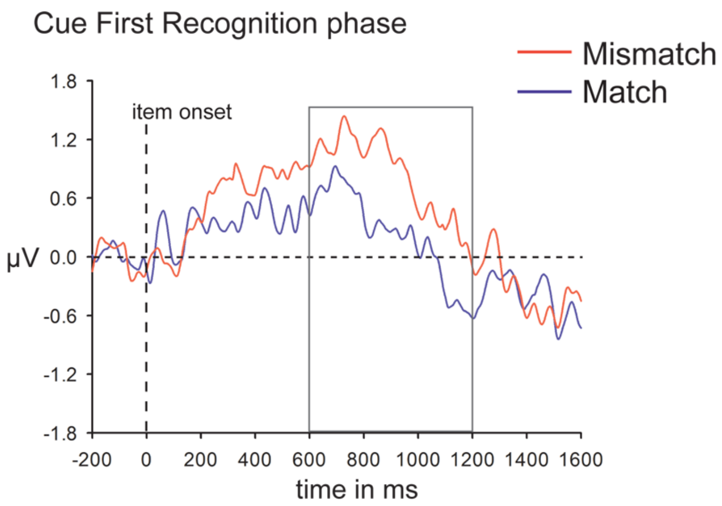
However, an alternative to the source monitoring account is that the specificity of the expectancy violation response to “old” decisions reflects the greater goal emphasis imparted towards them by aspects of the test environment. The influence of such unconstrained sources of goal emphasis was highlighted by the behavioural findings of Experiments 1-3 (see Section 7.2.), which documented shifts in criterion bias that instilled disconfirmatory strategies towards emphasised decisions. Whilst the propensity to instil goal emphasis biases was minimized in Experiment 7 by the complete removal of a question prompt, implicit emphasis might nevertheless have been imparted by the fixed assignment of “old” decisions to the “yes” response category. Such “response emphasis” effects were speculated to account for the bias asymmetries observed in Experiment 5 (see Section 7.3.), in which goal emphasis-driven adjustments in performance impacted primarily on how “old” decisions were evaluated. Indeed, a recent pupillometry study provides support for the involvement of a “yes” emphasis effect, by reporting that PD response during a visual discrimination task elevated exclusively for those participants who countermanded a disconfirmatory bias against responding “yes” (i.e. a “counter-yes” bias; de Gee et al., 2014). Hence, adjudicating between the competing source monitoring and “old” decision salience accounts of the unexpected familiarity effect would require future pupillometry studies to concurrently manipulate goal emphasis and cued expectation (as in Experiment 5), so as to ascertain whether imparting goal salience to “new” decisions leads to a reversal in expectancy effects on the PD.

Further analyses of the pupillometry data aimed to segregate the contribution of functionally distinct memory control processes to the evoked PD response via a multi-level modelling approach (MLM). The MLM analyses yielded a clear functional segregation of the trialwise dilation response into an intercept component, which exclusively predicted the experience of unexpected familiarity (i.e. with greater intercept predicting an increased likelihood of responding “old” to a “likely new” cue), and slope component, which predicted various indices of general decision uncertainty (i.e. with greater slope positivity correlating with slower RT, erroneous responding and the presence of cue-item conflict). The isolation of this latter autonomic correlate corroborates the significant mediating influence of decision uncertainty on aspects of memory control, as highlighted by the findings of Experiment 2 (see Section 7.2.) and Experiment 4 (see Section 7.3.). The observed link between PD slope and uncertainty also serves to address a debate extending beyond the episodic memory domain as to whether uncertainty is indexed in any form by the PD response (Nassar et al., 2012; Schacht, Dimigen & Sommer, 2010; see Section 6.4. for further details). As such, the finding that separate morphological components of the same task-evoked pupillary dilation represent functionally distint cognitive processes may have wide relevance for pupillometry research.

It is also worthwhile highlighting the morphological similarities between the mean PD response evoked in Experiment 7 and the grand averaged late positive ERP component identified in Experiment 6 (see Figure 7.2.), which both demonstrate a slow-wave positivity that is elevated by heightened control demands. This correspondence is not without precedent, given prior speculation that the task-evoked PD response and the P3 ERP complex (with which the observed LPC component is likely affiliated, see Section 7.4.) are both linked as respective autonomic and cortical correlates of a unified cognitive control process (Nieuwenhuis et al., 2010). The findings of Experiments 6 and 7 suggest grounds for further refinement of this link, such that the amplitude of the *specific* LPC sub-component of the P3 and the *specific* trailing slope of the PD response might reflect a common heightening of memory control under conflict and associated uncertainty. Future pupillometry studies spanning cognitive domains beyond episodic memory will be necessary to determine whether the implied overlap between the LPC and PD slope extends to a “cross-domain” control function.

Overlap between ERP and pupillary effects

a. Averaged mismatch control LPC component



b. Averaged invalid cueing PD response

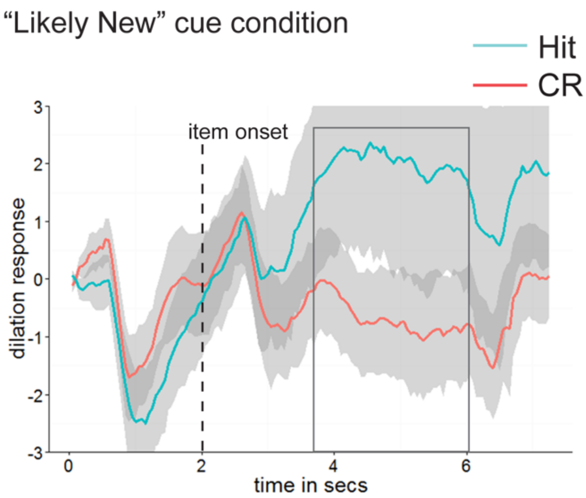


Figure 7.2. Comparison of morphological aspects of presented ERP and pupillometry effects. **a.** Mean item-locked waveform of the late positive ERP component (LPC) sensitive to cross-task mismatch control effects in Experiment 6, focusing on the LPC evoked in the "cue first" recognition phase. **b.** Mean cue-locked waveform of the pupil dilation response (PD) evoked by unexpected familiarity in Experiment 7 i.e. "old" decisions made under "Likely New" cues. Dashed lines represent the point of probe onset in each plot, and grey rectangles highlight the common slow-wave positivity underlying heightened control effects in each imaging modality. "Hits", correct "old" decisions; "CRs", correct "new" decisions.

The overlap between findings of the presented pupillometry and EEG-fMRI experiments is furthered by prior suggestion that the resolution of cue-item conflict is critically underpinned by the activation of the dorsal attention network, of which the intraparietal sulcus region identified in the cross-task analyses of Experiment 6 is considered a key component (Corbetta & Shulman, 2002; Corbetta et al., 2008). Crucially, the mechanisms by which the dorsal attention network is capable of resolving conflict has been associated with a neuromodulatory “reset” signal output by the pupil-linked LC-NE system, which serves to reconfigure the expected response maintained by brain regions in the dorsal attention network in light of conflicting information (Bouret & Sarah, 2005; Corbetta et al., 2008). Furthermore, the rostral cingulate zone region isolated as the other “cross-task” control region in Experiment 6 projects directly to the LC-NE system (Aston-Jones & Cohen, 2005; Nieuwenhuis et al., 2010).

Collectively, the overlap between the EEG-fMRI and pupillometry effects can be captured by an anatomical pathway suggested to regulate the controlled resolution of conflict, wherein a “top-down” signal generated by the RCZ projects to the LC-NE brainstem system, which in turn releases norepinephrine to boost the selection of an appropriate response represented in the IPS. In this model, the greater positivity of the PD slope reflects a longer period of norepinephrine gain modulation, and hence likely greater co-activation of the RCZ and IPS, as engendered by the greater presence of conflict and associated heightening of control. This suggested neural activation stream could be more directly interrogated by multi-modal approaches involving simultaneous pupillometry-fMRI or simultaneous pupillometry-EEG methods. The former would be capable of linking the task-evoked PD response to activity changes in spatially precise cortical regions, whereas the latter might clarify the temporal ordering by which pupil-linked arousal systems and EEG-fMRI-linked cortical systems interact (i.e. whether in concurrent or sequential fashion).

Overall, the pupillometry results of Experiment 7 provide further insight into the adaptive processing that underlies memory control. Despite the need for future corroboration, the collective findings of Experiments 6 and 7 favour the definition of an anatomically and functionally plausible neural pathway, by which controlled responses under evaluative conflict are generated by complementary cortical and autonomic processes.

7.6. Conclusion

The focus of this discussion chapter has been on summarising the major findings of my PhD research, as well as describing implications for prevailing theories of memory and cognitive control, and suggesting some avenues for future research. To highlight the significance of the presented empirical findings, let us re-consider the real-world scenarios introduced in Chapter 1 (see Section 1.1.).

Beginning in this instance with the goal emphasis scenarios (reprised here in Figure 7.3.), one can now furnish additional details of the core memory control processes driving the different evaluative outcomes in the supermarket and eyewitness testing environments. Whilst individuals under evaluation in both real-world environments were stated to evoke an equivalently “weak” level of memory evidence (likely characterised by reduced signal detection sensitivity), the greater perceived consequences of making “old” decisions in the eyewitness compared to the supermarket scenario was assumed to heighten goal emphasis towards this decision category. Considering the findings of Experiments 1-3 and Experiment 5 (summarised in Section 7.2. of this chapter), it is likely that the heightening of goal emphasis in the eyewitness testing scenario is underpinned by instilment of a “counter-emphasis” bias, indexed by a positive shift in signal detection criterion, and leading to a strategic reduction in the likelihood of recognising the individual as “old”. This disconfirmatory bias against responding “old” is likely to facilitate improved accuracy and speed of endorsement for those “old” decisions that are acted upon, in keeping with the heightening of evaluative caution in the eyewitness environment. These goal emphasis strategies will be especially pronounced if the evaluation of “old” decisions is selectively emphasised by aspects of the test environment, such as the framing of questions posed to the observer by the supervising police officer. These updated scenarios should highlight the improved understanding of goal- and reward-related processing involved in the regulation of real-world memory control.

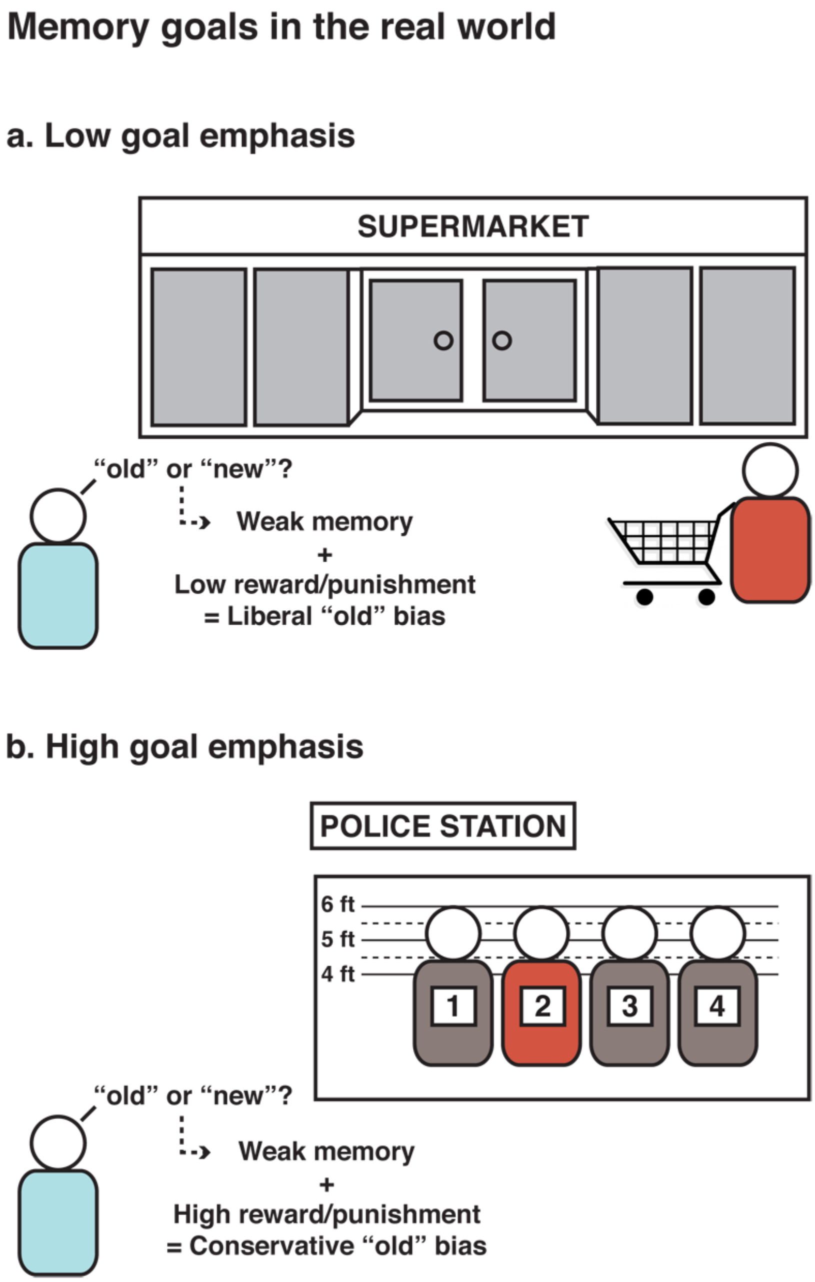


Figure 7.3. Schematic representation of the two scenarios used to demonstrate the influence of environmentally emphasised goals on real world memory evaluation. **a.** Supermarket scenario in which the scope for reward and punishment is reduced (i.e. low goal emphasis). **b.** Eyewitness scenario in which the scope for reward and punishment is increased (i.e. high goal emphasis). In both scenarios, a light blue observer is evaluating a recognised 'old' or unrecognised 'new' decision for an individual in red for whom available memory evidence is equivalently weak. The variation in goal emphasis information across the two scenarios disposes the observer towards different strategies - an increased likelihood of responding 'old' in the supermarket scenario (liberal 'old' bias) and a reduced likelihood of responding 'old' in the eyewitness scenario (conservative 'old' bias).

Turning now to the bus recognition scenario depicted in Figure 7.4., the findings of Experiments 5-7 clarify how environmentally cued expectations interact with the assessment of memory evidence in diagnosing memory decision outcomes. Considering Passenger A, who evoked in the observer a “weak” level of memory evidence and the absence of an expectation of being encountered in the bus environment, the resultant memory evaluation is likely to involve a high degree of uncertainty. This uncertainty is putatively reflected in poor signal detection sensitivity and the lack of a confirmatory “old” criterion shift, leading to a characteristic slowing in reaction time and reduction in decision accuracy. Although it is unclear whether the cross-task memory control substrates of Experiment 6 are engaged in the evaluation of Passenger A (given that these were linked with the resolution of uncertainty rather than uncertainty per se), an increasingly positive slope of the observer’s pupil dilation response can be more confidently asserted during this period of uncertain evaluation. Conversely, the overall “certainty” of the evaluation of Passenger B, who evoked a strong level of memory evidence and a prior expectation, is likely reflected in both greater sensitivity and a confirmatory “pro-expectation” criterion shift, leading to an increase in the accuracy and speed of diagnosing an “old” decision. These overt behavioural characteristics are underpinned by the reduced amplitude of the late positive ERP, reduced activity in the fMRI recognition control network, and reduced amplitude of the evoked pupil dilation response, consistent with the commission of an expected “old” decision.

Passenger C evoked in the observer a strong level of memory evidence in the absence of any expectation of meeting on the bus. This recognition scenario therefore involves a degree of conflict between the decision suggested by the environmental expectation (“not old” or “new”) and that suggested by the retrieved evidence (“old”). Hence, the greater sensitivity of the observer in this particular evaluative event overcomes the absence of any clear criterial tendency to respond “old”. Furthermore, the resolution of evidence-expectation mismatch is likely to engage the LPC and the fMRI recognition control network, as well as generate an amplitude elevation of the dilation response in light of the experience of unexpected familiarity. The final evaluation of Passenger D (the bus conductor) is also characterised by a degree of conflict, albeit one that eventually concludes with endorsement of a prior expectation. Hence, poor sensitivity is overcome by a strong “pro-expectation” criterion shift, leading to the endorsement of an “old” decision. As with passenger B, activation of the cortical correlates of memory control will likely be reduced in this scenario, although the slope of the pupil dilation response might display an increased positivity given the uncertain memory evidence available.

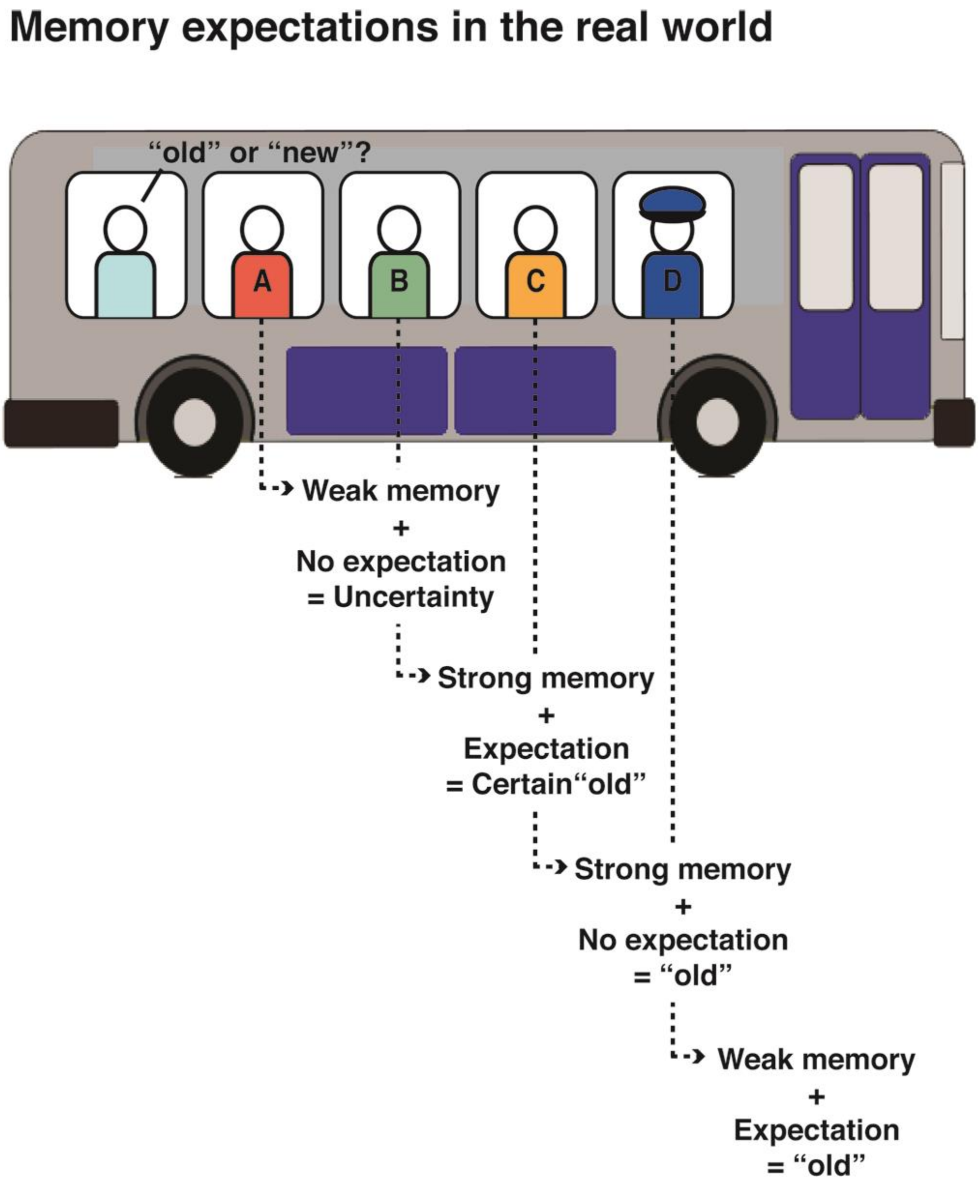


Figure 7.4. Schematic representation of the bus scenario used to demonstrate the influence of environmentally cued expectations on real-world memory evaluation. The passenger in light blue is evaluating recognised 'old' or unrecognised 'new' decisions for passengers A, B, C and D; each of whom is associated with a different combination of evoked memory strength and the presence/absence of expectations of making 'old' decisions, with both factors driving different memory decision outcomes.

Overall, the above augmentation of the introductory real-world examples should highlight the focus of this thesis in clarifying how retrieved memory evidence and various forms of environmental strategies interact in the evaluation of past experience. Whilst both real-world scenarios centre on evaluation in the episodic domain, findings from the empirical chapters should clarify that the assimilation of environmental goals and expectations into the ongoing analysis of memory evidence is mediated to a large extent by cognitive control processes with cross-domain properties. Collectively, the conducted experiments have illumined the workings of this adaptive control system, evidenced through behaviour and functional neuroimaging indices of central and autonomic nervous system activity, which is capable of impacting on episodic memory and beyond. Whilst future research will undoubtedly be necessary to clarify the mechanics of this system, the evidence provided in support of its mere existence nevertheless provides an insight into integrative and domain-general processes that likely perform a fundamental role in upholding our conscious experience amidst changing real-world environments.

References

Abercrombie, E. D., Keller, R. W., & Zigmond, M. J. (1988). Characterization of hippocampal norepinephrine release as measured by microdialysis perfusion: Pharmacological and behavioral studies. *Neuroscience, 27*(3), 897-904. doi: 10.1016/0306-4522(88)90192-3

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *J Exp Psychol Hum Percept Perform, 14*(3), 428-443.

Ahern, S.K., & Beatty, J. (1981). Physiological evidence that demand for processing capacity varies with intelligence. In M. Friedman, J. P. Dos & N. O'Connor (Eds.), *Intelligence and learning*. New York: Plenum Press.

Ahlfors, S. P., Han, J., Lin, F. H., Witzel, T., Belliveau, J. W., Hamalainen, M. S., & Halgren, E. (2010). Cancellation of EEG and MEG signals generated by extended and distributed sources. *Hum Brain Mapp, 31*(1), 140-149. doi: 10.1002/hbm.20851

Allen, P. J., Josephs, O., & Turner, R. (2000). A method for removing imaging artifact from continuous EEG recorded during functional MRI. *Neuroimage, 12*(2), 230-239. doi: 10.1006/nimg.2000.0599

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., . . . Miller, M. B. (2012). Individual differences in shifting decision criterion: a recognition memory study. *Mem Cognit, 40*(7), 1016-1030. doi: 10.3758/s13421-012-0204-6

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience, 27*(14), 3743-3752. doi: 10.1523/Jneurosci.0519-07.2007

Aston-Jones, G., & Bloom, F. E. (1981). Norepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. *J Neurosci, 1*(8), 887-900.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci, 28*, 403-450. doi: 10.1146/annurev.neuro.28.061604.135709

Baddeley, A., & Della Sala, S. (1996). Working memory and executive control. *Philos Trans R Soc Lond B Biol Sci, 351*(1346), 1397-1403; discussion 1403-1394. doi: 10.1098/rstb.1996.0123

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?

Nat Rev Neurosci, 10(9), 659-669. doi: 10.1038/nrn2667

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*(6), 907-918. doi: 10.1016/j.neuron.2005.07.023

Bailenson, J.N., Davies, A., Beall, A.C., Blascovich, J., Guadagno, R.E., & McCall, C. (2008). The Effects of Witness Viewpoint Distance, Angle, and Choice on Eyewitness Accuracy in Police Lineups Conducted in Immersive Virtual Environments. *PRESENCE: Teleoperators and Virtual Environments, 17*(3), 242-255.

Ball, T., Schreiber, A., Feige, B., Wagner, M., Lucking, C. H., & Kristeva-Feige, R. (1999). The role of higher-order motor areas in voluntary movement as revealed by high-resolution EEG and fMRI. *Neuroimage, 10*(6), 682-694. doi: 10.1006/nimg.1999.0507

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior research methods 39*(3), 445-459.

Banich, M. T., Milham, M. P., Atchley, R. A., Cohen, N. J., Webb, A., Wszalek, T., . . . Brown, B. (2000). Prefrontal regions play a predominant role in imposing an attentional 'set': evidence from fMRI. *Cognitive Brain Research, 10*, 1-9.

Barredo, J., Oztekin, I., & Badre, D. (2013). Ventral Fronto-Temporal Pathway Supporting Cognitive Control of Episodic Memory Retrieval. *Cerebral Cortex*. doi: 10.1093/cercor/bht291

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: a study of the effects of test format and aging. *Neuropsychology, 17*(1), 14-24.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull, 91*(2), 276-292.

Benar, C., Aghakhani, Y., Wang, Y., Izenberg, A., Al-Asmi, A., Dubeau, F., & Gotman, J. (2003). Quality of EEG in simultaneous EEG-fMRI for epilepsy. *Clin Neurophysiol, 114*(3), 569-580.

Benjamin, A. S. (2007). Memory Is More Than Just Remembering: Strategic Control of Encoding, Accessing Memory, and Making Decisions. *Skill and Strategy in Memory Use, 48*, 175-223. doi: 10.1016/S0079-7421(07)48005-7

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language, 51*(2), 159-172. doi: 10.1016/j.jml.2004.04.001

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: applications to recognition memory. *Psychol Rev, 116*(1), 84-115. doi:

10.1037/a0014351

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

Berridge, Craig W., & Waterhouse, Barry D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews, 42*(1), 33-84. doi: 10.1016/s0165-0173(03)00143-7

Birnbaum, M. H. (2000). *Psychological experiments on the internet*. USA: Academic Press.

Blackwell, H. R. (1963). Neural theories of simple visual discriminations. *J Opt Soc Am, 53*, 129-160.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn Affect Behav Neurosci, 7*(4), 356-366.

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.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci, 8*(12), 539-546. doi: 10.1016/j.tics.2004.10.003

Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci, 28*(11), 574-582. doi: 10.1016/j.tins.2005.09.002

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brass, M., Ullsperger, M., Knoesche, T. R., von Cramon, D. Y., & Phillips, N. A. (2005). Who comes first? The role of the prefrontal and parietal cortex in cognitive control. *J Cogn Neurosci, 17*(9), 1367-1375. doi: 10.1162/0898929054985400

Brass, M., & von Cramon, D. Y. (2002). The role of the frontal cortex in task preparation. *Cerebral Cortex, 12*(9), 908-914.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn Sci, 16*(2), 106-113. doi: 10.1016/j.tics.2011.12.010

Bruno, D, & Rutherford, A. (2010). How many response options? A study of remember-know testing procedures. *Acta Psychologica, 134*(2), 125-129.

Brutkowski, S. (1964). Prefrontal cortex and drive inhibition. In J. M. Warren & K. Akert (Eds.), *The frontal granular cortex and behavior* (pp. 242-270). New York: McGraw-Hill.

Buchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage, 8*(2), 140-148. doi: 10.1006/nimg.1998.0351

Buckner, R. L. (2003). Functional-anatomic correlates of control processes in memory. *Journal of Neuroscience, 23*(10), 3999-4004.

Bunge, S. A., Dudukovic, N. M., Thomason, M. E., Vaidya, C. J., & Gabrieli, J. D. E. (2002). Development of frontal lobe contributions to cognitive control in children: Evidence from fMRI. *Neuron, 33*, 301-311.

Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *Neuroimage, 17*(3), 1562-1571.

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., & Rosen, B. R. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proc Natl Acad Sci U S A, 99*(1), 523-528. doi: 10.1073/pnas.012470999

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and recognition memory: an attentional account. *Nat Rev Neurosci, 9*, 613-625.

Carver, C. S., & White, T. L. (1994). Behavioral-Inhibition, Behavioral Activation, and Affective Responses to Impending Reward and Punishment - the Bis Bas Scales. *J Pers Soc Psychol, 67*(2), 319-333. doi: 10.1037/0022-3514.67.2.319

Castel, A. D., Benjamin, A. S., Craik, F. I., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Mem Cognit, 30*(7), 1078-1085.

Cavada, C., & Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *The Journal of comparative neurology, 287*(4), 422-445. doi: 10.1002/cne.902870403

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. *J Neurophysiol, 79*(6), 2919-2940.

Chafee, M. V., & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *J Neurophysiol, 83*(3), 1550-1566.

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

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.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324. doi: 10.1016/j.neuron.2008.04.017

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215. doi: 10.1038/nrn755

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes, 13*(1), 21-58. doi: 10.1080/016909698386582

Cox, J. C., & Dobbins, I. G. (2011). The striking similarities between standard, distractor-free, and target-free recognition. *Memory & Cognition, 39*(6), 925-940. doi: 10.3758/s13421-011-0090-3

Craik, Fergus I. M., & Lockhart, Robert S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. doi: 10.1016/s0022-5371(72)80001-x

Damasio, A. (1994). *Descartes' error: emotion, reason, and the human brain*. New York: Avon Books.

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proc Natl Acad Sci U S A, 111*(5), E618-625. doi: 10.1073/pnas.1317557111

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., Finnigan, S., & Humphreys, M. S. (2005). fMRI evidence of word frequency and strength effects in recognition memory. *Brain Res Cogn Brain Res, 24*(3), 587-598. doi: 10.1016/j.cogbrainres.2005.03.009

Debener, S., Mullinger, K. J., Mazy, R. K., & Bowtell, R. W. (2008). Properties of the ballistocardiogram artefact as revealed by EEG recordings at 1.5, 3 and 7 T static magnetic field strength. *International Journal of Psychophysiology, 67*(3), 189-199. doi: 10.1016/j.ijpsycho.2007.05.015

Delgado, M. R., Stenger, V. A., & Fiez, J. A. (2004). Motivation-dependent responses in the human caudate nucleus. *Cereb Cortex, 14*(9), 1022-1030. doi: 10.1093/cercor/bhh062

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods, 134*(1), 9-21. doi: 10.1016/j.jneumeth.2003.10.009

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual-Attention. *Annu Rev Neurosci, 18*, 193-222. doi: 10.1146/annurev.neuro.18.1.193

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron, 35*(5), 989-996.

Dobbins, I. G., & McCarthy, D. (2008). Cue-framing effects in source remembering: a memory misattribution model. *Mem Cognit, 36*(1), 104-118.

Dobbins, I. G., & Wagner, A. D. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cereb Cortex, 15*(11), 1768-1778. doi: 10.1093/cercor/bhi054

Dodson, C. S., & Johnson, M. K. (1993). Rate of False Source Attributions Depends on How Questions Are Asked. *American Journal of Psychology, 106*(4), 541-557. doi: 10.2307/1422968

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Mem Cognit, 24*(4), 523-533.

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

Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat Neurosci, 3*(3), 277-283. doi: 10.1038/72991

Duffy, E. (1962). *Activation and behavior*. New York: Wiley.

Dum, R. P., & Strick, P. L. (1991). The origin of corticospinal projections from the premotor areas in the frontal lobe. *J Neurosci, 11*(3), 667-689.

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science, 337*(6093), 485-487. doi: 10.1126/science.1221936

Eichele, T., Rachakonda, S., Brakedal, B., Eikeland, R., & Calhoun, V. D. (2011). EEGIFT: group independent component analysis for event-related EEG data. *Comput Intell Neurosci, 2011*, 129365. doi: 10.1155/2011/129365

Einhauser, W., Koch, C., & Carter, O. L. (2010). Pupil dilation betrays the timing of decisions. *Front Hum Neurosci, 4*, 18. doi: 10.3389/fnhum.2010.00018

Eriksen, BA, Eriksen, CW. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception and Psychophysics, 16*, 143-149.

Estes, W. K., & Maddox, W. T. (1995). Interactions of stimulus attributes, base rates, and feedback in recognition. *J Exp Psychol Learn Mem Cogn, 21*(5), 1075-1095.

Fiedler, S., & Glockner, A. (2012). The dynamics of decision making in risky choice: an eye-tracking analysis. *Front Psychol, 3*, 335. doi: 10.3389/fpsyg.2012.00335

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new' effects: memory strength and decisional factor(s). *Neuropsychologia, 40*(13), 2288-2304. doi: Pii S0028-3932(02)00113-6

Doi 10.1016/S0028-3932(02)00113-6

Fleck, M. S., Daselaar, S. M., Dobbins, I. G., & Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex, 16*(11), 1623-1630. doi: 10.1093/cercor/bhj097

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philos Trans R Soc Lond B Biol Sci, 367*(1594), 1338-1349. doi: 10.1098/rstb.2011.0417

Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci, 32*(18), 6117-6125. doi: 10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*(5998), 1541-1543. doi: 10.1126/science.1191883

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152-170. doi: 10.1111/j.1469-8986.2007.00602.x

Fox, M. D., Snyder, A. Z., Zacks, J. M., & Raichle, M. E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nat Neurosci, 9*(1), 23-25. doi: 10.1038/nn1616

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev, 25*(4), 355-373.

Friedman, D., Hakerem, G., Sutton, S., & Fleiss, J. L. (1973). Effect of stimulus uncertainty on the pupillary dilation response and the vertex evoked potential. *Electroencephalogr Clin Neurophysiol, 34*(5), 475-484.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2*(1), 45-53. doi: 10.1006/nimg.1995.1007

Fuster, J. M. (1980). *The Prefrontal Cortex*. New York: Raven.

Fuster, J. M. (1985). The prefrontal cortex, mediator of cross-temporal contingencies. *Hum Neurobiol, 4*(3), 169-179.

Gardner, R.M., Mo, S.S., & Borrego, R. (1974). Inhibition of pupillary orienting reflex by novelty in conjunction with recognition memory. *Bulletin of the Psychonomic Society, 3*, 237-238.

Gillebert, C. R., Mantini, D., Thijs, V., Sunaert, S., Dupont, P., & Vandenberghe, R. (2011). Lesion evidence for the critical role of the intraparietal sulcus in spatial attention. *Brain, 134*(Pt 6), 1694-1709. doi: 10.1093/brain/awr085

Gillund, G., & Shiffrin, R. M. (1984). A Retrieval Model for Both Recognition and Recall. *Psychological Review, 91*(1), 1-67. doi: 10.1037/0033-295x.91.1.1

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Mem Cognit, 13*(1), 8-20.

Goldinger, S. D., & Papesh, M. H. (2012). Pupil Dilation Reflects the Creation and Retrieval of Memories. *Current Directions in Psychological Science, 21*(2), 90-95. doi: 10.1177/0963721412436811

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of Physiology: The Nervous System* (pp. 373-417). Bethesda, MD.: American Physiological Society.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*(4), 404-411.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Groninger, L. D. (1976). Predicting recognition during storage: The capacity of the memory system to evaluate itself. *Bulletin of the Psychonomic Society, 7*, 425-428.

Grossman, M., Koenig, P., Kounios, J., McMillan, C., Work, M., & Moore, P. (2006). Category-specific effects in semantic memory: category-task interactions suggested by fMRI. *Neuroimage, 30*(3), 1003-1009. doi: 10.1016/j.neuroimage.2005.10.046

Hagberg, G. E., Zito, G., Patria, F., & Sanes, J. N. (2001). Improved detection of event-related functional MRI signals using probability functions. *Neuroimage, 14*(5), 1193-1205. doi: 10.1006/nimg.2001.0880

Hagoort, P., Brown, C., & Groothusen, J. (1993). The Syntactic Positive Shift (Sps) as an Erp Measure of Syntactic Processing. *Language and Cognitive Processes, 8*(4), 439-483. doi: 10.1080/01690969308407585

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of cognitive neuroscience, 11*(2), 194-205.

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review, 16*(3), 469-474. doi: 10.3758/Pbr.16.3.469

Han, S., Huettel, S. A., Raposo, A., Adcock, R. A., & Dobbins, I. G. (2010). Functional significance of striatal responses during episodic decisions: recovery or goal attainment? *J Neurosci, 30*(13), 4767-4775. doi: 10.1523/JNEUROSCI.3077-09.2010

Harris, D. C. (1998). Nonlinear least-squares curve fitting with Microsoft excel solver. *Journal of Chemical Education, 75*(1), 119-121.

Hartung, C. M., Milich, G., Lynam, D. R., & Martin, C. A. (2002). Understanding the relations among gender, disinhibition, and disruptive behavior in adolescents. *Journal of Abnormal Psychology, 111*(4), 659-664. doi: 10.1037//0021-843x.111.4.659

Hayes, Andrew F. (2006). A Primer on Multilevel Modeling. *Human Communication Research, 32*(4), 385-410. doi: 10.1111/j.1468-2958.2006.00281.x

Hazeltine, E., Poldrack, R., & Gabrieli, J. D. (2000). Neural activation during response competition. *J Cogn Neurosci, 12 Suppl 2*, 118-129. doi: 10.1162/089892900563984

Healy, A. F., & Kubovy, M. (1978). The effects of payoff and prior probabilities on indices of performance and cutoff location in recognition memory. *Mem Cognit, 6*(5), 544-553. doi:: 10.3758/BF03198243

Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory, 19*(4), 398-405. doi: 10.1080/09658211.2011.575788

Henson, R. N., Rugg, M. D., Shallice, T., & Dolan, R. J. (2000). Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. *Journal of cognitive neuroscience, 12*(6), 913-923.

Herron, J. E., Henson, R. N., & Rugg, M. D. (2004). Probability effects on the neural correlates of retrieval success: an fMRI study. *Neuroimage, 21*(1), 302-310.

Herron, J. E., Quayle, A. H., & Rugg, M. D. (2003). Probability effects on event-related potential correlates of recognition memory. *Brain Res Cogn Brain Res, 16*(1), 66-73.

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science, 143*(3611), 1190-1192. doi: 10.1126/science.143.3611.1190

Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review, 6*(1), 117-122. doi: 10.3758/Bf03210818

Himberg, J., Hyvarinen, A., & Esposito, F. (2004). Validating the independent components of neuroimaging time series via clustering and visualization. *Neuroimage, 22*(3), 1214-1222. doi: 10.1016/j.neuroimage.2004.03.027

Hintzman, D. L. (1988). Judgments of Frequency and Recognition Memory in a Multiple-Trace Memory Model. *Psychological Review, 95*(4), 528-551. doi: 10.1037/0033-295x.95.4.528

Hirshman, E. (1995). Decision processes in recognition memory: criterion shifts and the list-strength paradigm. *J Exp Psychol Learn Mem Cogn, 21*(2), 302-313.

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*(3), 401-414. doi: 10.1016/S0191-8869(01)00032-0

Ishihara, S. (1917). *Tests for color-blindness*. Handaya, Tokyo: Hongo Harukicho.

Jacoby, L. L. (1991). A Process Dissociation Framework - Separating Automatic from Intentional Uses of Memory. *Journal of Memory and Language, 30*(5), 513-541. doi: 10.1016/0749-596x(91)90025-F

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope (Eds.), *Dual-process* *theories in social psychology* (Vol. xii, pp. 383-400). New York, NY, US: Guilford Press.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: depth of retrieval. *Psychon Bull Rev, 12*(5), 852-857.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating Conscious and Unconscious Influences of Memory - Measuring Recollection. *J Exp Psychol Gen, 122*(2), 139-154. doi: 10.1037//0096-3445.122.2.139

Jaeger, A., Cox, J. C., & Dobbins, I. G. (2012). Recognition confidence under violated and confirmed memory expectations. *Journal of experimental psychology. General, 141*(2), 282-301. doi: 10.1037/a0025687

Jaeger, A., Konkel, A., & Dobbins, I. G. (2013). Unexpected novelty and familiarity orienting responses in lateral parietal cortex during recognition judgment. *Neuropsychologia, 51*(6), 1061-1076. doi: 10.1016/j.neuropsychologia.2013.02.018

John, O.P., Donahue, E.M., & Kentle, R.L. (1991). *The Big five inventory - versions 4a and 54*. Berkeley, CA: University of California, Berkeley.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin, 114*(1), 3-28.

Johnson, M. K., Kounios, J., & Reeder, J. A. (1994). Time-course studies of reality monitoring and recognition. *J Exp Psychol Learn Mem Cogn, 20*(6), 1409-1419.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163-178.

Kahneman, D. (1973). *Attention and effort*. New Jersey: Prentice Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583-1585.

Kahneman, D., & Beatty, J. (1967). Pupillary responses in a pitch-discrimination task. *Perception and Psychophysics, 2*, 101-105.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases*. New York, NY: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237-251.

Kapucu, A., Rotello, C. M., Ready, R. E., & Seidl, K. N. (2008). Response bias in "Remembering" emotional stimuli: A new perspective on age differences. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(3), 703-711. doi: 10.1037/0278-7393.34.3.703

Karayanidis, F., Andrews, S., Ward, P. B., & McConaghy, N. (1991). Effects of inter-item lag on word repetition: an event-related potential study. *Psychophysiology, 28*(3), 307-318.

Kawashima, R., Roland, P. E., & O'Sullivan, B. T. (1995). Functional anatomy of reaching and visuomotor learning: a positron emission tomography study. *Cereb Cortex, 5*(2), 111-122.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., 3rd, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*(5660), 1023-1026. doi: 10.1126/science.1089910

Khoe, W., Kroll, N. E., Yonelinas, A. P., Dobbins, I. G., & Knight, R. T. (2000). The contribution of recollection and familiarity to yes-no and forced-choice recognition tests in healthy subjects and amnesics. *Neuropsychologia, 38*(10), 1333-1341.

Kim, H. (2013). Differential neural activity in the recognition of old versus new events: an activation likelihood estimation meta-analysis. *Hum Brain Mapp, 34*(4), 814-836. doi: 10.1002/hbm.21474

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci, 21*(16), RC159.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissocation of reward anticipation and outcome with event-related fMRI. *Neuroreport, 12*, 3683-3687.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *J Neurosci, 25*(19), 4806-4812. doi: 10.1523/JNEUROSCI.0642-05.2005

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399*(6732), 148-151. doi: 10.1038/20178

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302*(5648), 1181-1185. doi: 10.1126/science.1088545

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology, 38*(3), 557-577. doi: 10.1017/S0048577201990559

Konishi, S., Wheeler, M. E., Donaldson, D. I., & Buckner, R. L. (2000). Neural correlates of episodic retrieval success. *Neuroimage, 12*(3), 276-286. doi: 10.1006/nimg.2000.0614

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *J Exp Psychol Gen, 123*(3), 297-315.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol Rev, 103*(3), 490-517.

Kouneiher, F., Charron, S., & Koechlin, E. (2009). Motivation and cognitive control in the human prefrontal cortex. *Nat Neurosci, 12*(7), 939-945. doi: 10.1038/nn.2321

Kunimoto, C., Miller, J., & Pashler, H. (2001). Confidence and accuracy of near-threshold discrimination responses. *Conscious Cogn, 10*(3), 294-340. doi: 10.1006/ccog.2000.0494

Lau, H. (2010). Are we studying consciousness yet? In L. Weiskrantz & M. Davies (Eds.), *Frontiers of consciousness: Chichele lectures*. Oxford, UK: Oxford University Press.

LeCompte, D. C. (1995). Recollective experience in the revelation effect: separating the contributions of recollection and familiarity. *Mem Cognit, 23*(3), 324-334.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Mem Cognit, 17*(3), 349-358.

Lockhart, R. S., Craik, F. I., & Jacoby, L. L. (1976). Depth of processing, recognition and recall: some aspects of a general memory system. In J. Brown (Ed.), *Recall and Recognition*. London: Wiley.

Loftus, E. F. (1996). *Eyewitness testimony*. Cambridge, Mass.: Harvard University Press.

Lopes da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr Clin Neurophysiol, 79*(2), 81-93.

Lundstrom, B. N., Ingvar, M., & Petersson, K. M. (2005). The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage, 27*(4), 824-834. doi: 10.1016/j.neuroimage.2005.05.008

MacDonald, A. W., 3rd, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*(5472), 1835-1838.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (Second ed.)*. New York: Lawrence Erlbaum.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *J Exp Psychol Learn Mem Cogn, 23*(3), 539-559.

Mandler, G. (1980). Recognizing - the Judgment of Previous Occurrence. *Psychological Review, 87*(3), 252-271. doi: 10.1037//0033-295x.87.3.252

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology-Learning Memory and Cognition, 24*(5), 1137-1151. doi: 10.1037//0278-7393.24.5.1137

Mill, R. D., & O'Connor, A. R. (2014). Question format shifts bias away from the emphasised response in tests of recognition memory. *Conscious Cogn, 30C*, 91-104. doi: 10.1016/j.concog.2014.09.006

Mill, R.D., & O'Connor, A. R. (under review). Differentiating the functional contributions of resting connectivity networks to memory decision-making: fMRI support for multi-stage control processes.

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

Moeser, S. D. (1974). Memory for Meaning and Wording in Concrete and Abstract Sentences. *Journal of Verbal Learning and Verbal Behavior, 13*(6), 682-697.

Monsell, S. (2003). Task switching. *Trends Cogn Sci, 7*(3), 134-140.

Moray, N. (1967). Where is capacity limited? A survey and a model. *Acta Psychologica, 27*, 84-92.

Morrell, H. E. R., Gaitan, S., & Wixted, J. T. (2002). On the nature of the decision axis in signal-dection-based models of recognition memory. *J Exp Psychol Learn Mem Cogn, 19*, 851-864.

Mozer, M.C., Kinoshita, S., & Shettel, M. (2007). Sequential Dependencies in Human Behavior Offer Insights into Cognitive Control. In W. D. Gray (Ed.), *Integrated models of cognitive systems*: Oxford Scholarship Online.

Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *Neuroimage, 37*(4), 1354-1361. doi: 10.1016/j.neuroimage.2007.06.018

Näätänen, R., & Gaillard, A. W. K. (1983). The N2 deflection of the ERP and the orienting reflex. In A. W. K. Gaillard & W. Ritter (Eds.), *Tutorials in event-related potential research: endogenous components.* Amsterdam, North Holland.

Naber, M., Frassle, S., Rutishauser, U., & Einhauser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *J Vis, 13*(2), 11. doi: 10.1167/13.2.11

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat Neurosci, 15*(7), 1040-1046. doi: 10.1038/nn.3130

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1-25). Cambridge, Mass.: MIT Press.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-Related Brain Potentials during Initial Encoding and Recognition Memory of Congruous and Incongruous Words. *Journal of Memory and Language, 25*(1), 75-92. doi: 10.1016/0749-596x(86)90022-7

Newman, J. P., Widom, C. S., & Nathan, S. (1985). Passive avoidance in syndromes of disinhibition: psychopathy and extraversion. *J Pers Soc Psychol, 48*(5), 1316-1327.

Niazy, R. K., Beckmann, C. F., Iannetti, G. D., Brady, J. M., & Smith, S. M. (2005). Removal of FMRI environment artifacts from EEG data using optimal basis sets. *Neuroimage, 28*(3), 720-737. doi: 10.1016/j.neuroimage.2005.06.067

Nieuwenhuis, I. L., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural brain research, 218*(2), 325-334. doi: 10.1016/j.bbr.2010.12.009

Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2010). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*. doi: 10.1111/j.1469-8986.2010.01057.x

Nilsson, L. G. (Motivated memory: dissociation between performance data and subjective reports). 1987. *Psychological Research, 49*, 183-188.

Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behaviour. In R. Davidson, R. G. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1-18). New York: Plenum Press.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia, 41*(3), 371-377.

O'Connor, A. R., Guhl, E. N., Cox, J. C., & Dobbins, I. G. (2011). Some memories are odder than others: Judgments of episodic oddity violate known decision rules. *Journal of Memory and Language, 64*(4), 299-315. doi: 10.1016/j.jml.2011.02.001

O'Connor, A. R., Han, S., & Dobbins, I. G. (2010). The inferior parietal lobule and recognition memory: expectancy violation or successful retrieval? *The Journal of Neuroscience, 30*(8), 2924-2934. doi: 10.1523/JNEUROSCI.4225-09.2010

Ohrt, D. D., & Gronlund, S. D. (1999). The list-length effect and continuous memory: confounds and solutions. In C. Izawa (Ed.), *On Human Memory: Evolution, Progress and Reflections on the 30th anniversary of the Atkinson-Shiffrin Model*. Mahwah, NJ: Lawrence Erlbaum.

Osterhout, L., & Holcomb, P. J. (1992). Event-Related Brain Potentials Elicited by Syntactic Anomaly. *Journal of Memory and Language, 31*(6), 785-806. doi: 10.1016/0749-596x(92)90039-Z

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology, 48*(10), 1346-1353. doi: 10.1111/j.1469-8986.2011.01217.x

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *Int J Psychophysiol, 83*(1), 56-64. doi: 10.1016/j.ijpsycho.2011.10.002

Passingham, R. E. (1993). *The Frontal Lobes and Voluntary Action.* Oxford UK: Oxford University Press.

Paul, D. D., & Sutton, S. (1972). Evoked potential correlates of response criterion in auditory signal detection. *Science, 177*(4046), 362-364.

Peterson, W. W., Birdsall, T. G., & Fox, W. C. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory, 4*, 171-212.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J Appl Psychol, 88*(5), 879-903. doi: 10.1037/0021-9010.88.5.879

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128-2148. doi: 10.1016/j.clinph.2007.04.019

Posner, MI, Snyder, CRR, Davidson, BJ. (1980). Attention and the detection of signals. *Journal of experimental psychology. General., 109*, 160-174.

Preuschoff, K., t Hart, B. M., & Einhauser, W. (2011). Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. *Front Neurosci, 5*, 115. doi: 10.3389/fnins.2011.00115

Pritchard, W.S. (1981). Psychophysiology of P300: A review. *Psychological Bulletin, 89*, 504-540.

Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1993). Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. *Society for Neuroscience Abstracts, 19*, 974.

Ranganath, C., Johnson, M. K., & D'Esposito, M. (2003). Prefrontal activity associated with working memory and episodic long-term memory. *Neuropsychologia, 41*(3), 378-389.

Ratcliff, Roger. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59-108. doi: 10.1037/0033-295x.85.2.59

Ratcliff, R., Gronlund, S. D., & Sheu, C. F. (1992). Testing Global Memory Models Using Roc Curves. *Psychological Review, 99*(3), 518-535. doi: 10.1037//0033-295x.99.3.518

Ratcliff, Roger, & Murdock, Bennet B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83*(3), 190-214. doi: 10.1037/0033-295x.83.3.190

Reder, L. M., & Wible, C. (1984). Strategy Use in Question-Answering - Memory Strength and Task Constraints on Fan Effects. *Memory & Cognition, 12*(4), 411-419. doi: 10.3758/Bf03198302

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: effects of base rate, awareness, and feedback. *J Exp Psychol Learn Mem Cogn, 33*(2), 305-320. doi: 10.1037/0278-7393.33.2.305

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*(5695), 443-447. doi: 10.1126/science.1100301

Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*(2), 129-140. doi: 10.1016/j.bandc.2004.09.016

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cereb Cortex, 15*(11), 1779-1790. doi: 10.1093/cercor/bhi055

Rubin, S. R., Van Petten, C., Glisky, E. L., & Newberg, W. N. (1999). Memory conjunction errors in younger and older adults: Event-related potential and neuropsychological data. *Cognitive Neuropsychology, 16*(3-5), 459-488.

Ruchkin, D.S., & Sutton, S. (1983). Positive slow wave and P300: association and disassociation. *Advances in Psychology, 10*, 233-250.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn Sci, 11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., & Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature, 392*(6676), 595-598. doi: 10.1038/33396

Rushworth, M. F., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci, 8*(9), 410-417. doi: 10.1016/j.tics.2004.07.009

Sanquist, T. F., Rohrbaugh, J. W., Syndulko, K., & Lindsley, D. B. (1980). Electrocortical signs of levels of processing: perceptual analysis and recognition memory. *Psychophysiology, 17*(6), 568-576.

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: evidence from pupillometry and event-related fMRI. *Neuroimage, 37*(3), 1017-1031. doi: 10.1016/j.neuroimage.2007.04.066

Schacht, A., Dimigen, O., & Sommer, W. (2010). Emotions in cognitive conflicts are not aversive but are task specific. *Cogn Affect Behav Neurosci, 10*(3), 349-356. doi: 10.3758/CABN.10.3.349

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science, 249*(4971), 892-895.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory.

Psychological Review, 84(2), 127-190.

Small, D. M., Gitelman, D., Simmons, K., Bloise, S. M., Parrish, T., & Mesulam, M. M. (2005). Monetary incentives enhance processing in brain regions mediating top-down control of attention. *Cereb Cortex, 15*(12), 1855-1865. doi: 10.1093/cercor/bhi063

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of experimental psychology. General, 117*(1), 34-50.

Sokolov, E. N. (1963). Higher nervous functions; the orienting reflex. *Annu Rev Physiol, 25*, 545-580. doi: 10.1146/annurev.ph.25.030163.002553

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *J Exp Psychol Learn Mem Cogn, 26*(1), 204-221.

Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Conscious Cogn, 20*(4), 1787-1792. doi: 10.1016/j.concog.2010.12.011

Spaniol, J., Davidson, P. S. R., Kim, A. S. N., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia, 47*(8-9), 1765-1779. doi: 10.1016/j.neuropsychologia.2009.02.028

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nat Rev Neurosci, 8*(11), 872-883. doi: 10.1038/nrn2154

Squires, K. C., Wickens, C., Squires, N. K., & Donchin, E. (1976). The effect of stimulus sequence on the waveform of the cortical event-related potential. *Science, 193*(4258), 1142-1146.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol, 38*(4), 387-401.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *J Exp Psychol Learn Mem Cogn, 24*(6), 1397-1410. doi: 10.1037//0278-7393.24.6.1397

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *J Exp Psychol Learn Mem Cogn, 24*(6), 1379-1396. doi: 10.1037//0278-7393.24.6.1379

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi: 10.1037/0096-3445.121.1.15

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *J Exp*

Psychol Hum Percept Perform, 13(1), 89-103.
Swets, J., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychol Rev, 68*, 301-340.
Swets, J. A. (2014). *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition, 32*(5), 742-751.
Tanner, W. P. (1956). A theory of recognition. *Journal of the Acoustic Society of America, 28*, 922-928.
Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol Rev, 61*(6), 401-409.
Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *J Exp Psychol Hum Percept Perform, 8*(6), 834-854.
Tryon, Warren W. (1975). Pupillometry: A Survey of Sources of Variation. *Psychophysiology, 12*(1), 90-93. doi: 10.1111/j.1469-8986.1975.tb03068.x
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory*. New York: Academic Press.
van Olst, E.H., Heemstra, M.L., & Ten Kortenaar, T. (1979). Stimulus significance and the orienting reaction. In H. D. Kimmel, E. H. van Olst & J. F. Orlebeke (Eds.), *The orienting reflex in humans*. New York: Lawrence Erlbaum Associates.
Van Zandt, Trisha. (2000). ROC curves and confidence judgments in recognition memory. *J Exp Psychol Learn Mem Cogn, 26*(3), 582-600. doi: 10.1037/0278-7393.26.3.582
Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *Journal of Neurophysiology, 100*(6), 3328-3342. doi: 10.1152/jn.90355.2008
Vo, M. L., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: introducing the pupil old/new effect. *Psychophysiology, 45*(1), 130-140. doi: 10.1111/j.1469-8986.2007.00606.x
Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage, 18*(2), 293-309. doi: 10.1016/S1053-8119(02)00046-0
Wagner, A. D., Maril, A., Bjork, R. A., & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal

cortex. *Neuroimage, 14*(6), 1337-1347. doi: 10.1006/nimg.2001.0936

Wagner, A. D., Shannon, B. J., Kahn, I., & Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn Sci, 9*(9), 445-453. doi: 10.1016/j.tics.2005.07.001

Warren, L. R. (1980). Evoked potential correlates of recognition memory. *Biol Psychol, 11*(1), 21-35.

Weber, N., & Perfect, T. J. (2012). Improving Eyewitness Identification Accuracy by Screening Out Those Who Say They Don't Know. *Law and Human Behavior, 36*(1), 28-36. doi: 10.1037/H0093976

Wheeler, M. E., & Buckner, R. L. (2003). Functional dissociation among components of remembering: Control, perceived oldness, and content. *Journal of Neuroscience, 23*(9), 3869-3880.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *J Exp Psychol Learn Mem Cogn, 40*(2), 385-398. doi: 10.1037/a0034851

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *J Vis, 8*(3), 2 1-15. doi: 10.1167/8.3.2

Whittlesea, B. W. A. (2002). Two routes to remembering (and another to remembering not). *J Exp Psychol Gen, 131*, 325-348.

Whittlesea, B. W. A. (2004). The perception of integrality: Remembering through the validation of expectation. *Journal of Experimental Psychology-Learning Memory and Cognition, 30*(4), 891-908. doi: 10.1037/0278-7393.30.4.891

Wirsich, J., Benar, C., Ranjeva, J. P., Descoins, M., Soulier, E., Le Troter, A., . . . Guye, M. (2014). Single-trial EEG-informed fMRI reveals spatial dependency of BOLD signal on early and late IC-ERP amplitudes during face recognition. *Neuroimage, 100*, 325-336. doi: 10.1016/j.neuroimage.2014.05.075

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron, 23*(4), 747-764.

Wood, E. R., Dudchenko, P. A., & Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature, 397*(6720), 613-616. doi: 10.1038/17605

Woodruff, C. C., Hayama, H. R., & Rugg, M. D. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain Res, 1100*(1), 125-135. doi: 10.1016/j.brainres.2006.05.019

Worsley, K. J., & Friston, K. J. (1995). Analysis of Fmri Time-Series Revisited - Again. *Neuroimage, 2*(3), 173-181. doi: 10.1006/nimg.1995.1023

Wright, W. D., & Pitt, F. H. G. (1934). Hue-discrimination in normal colour-vision. *Proceedings of the Physical Society, 46*, 459-473. doi: 10.1088/0959-5309/46/3/317

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(3), 441-517. doi: 10.1006/jmla.2002.2864

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Conscious Cogn, 5*(4), 418-441. doi: 10.1006/ccog.1996.0026

Zechmeister, E. B. (1972). Orthographic Distinctiveness as a Variable in Word Recognition. *American Journal of Psychology, 85*(3), 425-430. doi: 10.2307/1420843

Appendix A

Question format shifts bias away from the emphasised response in tests of recognition memory

Ravi D. Mill, Akira O'Connor

Consciousness and cognition 30 (2014) 91-104

http://dx.doi.org/10.1016/j.concog.2014.09.006

Owing to copyright restrictions, the electronic version of this thesis does not contain the text of this article